

Índice general

1. Estadística descriptiva	9
1.1. Introducción a la estadística descriptiva	10
1.1.1. Agrupamiento en clases	15
1.2. Representaciones gráficas	20
1.2.1. Diagrama de barras	20
1.2.2. Histogramas	21
1.2.3. Diagrama de Pareto	21
1.2.4. Polígonos o curvas de frecuencias	22
1.2.5. Series temporales	23
1.3. Medidas características	24
1.3.1. Medidas de centralización	26
1.3.2. Medidas de dispersión	31
1.3.3. Momentos de una distribución de datos	37
1.3.4. Medidas de asimetría	38
1.3.5. Medidas de apuntamiento o curtosis	39
1.4. Transformaciones de la variable aleatoria	40
1.5. Muestras multivariantes	44
1.5.1. Distribuciones de frecuencia	44
1.5.2. Distribuciones marginales de frecuencia	45
1.5.3. Distribuciones de frecuencias condicionadas	46
1.5.4. Representaciones gráficas	49
1.5.5. Medidas características de la distribución de frecuencias	50
1.5.6. Covarianza y recta de regresión	55
1.6. Cuestiones y problemas	60
2. Teoría de probabilidades	65
2.1. Introducción a la teoría de probabilidades	65
2.2. Probabilidad condicionada	70
2.2.1. Teorema de Bayes	73
2.2.2. Entropía estadística. Principio de máxima entropía (PME).	75
2.3. Distribuciones de probabilidad	79
2.3.1. Variables aleatorias discretas y continuas	79
2.3.2. Variable aleatoria discreta. Función de probabilidad y función de distribución de probabilidad	80

2.3.3.	Variable aleatoria continua. Función de densidad de probabilidad y función de distribución de probabilidad. . .	81
2.3.4.	Medidas características de una distribución de probabilidad.	86
2.4.	Función característica	90
2.4.1.	Función generatriz de momentos	92
2.4.2.	Cumulantes. Función generatriz de cumulantes.	94
2.5.	Transformación de variables aleatorias.	94
2.6.	Algunas distribuciones de probabilidad relevantes	99
2.6.1.	Distribución binomial	99
2.6.2.	Distribución geométrica o de Pascal	105
2.6.3.	Distribución de Poisson	106
2.6.4.	Distribución exponencial	110
2.6.5.	Distribución normal o gaussiana	111
2.6.6.	Distribución log-normal	121
2.6.7.	Distribución de Weibull	122
2.6.8.	Procesos aleatorios estables. Distribuciones de Lévy. Distribución de Pareto.	125
2.7.	Distribuciones de probabilidad multidimensionales	130
2.7.1.	Medidas características de una distribución de probabilidad multivariante	134
2.7.2.	Distribuciones multinomial y normal n-dimensional . . .	138
2.8.	Cuestiones y problemas	141
3.	Métodos estadísticos	151
3.1.	Estimación de parámetros	152
3.1.1.	Estimación puntual de la media poblacional	155
3.1.2.	Estimación puntual de la varianza poblacional	159
3.1.3.	Distribución de \bar{x} para poblaciones madre normales . .	167
3.2.	Intervalos de confianza	170
3.2.1.	Intervalos de confianza para la media poblacional	171
3.2.2.	Intervalo de confianza para la varianza σ^2 de una distribución normal	178
3.3.	Test de hipótesis estadísticas	180
3.3.1.	Test de la media de una población normal	182
3.3.2.	Contraste de la varianza de una población normal . . .	184
3.3.3.	Test de χ^2 de una distribución de probabilidad	185
3.3.4.	Test de χ^2 de la bondad de un ajuste	188
3.3.5.	Test de hipótesis para dos muestras	191
3.4.	Método de máxima verosimilitud	193
3.4.1.	Estimador de la media poblacional	194
3.4.2.	Ajuste por el método de mínimos cuadrados	197
3.5.	Cuestiones y problemas	222

4. Expresión de la incertidumbre de la medida	227
4.1. Introducción. Error e incertidumbre	227
4.2. Unidades y expresión de la medida	229
4.2.1. Sistema Internacional de Unidades	230
4.2.2. Múltiplos y submúltiplos	233
4.2.3. Reglas prácticas de escritura	233
4.3. Cifras significativas y redondeo	234
4.3.1. Reglas para contar el número de cifras significativas	236
4.3.2. Reglas de redondeo	237
4.4. Incertidumbres de tipo A y de tipo B	238
4.4.1. Evaluación de incertidumbres de tipo A	239
4.4.2. Evaluación de incertidumbres de tipo B	240
4.5. Incertidumbre combinada. Propagación de incertidumbres.	244
4.5.1. Magnitudes de entrada no correlacionadas.	246
4.5.2. Magnitudes de entrada correlacionadas.	249
4.5.3. Incertidumbre expandida. Factor de cobertura.	252
4.6. Cuestiones y problemas	254
A. Tablas estadísticas	257
A.1. Distribución normal estándar	258
A.2. Percentiles más útiles de la χ^2 de Pearson	261
A.3. Percentiles más útiles de la t de Student	263
A.4. Percentiles más útiles de la F de Fisher	264
Bibliografía	273

Introducción

El presente texto de Tratamiento de Datos aborda la tarea de realizar una exposición actualizada del estado del arte en los conceptos y tratamiento de datos experimentales. Aunque resulta imposible exponer todas las técnicas de tratamiento de datos dentro de las disciplinas experimentales, este libro tiene como objetivo fundamental establecer las bases del análisis de datos en la ciencia experimental actual, especialmente en el ámbito de las Ciencias Físicas. Aunque se trata de una obra esencialmente introductoria a esta disciplina, no evita la introducción de conceptos y técnicas de nivel intermedio e incluso avanzadas. Así, el texto comienza con una exposición de los fundamentos de la estadística descriptiva (concepto de variable aleatoria, distribuciones de probabilidad y sus propiedades, incluyendo una exposición detallada de la estadística de procesos estables por primera vez en un texto de estas características y nivel, etc.). A continuación se abordan diferentes aspectos relacionados con la inferencia estadística (muestreo, estimadores) y, finalmente, se exponen algunos de los métodos estadísticos más comunes en el estudio de los datos (test de hipótesis estadísticas, método de máxima verosimilitud).

Uno de los aspectos novedosos de este libro consiste en vincular la exposición de los métodos estadísticos con la evaluación de la *incertidumbre* de la medida (diferenciándola del concepto de *error* usado hasta ahora de modo poco riguroso). El texto introduce aquí muy brevemente el sistema internacional de unidades y las reglas actualizadas de nomenclatura (ISO31-0), para proseguir con una exposición de los dos métodos de evaluación de incertidumbres (tipo A y B). Finalmente, y de nuevo de un modo plenamente vinculado a la estadística, se establecen las reglas de propagación de incertidumbres y la expresión final de la medida (valor estimado del mensurando e incertidumbre) definiendo el factor de cobertura.

Aunque desde el año 1981 existe una recomendación del Comité Internacional de Pesas y Medidas (CI-1981) sobre cómo se debe expresar el resultado de una medida experimental, ésta es poco conocida dentro del ámbito académico y científico. En España la nomenclatura y metodología básicas asociadas a la expresión de la medida están recogidas en la "Guía para la expresión de la incertidumbre de la medida" editada en el año 2000 por el Centro Español de Metrología. Ésta es una traducción de la popularmente conocida como *GUM* en el ámbito internacional ("Guide for the expression of the Uncertainty of Measurement") publicada por la International Organization for Standardization. A pesar de su conocimiento limitado dentro de la comunidad universitaria, es-

ta metodología está avalada por las organizaciones científicas y metroológicas más importantes en el mundo, como el Comité Internacional de Pesas y Medidas, la Comisión Electrotécnica Internacional, la International Organization for Standardization, la Organización Internacional de Metrología Legal, la Unión Internacional de Química Pura y Aplicada o la Unión Internacional de Física Pura y Aplicada. Es fundamental destacar que la metodología actualmente propuesta establece una profunda conexión conceptual entre la estadística y la evaluación de la incertidumbre de la medida experimental.

Este Manual en español conjuga todas estas vertientes (estadística, análisis de datos y expresión y cálculo de la incertidumbre de medida) en un sólo texto. De aquí por un lado la relevancia de la propuesta. Pero además cabe señalar la trascendencia del texto ya que su objetivo último es proporcionar – a quienquiera que desee medir – los conceptos y herramientas actualizados para hacerlo. Por tanto, aunque concebido para estudiantes universitarios, el texto es de interés para cualquier investigador o docente experimental, e incluso también, para la comunidad industrial sujeta a la Metrología Legal.

Santiago, febrero de 2010

Luis Miguel Varela Cabo

Prof. Titular del Departamento de Física de la Materia Condensada

Faustino Gómez Rodríguez

Prof. Titular del Departamento de Física de Partículas

Jesús Carrete Montaña

Licenciado en Física

Universidad de Santiago de Compostela

Capítulo 1

Estadística descriptiva

Las ciencias experimentales se basan en la contrastación de hipótesis con las medidas empíricas. Es sorprendente descubrir que incluso en la determinación del valor de una magnitud física subsiste de modo inevitable cierto nivel de variabilidad o imprecisión. La naturaleza de esta dificultad puede residir en diferentes fuentes: bien debido a la imperfección de nuestros instrumentos, a la propia fluctuación del mensurando o a la existencia de perturbaciones de la medida que no podemos controlar. Baste por el momento con reconocer que en cualquier conjunto de medidas de una propiedad física realizadas con instrumentos de calidad suficiente existirá inevitablemente variación entre los valores de la magnitud obtenidos en los sucesivos ensayos.

Consideremos, por ejemplo, la medida de la masa de un objeto mediante un experimento formado por cinco ensayos sucesivos. En el caso de que deseemos conocer la masa de un determinado objeto, es evidente que no tenemos más remedio que realizar su comparación con un patrón de masa previamente establecido, operación para la cual debemos utilizar un instrumento calibrado con dicho patrón que denominamos balanza. Cabría esperar que, si nuestro instrumento funcionase suficientemente bien, obtendríamos un valor igual al valor real de la masa del cuerpo y que nuestra lectura se mantendría constante en sucesivas mediciones. En realidad, al realizar el experimento de determinación de la masa del cuerpo con una balanza de la precisión suficiente obtendremos un conjunto de medidas como, por ejemplo, el siguiente:

$$m \text{ (kg)} = \{1,992035; 1,992036; 1,992037; 1,992033; 1,992034\}$$

El hecho más relevante no es que la masa del cuerpo sea una u otra, ni siquiera que podamos adivinar que ésta está situada en torno a un valor concreto, sino el de que exista una variación entre observaciones sucesivas del valor de la masa del cuerpo y que el resultado de la medición no sea un valor constante y repetido en cada ensayo individual. Observemos que si hubiésemos realizado nuestro experimento con una balanza cuya precisión fuese únicamente de 10^{-5} kg, no hubiésemos obtenido una variación apreciable en la magnitud medida o mensurando, lo cual lógicamente no querría decir que no existiese, sino únicamente que estaría fuera del alcance de nuestra capacidad de detección (sensibilidad)

experimental. A la vista de los resultados anteriores la pregunta surge de modo inmediato: ¿cuál es el valor que podemos asignar a la masa del cuerpo? Y en cualquier caso, ¿coincide este valor asignado a la masa con el valor *real* de la masa del cuerpo, magnitud que siempre suponemos implícitamente que existe? ¿Cuál es la discrepancia entre ambos valores? ¿Existe algún método para evaluarla? ¿Con qué nivel de certidumbre podemos concluir que conocemos el valor de la masa del cuerpo a partir del experimento anterior?

El origen de esta variabilidad intrínseca de la determinación experimental de las magnitudes físicas -si es inherente a la naturaleza de los objetos, imputable a limitaciones en nuestra capacidad de observación (precisión de los instrumentos, deficiencias del método experimental, etc.) o a ambas- es un tema que va más allá del ámbito del presente texto y, por tanto, no debe ocuparnos en este punto. Baste con reconocer que en la determinación experimental de magnitudes físicas subyace una incertidumbre de carácter estadístico. El objeto del presente capítulo es la introducción de algunas de las herramientas matemáticas desarrolladas en el ámbito de la Estadística matemática para la descripción de la información contenida en los datos y su posterior utilización para la evaluación de la incertidumbre asociada al proceso experimental. Concretamente, nos ocuparemos en el presente tema de los métodos de descripción de los conjuntos de datos obtenidos del colectivo objeto de estudio mediante algún experimento (aleatorio) y que posteriormente pretende analizarse. Como veremos, la simple descripción de las muestras proporcionará información relevante acerca del fenómeno en cuestión, que se complementa con el uso de métodos procedentes de la teoría de probabilidades, o con la inferencia de comportamientos generales sobre la base de esta información recogida en el experimento aleatorio.

1.1. Introducción a la estadística descriptiva

En el análisis de una determinada propiedad física o característica del sistema en la que estamos interesados debemos, en primer lugar y con anterioridad a la celebración del experimento (aleatorio), especificar el conjunto de todos los elementos, individuos o unidades estadísticas que manifiestan dicha cualidad, conjunto al que denominaremos *población*. Estas poblaciones pueden ser finitas o infinitas. En raras ocasiones dispondremos del conjunto de la población de objetos con la propiedad estudiada para realizar un experimento aleatorio, debiendo conformarnos con utilizar subconjuntos más o menos reducidos de estas poblaciones que denominamos *muestras*. Denominaremos tamaño de la muestra al número de elementos por los que está formada.

Ejemplo 1.1

Si nuestra magnitud de interés es la vida media del ${}_{92}^{238}\text{U}$ nos interesaría conocer todas las desintegraciones radiactivas de este isótopo a lo largo del tiempo. Sin embargo, estudiaremos solamente una cantidad limitada de este isótopo durante un intervalo temporal finito, lo que constituye nuestra muestra.

Ejemplo 1.2

En el estudio de una determinada enfermedad en una determinada especie botánica o zoológica en una región geográfica, la población está constituida por la totalidad de los individuos integrantes de la especie en la región, mientras que una muestra es un grupo de individuos seleccionados por medio de algún procedimiento.

Ejemplo 1.3

Cuando estamos interesados en el comportamiento electoral de un determinado país, la población es el conjunto de todos los ciudadanos con derecho al voto en dicho país (supongamos 32 millones de individuos como en el caso español en la actualidad), y una muestra puede ser un subconjunto de 2000 votantes seleccionados.

Ejemplo 1.4

Respecto al análisis de la distribución de renta en una sociedad, la población está constituida por el conjunto de los individuos que forman la sociedad en un instante determinado y una muestra está formada por cualquier subgrupo de estos, seleccionados con cualquier criterio.

En general, debido a la dificultad práctica de analizar el conjunto de todos los elementos de la población, es en las muestras donde se realizan los experimentos aleatorios de medición de las diferentes cantidades de interés. A partir de la información suministrada por los experimentos sobre muestras (que deben ser representativas de la población) podremos obtener propiedades de las poblaciones, mediante las técnicas que introduciremos en el capítulo 3 de esta obra.

Para la evaluación del experimento aleatorio, asignaremos a cada magnitud o característica de los elementos estudiados un determinado valor. Estos valores asignados pueden ser de tipo *cualitativo* o de tipo *cuantitativo*.

Las *variables cualitativas* asignan valores no numéricos a los sucesos aleatorios. Sin embargo, en las *variables cuantitativas* o *estadísticas* los valores asignados a las características de estudio son numéricos. A su vez estas últimas pueden ser *discretas*, cuando pueden tomar un número finito o infinito numerable de valores (número de veces que ocurre un suceso determinado, por ejemplo), o *continuas*, cuando es posible que adopten cualquier valor dentro de un intervalo de la recta real (intervalo de tiempo entre dos colisiones consecutivas de una partícula en el seno de un fluido, altura de una persona, etc.). Una vez que ya hemos elegido la muestra y sabemos qué característica cuantitativa queremos analizar, estamos en condiciones de realizar el *experimento aleatorio*, que cumplirá los siguientes requisitos:

1. Todos los posibles resultados de los ensayos del experimento son conocidos con anterioridad a la realización del mismo.
2. No se puede predecir el resultado de los ensayos concretos del experimento aleatorio.
3. Los diferentes ensayos que componen el experimento pueden repetirse en condiciones idénticas¹.

Ejemplos habituales de un experimento aleatorio son el lanzamiento de una moneda o el número resultante en la tirada de una ruleta. Es menos habitual incluir dentro de la categoría de experimento aleatorio la medición de una magnitud física, pero, como hemos venido comentando, la variabilidad intrínseca del resultado experimental de este proceso obliga a considerarlo como tal. En los casos de los ejemplos anteriormente citados, los experimentos aleatorios serían, respectivamente, la determinación del número de desintegraciones en intervalos de tiempo consecutivos mediante un contador Geiger, la realización de un determinado análisis diagnóstico de ocurrencia o no de la enfermedad en la muestra de seres vivos, la realización de una encuesta de opinión electoral o el estudio de la renta o de los hábitos de consumo de los individuos seleccionados mediante algún procedimiento demoscópico.

Consideremos un experimento aleatorio consistente en la realización de N ensayos de medición de una determinada característica a la que se le asocia una variable aleatoria² X que consideraremos, por simplicidad, que es discreta y puede tomar k valores diferentes. Tras la realización del experimento aleatorio de medición de la variable X -bien en la muestra bien, en algunos casos concretos, en la población- se habrá obtenido un conjunto o muestra de N datos

$$\{x_1, x_2, \dots, x_N\}$$

correspondientes a las N realizaciones concretas de la variable estadística X . Los valores x_i pueden corresponder a una cantidad unidimensional, en el caso en

¹Al menos en el sentido de las magnitudes de influencia, es decir, aquellas magnitudes conocidas que no son el mensurando y afectan al resultado de la medida.

²Para una adecuada definición del concepto de variable aleatoria, remitimos a lector al capítulo 2 del presente texto.

el que en la medición se analiza una única característica, o a una cantidad multidimensional en caso de que se analicen simultáneamente varias características en la medición.

Ejemplo 1.5

En el experimento de determinación de la masa del cuerpo citado anteriormente la variable estadística analizada X ="masa del cuerpo" es una variable unidimensional continua y los resultados del experimento son de la forma $x_i = m_i$ (el resultado de cada medida de la masa).

Ejemplo 1.6

En un experimento de determinación de la ecuación de estado $P = P(V, T)$ de un gas en el que se miden simultáneamente la presión y el volumen del mismo a una temperatura que suponemos fija, la variable estadística analizada $\vec{X} = (V, P)$ es bidimensional, por lo que los resultados del experimento serán de la forma $\vec{x}_i = (V_i, P_i)$.

Ejemplo 1.7

Una variable tridimensional sería, por ejemplo, la correspondiente a un experimento aleatorio en el que se trate de determinar la relación entre la edad, sexo y hábitos de consumo de un determinado producto, en el cual obviamente la variable analizada sería de la forma $\vec{X} = (E, S, P)$, siendo E la edad de cada individuo, S su sexo y P la cantidad de producto consumida en un determinado período de tiempo.

Una vez que hemos obtenido los resultados del experimento aleatorio llegamos a la fase crucial del proceso de descripción del fenómeno en cuestión, puesto que debemos proceder a organizar la información obtenida. En algunos casos podremos considerar que el experimento aleatorio consiste en un simple suministro de observaciones directas (empíricas) acerca del fenómeno. Sin embargo en otros casos (e.g. medida física) se trata de un proceso complejo que involucra al mensurando, al instrumento de medida, al observador y al entorno.

A partir de la muestra de N datos, lo único que podemos afirmar es cuántas veces hemos obtenido cada uno de los valores posibles de la variable aleatoria analizada. Este número nos conduce al concepto estadístico fundamental de *distribución de frecuencia*. Denominaremos *frecuencia absoluta* (n_i) del valor i de la variable aleatoria al número de veces que se observa dicho valor en el total de las observaciones realizadas en el experimento, magnitud que verifica trivialmente

$$\sum_{i=1}^k n_i = N \quad (1.1)$$

y denominamos *distribución de frecuencias absolutas* al conjunto de los valores de la frecuencia absoluta de los distintos valores de la variable estadística, $\{n_i\}_{i=1}^k$.

De la misma manera, podemos definir la *frecuencia relativa* del valor i -ésimo de la variable estadística (f_i) a la fracción de veces que se ha obtenido dicho valor respecto al total de las observaciones realizadas, $f_i = n_i/N$, que obviamente verifica

$$\sum_{i=1}^k f_i = 1 \quad (1.2)$$

De manera análoga a lo que sucedía con la distribución de frecuencias absolutas, la distribución de frecuencias relativas corresponde al conjunto de las frecuencias relativas de todos los valores de la variable estadística obtenidos en el experimento, $\{f_i\}_{i=1}^k$.

Las frecuencias anteriores, referidas a cada valor de la variable, se denominan distribuciones de frecuencia *ordinarias*. En el caso de que consideremos las frecuencias absoluta o relativa con la que aparecen los valores inferiores a uno dado hablaremos de *distribuciones acumuladas de frecuencia*. Así, la frecuencia absoluta acumulada hasta el valor i -ésimo de la variable estadística es

$$N_i = \sum_{j=1}^i n_j \quad (1.3)$$

y la correspondiente frecuencia relativa acumulada es

$$F_i = \sum_{j=1}^i f_j = \frac{N_i}{N} \quad (1.4)$$

Entre las propiedades de las frecuencias definidas anteriormente podemos destacar las siguientes:

i)

$$0 \leq n_i \leq N; 0 \leq f_i \leq 1$$

ii)

$$\sum_{i=1}^k n_i = N \Leftrightarrow \sum_{i=1}^k f_i = 1$$

iii)

$$N_k = \sum_{i=1}^k n_i = N \Leftrightarrow F_k = 1$$

iv)

$$N_i = N_{i-1} + n_i; F_i = F_{i-1} + f_i$$

1.1.1. Agrupamiento en clases

Para una variable aleatoria discreta con un número pequeño de valores, las distribuciones de frecuencia pueden obtenerse en función de todos los valores discretos de la variable. Sin embargo, en los casos en los que el número de posibles valores observables sean comparables o superiores al número de observaciones, las distribuciones de frecuencia pierden toda su utilidad, puesto que en ese caso $n_i \rightarrow 0$ para prácticamente todos los valores posibles. Esto sucede en el caso de variables discretas con un elevado número de posibles valores y en el caso de variables continuas.

En estos supuestos, se hace necesario proceder a un agrupamiento de los datos obtenidos en clases. Mediante este procedimiento se divide el recorrido de la variable estadística analizada (la diferencia entre el mayor y el menor valor observados) en una serie de subintervalos disjuntos tal que la unión de todos ellos cubra la totalidad del recorrido de la variable, asociando a todos sucesos dentro de cada *clase* un único valor concreto o *marca de clase* idéntico.

La frecuencia absoluta de observación de cada marca de clase será el número de datos originales que pertenezcan a su clase asociada. Es evidente que esta operación conlleva inevitablemente una pérdida de información.

Aunque en principio la operación de agrupamiento en clases de los valores de la variable aleatoria no está sometida necesariamente a ninguna norma particular, sí que existen reglas para optimizarla. Supongamos que hemos obtenido una muestra de datos de una variable estadística continua cuyo recorrido está contenido en $[a_0, a_k]$. Las operaciones para proceder al agrupamiento de los datos obtenidos son:

- i) Redondear los datos a dos o a lo sumo tres cifras significativas³.
- ii) El número de clases no debe ser demasiado grande (para no obtener frecuencias de clase muy bajas) ni demasiado pequeño. Se suele recomendar un número de clases, r , en torno al entero más próximo a \sqrt{N} (siendo N el número de sucesos).
- iii) Tomar todas las clases de la misma amplitud -aunque es posible que en un problema determinado lo aconsejable sea tomar amplitudes variables- y sin ambigüedad en la asignación de los datos a las diferentes clases. Esto último se logra mediante la utilización de intervalos semiabiertos para que no haya duda alguna en la asignación de los valores límite de los intervalos. Es aceptable extender el recorrido de la variable más allá del rango de los

³Véase la definición de cifra significativa en el capítulo 4.

valores observados para obtener clases de una longitud adecuada según el número de clases. Así, el recorrido elegido para la variable estadística elegida aparecerá dividido en un conjunto de subintervalos (clases) de la forma:

$$\{[a_0, a_1), [a_1, a_2), [a_2, a_3), \dots, [a_{k-1}, a_k]\}$$

a los que se asocian unas determinadas marcas de clase $\{x_1, \dots, x_k\}$ que constituyen los valores posibles de una nueva variable aleatoria discreta construida mediante el proceso de agrupamiento en clases. Suelen tomarse como marcas de las clases los centros de los diferentes subintervalos. Así, la marca asociada a los datos pertenecientes a la clase i -ésima será:

$$x_i = \frac{a_i + a_{i-1}}{2}; \quad i = 1, \dots, k$$

- iv) Finalmente, se procede al recuento del número de datos crudos originales que pertenecen a las diferentes clases, siendo $\{n_1, \dots, n_k\}$ los valores de las frecuencias absolutas asociadas a las marcas de clases correspondientes, y que denominaremos frecuencias de clase. El procedimiento supone establecer la equivalencia entre el experimento aleatorio original y un experimento realizado con una variable discreta $\{x_1, \dots, x_k\}$ en el que se hubiese obtenido esta última distribución de frecuencias. Nótese que únicamente estamos interesados en el número de datos que caen dentro de una clase determinada y no en su distribución concreta dentro de los intervalos de clase, estableciendo la hipótesis de que están distribuidos homogéneamente en cada intervalo.

Ilustremos los conceptos anteriores acerca de las distribuciones de frecuencia con algunos ejemplos.

Ejemplo 1.8

Datos de desintegración radiactiva. En un experimento de determinación de la actividad de un radionúclido se han realizado, mediante un contador Geiger, 300 mediciones consecutivas de 30 segundos cada una en las que se contaron las partículas emitidas por la fuente de la que se dispone, y los resultados fueron los que se muestran en la tabla siguiente. Evidentemente, las lecturas de esta tabla proporcionan amplia información sobre el comportamiento de la muestra radiactiva, pero para extraerla debemos proceder previamente a su organización.

7	5	3	6	8	4	5	7	6	4	6	3	4	4	5	7	5	4	3	4
9	8	9	7	0	3	5	8	7	8	4	6	5	5	7	4	7	3	5	2
3	5	3	8	4	9	4	10	3	5	5	8	6	7	6	5	6	6	2	9
8	9	5	9	6	5	5	7	3	7	6	4	6	9	4	5	7	6	5	8
4	8	4	5	7	8	7	6	5	4	5	7	8	3	9	6	1	6	1	5
7	5	3	9	8	1	6	4	7	8	5	6	11	9	7	4	5	10	7	4
6	4	6	10	7	6	2	13	3	6	0	8	1	6	8	1	11	6	8	3
5	6	9	4	10	7	6	7	9	6	3	7	5	12	7	8	6	3	5	6
2	7	5	6	7	5	5	2	4	6	9	2	5	10	2	9	5	5	7	4
2	6	7	8	4	5	7	6	6	7	5	4	3	2	6	8	7	1	6	5
10	8	3	2	8	4	6	3	3	8	4	5	6	7	8	6	9	8	3	2
11	2	6	5	5	7	9	8	5	2	4	6	6	3	5	4	6	4	4	5
7	5	6	7	4	10	6	7	4	5	8	7	5	5	4	6	3	8	6	6
12	10	5	6	12	3	11	4	10	4	5	4	9	8	3	6	8	7	5	2
3	5	10	7	9	6	7	4	11	7	6	1	11	2	5	9	4	8	5	6

Tabla 1.1: Resultados del experimento de contaje cada 30 s mediante un contador Geiger para una muestra radiactiva.

La variable estadística que estamos analizando es $X =$ "número de cuentas (desintegraciones) procedentes de la muestra en un intervalo de 30 segundos". Observamos además que las lecturas del contador Geiger son números enteros por lo que estamos ante una variable estadística discreta. Además únicamente presentan 14 valores posibles, de 0 a 13. Una buena forma -i.e. no sesgada- de organizar la información es proceder al contaje del número de veces que obtenemos cada uno de los valores posibles de la variable, esto es, a la determinación de las distribuciones de frecuencia correspondientes. En nuestro caso tenemos unas distribuciones de frecuencias absoluta y relativa, ordinarias y acumuladas, dadas por la siguiente tabla:

x_i	n_i	f_i	N_i	F_i
0	2	0,0067	2	0,0067
1	7	0,0233	9	0,0300
2	15	0,0500	24	0,0800
3	25	0,0833	49	0,1633
4	38	0,1267	87	0,2900
5	52	0,1733	139	0,4633
6	52	0,1733	191	0,6367
7	40	0,1333	231	0,7700
8	30	0,1000	261	0,8700
9	19	0,0633	280	0,9333
10	10	0,0333	290	0,9667
11	6	0,0200	296	0,9867
12	3	0,0100	299	0,9967
13	1	0,0033	300	1,0000

$$\sum_i n_i = 300$$

$$\sum_i f_i = 1$$

Tabla 1.2 Tabla de frecuencias correspondiente a la muestra de datos de la Tabla 1.1

Ejemplo 1.9

Medidas de la densidad del agua a 4 °C. Con el fin de determinar la pureza de una muestra de agua, se han realizado con un densímetro de tubo vibrante 50 medidas de la densidad del agua a 277,15 K -temperatura que podemos considerar constante por ejemplo por estar fijada mediante una cabeza termostática de alta precisión- y los valores obtenidos se muestran en la tabla siguiente.

1,000001	1,000002	1,000000	0,999997	0,999999	1,000000	1,000004	0,999999	1,000001	0,999997
0,999998	1,000000	1,000002	0,999997	1,000000	1,000001	1,000001	1,000004	0,999998	1,000003
1,000000	1,000003	0,999998	1,000000	0,999999	1,000000	1,000003	0,999998	1,000000	1,000001
1,000001	1,000000	1,000003	1,000000	1,000002	1,000000	1,000001	0,999998	1,000001	1,000000
1,000001	0,999999	0,999999	0,999997	1,000002	1,000004	1,000001	1,000002	1,000000	0,999999

Tabla 1.3: Medidas de la densidad de una muestra de agua a 4 °C (en g cm^{-3}).

Aunque en este caso estamos ante una variable estadística continua, el hecho de que únicamente se repitan lecturas de siete valores diferentes nos permite una descripción discreta de los resultados del experimento. La tabla de frecuencias correspondientes a los resultados del experimento anterior es:

ρ_i (g cm^{-3})	n_i	f_i	N_i	F_i
0,999997	4	0.08	4	0.08
0,999998	5	0.10	9	0.18
0,999999	6	0.12	15	0.30
1,000000	13	0.26	28	0.56
1,000001	10	0.20	38	0.76
1,000002	5	0.10	43	0.86
1,000003	4	0.08	47	0.94
1,000004	3	0.06	50	1.00

Tabla 1.4: Distribuciones de frecuencia asociadas a las medidas de la densidad de una muestra de agua a 4°C (en g cm^{-3}).

Ejemplo 1.10

Los valores de la carga eléctrica medida en un experimento de la gota de Millikan fueron los que se muestran en la tabla siguiente:

3.8725	4.3006	4.7188	5.8520	6.0101	6.3232
6.5963	6.9390	7.0302	7.2755	7.3768	8.2171
8.9483	9.2250	9.7594	9.9249	10.126	10.266
11.217	12.105	13.502	16.191	16.648	16.970

Tabla 1.5 Valores de la carga eléctrica obtenida en un experimento de la gota de Millikan (en unidades de 10^{-19} C).

Como vemos, estamos ante una variable estadística de tipo continuo, por lo que antes de proceder a la obtención de las distribuciones de frecuencia hemos de aplicar el procedimiento de agrupamiento en clases. En primer lugar, y dado que tenemos un total de 24 datos, el número óptimo de clases es el entero más próximo a $\sqrt{24}$ que es 5. Puesto que el recorrido de los datos es de 3×10^{-19} C a 17×10^{-19} C, consideraremos clases de anchura 3×10^{-19} C. Los datos redondeados a tres cifras significativas serán (en unidades de 10^{-19} C).

3,87	4,30	4,72	5,85	6,01	6,32
6,60	6,94	7,03	7,28	7,38	8,22
8,95	9,23	9,76	9,92	10,13	10,27
11,22	12,11	13,50	16,19	16,66	16,97

Tabla 1.6 Valores de la carga eléctrica contenidos en la Tabla 1.5 redondeados a dos dígitos decimales (en unidades de 10^{-19} C).

Las clases, marcas de clase y frecuencias correspondientes son las que se muestran a continuación:

Clase (10^{-19} C)	marca de clase (10^{-19} C)	n_i	f_i	N_i	F_i
[3,6)	4,5	4	0,17	4	0,17
[6,9)	7,5	9	0,37	13	0,54
[9,12)	10,5	6	0,25	19	0,79
[12,15)	13,5	2	0,08	21	0,87
[15,18]	16,5	3	0,13	24	1,00

Tabla 1.7 Tabla de intervalos de clase y frecuencias de los valores de la carga eléctrica contenidos en la Tabla 1.5 redondeados a dos dígitos decimales

Como veremos posteriormente, la información contenida en los datos de los ejemplos anteriores puede ser analizada a través de las tablas de frecuencias absoluta y relativa, obteniendo las denominadas medidas características de la distribución. Sin embargo, la representación gráfica de esta información presenta también indudables ventajas a la hora de analizar la información contenida en los datos experimentales.

1.2. Representaciones gráficas

Disponer de una representación gráfica permite, en la mayoría de las situaciones prácticas, una mejor y más intuitiva interpretación de la información que del fenómeno nos proporcionan las distribuciones de frecuencia.

1.2.1. Diagrama de barras

Esta es la representación adecuada para datos asociados a variables discretas. A cada uno de los valores posibles de la variable se le asocia una barra de altura igual a la frecuencia (absoluta o relativa) de dicho dato en el experimento aleatorio realizado sobre la muestra. Por esta vía podemos obtener los diagramas de frecuencias ordinarias y acumuladas. Como ejemplo, en la Fig. 1.1 se muestra la representación en diagramas de barras de las distribuciones de frecuencia de la Tabla 1.2 correspondientes a los datos de la desintegración radiactiva del ejemplo 1.8.

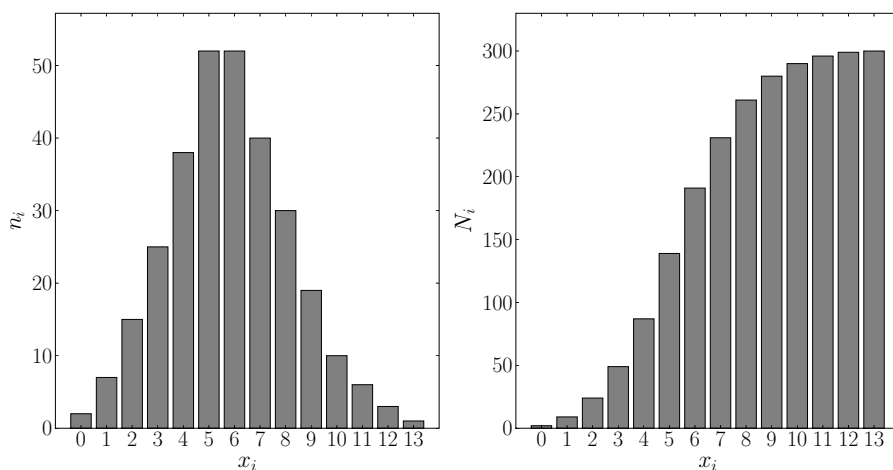


Figura 1.1: Diagramas de barras de las frecuencias absolutas ordinaria y acumulada de la tabla 1.2. Las barras de frecuencias relativas serían, obviamente, proporcionales.

De la mera observación directa de los diagramas de barras anteriores, ya podemos deducir una importante cantidad de información acerca de la conducta

de la variable estadística "cuentas por unidad de tiempo" en nuestra muestra: cuál es el número de cuentas que se observa con mayor frecuencia, en qué valor se encuentra centrada la distribución, si existe poca o mucha desviación en las observaciones respecto a este valor central... Estas propiedades merecen ser cuantificadas debidamente mediante la asignación de parámetros concretos, lo que haremos al introducir las denominadas medidas características de la distribución de frecuencia.

1.2.2. Histogramas

Esta representación es particularmente adecuada para variables estadísticas agrupadas, y es la que se utiliza con mayor frecuencia. Se obtiene asociando a cada clase de la variable un rectángulo cuya base sea igual a la amplitud del intervalo y su altura proporcional a la frecuencia correspondiente a cada clase. Como ejemplo, la Fig. 1.2 muestra el histograma de las frecuencias ordinarias y acumuladas correspondientes al experimento de la gota de Millikan presentado en la sección anterior (1.1).

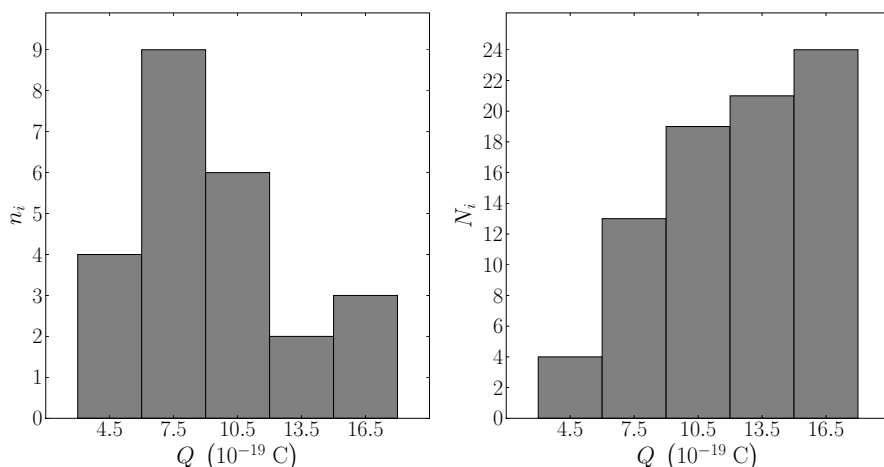


Figura 1.2: Histogramas de frecuencias absolutas ordinaria y acumulada de la tabla 1.7.

Los histogramas constituyen con mucho el método de representación de datos estadísticos más frecuente y, al igual que sucedía con el caso de los diagramas de barras, la mera inspección de un histograma proporciona una información bastante exhaustiva acerca del comportamiento de la variable estadística en cuestión.

1.2.3. Diagrama de Pareto

Se trata de una representación estadística frecuentemente utilizada para el estudio gráfico de variables cuantitativas. Este diagrama lo construiremos mediante:

- i) la ordenación de las categorías o clases en frecuencia de aparición descendente.
- ii) A cada propiedad se le asocia un rectángulo cuya altura es igual a su frecuencia relativa.

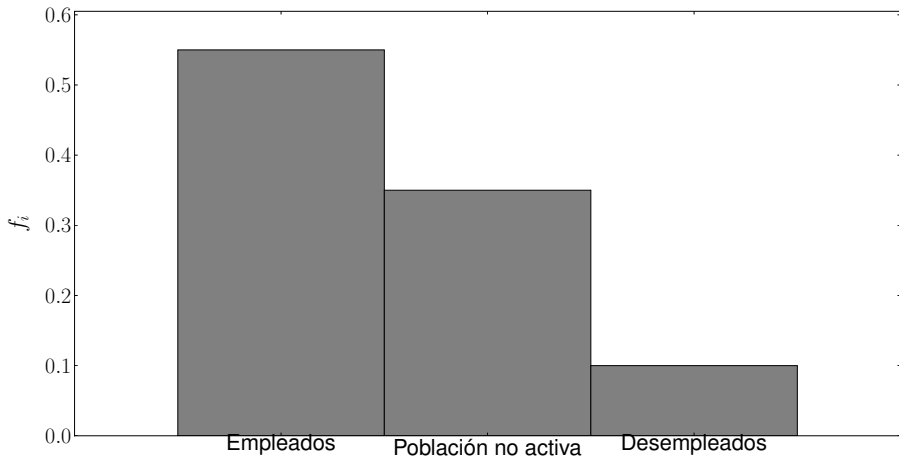


Figura 1.3: Diagrama de Pareto de la estructura de población de un país según su situación activa. Es habitual acompañarlo de un polígono de frecuencias relativas acumuladas.

Esta forma de representar los datos recibe su nombre en honor del sociólogo y economista italiano Wilfredo Pareto. Esta representación suele poner de manifiesto la llamada "ley de Pareto" descubierta por éste en estudios de distribución de renta en las sociedades de finales del siglo XIX, según la cual la mayoría de los individuos poseían rentas bajas y unos pocos disfrutaban de rentas elevadas. Este fenómeno es un caso particular de una propiedad muy general que presentan determinados sistemas en los cuales pueden ocurrir sucesos aleatorios de cualquier tamaño (e.g. rentas extraordinariamente altas) con una frecuencia apreciable. La ley de Pareto se ha extendido a las ciencias naturales, a las ciencias sociales, a la lingüística, etc., ámbitos en los que recibe nombres diferentes, y para la adecuada comprensión de su origen nos referimos al tema de teoría de probabilidades.

1.2.4. Polígonos o curvas de frecuencias

Cuando analizamos variables estadísticas continuas es frecuente representar las distribuciones de frecuencia ordinarias o acumuladas en forma de curvas que se generan de dos modos diferentes:

- i) cuando se trata de una variable no agrupada se trazan líneas rectas que unan los extremos superiores de las barras del diagrama de frecuencias.

- ii) Si tenemos los datos agrupados y representados en un histograma, se unen los centros de las bases superiores de los rectángulos del histograma.

Como ejemplo, se incluye la representación de la curva de frecuencias para el experimento de la gota de Millikan en la Fig. 1.4. Es evidente que cuanto mayor sea el número de clases, tanto más cerca estará el polígono de frecuencias de convertirse en una curva continua.

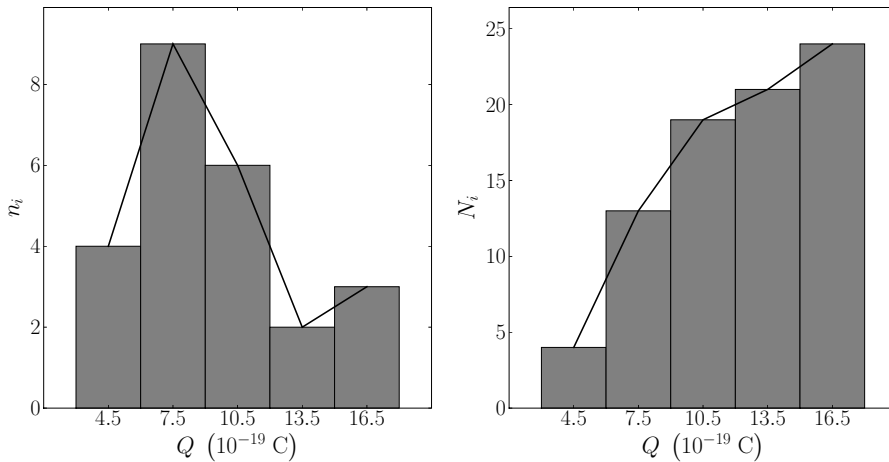


Figura 1.4: Histograma y polígonos de frecuencias de las distribuciones de la Tabla 1.7 correspondientes al experimento de la gota de Millikan.

1.2.5. Series temporales

Aunque ya no se trate de una forma de representación de distribuciones de frecuencia, antes de terminar esta sección debemos citar una forma de representación gráfica que, por su frecuente utilización en métodos estadísticos, merece ser objeto de estudio independiente. Si las muestras o series de datos de la variable estadística se toman a lo largo de un cierto intervalo de tiempo por ser de interés la evolución temporal de aquella, estamos ante lo que se denomina una serie temporal, en la cual la variable independiente la constituye el tiempo y en el eje de ordenadas se representan los sucesivos valores de la variable estadística (ver Figs. 1.5-1.7)

Señalemos finalmente que, además de los anteriormente citados, existe un buen número de formas de representación alternativas, aunque corresponden mayoritariamente a formas de tratamiento de variables cualitativas. Pese a que, debido al objeto del presente texto, no tienen un interés directo para nuestros propósitos, citaremos aquí los diagramas de rectángulos, diagramas de sectores, perfiles ortogonales y radiales, pictogramas, cartogramas, etc., para cuyo estudio remitimos al lector a cualquier texto de introducción a la Estadística.

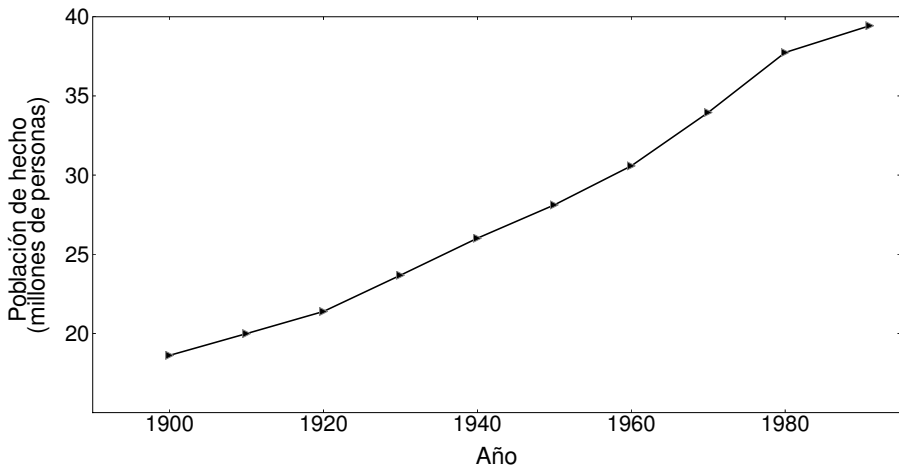


Figura 1.5: Serie temporal de la evolución de la población española a lo largo del siglo XX (Fuente: Instituto Nacional de Estadística – series históricas de población).

1.3. Medidas características de una distribución de frecuencias

Las medidas características de una variable aleatoria son una serie de parámetros que resumen las propiedades generales de la variable aleatoria. Hay que mencionar que las medidas que introduciremos a continuación proporcionan información relevante en el caso de muestras homogéneas de datos, mientras que pueden ser engañosas en el caso de muestras heterogéneas procedentes de distintas poblaciones.

En general, se suelen definir cuatro tipos de medidas características fundamentales de la muestra de datos:

1. *Medidas de posición central:*

Nos indican cuál es la tendencia central de los datos, en torno a qué valor concreto de la variable estadística aparecen distribuidos los datos de la muestra.

2. *Medidas de dispersión:*

Dan cuenta de la variabilidad inherente a los datos muestrales, i.e. en qué medida se desvían de su valor central.

3. *Medidas de asimetría:*

Proporcionan información acerca de la forma concreta de la distribución de datos, si predominan los datos menores o mayores que los valores centrales o si la distribución de datos es simétrica respecto al valor medio.

4. *Medidas de concentración o apuntamiento:*

Nos informan acerca del grado de concentración de los datos, esto es,

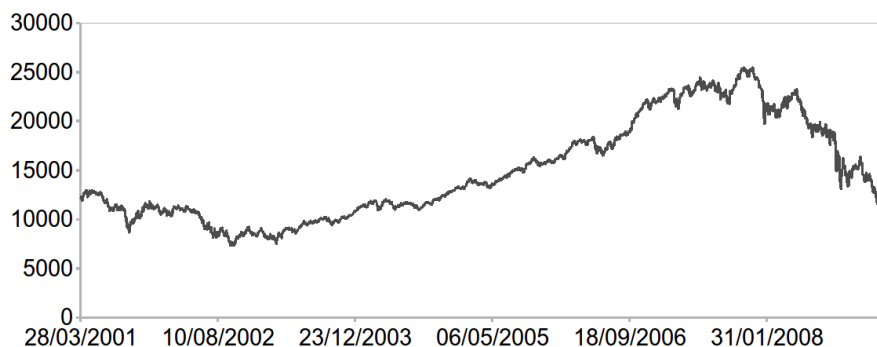


Figura 1.6: Serie temporal de la evolución del IBEX35 a lo largo de nueve años.

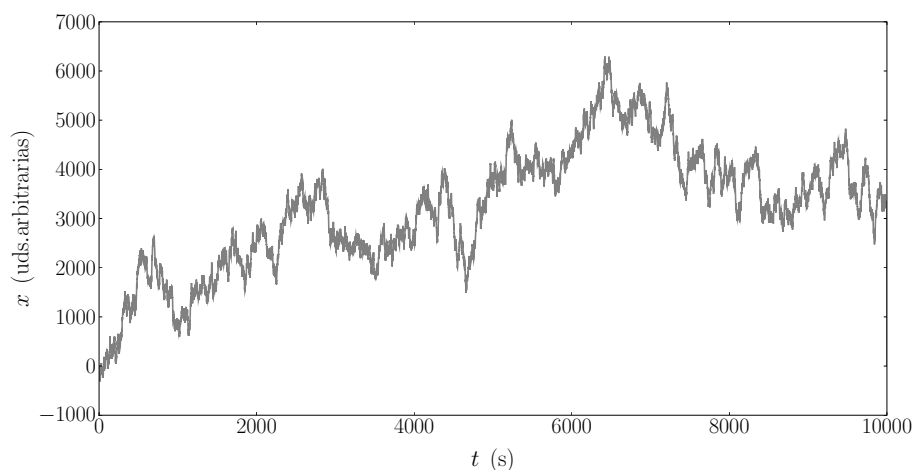


Figura 1.7: Serie temporal correspondiente a un movimiento browniano en una dimensión.

de cómo se reparte la frecuencia relativa entre los valores próximos a la media y los valores más extremos o alejados de esta.

Estudiemos detalladamente cada uno de estos tipos de medidas en las secciones siguientes. Como hemos venido haciendo con anterioridad, en todo lo que sigue consideraremos que la muestra de datos es $\{x_1, \dots, x_N\}$, correspondiente a una variable estadística X con k posibles valores,⁴ y que las frecuencias absoluta y relativa asociadas al dato i -ésimo son respectivamente n_i y f_i . En general, formularemos las expresiones de los coeficientes en términos de las frecuencias relativas⁵, siendo inmediata la obtención de los correspondientes

⁴Estos pueden corresponder a los diferentes valores de una variable estadística discreta o a las marcas de clase de la variable discreta resultante de un proceso de agrupamiento en clases.

⁵Adoptamos aquí esta opción por ser la más adecuada para la posterior generalización a

resultados para las frecuencias absolutas.

1.3.1. Medidas de centralización

Existen diferentes medidas que proporcionan información acerca de la tendencia de los datos muestrales a distribuirse en torno a un valor o valores determinados. Entre estas destacan:

1. *Moda (Md):*

La moda de la muestra de datos corresponde al valor más frecuente de la distribución:

$$Md = \{x_i \mid f_i = \max\{f_j\}_{j=1}^N\} \quad (1.5)$$

Una muestra de datos puede tener una única moda (distribución unimodal) o varias (multimodal), lo que limita la utilidad de este parámetro como expresión del punto en torno al que se distribuyen los datos muestrales. Además, la obtención de la moda implica el manejo de un único dato de la distribución o, en el mejor de los casos, de un número pequeño de ellos en distribuciones multimodales, siendo irrelevante el valor que tomen los demás datos.

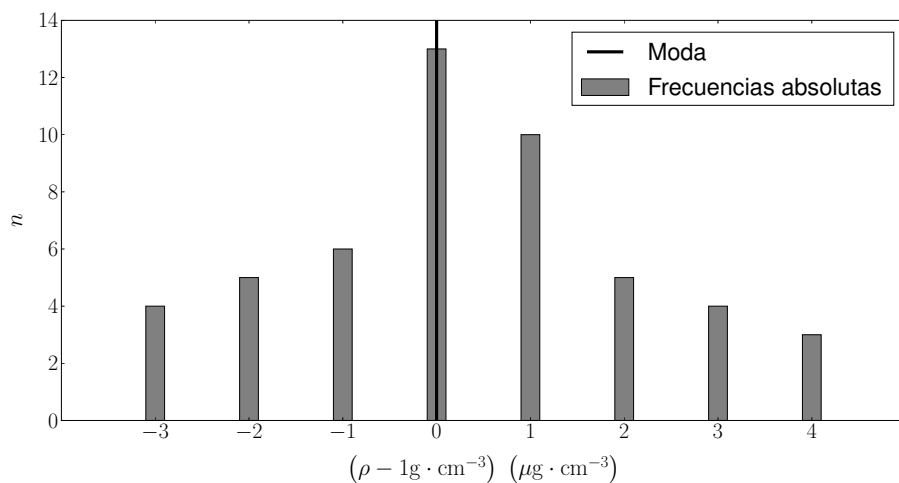


Figura 1.8: Moda de la muestra de datos de densidad de una muestra de agua a 4 °C.

2. *Percentiles:*

Se define el percentil q -ésimo de la muestra de datos siendo $0 \leq q \leq 1$ como el valor de la variable aleatoria P_q tal que la frecuencia de todos los valores menores o iguales que P_q es igual a q :

$$f(X \leq P_q) = q \quad (1.6)$$

distribuciones de probabilidad.

A algunos percentiles de uso frecuente se les da un nombre propio. Es el caso de la mediana $Me = P_{1/2}$, que se menciona expresamente a continuación y de los cuartiles primero ($P_{1/4}$) y tercero ($P_{3/4}$), así como de los deciles, que se corresponden con $q = \frac{1}{10}, \frac{2}{10}, \dots, \frac{9}{10}$. Los percentiles nos permiten describir cómo se distribuyen los datos muestrales a lo largo de todo su recorrido.

El cálculo de los percentiles se llevará a cabo de distinta manera según estemos tratando con una variable aleatoria discreta o continua:

Variable discreta:

Comenzaremos por ordenar los posibles valores de los datos en orden creciente según su magnitud ($x_1 < x_2 < \dots < x_N$) y construir la tabla de frecuencias absolutas acumuladas. A continuación calcularemos la fracción q del número de observaciones. Si es un entero y coincide con alguna de las frecuencias absolutas acumuladas, tomaremos como valor del percentil buscado $P_q = (x_{i+1} + x_i) / 2$, siendo x_i el valor de la variable correspondiente a tal frecuencia. En caso contrario estará comprendido entre las dos frecuencias asociadas a los valores x_i y x_{i+1} de la variable, y se tomará como valor del percentil el mayor de los dos, $P_q = x_{i+1}$.

Variable aleatoria continua:

Los datos se hallarán agrupados en clases, cuyas marcas de clase en orden creciente serán $x_1 < x_2 < \dots < x_N$. Para proceder se construye, como en el caso discreto, la tabla de frecuencias absolutas acumuladas y se calcula $q \cdot N$. Si éste valor pertenece a la tabla, se tomará como percentil q -ésimo el extremo superior de la clase correspondiente. De lo contrario se procederá por interpolación lineal entre las dos marcas de clase, x_i y x_{i+1} , cuyas frecuencias absolutas acumuladas dejen en medio el valor buscado:

$$\frac{N_{i+1} - N_i}{x_{i+1} - x_i} = \frac{q \cdot N - N_i}{P_q - x_i}. \quad (1.7)$$

3. Mediana (Me):

Es aquel valor de la variable para el cual los datos de valor inferior son tan frecuentes como los de valor superior:

$$f(X \leq Me) = f(X > Me) = \frac{1}{2} \quad (1.8)$$

Es evidente que la mediana, como ha quedado dicho, no es más que un caso particular de percentil, con $q = \frac{1}{2}$. En consecuencia, para su cálculo se empleará el método visto anteriormente.

4. Media:

La media de una muestra de datos es, con mucho, la medida de centralización más importante que podemos determinar acerca de la muestra. Introduciremos a continuación diferentes medias que pueden definirse para la muestra de datos, aclarando ya desde el primer momento que la

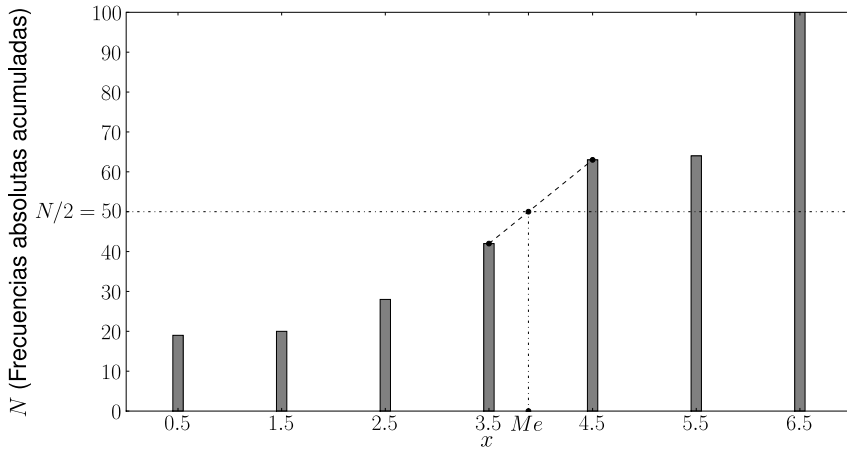


Figura 1.9: Cálculo de la mediana de una variable agrupada.

más importante de todas ellas es la media aritmética, si bien las restantes pueden presentar cierto interés práctico en aplicaciones particulares.

a) Media aritmética (\bar{x}):

Se define la media aritmética de la muestra de datos como

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (1.9)$$

donde el sumatorio se extiende a todos los datos de la muestra. En el caso de datos discretos agrupados como en el caso de la Tabla 1.2, con un número k de valores diferentes, tendremos⁶

$$\bar{x} = \sum_{i=1}^k f_i x_i. \quad (1.10)$$

En el caso de que nuestra muestra corresponda a una variable aleatoria continua agrupada en clases, la ecuación anterior sigue siendo válida, pero las x_i representan entonces las marcas de clase correspondientes y las f_i las frecuencias relativas de los datos de la clase i -ésima.

⁶Notemos que la expresión de la media sería completamente idéntica en el caso de que usásemos las frecuencias absolutas en lugar de las relativas para su cálculo:

$$\bar{x} = \sum_{i=1}^k \frac{n_i}{N} x_i$$

y que en ambos casos la expresión es idéntica a la suma de la ecuación (1.9) para el conjunto de los datos de una muestra de datos de una variable aleatoria discreta.

La definición se puede extender de manera directa a la media aritmética de una función $g(x)$ definida sobre para los datos, ya que ésta es a su vez una variable aleatoria:

$$\overline{g(x)} = \sum_{i=1}^N f_i g(x_i) \quad (1.11)$$

Por ejemplo, la media aritmética de la variable aleatoria al cuadrado es

$$\overline{x^2} = \sum_{i=1}^N f_i x_i^2.$$

Es importante advertir que $\overline{g(x)} \neq g(\bar{x})$ en general.

b) Media geométrica (\bar{x}_g):

Se define la media geométrica de la muestra de datos como

$$\bar{x}_g = \left(\prod_{i=1}^N x_i \right)^{1/N} \quad (1.12)$$

o en el caso de datos agrupados en k clases⁷:

$$\bar{x}_g = \left(\prod_{i=1}^k x_i^{n_i} \right)^{1/N}, \quad (1.13)$$

La forma más cómoda de calcular en la práctica la media geométrica de una muestra de datos es tomar logaritmos en ambos miembros de la ecuación anterior, en cuyo caso tendremos:

$$\begin{aligned} \ln \bar{x}_g &= \frac{1}{N} \sum_{i=1}^k n_i \ln x_i \\ &= \sum_{i=1}^k f_i \ln x_i \end{aligned} \quad (1.14)$$

donde nuevamente el sumatorio se extiende a todos los valores diferentes de la variable o a todas las clases en las que previamente hayamos agrupado los datos muestrales. Podría decirse que el logaritmo de la media geométrica es la media aritmética de una muestra cuyos datos fuesen iguales a los logaritmos de los datos de la muestra original y tuviesen idéntica distribución de frecuencias. Si uno de los valores posibles de la variable aleatoria es cero, la media geométrica de los datos también lo es, con lo cual en este caso resulta inútil esta medida de centralización.

⁷A partir de este punto y siempre que no se especifique lo contrario supondremos que los datos de la muestra han sido agrupados en k valores posibles o k clases.

c) Media cuadrática (\bar{x}_q):

La media cuadrática de la muestra de datos se define como:

$$\bar{x}_q = \sqrt{\sum_{i=1}^k f_i x_i^2}. \quad (1.15)$$

d) Media armónica (\bar{x}_a):

La media armónica de una muestra de datos en el que ninguno de los posibles valores (o marcas de clase) sea igual a cero es:

$$\frac{1}{\bar{x}_a} = \sum_{i=1}^k \frac{f_i}{x_i}. \quad (1.16)$$

En general, se verificará que $\bar{x}_a \leq \bar{x}_g \leq \bar{x} \leq \bar{x}_q$.

Ejemplo 1.11

Consideremos la muestra de datos procedentes del experimento de desintegración radiactiva cuya tabla de frecuencias se muestra en la Tabla 1.1. En este caso las medidas de centralización serán:

$$\begin{aligned} Md_1 &= 5 \text{ cuentas}; Md_2 = 6 \text{ cuentas} \text{ (distribución bimodal)} \\ \frac{N}{2} &= 150; Me = 6 \text{ cuentas} \\ \frac{N}{4} &= 75; P_{1/4} = 4 \text{ cuentas} \\ \frac{3N}{4} &= 225; P_{3/4} = 7 \text{ cuentas} \\ \bar{x} &= \sum_{i=0}^{13} f_i x_i = 5,81 \text{ cuentas}; \bar{x}_q = 6,28 \text{ cuentas} \end{aligned}$$

Obviamente, en este caso, al ser cero uno de los posibles valores de la variable aleatoria, la media geométrica es nula y la media armónica no se encuentra definida.

Ejemplo 1.12

Es un buen ejercicio que el lector compruebe que las medidas características de centralización de la muestra de datos obtenida en el experimento de la gota

de Millikan son:

$$\begin{aligned}
 Md &= 7,5 \cdot 10^{-19} \text{ C (distribución unimodal)} \\
 \frac{N}{2} &= 12,5; \quad Me = 7,5 \cdot 10^{-19} \text{ C} \\
 \frac{N}{4} &= 6,25; \quad P_{1/4} = 7,5 \cdot 10^{-19} \text{ C} \\
 \frac{3N}{4} &= 18,75; \quad P_{3/4} = 10,5 \cdot 10^{-19} \text{ C} \\
 \bar{x} &= 9,375 \cdot 10^{-19} \text{ C}; \quad \bar{x}_g = 8,684 \cdot 10^{-19} \text{ C} \\
 \bar{x}_q &= 10,062 \cdot 10^{-19} \text{ C}; \quad \bar{x}_a = 8,026 \cdot 10^{-19} \text{ C}
 \end{aligned}$$

Aunque todas las medidas de centralización introducidas anteriormente nos dan una idea de en torno a qué valor o valores se distribuyen los datos de las muestras, son sin embargo diferentes en lo que a contenido y robustez frente a posibles cambios en la muestra. Así, mientras la existencia de un sólo dato extremo puede no modificar sensiblemente la posición de la mediana, normalmente variará considerablemente el valor de la media aritmética. Por ello, las medias anteriormente definidas son más sensibles a variaciones en la muestra o la existencia de datos atípicos⁸ (baja robustez), mientras que la mediana (o los percentiles) son mucho más robustos frente a variaciones en la muestra. En general, es conveniente calcular todos los valores y compararlos, puesto que las divergencias entre moda, mediana y media suelen indicar asimetrías en la distribución o heterogeneidades en la muestra de datos (mezcla de poblaciones, diferentes procedimientos de obtención...).

1.3.2. Medidas de dispersión

Las medidas de dispersión reflejan las fluctuaciones de los valores de la variable aleatoria. Mientras que las medidas de centralización nos indican los valores en torno a los que se distribuye la variable, estas nuevas medidas de dispersión nos dan una idea de la anchura de la distribución de la variable aleatoria en torno a los valores centrales.

Ejemplo 1.13

Consideremos las dos muestras de datos correspondientes a secuencias de 20 tiradas de dos dados diferentes aunque en principio de comportamiento esperablemente idéntico:

⁸Frecuentemente, se considera dato atípico leve aquel que está a más de 1.5 veces el rango intercuartílico ($P_{3/4} - P_{1/4}$) de los cuartiles $P_{1/4}$ y $P_{3/4}$. Un dato atípico extremo es el que está a más de tres veces el rango intercuartílico de estos valores.

Dado 1	1	2	4	4	3	4	5	4	6	4	4	5	4	4	3	4	4	4	1	6
Dado 2	2	2	4	3	5	4	6	6	1	4	6	3	4	2	3	2	4	5	5	6

Tabla 1.7 Muestras de datos correspondientes a las tiradas de dos dados.

Las tablas de frecuencias correspondientes a ambas muestras se muestran a continuación junto con los diagramas de barras correspondientes.

Dado 1			Dado 2	
x_i	n_i	f_i	n_i	f_i
1	2	0.10	1	0.05
2	1	0.05	4	0.20
3	2	0.10	3	0.15
4	11	0.55	5	0.25
5	2	0.10	3	0.15
6	2	0.10	4	0.20

Tabla 1.8 Tabla de frecuencias correspondiente a los datos de la Tabla 1.7.

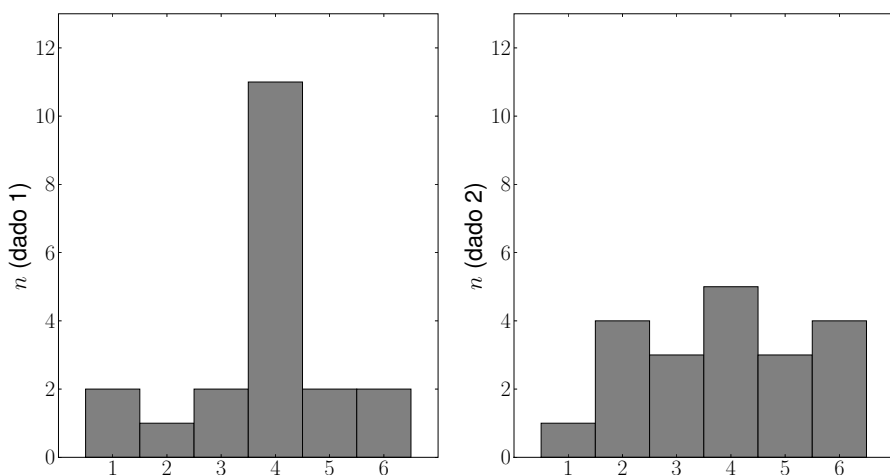


Figura 1.10: Diagramas de barras de las muestras correspondientes a las tiradas de dos dados diferentes que se muestran en la Tabla 1.7.

Es evidente que aunque en principio esperaríamos comportamientos muy similares en ambos dados -y que estos aún podrían obtenerse de seguir aumentando el número de tiradas- las muestras obtenidas nos dicen que los dados elegidos tienen conductas significativamente diferentes. Aunque las medias de los valores obtenidos son prácticamente idénticas en ambos casos ($\bar{x}_1 = 3,80$ y $\bar{x}_2 = 3,85$) de la mera inspección de los diagramas observamos que la dispersión en los valores de ambas muestras es muy diferente. Así, mientras el dado 2 registra una frecuencia razonablemente similar de obtención de todos los posibles valores, el dado 1 exhibe una sospechosa intensidad de realización del valor 4 de la variable, lo que puede indicar un cierto sesgo de la unidad 1 ausente en la 2.

Para construir una forma de cuantificar adecuadamente las diferencias entre estas distribuciones, hemos de preguntarnos en primer lugar ¿en qué consiste la dispersión de una muestra de datos? La respuesta es que la dispersión de los datos se manifiesta en la separación de las diferentes observaciones respecto a un valor concreto que tomamos como referencia y que es siempre alguna de las medidas de centralización de la muestra. Dependiendo del punto de referencia elegido y de la manera concreta de ponderar la desviación de los datos muestrales de ese punto fijo, obtendremos una u otra medida característica de dispersión.

1. *Varianza:*

La varianza s^2 de la muestra de datos se define como

$$s^2 = \sum_{i=1}^k f_i (x_i - \bar{x})^2, \quad (1.17)$$

es decir, la suma de las desviaciones cuadráticas de cada uno de los datos de la muestra respecto al valor central dado por la media aritmética, ponderadas dichas contribuciones por la correspondiente distribución de frecuencia⁹. La varianza es, sin duda alguna, la medida de dispersión más importante y, debido a que se trata de una suma de sumandos positivos, ella misma será una cantidad definida positiva en todo caso.

La varianza es el promedio con la distribución de frecuencias de la variable $(x - \bar{x})^2$, esto es del cuadrado de las desviaciones de los datos respecto a la media muestral. Desarrollando el cuadrado en el sumatorio de la Ec. (1.17) obtenemos:

$$\begin{aligned} s^2 &= \sum_{i=1}^k f_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^k f_i x_i^2 - 2\bar{x} \sum_{i=1}^k f_i x_i + \bar{x}^2 \sum_{i=1}^k f_i \\ &= \overline{x^2} - 2\bar{x}^2 + \bar{x}^2 = \overline{x^2} - \bar{x}^2 \end{aligned} \quad (1.18)$$

⁹Notemos que podríamos haber elegido para la medición de la dispersión de los datos de la muestra la expresión

$$\sum_{i=1}^k f_i (x_i - \bar{x})$$

que presentaría la indudable ventaja de ser una expresión lineal en las distancias a la media. Sin embargo, no tardamos en darnos cuenta de que esta expresión es idénticamente nula, dado que

$$\sum_{i=1}^k f_i = 1 \quad ; \quad \sum_{i=1}^k f_i x_i = \bar{x},$$

propiedad de la media que la constituye en el centro de gravedad de la muestra, puesto que la suma de las distancias a la media de los datos inferiores a esta es igual a la suma de las distancias de los superiores.

relación que en ocasiones se usa como definición de la varianza muestral. Sin embargo, puede ser numéricamente problemática cuando $\overline{x^2}$ y \bar{x}^2 son mucho mayores que la diferencia entre ambos. En ese caso, el cálculo de la varianza por este método puede llevar a una gran pérdida de precisión.

2. *Desviación típica o estándar:*

Se define como la raíz cuadrada de la varianza

$$s = \sqrt{\sum_{i=1}^k f_i (x_i - \bar{x})^2},$$

La llamada *desigualdad de Tchebychev* nos permite establecer, para una distribución genérica, la frecuencia de los datos que están más alejados de la media que un cierto número de veces la desviación típica.

Desigualdad de Tchebychev: Para una muestra de datos (x_1, \dots, x_N) la frecuencia de los datos que distan de la media muestral una cantidad superior a α veces la desviación típica de la muestra está acotada en la forma:

$$f(x_i \mid |x_i - \bar{x}| > \alpha \cdot s) < \frac{1}{\alpha^2}; \quad \alpha > 0 \quad (1.19)$$

La demostración de la desigualdad anterior es inmediata si dividimos el conjunto de los datos muestrales en dos conjuntos disjuntos de la forma:

$$\begin{aligned} A_1 &= \{x_i \mid |x_i - \bar{x}| \leq \alpha \cdot s\} \\ A_2 &= \{x_i \mid |x_i - \bar{x}| > \alpha \cdot s\} \end{aligned}$$

de tal manera que podemos escribir la varianza de la distribución de datos como

$$\begin{aligned} s^2 &= \sum_{i=1}^k f_i (x_i - \bar{x})^2 = \sum_{x_i \in A_1} f_i (x_i - \bar{x})^2 + \sum_{x_i \in A_2} f_i (x_i - \bar{x})^2 \\ &\geq \sum_{x_i \in A_2} f_i (x_i - \bar{x})^2 > \sum_{x_i \in A_2} f_i (\alpha \cdot s)^2 \\ &= (\alpha \cdot s)^2 f(x_i \mid |x_i - \bar{x}| > \alpha \cdot s) \end{aligned}$$

lo que demuestra la desigualdad.

A partir de la desigualdad de Tchebychev anterior, podemos extraer conclusiones acerca de la frecuencia de los datos en diferentes intervalos entorno a la media. Así:

- Los datos en el intervalo $(\bar{x} - 2s, \bar{x} + 2s)$ ($\alpha = 2$) son al menos el 75 % del total de los que componen la muestra ya que

$$f(x_i \mid |x_i - \bar{x}| > 2 \cdot s) < \frac{1}{2^2} = 0,25$$

- Los datos en el intervalo $(\bar{x} - 3s, \bar{x} + 3s)$ ($\alpha = 3$) son al menos el 89% del total de los que componen la muestra ya que

$$f(x_i \mid |x_i - \bar{x}| > 3 \cdot s) < \frac{1}{3^2} = 0,11$$

y así sucesivamente. Este resultado es aplicable a cualquier distribución de datos de varianza finita, por lo que podemos considerarlo también como una garantía de que los datos de cualquier experimento aleatorio van a aparecer concentrados en torno a un determinado valor centralizador, es decir, podemos hacer afirmaciones sobre la probabilidad de encontrar cierto valor de la variable en un determinado intervalo aun cuando desconozcamos la naturaleza de la distribución estadística que la describe.

3. *Desviaciones medias con respecto a la media y con respecto a la mediana:* se definen como

$$DM_{\bar{x}} = \sum_{i=1}^k f_i |x_i - \bar{x}|$$

y

$$DM_{Me} = \sum_{i=1}^k f_i |x_i - Me|$$

respectivamente. Son conceptualmente semejantes a la varianza porque promedian desviaciones con respecto a un punto medio. Sin embargo, las peores propiedades analíticas de la función valor absoluto con respecto a la función cuadrática hacen que sean menos empleadas.

Tanto la varianza como las desviaciones medias dependen de las unidades de la variable aleatoria correspondiente. De hecho, en el caso de que la variable estadística estudiada tenga unidades, la varianza tendrá esas mismas unidades al cuadrado y la desviación típica tendrá las unidades de la variable analizada. Así, es obvio que el hecho de que determinemos una longitud de un metro con una dispersión (incertidumbre) de un milímetro, no es en términos absolutos lo mismo que realizar la medición de una longitud media de un kilómetro con una dispersión en los datos muestrales de un metro. Y sin embargo en términos relativos ambas mediciones son análogas. Es por ello que resulta práctico introducir medidas relacionadas con las anteriores pero que no sean dependientes de la escala del problema concreto, esto es, que sean adimensionales.

4. *Coficiente de variación de Pearson:*

el coeficiente de variación de Pearson CV de una distribución de datos se define como

$$CV = \frac{s}{|\bar{x}|} \quad (1.20)$$

Esta es una medida relativa (adimensional) de variabilidad de la muestra. En Ingeniería se utiliza de modo análogo el cociente señal-ruido \bar{x}/s . Si consideramos (lo que será objeto de análisis detallado en el tema correspondiente) que la desviación típica es una medida de la incertidumbre de la medida y que el mejor valor que podemos atribuir al mensurando es el de la media muestral, entonces el coeficiente de variación de Pearson estará relacionado con la incertidumbre relativa de la medida.

5. *Coefficientes de variación media:*

de la misma manera en que se ha introducido el coeficiente de variación de Pearson es posible definir el coeficiente de variación media respecto a la media

$$CV_{M_{\bar{x}}} = \frac{DM_{\bar{x}}}{|\bar{x}|}, \quad (1.21)$$

y el coeficiente de variación media respecto a la mediana

$$CV_{M_{Me}} = \frac{DM_{Me}}{|Me|}. \quad (1.22)$$

Ejemplo 1.14

En el caso de la muestra de datos de desintegración radiactiva, la varianza, desviación típica y coeficiente de variación de Pearson se obtienen a partir de los datos de la Tabla 1.2 de manera directa como

$$\begin{aligned} s^2 &= \sum_{i=0}^{13} f_i(x_i - 5,81)^2 = 5,68 \text{ (cuentas)}^2 \\ s &= \sqrt{s^2} = 2,38 \text{ (cuentas)} \\ CV &= \frac{s}{\bar{x}} = \frac{2,38}{5,81} = 0,41 \end{aligned}$$

Dejamos al lector como ejercicio la obtención de los coeficientes de variación media respecto a la media y respecto a la mediana.

6. *Recorrido:*

Este parámetro de dispersión de una muestra de datos se define como la diferencia entre el valor máximo y el valor mínimo de la variable estadística:

$$R = \text{máx}(x) - \text{mín}(x) \quad (1.23)$$

obtenidas en la muestra.

7. *Recorrido intercuartílico:*

Es la diferencia entre el tercer y el primer cuartil:

$$R_I = P_{3/4} - P_{1/4}. \quad (1.24)$$

En ocasiones se define también el recorrido semi-intercuartílico como la mitad del recorrido intercuartílico, $R_{SI} = R_I/2$.

1.3.3. Momentos de una distribución de datos

Se define el momento de orden j de la distribución de frecuencias respecto al punto c de la forma:

$$m_j(c) = \sum_{i=1}^k f_i (x_i - c)^j = \overline{(x_i - c)^j} \quad (1.25)$$

Cuando $c = 0$ hablamos de momentos respecto al origen, $m_j(0)$, y cuando $c = \bar{x}$ los momentos se denominan centrales o respecto a la media.

La definición anterior nos permite sistematizar algunas de las medidas características de la distribución de datos que hemos introducido hasta el momento. Así, la media aritmética no es sino el momento de primer orden respecto al origen:

$$m_1(0) = \sum_{i=1}^k f_i (x_i - 0)^1 = \sum_{i=1}^k f_i x_i = \bar{x},$$

y la varianza es el momento central de segundo orden:

$$m_2(\bar{x}) = \sum_{i=1}^k f_i (x_i - \bar{x})^2 = s^2,$$

El lector puede apreciar que los momentos de primer orden están asociados a medidas de centralización, mientras que los de segundo orden involucran medidas de dispersión. Esta regla se confirmará en lo que sigue al comprobar cómo los momentos de tercer orden nos proporcionan una medida de la asimetría de la distribución y los de cuarto orden de su apuntamiento. Por otro lado, los momentos de orden cero son iguales a la unidad con independencia del punto de referencia.

Antes de proseguir con las aplicaciones de los momentos de la distribución de datos en la obtención de medidas características, observemos que, aplicando la fórmula del binomio de Newton, los diferentes momentos centrales pueden expresarse en términos de los momentos respecto al origen de la forma:

$$\begin{aligned} m_j(\bar{x}) &= \sum_{i=0}^j (-1)^i \binom{j}{i} m_i(0) [m_1(0)]^{j-i} \\ m_2(\bar{x}) &= m_2(0) - [m_1(0)]^2 \\ m_3(\bar{x}) &= m_3(0) - 3m_2(0)m_1(0) - 2[m_1(0)]^3 \\ &\dots \end{aligned} \quad (1.26)$$

1.3.4. Medidas de asimetría

Otra información interesante de una distribución de datos es si esta es simétrica en puntos equidistantes respecto al valor central, esto es, si los valores inferiores a la media que distan una cierta cantidad de esta son más, menos o igual de frecuentes que los valores superiores que se sitúan a la misma distancia. Debemos destacar que esta propiedad no se encuentra descrita por las medidas características introducidas con anterioridad, por lo que su medición nos obliga a introducir nuevos coeficientes.

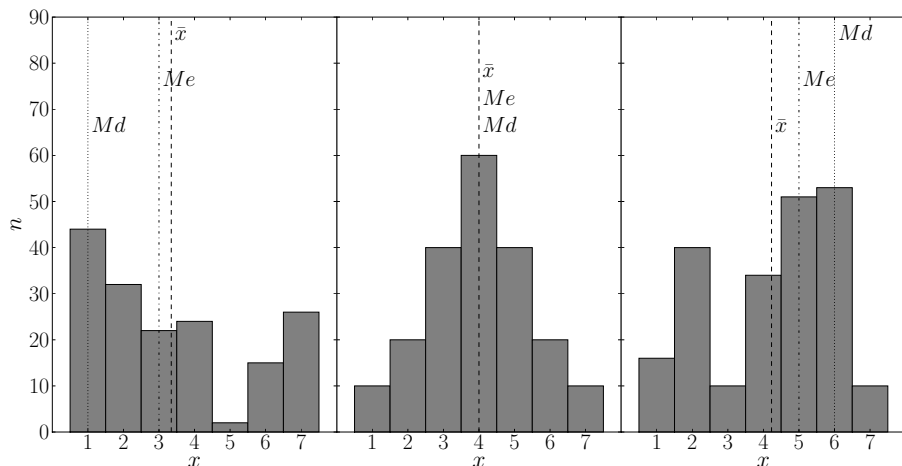


Figura 1.11: (a) Distribución de datos asimétrica a la derecha o positiva ($Md < Me \leq \bar{x}$). (b) Distribución simétrica ($Md = Me = \bar{x}$). (c) Distribución asimétrica a la izquierda o negativa ($Md \geq Me \geq \bar{x}$).

Los principales coeficientes que se utilizan para la caracterización de la simetría de la distribución de datos son:

1. *Coefficiente de asimetría de Pearson:*

Este coeficiente adimensional se define de la forma¹⁰:

$$A_P = \frac{\bar{x} - Md}{s} \quad (1.27)$$

En el caso de que $A_P > 0$ estaremos ante una distribución asimétrica positiva, mientras que $A_P < 0$ refleja la existencia de un sesgo hacia la izquierda o negativo en la muestra de datos. Lógicamente, el caso $A_P \simeq 0$ corresponde a una distribución aproximadamente simétrica. Debemos tener en cuenta que el valor de A_P nos da una idea de la asimetría o sesgo

¹⁰Otra medida de asimetría, aunque menos usada que esta, se obtiene tomando como valor de referencia central la mediana en lugar de la moda de la distribución:

$$\tilde{A}_P = \frac{\bar{x} - Me}{s}$$

de la distribución teniendo en cuenta únicamente la distancia entre la moda y la media aritmética, por lo que su contenido en información acerca de la distribución de datos es bajo. Además, se trata de un coeficiente inadecuado para distribuciones multimodales.

2. *Coefficiente de Bowley-Yule:*

Esta medida de asimetría se define a partir de los cuartiles de la distribución de datos como:

$$A_{BY} = \frac{(P_{3/4} - Me) - (Me - P_{1/4})}{P_{3/4} - P_{1/4}} = \frac{P_{3/4} + P_{1/4} - 2Me}{P_{3/4} - P_{1/4}}. \quad (1.28)$$

Cuando la distribución es simétrica, la distancia entre el primer cuartil y la mediana será igual a la existente entre ésta y el tercer cuartil, dando un coeficiente nulo. Cuando A_{BY} es distinto de cero, en cambio, una de éstas dos distancias es mayor que la otra y la distribución es asimétrica, con el mismo signo que el coeficiente.

3. *Coefficiente de asimetría de Fisher:*

El coeficiente de Fisher de la muestra de datos se define a partir del momento central de tercer orden de la distribución:

$$A_F = \frac{m_3(\bar{x})}{s^3} = \frac{1}{s^3} \sum_{i=1}^k f_i (x_i - \bar{x})^3. \quad (1.29)$$

La información que contiene el coeficiente adimensional de Fisher es notablemente superior a la de los dos coeficientes anteriores, siendo de hecho el más fiable de los tres en el sentido de que tiene en cuenta la contribución del conjunto de los datos en el cálculo de la asimetría de la muestra. Evidentemente, si $A_F > 0$ (< 0) estaremos ante una distribución asimétrica positiva (negativa) y el caso $A_F = 0$ corresponde a la distribución simétrica.

1.3.5. Medidas de apuntamiento o curtosis

Dos distribuciones de datos pueden tener idénticas medias y desviaciones típicas y tener no obstante formas diferentes debido a la diferente concentración de los datos. La medida característica que nos proporciona una estimación de cómo se reparte la frecuencia entre los datos extremos y los datos concentrados en torno a la media es el denominado coeficiente de apuntamiento o curtosis

$$g = \frac{m_4(\bar{x})}{s^4} = \frac{1}{s^4} \sum_{i=1}^k f_i (x_i - \bar{x})^4. \quad (1.30)$$

Demostremos posteriormente (ver capítulo 2) que la distribución normal o gaussiana presenta un valor de este coeficiente exactamente igual a 3, valor que

se toma como referencia para el estudio de la curtosis de las distribuciones estadísticas¹¹. Las distribuciones con una alta frecuencia para los datos extremos o un elevado apuntamiento se denominan leptocúrticas y se caracterizan por tener un valor del coeficiente de apuntamiento mayor que el de la distribución normal $g > 3$, mientras que las de bajo apuntamiento o alta concentración de los datos en torno al valor central reciben el nombre de platocúrticas ($g < 3$).

A la curtosis contribuyen dos factores: el apuntamiento y las colas de la distribución. Para evaluarlos de manera separada pueden emplearse las siguientes medidas basadas en distancias entre percentiles, debidas a Schmid y Tiede:

$$CP = \frac{P_{0,875} - P_{0,125}}{P_{0,075} - P_{0,025}} \quad (1.31)$$

$$CT = \frac{P_{0,975} - P_{0,025}}{P_{0,875} - P_{0,025}} \quad (1.32)$$

respectivamente. Sus valores para la distribución normal son aproximadamente $CP = 1,7055$ y $CT = 1,7038$. Una distribución de frecuencias con CP mayor que el de la normal es más apuntada que ésta. Del mismo modo, si tiene CT mayor que el de la gaussiana se hablará de que sus colas, conformadas por los datos muy alejados de la parte central, son más relevantes que las de la normal.

Como se verá en el capítulo 2, existe una clase importante de distribuciones leptocúrticas con colas gruesas, cuyo comportamiento lejos de la parte central se puede describir por una ley potencial del tipo $f_i = x_i^{-\alpha}$. Un importante ejemplo real de este tipo de distribuciones es el tamaño de los dominios de magnetización en un material ferromagnético en las inmediaciones de su punto crítico, la denominada ley de Pareto ya introducida de distribución de la renta, las distribuciones de frecuencia de variaciones de precios de activos en los mercados financieros y, en general, las distribuciones de probabilidad asociadas a sistemas complejos.

1.4. Transformaciones de la variable aleatoria

En muchas ocasiones estamos interesados en realizar operaciones con los datos procedentes de los experimentos aleatorios, bien porque en el caso concreto sea más conveniente su manipulación para facilitar su tratamiento, bien porque sea necesaria su transformación para poder acceder a una nueva variable aleatoria de interés.

a) *Transformaciones lineales:*

Consideremos una muestra $\{x_1, \dots, x_N\}$ de realizaciones de la variable estadística

¹¹En ocasiones se define el coeficiente de apuntamiento respecto a la distribución gaussiana como

$$g = \frac{m_4(\bar{x})}{s^4} - 3,$$

en cuyo caso esta distribución presentaría un valor $g = 0$.

ca X con los posibles valores $\{x_1, \dots, x_k\}$ y media aritmética

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \sum_{i=1}^k f_i \cdot x_i$$

Transformemos esta variable a una nueva variable $Y = a + bX$ con valores posibles $\{a + bx_1, \dots, a + bx_k\}$. Analicemos el impacto de esta transformación en las principales medidas características de la muestra de datos original. En primer lugar, la media aritmética de la nueva variable aleatoria es

$$\begin{aligned} \bar{y} &= \sum_{i=1}^k f_i y_i = \sum_{i=1}^k f_i (a + bx_i) \\ &= a + b \sum_{i=1}^k f_i x_i = a + b\bar{x}, \end{aligned} \quad (1.33)$$

Como vemos la media aritmética se transforma de la misma manera que la propia variable, una consecuencia de la linealidad de ambas operaciones. Por su parte, la varianza de la nueva variable aleatoria será

$$\begin{aligned} s_y^2 &= \sum_{i=1}^k f_i (y_i - \bar{y})^2 \\ &= \sum_{i=1}^k f_i (a + bx_i - a - b\bar{x})^2 \\ &= b^2 \sum_{i=1}^k f_i (x_i - \bar{x})^2 = b^2 s^2 \end{aligned} \quad (1.34)$$

Observemos, en particular, que en el caso de que tengamos únicamente un cambio de origen o un desplazamiento global de la variable ($b = 1$) la varianza de la nueva variable será la misma que la de la variable no transformada. Otra transformación lineal importante cuyas implicaciones en la transformación de los datos de la variable anterior podemos analizar a partir de las ecuaciones anteriores es un cambio de unidades en la magnitud de estudio, lo que equivale a una transformación de escala.

Ejemplo 1.15

En el caso de los datos de densidad del agua que presentamos en la Tabla 1.3, la sencillez de manejo matemático sugiere una transformación de la forma:

$$\rho \rightarrow \rho' = 10^6 (\rho - 1 \text{ g cm}^{-3})$$

variable que presentará un menor número de dígitos y que se representa en la Fig. 1.8. Esta transformación nos conducirá a una tabla para la nueva variable estadística de la forma:

ρ'_i ($\mu\text{g cm}^{-3}$)	n_i	f_i	N_i	F_i
-3	4	0.08	4	0.08
-2	5	0.10	9	0.18
-1	6	0.12	15	0.30
0	13	0.26	28	0.56
1	10	0.20	38	0.76
2	5	0.10	43	0.86
3	4	0.08	47	0.94
4	3	0.06	50	1.00

Como vemos, las distribuciones de frecuencia permanecen inalteradas, un hecho atribuible a la linealidad de la transformación considerada.

b) *Transformaciones no lineales:*

En muchas ocasiones nos interesa estudiar una magnitud que depende de la variable estadística que controlamos en nuestro experimento de manera no lineal. Algunas transformaciones usuales son:

$$Y = X^p, Y = \ln X, Y = \sqrt[p]{X} \dots$$

Las transformaciones no lineales, a diferencia de sus homólogas lineales, suelen variar la forma de las distribuciones originales de frecuencias, lo que nos obliga a ser cuidadosos en el análisis de las implicaciones de la transformación en las diferentes medidas características. Por ejemplo, la transformación $Y = X^2$ desplaza los histogramas de frecuencia hacia valores superiores, al comprimir la escala de los valores bajos y expandir los valores altos de la variable. Por su parte, la transformación $Y = \ln X$ suele asimetrizar distribuciones originalmente simétricas.

Veamos a continuación cuáles son los efectos de este tipo de transformaciones en algunas medidas características de las distribuciones de datos.

a) *Percentiles:*

En lo que respecta a las medidas características que tienen en cuenta únicamente el número de datos y su orden, las transformaciones $y = h(x)$ que preservan dicho orden (monótonas crecientes),

$$x_1 < x_2 \Leftrightarrow h(x_1) < h(x_2)$$

transforman los percentiles de la distribución de manera directa:

$$P_q(y) = h[P_q(x)] \quad (1.35)$$

b) *Media aritmética:*

La transformación de la media aritmética mediante una transformación

no lineal $y = h(x)$ no es trivial y, en general, no verifica $\bar{y} = h(\bar{x})$. La media aritmética de la nueva variable está dada por la expresión

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} \sum_{i=1}^N h(x_i) = \sum_{i=1}^k f_i \cdot h(x_i) \quad (1.36)$$

Realizando un desarrollo de Taylor a segundo orden alrededor de la media aritmética de la variable original, tendremos

$$\begin{aligned} \bar{y} &\simeq \frac{1}{N} \sum_{i=1}^N \left[h(\bar{x}) + h'(\bar{x})(x_i - \bar{x}) + \frac{1}{2!} h''(\bar{x})(x_i - \bar{x})^2 \right] \\ \bar{y} &\simeq h(\bar{x}) + \frac{1}{2} h''(\bar{x}) s^2 \end{aligned} \quad (1.37)$$

donde hemos usado

$$\sum_{i=1}^N (x_i - \bar{x}) = 0$$

c) Desviación típica:

A partir del resultado anterior para la media aritmética es inmediato obtener una expresión aproximada para la varianza de la variable transformada. Así

$$\begin{aligned} s_y^2 &= \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 \\ &\simeq \frac{1}{N} \sum_{i=1}^N \left[h(\bar{x}) + h'(\bar{x})(x_i - \bar{x}) + \frac{1}{2!} h''(\bar{x})(x_i - \bar{x})^2 - \bar{y} \right]^2 \\ &\simeq \frac{1}{N} \sum_{i=1}^N \left[h(\bar{x}) + h'(\bar{x})(x_i - \bar{x}) + \frac{1}{2!} h''(\bar{x})(x_i - \bar{x})^2 - h(\bar{x}) \right. \\ &\quad \left. - \frac{1}{2!} h''(\bar{x}) s^2 \right]^2 \simeq \frac{1}{N} \sum_{i=1}^N [h'(\bar{x})(x_i - \bar{x})]^2 = [h'(\bar{x})]^2 s^2 \end{aligned} \quad (1.38)$$

Si se cumple $h''(\bar{x}) \cdot s^2 \rightarrow 0$, a orden más bajo en el desarrollo de Taylor la media aritmética de la variable transformada \bar{y} es aproximadamente igual a $h(\bar{x})$. La varianza, sin embargo, presenta un factor $[h'(\bar{x})]^2$ que cambia su magnitud en transformaciones de la variable aleatoria.

Ejemplo 1.16

Mercado de valores: Cuando se toman valores de la evolución del precio de las acciones de una determinada compañía en la Bolsa de valores se obtiene una serie temporal $\{Y_j = Y(t_j)\}_{j=1}^n$. $Y(t)$ es la variable aleatoria por excelencia en el estudio de mercados financieros, aunque razones de conveniencia llevan a introducir como variables alternativas de estudio:

- a) Una transformación lineal que corresponde a la variación de precios en un instante determinado $\Delta Y(t) = Y(t + \Delta t) - Y(t)$.
- b) Los retornos, o incrementos relativos de los precios, que se construyen mediante la transformación no lineal:

$$R(t) = \frac{Y(t + \Delta t) - Y(t)}{Y(t)}$$

- c) Las diferencias del logaritmo de los precios entre dos instantes consecutivos:

$$S(t) = \ln Y(t + \Delta t) - \ln Y(t)$$

que coincide con el retorno para datos de alta frecuencia ($\Delta t \rightarrow 0$):

$$S(t) = \ln \frac{Y(t + \Delta t)}{Y(t)} = \ln \left(1 + \frac{\Delta Y(t)}{Y(t)} \right) \simeq \frac{\Delta Y(t)}{Y(t)}$$

Volveremos sobre este tema al hablar de transformación de distribuciones de probabilidad en el capítulo siguiente.

1.5. Muestras multivariantes

El establecimiento de correlaciones entre magnitudes hace que, con frecuencia, más que el análisis de una sola variable aleatoria sea el estudio conjunto de dos o más variables aleatorias el que centre nuestro interés. Por ejemplo, para analizar la relación entre la corriente eléctrica I que circula por un circuito sometido a un potencial eléctrico V , observaremos la variación de I a medida que variamos V . Este experimento lo podemos considerar como fuente de una variable aleatoria bidimensional (I, V) que luego analizaremos para describir el grado de relación existente entre las mismas.

1.5.1. Distribuciones de frecuencia

En lo que sigue, por simplicidad notacional y por razones de representación práctica, trabajaremos con muestras de dos variables, aunque la generalización a muestras de más variables es inmediata. Consideremos así una muestra de datos de dos variables estadísticas X e Y con una serie de valores posibles para la variable estadística X , $\{x_1, \dots, x_k\}$, y para la variable Y , $\{y_1, \dots, y_l\}$. En este caso el experimento aleatorio suministrará una muestra de N datos formada por pares de la forma (x_i, y_j) , con una determinada frecuencia que denominaremos n_{ij} , que representa el número de veces que se ha obtenido el par (x_i, y_j) en el experimento aleatorio. Las tablas de frecuencias en este caso serán bidimensionales, tal y como muestra la Tabla 1.9 para la distribución de frecuencias absolutas.

YX	y_1	y_2	...	y_j	...	y_l
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1l}
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2l}
...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{il}
...
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{kl}

Tabla 1.9 Tabla de frecuencias absolutas de una variable aleatoria bidimensional.

A partir de la distribución de frecuencias absolutas podemos obtener la de frecuencias relativas en la manera usual:

$$f_{ij} = \frac{n_{ij}}{N} \quad (1.39)$$

Evidentemente, las distribuciones de frecuencias anteriores deben verificar las propiedades:

$$\begin{aligned} \sum_{i=1}^k \sum_{j=1}^l n_{ij} &= N \\ \sum_{i=1}^k \sum_{j=1}^l f_{ij} &= 1 \end{aligned} \quad (1.40)$$

de la misma manera que lo hacía sus homólogas unidimensionales.

1.5.2. Distribuciones marginales de frecuencia

En el caso de muestras multidimensionales podemos plantearnos la pregunta de en cuántas ocasiones obtenemos un determinado valor de la variable estadística Y con independencia del valor de X con el que este venga acompañado (o viceversa). Esta pregunta, sin analogía posible en el caso unidimensional, es la que nos conduce a la introducción de las distribuciones marginales de frecuencia, entendiéndolo por tales las obtenidas al estudiar dicha variable aisladamente, independientemente del resto de las variables.

A partir de la distribución de frecuencias conjunta n_{ij} , se define la frecuencia absoluta marginal para el valor x_i como:

$$n_{x_i} = \sum_{j=1}^l n_{ij} = n_{i1} + n_{i2} + \dots + n_{il} \quad (1.41)$$

siendo la serie de frecuencias $\{n_{x_i}\}_{i=1}^k$ la denominada distribución marginal de frecuencias absolutas de la variable X . De manera análoga podemos definir la frecuencia absoluta marginal del valor y_j como

$$n_{y_j} = \sum_{i=1}^k n_{ij} = n_{1j} + n_{2j} + \dots + n_{kj} \quad (1.42)$$

siendo la serie de frecuencias $\{n_{y_j}\}_{j=1}^l$ la denominada distribución marginal de frecuencias absolutas de la variable Y .

La definición de las distribuciones de frecuencias relativas marginales es inmediata a partir de sus homólogas absolutas. Así, las frecuencias relativas marginales de los valores x_i e y_j respectivamente se definen como:

$$\begin{aligned} f_{x_i} &= \frac{n_{x_i}}{N} = \sum_{j=1}^l \frac{n_{ij}}{N} = \sum_{j=1}^l f_{ij} \\ f_{y_j} &= \frac{n_{y_j}}{N} = \sum_{i=1}^k \frac{n_{ij}}{N} = \sum_{i=1}^k f_{ij} \end{aligned} \quad (1.43)$$

Es obvio que las distribuciones de frecuencias marginales que acabamos de introducir presentan, entre otras, las siguientes propiedades:

$$\begin{aligned} \sum_{i=1}^k n_{x_i} &= N ; \sum_{j=1}^l n_{y_j} = N \\ \sum_{i=1}^k f_{x_i} &= 1 ; \sum_{j=1}^l f_{y_j} = 1 \end{aligned} \quad (1.44)$$

Ejemplo 1.17

Consideremos una distribución de frecuencias absolutas como la que se refleja en la Tabla 1.10. Las distribuciones marginales se obtienen sumando (y anotando al margen, de ahí su nombre) los valores de cada una de las filas y columnas de la tabla.

$X \setminus Y$	0	1	2	3	$n_{x_i} = \sum_{j=1}^l n_{ij}$
10	25	5	3	0	33
15	10	12	13	1	36
18	0	2	5	7	14
25	1	0	0	1	2
$n_{y_j} = \sum_{i=1}^k n_{ij}$	36	19	21	9	85

Tabla 1.10 Tabla de frecuencias absolutas y marginales de la variable aleatoria bidimensional mostrada en la Tabla 1.10.

1.5.3. Distribuciones de frecuencias condicionadas

Resulta muy conveniente en muchas situaciones conocer con qué frecuencia se observa un determinado valor de una variable estadística cuando la otra variable presenta un cierto valor de interés.

Se define la frecuencia absoluta con la que observamos el valor x_i de la variable X condicionada a que se observe el valor y_j de la variable Y como

$$n(x_i | Y = y_j) = n(x_i | y_j) = n_{ij}. \quad (1.45)$$

Obviamente, una expresión similar adopta la frecuencia absoluta con la que observamos el valor y_j de la variable Y condicionada a que se observe el valor x_i de la variable X :

$$n(y_j | X = x_i) = n(y_j | x_i) = n_{ji}. \quad (1.46)$$

A partir de las definiciones anteriores es inmediato introducir las distribuciones de frecuencias relativas condicionadas

$$\begin{aligned} f(x_i | Y = y_j) = f(x_i | y_j) &= \frac{n(x_i | Y = y_j)}{n_{y_j}} = \frac{n_{ij}}{n_{y_j}} = \frac{f_{ij}}{f_{y_j}} \\ f(y_j | X = x_i) = f(y_j | x_i) &= \frac{n(y_j | X = x_i)}{n_{x_i}} = \frac{n_{ji}}{n_{x_i}} = \frac{f_{ij}}{f_{x_i}} \end{aligned} \quad (1.47)$$

donde, como vemos en las expresiones anteriores, la frecuencia condicionada $f(x_i | y_j)$ corresponde a la frecuencia relativa de $X = x_i$ en el conjunto de datos restringido a aquellos donde $Y = y_j$ (no respecto al conjunto total de observaciones). Debido a esto tenemos también que, en general, $f(x_i | y_j) \neq f(y_j | x_i)$.

Ejemplo 1.18

Lanzamiento de dos dados. Supongamos que hemos realizado un experimento aleatorio consistente en lanzar dos dados simultáneamente habiéndose obtenido la siguiente muestra de datos de la variable bidimensional $\vec{X} = \{X, Y\}$, $X =$ "resultado lanzamiento dado 1" e $Y =$ "resultado lanzamiento dado 2":

$$\{(1, 3), (2, 1), (2, 3), (5, 4), (2, 6), (3, 4), (4, 3), (5, 2), (6, 2), (1, 6)\}$$

Si estamos interesados en calcular las distribuciones de frecuencias del valor 2 en el primer dado condicionadas a la observación del valor 3 en el segundo dado, debemos calcular previamente la tabla de frecuencias conjuntas y las distribuciones marginales:

$X \setminus Y$	1	2	3	4	5	6	$n_{x_i} = \sum_{j=1}^l n_{ij}$
1	0	0	1	0	0	1	2
2	1	0	1	0	0	1	3
3	0	0	1	0	0	0	1
4	0	0	1	0	0	0	1
5	0	1	0	1	0	0	2
6	0	1	0	0	0	0	1
$n_{y_j} = \sum_{i=1}^k n_{ij}$	1	2	4	1	0	2	10

Así,

$$f(X = 2 | Y = 3) = \frac{n(X = 2 | Y = 3)}{n_{(Y=3)}} = \frac{n_{23}}{n_{(Y=3)}} = \frac{1}{4}$$

lo que responde al hecho de que en la muestra concreta de las cuatro veces que ha aparecido el resultado 3 en el dado 2, sólo una lo ha hecho acompañado del valor 2 en el dado 1. Observemos que es diferente la distribución de frecuencias condicionadas del valor 3 en el dado 1 al valor 2 en el dado 2. En este caso,

$$f(Y = 3 | X = 2) = \frac{n(Y = 3 | X = 2)}{n_{(X=2)}} = \frac{n_{32}}{n_{(X=2)}} = \frac{1}{3}$$

puesto que ahora, de las tres veces que en el conjunto de la muestra se ha obtenido un resultado 2 en el dado 1, solamente una vez lo ha hecho acompañado del 3 en el dado 2.

Las distribuciones de frecuencias condicionadas verifican la propiedad de normalización:

$$\begin{aligned} \sum_{i=1}^k f(x_i | Y = y_j) &= \frac{\sum_{i=1}^k n(x_i | Y = y_j)}{n_{y_j}} = 1 \\ \sum_{j=1}^l f(y_j | X = x_i) &= \frac{\sum_{j=1}^l n(y_j | X = x_i)}{n_{x_i}} = 1. \end{aligned} \quad (1.48)$$

Por otro lado, a partir de la Ec. (1.47) se puede obtener las relaciones siguientes entre las distribuciones de frecuencias:

$$\begin{aligned} f_{ij} &= f(x_i | y_j) f_{y_j} \\ f_{ij} &= f(y_j | x_i) f_{x_i} \end{aligned} \quad (1.49)$$

Las frecuencias condicionadas presentan interés en el estudio de subconjuntos de la muestra total. Podemos entonces relacionar las frecuencias relativas f_{x_i} y f_{y_j}

$$f_{x_i} = \sum_{j=1}^l f_{ij} = \sum_{j=1}^l f(x_i | Y = y_j) f_{y_j} \quad (1.50)$$

$$f_{y_j} = \sum_{i=1}^k f_{ij} = \sum_{i=1}^k f(y_j | X = x_i) f_{x_i} \quad (1.51)$$

A partir del concepto de frecuencia condicionada podemos introducir la denominada *independencia estadística de variables aleatorias*. Diremos que dos variables estadísticas X e Y son estadísticamente independientes entre sí cuando la distribución de frecuencias del valor x_i de la variable X condicionada a que Y haya tomado el valor y_j es igual a la distribución marginal de x_i :

$$f(x_i | Y = y_j) = f_{x_i} \quad (1.52)$$

Análogamente, si X e Y son variables independientes

$$f(y_j | X = x_i) = f_{y_j} \quad (1.53)$$

De esta forma, usando la definición de las distribuciones condicionadas de frecuencia en la Ec. (1.47) obtenemos que, para variables independientes:

$$f_{ij} = f_{x_i} \cdot f_{y_j} \quad (1.54)$$

1.5.4. Representaciones gráficas

Una vez introducidas las distribuciones de frecuencias en muestras multivariantes podemos plantearnos representar gráficamente la información que éstas contienen. Existen dos tipos básicos de representaciones para muestras de datos bidimensionales:

a) *Diagramas de dispersión:*

Consisten en la representación en unos ejes coordenados del conjunto de datos $\{(x_i, y_j)\}_{i=1}^N$ que componen la muestra, siendo x_i los valores de las abscisas e y_j los de las ordenadas de cada punto (Fig. 1.12).

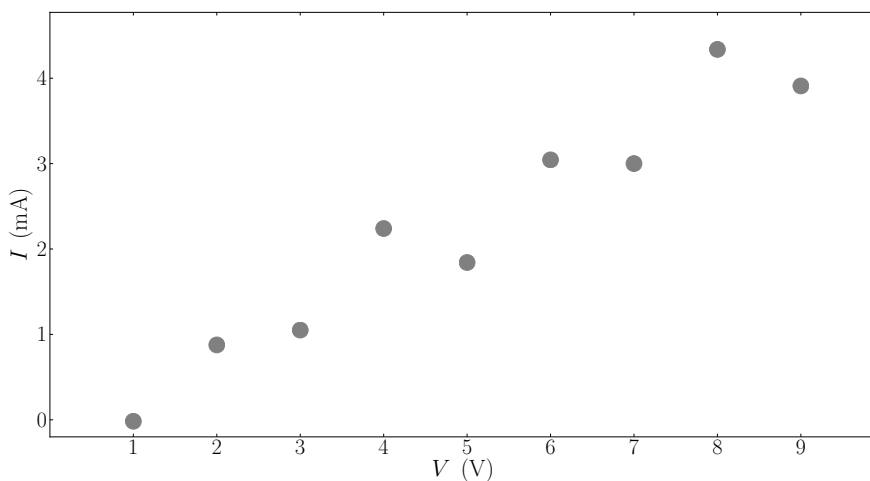


Figura 1.12: Diagrama de dispersión para los datos de voltaje aplicado e intensidad medidos en una resistencia eléctrica de $2 \text{ k}\Omega$.

b) *Diagramas de frecuencia:*

Las mismas representaciones que hemos introducido en el caso unidimensional para la representación de las distribuciones de frecuencia -una información que no se contiene en los diagramas de dispersión- son extrapolables al caso multidimensional. Si a cada par (x_i, y_j) de las variables aleatorias discretas X e Y se le asocia una barra cuya altura es la frecuencia absoluta o relativa estamos ante un diagrama de barras. Si los datos están agrupados y lo que asociamos

es un prisma de base rectangular de altura proporcional a la frecuencia de la clase (x_i, y_j) , estamos ante un histograma de la muestra de datos (Fig. 1.13). Los histogramas de variables bidimensionales pueden resultar confusos porque los prismas más altos pueden tapar los más bajos desde una determinada perspectiva. Por ello una alternativa es presentar un mapa bidimensional en que a cada (x_i, y_j) se le asocie un color proporcional a su frecuencia de una escala previamente especificada.

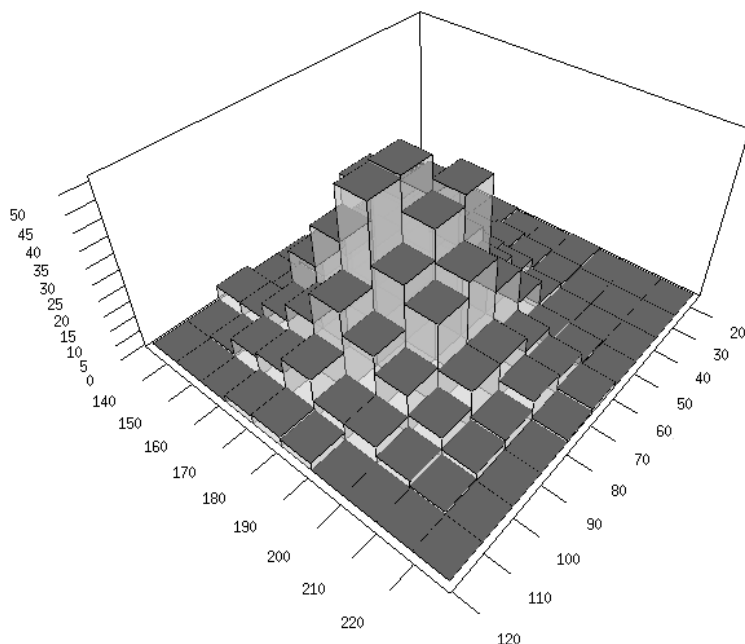


Figura 1.13: Ejemplo de histograma de frecuencias absolutas para dos variables.

1.5.5. Medidas características de la distribución de frecuencias

De manera análoga a lo que sucedía en el caso de las muestras unidimensionales, es conveniente introducir las medidas características de la distribución de frecuencias muestrales. Aunque es posible introducir otras medidas, analizaremos en esta obra esencialmente las medidas de centralización y de dispersión más importantes, la media aritmética y la varianza de la distribución, así como la covarianza, medida característica asociada a muestras multivariantes que nos permite medir el grado de correlación existente entre dos variables estadísticas.

Consideremos una muestra de datos $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ asociada a una variable estadística bidimensional $\vec{X} = (X, Y)$ formada por componentes discretas o agrupadas en clases de tal manera que los valores posibles

de X son (x_1, \dots, x_k) y los de Y son (y_1, \dots, y_l) . Denotaremos, como es habitual, la frecuencia del par (x_i, y_j) como f_{ij} .

Definimos el *momento de órdenes r y s respecto al punto (c, d)* de la distribución de frecuencias como:

$$m_{r,s}(c, d) = \sum_{i=1}^k \sum_{j=1}^l f_{ij} (x_i - c)^r (y_j - d)^s \quad (1.55)$$

Los momentos respecto a $(c, d) = (0, 0)$ se denominan momentos respecto al origen y los momentos respecto a las medias de las variables $(c, d) = (\bar{x}, \bar{y})$ reciben el nombre de momentos centrales, nuevamente de manera análoga a lo que sucedía en el caso de las muestras unidimensionales. Introduciremos por generalidad las principales medidas características de la distribución de datos usando los momentos de la distribución antes definidos.

a) *Medias aritméticas:*

Las medias aritméticas de las variables estadísticas que componen la variable multidimensional se definen como los momentos respecto al origen de primer orden en la variable promediada y de orden cero en las restantes. Para el caso de una muestra bidimensional:

$$\begin{aligned} \bar{x} &= m_{1,0}(0, d) = \sum_{i=1}^k \sum_{j=1}^l f_{ij} x_i (y_j - d)^0 = \sum_{i=1}^k \sum_{j=1}^l f_{ij} x_i \\ \bar{y} &= m_{0,1}(c, 0) = \sum_{i=1}^k \sum_{j=1}^l f_{ij} (x_i - c)^0 y_j = \sum_{i=1}^k \sum_{j=1}^l f_{ij} y_j \end{aligned} \quad (1.56)$$

la media se obtiene (de modo análogo al caso unidimensional) como la suma de los diferentes valores de la variable o función a promediar ponderados cada uno de ellos por la frecuencia que le corresponde según la distribución.

Usando la definición de las distribuciones marginales de frecuencia en la Ec. (1.43) obtenemos:

$$\begin{aligned} \bar{x} &= \sum_{i=1}^k \sum_{j=1}^l f_{ij} x_i = \sum_{i=1}^k f_{x_i} x_i \\ \bar{y} &= \sum_{i=1}^k \sum_{j=1}^l f_{ij} y_j = \sum_{j=1}^l f_{y_j} y_j \end{aligned} \quad (1.57)$$

resultados que, en ocasiones, reciben el nombre de medias marginales de las variables X e Y .

b) *Varianzas:*

Las varianzas de las variables estadísticas X e Y se obtienen a partir de

los momentos centrales de segundo orden como:

$$\begin{aligned} s_x^2 &= m_{2,0}(\bar{x}, d) = \sum_{i=1}^k \sum_{j=1}^l f_{ij} (x_i - \bar{x})^2 \\ s_y^2 &= m_{0,2}(c, \bar{y}) = \sum_{i=1}^k \sum_{j=1}^l f_{ij} (y_j - \bar{y})^2 \end{aligned} \quad (1.58)$$

que también pueden expresarse en términos de las distribuciones marginales de frecuencia correspondientes a X e Y como

$$\begin{aligned} s_x^2 &= \sum_{i=1}^k \sum_{j=1}^l f_{ij} (x_i - \bar{x})^2 = \sum_{i=1}^k f_{x_i} (x_i - \bar{x})^2 \\ s_y^2 &= \sum_{i=1}^k \sum_{j=1}^l f_{ij} (y_j - \bar{y})^2 = \sum_{j=1}^l f_{y_j} (y_j - \bar{y})^2 \end{aligned} \quad (1.59)$$

razón por la que también reciben el nombre de varianzas marginales. Lógicamente las desviaciones típicas correspondientes se obtendrán como la raíz cuadrada de los valores anteriores.

c) *Covarianza:*

Se define la covarianza de las variables estadísticas X e Y como el momento central de primer orden en las dos variables aleatorias de la distribución conjunta de frecuencias:

$$\text{cov}(x, y) = m_{1,1}(\bar{x}, \bar{y}) = \sum_{i=1}^k \sum_{j=1}^l f_{ij} (x_i - \bar{x})(y_j - \bar{y}) \quad (1.60)$$

Observemos que, desarrollando el producto en los sumandos del miembro de la derecha de la ecuación anterior, podemos expresar la covarianza de la forma:

$$\begin{aligned} \text{cov}(x, y) &= \sum_{i=1}^k \sum_{j=1}^l f_{ij} (x_i - \bar{x})(y_j - \bar{y}) \\ &= \sum_{i=1}^k \sum_{j=1}^l f_{ij} x_i y_j - \bar{x} \sum_{i=1}^k \sum_{j=1}^l f_{ij} y_j - \bar{y} \sum_{i=1}^k \sum_{j=1}^l f_{ij} x_i + \\ &\quad + \bar{x} \bar{y} \sum_{i=1}^k \sum_{j=1}^l f_{ij} = \sum_{i=1}^k \sum_{j=1}^l f_{ij} x_i y_j - \bar{x} \bar{y} = \overline{xy} - \bar{x} \bar{y} \end{aligned} \quad (1.61)$$

expresión muy importante que, en muchas ocasiones, se utiliza incluso como la propia definición de covarianza. Hemos usado en esta expresión

$$\overline{xy} = \sum_{i=1}^k \sum_{j=1}^l f_{ij} x_i y_j \quad (1.62)$$

que podemos extender a cualquier función de las dos variables aleatorias $g(x, y)$ de modo análogo a la definición de la ecuación (1.11), de forma que

$$\overline{g(x, y)} = \sum_{i=1}^k \sum_{j=1}^l f_{ij} g(x_i, y_j) \quad (1.63)$$

De la definición de la covarianza se deduce la propiedad de simetría $\text{cov}(x, y) = \text{cov}(y, x)$. Por otro lado, también se cumple que $\text{cov}(x, x) = s_x^2$. Observemos que en el caso de que las variables X e Y sean estadísticamente independientes,

$$\begin{aligned} \overline{xy} &= \sum_{i=1}^k \sum_{j=1}^l f_{ij} x_i y_j = \sum_{i=1}^k \sum_{j=1}^l f_{x_i} f_{y_j} x_i y_j \\ &= \sum_{i=1}^k f_{x_i} x_i \sum_{j=1}^l f_{y_j} y_j = \bar{x} \bar{y} \end{aligned} \quad (1.64)$$

y por tanto la covarianza de ambas variables es nula.

La covarianza es una medida cuya importancia dentro del formalismo estadístico radica en que proporciona información acerca del grado de correlación lineal que existe entre las dos variables estadísticas a las que se refiere. Así,

$$\text{cov}(x, y) = \begin{cases} > 0 & \text{correlación positiva} \\ = 0 & \text{ausencia de correlación} \\ < 0 & \text{correlación negativa} \end{cases}$$

Tomando como origen de coordenadas el punto (\bar{x}, \bar{y}) , si los datos se encuentran distribuidos entre el primer y tercer cuadrantes, entonces contribuyen con factores $(x_i - \bar{x})(y_j - \bar{y})$ positivos, de modo que $\text{cov}(x, y) > 0$. Si los datos están distribuidos entre el segundo y cuarto cuadrante, entonces los términos $(x_i - \bar{x})(y_j - \bar{y}) < 0$ de modo que $\text{cov}(x, y) < 0$. En cambio, si los datos se encuentran dispersos al azar en todas las direcciones esperamos $\text{cov}(x, y) \simeq 0$.

Por otro lado, como vemos en la Ec. (1.60), la covarianza no es una medida característica de la muestra que sea invariante ante cambios de escala y su dominio no está acotado (puede tomar cualquier valor posible) por lo que es conveniente introducir algún parámetro que, reteniendo esta misma información, corrija estas deficiencias. Este es el denominado coeficiente de correlación lineal o de covariación entre variables estadísticas dado por:

$$r(x, y) = \frac{\text{cov}(x, y)}{s_x s_y}, \quad (1.65)$$

que es un coeficiente adimensional¹² y cuyo dominio está acotado ya

¹² El coeficiente de correlación lineal permanece invariante ante cambios de escala o unidades en las variables x e y .

que se verifica $-1 \leq r(x, y) \leq +1$. Para probar esto, consideraremos la siguiente combinación de cuadrados:

$$\sum_{i=1}^k \sum_{j=1}^l f_{ij} [(x_i - \bar{x}) + \lambda (y_j - \bar{y})]^2$$

que es obviamente positiva para cualquier valor real de λ . Desarrollando todos los cuadrados, observaremos que esto equivale a escribir:

$$s_x^2 + \lambda^2 s_y^2 - 2\lambda \text{cov}(x, y) \geq 0 \quad \forall \lambda \in \mathbb{R}$$

Si $s_y = 0$, de esta desigualdad se desprende que $\text{cov}(x, y) = 0$, un caso trivial. En caso contrario, basta con elegir $\lambda = \frac{\text{cov}(x, y)}{s_y^2}$ para transformar la desigualdad en $s_x^2 - \frac{[\text{cov}(x, y)]^2}{s_y^2} \geq 0$, de la cual se despeja $|\text{cov}(x, y)| \leq s_x s_y$, es decir, $|r(x, y)| \leq 1$, como se quería probar.

Por otra parte, también podemos extender esta demostración para establecer la existencia de una correlación lineal perfecta cuando el coeficiente de correlación es $r(x, y) = \pm 1$. Para ello observemos que

$$s^2(x + \lambda y) = s_x^2 + \lambda^2 s_y^2 + 2\lambda \text{cov}(x, y) \quad \forall \lambda \in \mathbb{R}$$

Supongamos, sin pérdida de generalidad, que $r(x, y) = 1$, esto es equivalente a $\text{cov}(x, y) = s_x s_y$, por lo que en la expresión anterior podemos escribir

$$s^2(x + \lambda y) = s_x^2 + \lambda^2 s_y^2 + 2\lambda s_x s_y = [s_x + \lambda s_y]^2$$

Basta entonces tomar $\lambda = -\frac{s_x}{s_y}$ para obtener

$$s^2\left(x - \frac{s_x}{s_y} y\right) = \left[s_x - \frac{s_x}{s_y} s_y\right]^2 = 0$$

de donde se concluye la existencia de una relación lineal determinista (correlación perfecta) entre las variables aleatorias X e Y .

En términos del coeficiente de correlación tendremos:

$$r(x, y) = \begin{cases} +1 & \text{correlación lineal positiva perfecta} \\ = 0 & \text{ausencia de correlación lineal} \\ -1 & \text{correlación lineal negativa perfecta} \end{cases}$$

Observemos que el coeficiente de correlación no es sino una medida resumen del comportamiento de la muestra de datos, por lo que situaciones muy diferentes pueden conducir a valores similares de $r(x, y)$.

1.5.6. Covarianza y recta de regresión

Aunque posteriormente será objeto de un estudio detallado en el marco del método de máxima verosimilitud, mediante la introducción en este punto de las técnicas de regresión (lineal) podemos obtener una interpretación del significado de la covarianza entre dos variables. Supongamos a tal efecto que tenemos una muestra bidimensional de datos $\{(x_i, y_i)\}_{i=1}^N$ y que tenemos razones para suponer (o simplemente emitimos la hipótesis) que ambas variables estadísticas X e Y están relacionadas mediante una relación lineal de la forma:

$$Y = a + b X$$

En muchos casos las incertidumbres asociadas a los datos hacen que sea necesario estimar los parámetros a y b que corresponden a la recta de mejor ajuste. La adecuación de estos valores se suele caracterizar a través de alguna medida de discrepancia o función objetivo. Es habitual elegir a y b de modo que la distancia de la distribución de datos a la recta sea lo menor posible.

Tomaremos como distancia de un punto (x_i, y_i) a una recta $y = a + bx$, la distancia vertical entre el punto y la propia recta ¹³:

$$d_i = y_i - a - b x_i$$

tal como se muestra en la Fig. 1.14.

Para la obtención de la recta de mejor ajuste se suele definir una función de discrepancia $G(a, b)$ cuyo valor mínimo en $a=a_0$ $b=b_0$ se obtiene a partir de

$$\left. \frac{\partial G}{\partial a} \right|_{(a,b)=(a_0,b_0)} = 0; \quad \left. \frac{\partial G}{\partial b} \right|_{(a,b)=(a_0,b_0)} = 0 \quad (1.66)$$

corresponde a la mejor recta de regresión. Una forma de dicha función G que garantiza la existencia de un mínimo y la solución algebraica unívoca del mismo, consiste en tomar la suma de los cuadrados de las distancias verticales de los puntos de la muestra a la recta $Y = a + b X$, es decir:

$$G(a, b) = (y_1 - a - b x_1)^2 + (y_2 - a - b x_2)^2 + \dots + (y_N - a - b x_N)^2$$

o bien,

$$G(a, b) = \sum_{i=1}^N d_i^2 \quad (1.67)$$

Este método se conoce con el nombre de *método de ajuste por mínimos cuadrados*. Veremos en el capítulo 3 que puede justificarse de modo formal en

¹³ Podríamos definir la distancia de la distribución a la recta como la suma de la distancia a ésta de cada uno de sus puntos, entendida en el sentido de la geometría afín, i.e. como distancia a lo largo de una recta normal a la recta que pase por el punto en cuestión. Sin embargo, las ecuaciones que obtendríamos entonces para el método de regresión serían no lineales. Un inconveniente adicional de este método es que x e y no tienen las mismas unidades, por lo que esa hipotética distancia dependería de la escala de cada eje. Para la solución a este caso, véase el capítulo 3.

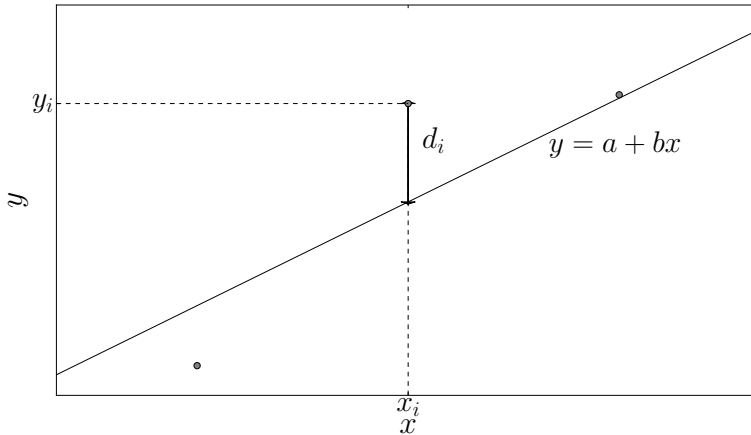


Figura 1.14: Cálculo de las distancias verticales de un punto de la muestra de datos a la recta de mejor ajuste.

términos estadísticos, ya que la minimización de $G(a, b)$ es equivalente a la maximización de la *verosimilitud*¹⁴ de haber obtenido la muestra de datos $\{(x_i, y_i)\}_{i=1}^N$ a partir de una población normal bajo a la hipótesis estadística $\bar{y}_i = a + b \bar{x}_i$.

Las Ecs. (1.66) y (1.67) nos conducen a un sistema de ecuaciones algebraicas lineales para los parámetros a_0 y b_0 :

$$\begin{aligned} \frac{\partial G}{\partial a} \Big|_{(a,b)=(a_0,b_0)} &= 2 \sum_{i=1}^N (y_i - a_0 - b_0 x_i) (-1) = 0 \\ \frac{\partial G}{\partial b} \Big|_{(a,b)=(a_0,b_0)} &= 2 \sum_{i=1}^N (y_i - a_0 - b_0 x_i) (-x_i) = 0 \end{aligned} \quad (1.68)$$

que podemos reexpresar de la forma

$$\frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} \sum_{i=1}^N a_0 + \frac{b_0}{N} \sum_{i=1}^N x_i \quad (1.69)$$

$$\frac{1}{N} \sum_{i=1}^N y_i x_i = \frac{1}{N} \sum_{i=1}^N a_0 x_i + \frac{b_0}{N} \sum_{i=1}^N x_i^2 \quad (1.70)$$

donde hemos multiplicado por $1/N$ ambos miembros de las dos ecuaciones. Utilizando la definición de promedio de una variable estadística, podemos escribir las Ecs. (1.69) y (1.70) como:

¹⁴ Probabilidad *a posteriori* de la muestra de datos si la hipótesis es cierta.

$$\bar{y} = a_0 + b_0 \bar{x} \quad (1.71)$$

$$\overline{xy} = a_0 \bar{x} + b_0 \overline{x^2} \quad (1.72)$$

Usando (1.71) podemos reescribir (1.72) de la forma

$$\overline{xy} = \bar{x}(\bar{x} - b_0 \bar{y}) + b_0 \overline{x^2} \Leftrightarrow \overline{xy} - \bar{x}\bar{y} = b_0 (\overline{x^2} - \bar{x}^2) \quad (1.73)$$

de tal manera que concluimos que los parámetros a_0 y b_0 de la recta de mejor ajuste -esto es, la que minimiza la suma de las distancias verticales al cuadrado de los puntos de la muestra de datos a la recta- son:

$$b_0 = \frac{\text{cov}(x, y)}{s_x^2} \quad (1.74)$$

$$\bar{y} = a_0 + b_0 \bar{x} \quad (1.75)$$

La primera de las ecuaciones anteriores nos proporciona una interpretación de la covarianza de las variables estadísticas X e Y relacionando ésta con la pendiente de la recta de mejor ajuste por mínimos cuadrados. Por otro lado, la ecuación (1.75) expresa el hecho de que la recta que hemos definido pasa siempre por el centro de gravedad de la muestra de datos, el punto (\bar{x}, \bar{y}) . Notemos finalmente que la aparición de s_x^2 , en el denominador de la ecuación de la pendiente de la recta de mejor ajuste, se debe a que hemos definido las distancias de los puntos de la muestra a la recta en la dirección vertical en lugar de la horizontal.

Finalmente, mencionaremos que en el caso de que nuestros experimentos aleatorios se refieran a características de más de dos dimensiones, los conceptos anteriores pueden ser fácilmente generalizados. Consideremos, por ejemplo, un experimento aleatorio en el que la variable estadística analizada es tridimensional $\vec{X} = (X, Y, Z)$, y supongamos que se ha obtenido una muestra $\{(x_i, y_i, z_i)\}_{i=1}^N$. Las medias de estas tres variables definen un vector, denominado vector de valores medios o vector de medias, de la forma

$$\vec{a} = \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} \quad (1.76)$$

donde como de costumbre

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i; \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i; \quad \bar{z} = \frac{1}{N} \sum_{i=1}^N z_i$$

De la misma manera, podremos calcular las varianzas marginales de cada una de las variables analizadas

$$s_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2; \quad s_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2; \quad s_z^2 = \frac{1}{N} \sum_{i=1}^N (z_i - \bar{z})^2 \quad (1.77)$$

y las covarianzas de los posibles pares de variables como:

$$\begin{aligned}\operatorname{cov}(x, y) &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \\ \operatorname{cov}(x, z) &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(z_i - \bar{z}) \\ \operatorname{cov}(y, z) &= \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})(z_i - \bar{z})\end{aligned}\quad (1.78)$$

Esto nos permite definir una matriz real y simétrica, denominada matriz de covarianzas, cuyos elementos diagonales son las varianzas de las diferentes variables estadísticas y los elementos del exterior de la diagonal las covarianzas:

$$M = \begin{pmatrix} s_x^2 & \operatorname{cov}(x, y) & \operatorname{cov}(x, z) \\ \operatorname{cov}(y, x) & s_y^2 & \operatorname{cov}(y, z) \\ \operatorname{cov}(z, x) & \operatorname{cov}(z, y) & s_z^2 \end{pmatrix}\quad (1.79)$$

En el caso de que nuestra variable estadística sea p -dimensional, los datos muestrales serán de la forma $\{(x_1^k, \dots, x_p^k)\}_{k=1}^N$, por lo que tendremos que las componentes del vector de medias serán

$$\bar{x}_i = \frac{1}{N} \sum_{k=1}^N x_i^k \quad i = 1, 2, \dots, p\quad (1.80)$$

y la matriz de covarianzas tendrá elementos de la forma:

$$m_{ij} = \operatorname{cov}(x_i, x_j) = \frac{1}{N} \sum_{k=1}^N (x_i^k - \bar{x}_i)(x_j^k - \bar{x}_j)\quad (1.81)$$

que definen una matriz real y simétrica ($p \times p$) de la forma:

$$M = \begin{pmatrix} s_{x_1}^2 & \operatorname{cov}(x_1, x_2) & \dots & \operatorname{cov}(x_1, x_p) \\ \operatorname{cov}(x_2, x_1) & s_{x_2}^2 & \dots & \operatorname{cov}(x_2, x_p) \\ \dots & \dots & \dots & \dots \\ \operatorname{cov}(x_p, x_1) & \operatorname{cov}(x_p, x_2) & \dots & s_{x_p}^2 \end{pmatrix}\quad (1.82)$$

Para el caso de muestras multidimensionales podemos definir una varianza generalizada s_g^2 , como una medida global de la variabilidad de la muestra, a partir del determinante de la matriz de covarianzas:

$$s_g^2 = \det(M)\quad (1.83)$$

Esta varianza generalizada está bien definida ya que $\det(M) > 0$ por construcción y su valor es una medida del volumen generalizado que ocupan los datos muestrales.

Ejemplo 1.19

En el caso de una muestra bidimensional, podemos escribir la matriz de covarianzas de la forma

$$\begin{aligned} M &= \begin{pmatrix} s_x^2 & \text{cov}(x, y) \\ \text{cov}(y, x) & s_y^2 \end{pmatrix} \\ &= \begin{pmatrix} s_x^2 & r s_x s_y \\ r s_x s_y & s_y^2 \end{pmatrix} \end{aligned}$$

por lo que $\det M = s_x^2 s_y^2 (1 - r^2)$, donde r es el coeficiente de correlación de las dos variables estadísticas analizadas. Para entender el significado del coeficiente de correlación, observemos que si $r = 0$, en virtud de la desigualdad de Tchebychev, sabemos que al menos un 89 % de los datos se encontrarán dentro de un volumen $6s_x 6s_y$, y en la situación $r \neq 1$, este volumen se verá reducido por el término $r^2 s_x^2 s_y^2$ que da cuenta del grado de agrupamiento de los datos respecto a la recta de regresión.

1.6. Cuestiones y problemas

1.1 En una clínica infantil se han ido anotando, durante un mes, el número de metros que andan sus niños, seguido y sin caerse, el primer día que comienzan a caminar. La tabla obtenida fue:

Número de niños	2	6	10	5	10	3	2	2
Número de metros	1	2	3	4	5	6	7	8

Se pide:

- Tabla de frecuencias: Diagrama de barras para frecuencias absolutas y relativas de la variable $X =$ "número de metros que el niño anda". Diagrama de frecuencias absolutas y relativas acumuladas.
 - Moda, mediana, cuartiles y deciles.
 - Rango intercuartílico. ¿Existe algún dato atípico? ¿Y algún dato atípico extremo? ¿Qué significado tendrían en el caso de que apareciesen?
 - Media aritmética, geométrica, cuadrática y armónica.
 - Analizar la dispersión de la distribución a través de la varianza, la desviación típica y el coeficiente de variación de Pearson.
 - Momentos respecto al origen de primer, segundo y tercer orden.
 - Momentos centrales (respecto a la media) de orden primero y tercero.
 - Estudiar la asimetría y la curtosis de la distribución.
- 1.2 En un determinado experimento de medición de la concentración de formación de agregados en una disolución, se obtuvieron los siguientes resultados (en mmol/l):

52	61	49	46	51	50	59	63	57	67
56	41	52	47	46	55	38	65	57	54

Calcular:

- Histograma y curva de frecuencias de la muestra de datos.
 - Medidas características de la muestra: medias (aritmética y geométrica), varianza y desviación típica, coeficiente de variación de Pearson, coeficientes de asimetría de Pearson y Fisher, y coeficiente de apuntamiento. ¿Cuál es la concentración a la que podemos decir que comienza la formación de los agregados objeto de estudio? ¿Con qué incertidumbre?
 - Analizar la existencia de datos atípicos. ¿Qué medida de centralización es más resistente a este tipo de datos?
- 1.3 En un determinado experimento de medición de la densidad de una disolución acuosa diluida a $25\text{ }^{\circ}\text{C}$ de temperatura mediante medidas de masa (m) y volumen (V), se obtuvieron los siguientes resultados:

m (g)	1,1	1,2	1,1	0,9	1,0	1,0	0,9	1,1	1,0	0,9
V (cm ³)	1,1	1,1	1,1	0,9	1,0	1,0	0,8	1,1	1,0	0,9

Se pide:

- Histograma de la densidad.
 - Medidas características de la muestra de densidades: medias (aritmética y geométrica), varianza y desviación típica, coeficiente de variación de Pearson, coeficientes de asimetría de Pearson y Fisher, y coeficiente de apuntamiento. ¿Cuál es la densidad que podemos atribuir a la disolución problema? ¿Con qué incertidumbre?
- 1.4 En un reclutamiento militar se ha tomado una muestra de dieciséis jóvenes, obteniéndose las siguientes estaturas (en cm.):

160,0	172,4	168,0	167,0	175,0	179,0	180,0	198,0
164,0	166,0	174,0	177,0	182,5	185,0	191,0	173,5

- ¿Qué tipo de variable aleatoria estamos analizando? ¿Cuál es el número óptimo de clases? Agrupar los datos en intervalos de amplitud constante.
 - Media aritmética, media geométrica y armónica.
 - Mediana y desviación media respecto a la mediana.
 - Coeficientes de variación media.
 - Coeficiente de variación de Pearson.
- 1.5 En un experimento de medida de la masa atómica de un determinado elemento químico se han obtenido los siguientes valores (g/mol):

134,56	134,89	133,99	133,56	135,03	134,65	135,10	137,20
134,25	134,78	134,29	133,62	135,23	134,99	135,56	136,65
134,34	134,32	134,05	134,78	133,25	133,01	132,76	132,85
134,66	134,89	134,20	133,82	135,67	136,02	133,78	133,98

- a) Histogramar los datos de frecuencias absoluta y relativa de la variable aleatoria $X =$ "masa atómica del elemento". ¿Cuántas y qué clases se han elegido?
- b) ¿Qué valor podemos asignar a la masa atómica del elemento?
- c) Si tomamos como medida de la incertidumbre experimental la desviación típica de la distribución (incertidumbre estadística o tipo A), ¿qué valor toma dicha incertidumbre? Usualmente, sin embargo, se prefiere como medida de esta incertidumbre estadística experimental la denominada desviación típica de la media, cuyo valor es:

$$s_A(\bar{x}) = \frac{s(x)}{\sqrt{N}}$$

donde N es el número de datos. Calcular dicho valor y comentar brevemente la definición anterior.

- d) Analizar la existencia de datos atípicos y atípicos extremos y posibles causa de los mismos.
- e) Obtener los momentos de primer, segundo, tercer y cuarto orden, respecto a la media y respecto al origen.
- f) Analizar la asimetría de la distribución, así como el apuntamiento de la misma.
- 1.6 En un determinado experimento de medición de la densidad de una disolución acuosa diluida a 25 °C de temperatura mediante medidas de masa (m) y volumen (V), se obtuvieron los siguientes resultados:

m (g)	1,1	1,2	1,1	0,9	1,0	1,0	0,9	1,1	1,0	0,9
V (cm ³)	1,1	1,1	1,1	0,9	1,0	1,0	0,8	1,1	1,0	0,9

Se pide:

- a) Histograma de la densidad.
- b) Medidas características de la muestra de densidades: medias (aritmética y geométrica), varianza y desviación típica, coeficiente de variación de Pearson, coeficientes de asimetría de Pearson y Fisher, y coeficiente de apuntamiento. ¿Cuál es la densidad que podemos atribuir a la disolución problema? ¿Con qué incertidumbre?
- 1.7 Demuéstrese que si construimos una variable z mezclando n_1 valores de la variable x y n_2 de la variable y , la media de z es:

$$\bar{z} = \frac{n_1}{n_1 + n_2} \bar{x} + \frac{n_2}{n_1 + n_2} \bar{y}$$

1.8 Consideremos la distribución de datos:

x	1	2	3	4	5	Obtégase la covarianza de las variables aleatorias anteriores.
y	0.34	0.70	1.08	1.43	1.70	

1.9 Considérense dos variables aleatorias x e y con varianzas iguales. Calcúlese la covarianza siguiente: $\text{cov}[(x + y), (x - y)]$.

1.10 Probar que si $q = q(x, y)$ es una función cualquiera de las variables x e y , y las desviaciones de la media de ambas variables son pequeñas, entonces se verifica:

$$s_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 s_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 s_y^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\text{cov}(x, y)$$

donde las derivadas deben entenderse evaluadas en los valores medios de x e y , y $\text{cov}(x, y)$ es la covarianza de las variables aleatorias x e y definida de la forma:

$$\text{cov}(x, y) = \sum_{i=1}^k f_i(x_i - \bar{x})(y_i - \bar{y})$$

1.11 Considérese la función $y = kx$ de la variable aleatoria (siendo k un entero positivo). Calcular la varianza y la desviación típica de y en función de la varianza y la desviación típica de x , s_x^2 y s_x respectivamente, así como la covarianza siguiente: $\text{cov}[(x + y), (x - y)]$. Consideremos ahora el caso de la variable aleatoria $y = \sum_{i=1}^k x_i$ donde $\{x_i\}_{i=1}^k$ es un conjunto de variables aleatorias independientes e idénticamente distribuidas, todas ellas con varianza s_x^2 . Repítanse los cálculos anteriores y compárense los resultados obtenidos.

1.12 Demostrar que si dos variables aleatorias están relacionadas de modo lineal $Y = a + bX$ entonces el coeficiente de correlación $r(x, y) = \pm 1$ (cuyo signo está asociado al signo de b). (N.B. Obsérvese que ya hemos visto que $s_x = b \cdot s_y$).

1.13 Considérense los resultados del experimento de medición de la masa atómica del elemento del ejercicio 3. Histogramar las distribuciones correspondientes a las siguientes variables aleatorias:

i) $y_i = x_i^2$; ii) $y_i = \ln x_i$ iii) $y_i = \sqrt{x_i}$

Calcular las medidas características de las anteriores distribuciones de manera exacta. Comparar la media y la desviación típica obtenidas con las resultantes de aplicar las aproximaciones de las transformaciones en torno a la media.

1.14 Dada la variable bidimensional $\vec{X} = (X, Y)$ con la tabla de frecuencias:

$X \backslash Y$	1	2	4	6
1	2	0	0	1
3	3	1	0	1
5	0	1	0	5

Se pide:

i) Calcular $\sum_i \sum_j n_{ij}$

ii) Frecuencias relativas, f_{ij} .

iii) Las distribuciones marginales n_{xi} y n_{yj} . Obtener las medias de las variables X e Y .

iii) Obtener: a) los momentos respecto al origen $M_{1,0}(0,0)$ y $M_{0,1}(0,0)$, b) el momento respecto a la media (central) $M_{1,1}(\bar{x}, \bar{y})$. Obsérvese que este momento corresponde a la covarianza de la distribución:

$$cov(x, y) = \sum_{i=1}^k \sum_{j=i}^l f_{ij}(x_i - \bar{x})(y_j - \bar{y})$$

Demostrar que la definición anterior es equivalente a la dada en el ejercicio 9.

- 1.15 En una determinada región vitivinícola se observaron durante algunos años el precio del kilogramo de uva y la cantidad de producción, obteniéndose la siguiente tabla:

x	21	19	29	36	31	29	37	31	33	35
y	100	140	120	110	200	200	110	160	160	200

donde x es el precio del kg en céntimos de euro e y la cantidad producida en miles de kg. Considerando la variable x agrupada en intervalos de amplitud constante y considerando el primero de ellos 15-25 se pide:

- Tabla bidimensional.
- Distribuciones marginales
- Distribución de x condicionada a la modalidad $y = 200$.
- Media, mediana, moda y cuartiles de x .
- Media, mediana, moda y cuartiles de y .
- Recorrido, desviación típica, desviación media (respecto a la media) y coeficiente de variación de Pearson de x e y .
- ¿Es simétrica la x ? ¿Y la y ?
- Curtosis de ambas distribuciones.
- ¿Qué media es menos representativa?
- Porcentaje de años en los que el precio del kg de uva fue inferior a 27 céntimos de euro.
- Rectas de regresión y coeficiente de correlación. Matriz de covarianzas.

Capítulo 2

Teoría de probabilidades

2.1. Introducción a la teoría de probabilidades

Como hemos visto en el capítulo anterior, nuestro conocimiento del mundo exterior se basa en la realización de lo que se denominan *experimentos aleatorios*, mediante los cuales obtenemos información acerca del comportamiento de características de los objetos que se presentan como *sucesos aleatorios* y que cuantificamos mediante la asignación a éstos de una *variable aleatoria*, concepto que definiremos formalmente en este capítulo. Estos experimentos aleatorios abarcan un elevado número de situaciones tales como la toma de datos por observación empírica en sistemas o sociales o la realización de mediciones experimentales en sistemas físicos. Como hemos visto, este tipo de experimentos aleatorios implica que todos los posibles resultados del experimento son conocidos con anterioridad a la realización del mismo, que no se puede predecir *a priori* el resultado de un experimento aleatorio y que el experimento puede repetirse en condiciones idénticas.

Ejemplos habituales de un experimento aleatorio son el lanzamiento de una moneda o el número resultante en la tirada de una ruleta. Es menos habitual incluir dentro de la categoría de experimento aleatorio la medida de una magnitud física, pero la variabilidad intrínseca del resultado experimental de este proceso hace que debamos considerarla como tal.

El conjunto de todos los posibles resultados que puede arrojar un experimento aleatorio, constituye lo que denominamos *espacio muestral*, Ω , un concepto debido a von Mises. El *espacio muestral* puede ser un conjunto finito o infinito de resultados según el experimento aleatorio que consideremos.

Ejemplo 2.1

-
1. Lanzamiento de una moneda dos veces

$$\Omega = \{cc, cx, xc, xx\} \tag{2.1}$$

$c = \text{cara}$ $x = \text{cruz}$

2. Lanzamiento único de dos monedas al aire.

$$\Omega = \{cc, cx, xx\} \quad (2.2)$$

3. Medida de la energía de un sistema físico en Mecánica Clásica: un sistema clásico puede tener cualquier energía posible, por lo que el espacio muestral de esta variable aleatoria será:

$$\Omega = [0, \infty)$$

4. A diferencia del ejemplo anterior, en Mecánica Cuántica los niveles de energía de sistemas ligados toman valores que están cuantizados. Por ejemplo, en el caso de un oscilador armónico, tendremos

$$\Omega = \left\{ \frac{1}{2}\hbar\omega, \frac{3}{2}\hbar\omega, \frac{5}{2}\hbar\omega, \dots \right\}$$

donde ω es la frecuencia propia del oscilador y \hbar es la constante de Planck dividida entre 2π .

5. Microestados de un sistema físico macroscópico. Todas las posibles situaciones microscópicas de un sistema compatibles con su situación macroscópica forman lo que se denomina la colectividad de microestados.

$$\Gamma = \{1 \mid 1 \text{ microestado del sistema}\}$$

Diremos que el espacio muestral Ω es *discreto* cuando se pueda poner en correspondencia uno a uno con el conjunto de los números enteros positivos o algún subconjunto suyo. El espacio muestral Ω se denominará *continuo* cuando se pueda establecer una aplicación biyectiva con uno o más intervalos de la recta real. Así en el ejemplo anterior los casos 1,2 y 4 corresponden a espacios muestrales discretos, mientras que el caso 3 es un ejemplo de espacio muestral continuo.

Dado un espacio muestral Ω , consideraremos que un *suceso aleatorio* A es cualquier subconjunto del espacio muestral Ω ¹. Un *suceso aleatorio* puede contener a uno o varios elementos del *espacio muestral*. En el caso de que contenga únicamente un elemento hablaremos de suceso aleatorio simple, y si más de un elemento está contenido en el suceso, lo denominaremos compuesto. Sobre un espacio muestral podemos definir diferentes sucesos aleatorios compuestos A , B , ... los cuales pueden tener elementos comunes entre sí.

¹ A veces se usa de modo indistinto la denominación de *suceso*, *evento* o *propiedad aleatoria*.

Ejemplo 2.2

1. Lanzamiento de una moneda dos veces:

$$\Omega = \{cc, cx, xc, xx\}. \quad (2.3)$$

Dependiendo de nuestros intereses, podríamos considerar los sucesos compuestos:

$A = \{cc, cx, xc\}$ Sucesos con al menos una cara.

$B = \{cx, xc, xx\}$ Sucesos con al menos una cruz.

$C = \{cc, xc\}$ Sucesos en los que el segundo lanzamiento es cara.

o el suceso simple $D = \{cc\}$, suceso con dos caras.

2. Niveles de energía de un oscilador armónico unidimensional en Mecánica Cuántica:

$$\Omega = \left\{ \frac{1}{2}\hbar\omega, \frac{3}{2}\hbar\omega, \frac{5}{2}\hbar\omega, \dots \right\} \quad (2.4)$$

$A = \left\{ \frac{1}{2}\hbar\omega \right\}$ Estado fundamental.

$B = \left\{ \frac{3}{2}\hbar\omega, \frac{5}{2}\hbar\omega, \dots \right\}$ Estados excitados .

El problema central de la teoría de probabilidades es la asignación a cada subconjunto A del espacio muestral de una cantidad que mida su tendencia a producirse cuando se realiza un determinado experimento aleatorio. Esta cantidad es la que se denomina probabilidad del suceso A , $p(A)$.

La interpretación de la *probabilidad* dista mucho de ser un asunto sencillo, pues como veremos son múltiples las interpretaciones de la misma incluso dentro del marco estricto de la teoría de probabilidades. Pero es que además, el término probabilidad se utiliza vulgarmente con un elevado número de acepciones que difieren de su significado formal, lo que genera aún mas confusión acerca del significado concreto de ésta.

Las interpretaciones principales del término probabilidad se clasifican convencionalmente en dos grandes categorías: interpretaciones objetivas y subjetivas. Las primeras son aquellas que derivan de la frecuencia asociada al suceso A en un determinado experimento aleatorio realizado sobre una muestra de los objetos que presentan la característica asociada objeto de estudio. Por el contrario, las interpretaciones subjetivas son aquellas que interpretan la probabilidad como una medida de nuestra información o certidumbre acerca de la ocurrencia del suceso aleatorio observado. Entre las primeras podemos distinguir principalmente dos interpretaciones ²:

² Karl R. Popper en su libro Teoría cuántica y el cisma en física (Popper, 1985), introduce una tercera interpretación, que el propio autor responsable de su formulación denomina propensivista y que considera como un "refinamiento de la interpretación clásica". Así, la probabilidad sería una medida de la propensión o tendencia de un suceso a producirse con la realización de un determinado experimento aleatorio.

- a) *Interpretación clásica* (Hume, Bernouilli, de Moivre, Laplace) según la cual la probabilidad se obtiene como razón entre el número de casos en los que se presenta la propiedad observada sobre el total de casos posibles. Esta interpretación conecta con el denominado principio de razón insuficiente (PRI). Efectivamente, en el caso de que tengamos sucesos elementales cuya naturaleza no permita asignarles probabilidades diferentes (e.g. lanzamiento de dados o monedas, extracción de un naipe), la probabilidad de cada uno de ellos debe ser:

$$p = \frac{1}{\mathcal{N}} \quad \text{Principio de razón insuficiente (PRI)} \quad (2.5)$$

donde \mathcal{N} es el número total de sucesos elementales posibles. Un suceso compuesto A , formado a partir de \mathcal{F} sucesos elementales tendrá una probabilidad asociada:

$$p(A) = \frac{\mathcal{F}}{\mathcal{N}} \quad (2.6)$$

enunciado que constituye la denominada regla de Laplace:

$$p(A) = \frac{\text{casos favorables } (\mathcal{F})}{\text{casos posibles } (\mathcal{N})} \quad (2.7)$$

Aunque se ha sugerido que esta era la definición de probabilidad de Laplace, se deduce claramente de sus escritos que era consciente de sus limitaciones. El principio de razón insuficiente -sobre el que se asienta en última instancia la regla de Laplace- se puede encontrar ya en Bernouilli y en Hume.

- b) *Interpretación frecuencial* (John Venn, George Helm, Richard von Mises) o estadística, según la cual la probabilidad de un suceso puede obtenerse a partir de la frecuencia de observación del mismo en experimentos aleatorios realizados sobre muestras.

Por lo que respecta a la *interpretación subjetiva (o bayesiana)* es aquella en la que la probabilidad mide el grado de certeza del observador acerca del acaecimiento de un determinado suceso, esto es, el grado de creencia acerca de que la reiteración de un experimento aleatorio que mida la ocurrencia o no de un determinado suceso producirá una frecuencia igual a la probabilidad que evaluamos en base a estimaciones. La culminación de este enfoque se encuentra en el teorema de Bayes y en el principio de máxima entropía de Jaynes, que analizaremos posteriormente.

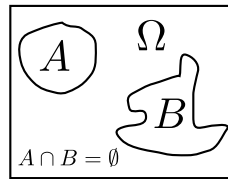
Cuando realizamos un experimento aleatorio, las frecuencias relativas de un determinado suceso A del espacio muestral tienden a estabilizarse alrededor de un número perteneciente al intervalo $[0, 1]$. Sea n el número de observaciones del suceso y N el número total de observaciones. La frecuencia relativa $f(A) = n/N$ del suceso A tiene un límite bien definido que denominamos probabilidad de ocurrencia del suceso A , $p(A)$:

$$\lim_{N \rightarrow \infty} f(A) = \lim_{N \rightarrow \infty} \frac{n}{N} = p(A) = \text{cte.} \quad (2.8)$$

La relación anterior constituye la formalización de lo que hemos denominado interpretación frecuencial de la probabilidad. Como vemos, mientras la frecuencia es una cantidad asociada a una determinada muestra de tamaño finito, la probabilidad se asocia a la población.

A continuación presentamos las propiedades de la probabilidad anteriormente definida. En ocasiones, estas propiedades se toman como una definición axiomática de la probabilidad misma (Kolmogorov, 1933)³

- a) La probabilidad de un determinado suceso $A \subset \Omega$ verifica $0 \leq p(A) \leq 1$, siendo $p(A) = 0$ la probabilidad del suceso imposible y $p(A) = 1$ la del suceso seguro.
- b) $p(\Omega) = 1$. La probabilidad de todo el espacio muestral es igual a 1, lo que significa que tendremos siempre algún resultado del experimento que pertenezca al espacio muestral.
- c) Si A y B son sucesos mutuamente excluyentes, $p(A \cup B) = p(A) + p(B)$.



A partir de los axiomas anteriores podemos obtener fácilmente los siguientes resultados:

³ Aunque en esta obra no pretendemos presentar el concepto de probabilidad con todo el rigor que permite la teoría de los espacios de medida, no podemos dejar de mencionar el concepto de *espacio de probabilidad*. En efecto, se denomina espacio de probabilidad a la terna formada por el espacio muestral, Ω , el conjunto de todos sus subconjuntos medibles, \mathcal{F} y la medida definida por

$$\begin{aligned} P : \Omega &\rightarrow R \\ A &\rightarrow P(A) \end{aligned} \quad (2.9)$$

donde $\sum_{A \in \Omega} P(A) = 1$. Este es un caso particular de espacio de medida donde la medida total del espacio es la unidad. Recordamos en este punto al lector que un espacio de medida o espacio medible es una terna (Ω, Σ, μ) tal que:

A) Σ es una σ -álgebra sobre Ω , i.e. Σ es un conjunto subconjuntos de X que tiene las propiedades siguientes:

A.1) $X \in \Sigma$

A.2) $A \in \Sigma \Leftrightarrow \bar{A} \in \Sigma$

A.3) Σ es cerrado bajo uniones contables,

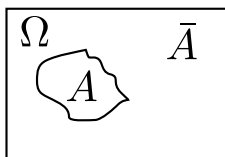
$$\{A_i\}_{i=1}^n \subset \Sigma \Rightarrow \bigcup_{i=1}^n A_i \in \Sigma \quad (2.10)$$

B) μ es una medida en Σ , i.e. una función que asigna un valor real a cada uno de los elementos de la σ -álgebra Σ :

$$\begin{aligned} \mu : \Sigma &\rightarrow \mathbb{R} \\ \sigma &\rightarrow \mu(\sigma) \end{aligned} \quad (2.11)$$

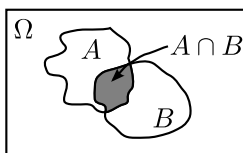
Para profundizar en esta materia referimos al lector a alguno de los múltiples manuales especializados disponibles.

i) $p(A \cup \bar{A}) = 1$, donde $\bar{A} \equiv \Omega \setminus A$ es el suceso complementario del suceso A



ii) $p(\bigcup_{i=1}^s A_i) = \sum_{i=1}^s p(A_i)$ si $A_i \cap A_k = \emptyset \ \forall i \neq k$

iii) Si A y B no son mutuamente excluyentes, i.e. si $A \cap B \neq \emptyset$, entonces:
 $p(A \cup B) = p(A) + p(B) - p(A \cap B)$



Veamos una prueba de la última propiedad enunciada. Sean A y B sucesos no mutuamente excluyentes, y llamemos n_{AB} , $n_{A\bar{B}}$ y $n_{\bar{A}B}$ al número de veces que aparecen A y B , A y no B , B y no A respectivamente. Entonces, se verifica:

$$\begin{cases} n_A &= n_{AB} + n_{A\bar{B}} \\ n_B &= n_{AB} + n_{\bar{A}B} \\ n_{A \cup B} &= n_{AB} + n_{A\bar{B}} + n_{\bar{A}B} \end{cases} \quad (2.12)$$

Obviamente tendremos:

$$n_{A \cup B} = n_A + n_B - n_{AB} \quad (2.13)$$

Dividiendo por el número total de observaciones:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \quad (2.14)$$

2.2. Probabilidad condicionada

Consideremos un experimento cuyos resultados pueden ser los sucesos A y B , no mutuamente excluyentes. Como hemos visto en el capítulo anterior la frecuencia relativa de B condicionada a la ocurrencia de A se define considerando únicamente los casos en los que aparece A , y viendo en cuántos de esos casos ocurre también el suceso B . Así, dicha frecuencia será el cociente del número de veces que han ocurrido A y B conjuntamente dividido por el número de veces que se observa la propiedad A

$$f(B|A) = \frac{n_{AB}}{n_A} = \frac{n_{AB}/N}{n_A/N} = \frac{f(A \cap B)}{f(A)} \quad (2.15)$$

Podemos plantear la cuestión de la siguiente forma. Si un experimento tiene como resultado A , ¿cuál es la probabilidad de que también tenga la propiedad

B ? Esto es, la probabilidad B bajo la condición A . A partir de la ecuación anterior podemos definir esta probabilidad como

$$p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{p(AB)}{p(A)} \quad (2.16)$$

que es la probabilidad de un suceso B condicionado a la ocurrencia de otro A .

Es muy importante diferenciar entre la probabilidad condicionada anterior $p(B|A)$ y la probabilidad de ocurrencia conjunta de los sucesos A y B (i.e. $p(A \cap B)$)⁴. Esta última indica la probabilidad de que A y B sucedan de manera simultánea, y es siempre menor que la probabilidad de que ocurran A ó B (i.e. $p(A \cup B)$). Sin embargo, $p(B|A)$ indica la probabilidad de que suceda B en los casos en los que ya ha ocurrido A , y puede ser mayor o menor que $p(A)$ o $p(B)$. En el primer caso nos movemos en el espacio muestral original completo Ω , mientras que en el segundo únicamente dentro del subespacio muestral de sucesos con la propiedad A .

Sucesos estadísticamente independientes:

Diremos que dos sucesos son estadísticamente independientes cuando la ocurrencia de uno no modifica la probabilidad de aparición del otro. Por tanto, A y B serán independientes si:

$$p(A|B) = p(A) \quad p(B|A) = p(B) \quad (2.17)$$

La condición anterior puede reexpresarse en una forma más conocida si usamos la definición de probabilidad condicionada:

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = p(A) \Rightarrow p(A \cap B) = p(A)p(B) \quad (2.18)$$

i.e. dos sucesos son estadísticamente independientes cuando la probabilidad de ocurrencia conjunta de ambos es igual al producto de las probabilidades individuales.

Ejemplo 2.3

En una encuesta electoral ante las elecciones generales en un cierto país el 80 % de la población suscribe opciones políticas consideradas de centro (40 % a la opción CD y 40 % a la opción CI , aproximadamente), mientras que el 20 % restante apoya opciones políticas diferentes (O). Si el porcentaje de mujeres en este país es del 51 %, el 47 % de las mujeres se declara de la opción CI y de los hombres encuestados en edad de votar un 52 % se declaran partidarios de la opción CD , ¿cuál es el porcentaje de mujeres votantes del centro derecha (CD)? ¿Y cuál la probabilidad de que un hombre sea votante de otras opciones políticas (O)?

Solución:

⁴Denotaremos indistintamente como equivalentes $A \cap B$ y AB .

CDH	CIH	MO
CDM	CIM	HO

Obviamente el espacio muestral del problema anterior es $\Omega = \{MCD, MCI, HCD, HCI, MO, HO\}$. Podemos representar dicho espacio muestral como un cuadrado de lado unidad, por lo que las probabilidades son las áreas de los rectángulos (ver figura). A partir de la definición de probabilidad condicionada, la probabilidad de que una mujer sea de CD la calcularemos como:

$$p(CD|M) = \frac{p(MCD)}{p(M)}$$

Teniendo en cuenta que, en virtud de la definición de probabilidad marginal, $p(CD) = p(MCD) + p(HCD)$, podemos escribir

$$\begin{aligned} p(MCD) &= p(CD) - p(HCD) \\ &= p(CD) - p(CD|H)p(H) \\ &= 0,40 - 0,52 \cdot 0,49 = 0,145 \end{aligned}$$

con lo que la probabilidad condicionada buscada es

$$\begin{aligned} p(CD|M) &= \frac{p(MCD)}{p(M)} = \frac{p(CD) - p(HCD)}{p(M)} \\ &= \frac{0,145}{0,4} = 0,363 \end{aligned}$$

Por otro lado, la probabilidad de que un hombre sea votante de otras opciones políticas (O) es

$$\begin{aligned} p(O|H) &= 1 - p(CD|H) - p(CI|H) \\ &= 1 - 0,52 - p(CI|H)p(H) \\ &= 0,48 - p(H) [1 - p(CIM)] \\ &= 0,48 - 0,49 [1 - p(CI|M)p(M)] \\ &= 0,48 - 0,49 (1 - 0,47 \cdot 0,51) \\ &= 0,11 \end{aligned}$$

2.2.1. Teorema de Bayes

Mediante el teorema de Bayes podemos construir probabilidades de sucesos sobre los que no disponemos de observaciones empíricas o experimentales a partir de las probabilidades conocidas de sucesos relacionados. Consideremos un experimento que se realiza en dos etapas. En la primera, los sucesos posibles son A_1, A_2, \dots, A_n , mutuamente excluyentes ($A_i \cap A_j = \emptyset$), con probabilidades conocidas, $p(A_i)$, de manera que:

$$\sum_{i=1}^n p(A_i) = 1 \quad (2.19)$$

En la segunda etapa, los resultados posibles, B_j con $j = 1, \dots, m$, dependen de los resultados de la primera, y se conocen las probabilidades condicionadas $p(B_j|A_i)$ de obtener cada B_j cuando en la primera fase aparece A_i . Supongamos que se realiza el experimento, pero no se conoce el resultado de la primera fase, aunque sí el de la segunda. El teorema de Bayes permite calcular las probabilidades $p(A_i|B_j)$ de los sucesos no observados en la primera etapa, dado el resultado observado en la segunda⁵. En concreto, el teorema de Bayes establece que la probabilidad condicionada de que haya aparecido en la primera fase A_i sabiendo que se ha realizado B_j en la segunda es:

$$p(A_i|B_j) = \frac{p(B_j|A_i)p(A_i)}{\sum_{k=1}^n p(B_j|A_k)p(A_k)} \quad (2.20)$$

Veamos la demostración de la relación anterior. Partiendo de la definición de probabilidad condicionada,

$$p(A_i|B_j) = \frac{p(A_i B_j)}{p(B_j)} = \frac{p(B_j|A_i)p(A_i)}{p(B_j)} \quad (2.21)$$

y usando que la probabilidad de obtener un determinado B_j es la suma de las probabilidades de que se presente acompañado de cualquiera de los A_i 's posibles (distribución marginal) obtenemos:

$$p(B_j) = \sum_{k=1}^n p(B_j A_k) = \sum_{k=1}^n p(B_j|A_k)p(A_k). \quad (2.22)$$

De la combinación de las dos ecuaciones anteriores, se obtiene de manera inmediata la relación buscada:

$$p(A_i|B_j) = \frac{p(B_j|A_i)p(A_i)}{\sum_{k=1}^n p(B_j|A_k)p(A_k)} \quad (2.23)$$

⁵Es por tanto un método para la construcción subjetiva de nuevas distribuciones de probabilidad a partir de información conocida.

Ejemplo 2.4

En una empresa se producen piezas mediante dos máquinas diferentes. La primera de ellas genera una pieza defectuosa cada 150 unidades producidas, mientras la segunda disminuye dicha tasa de fallo hasta una pieza cada 1.000 unidades. Si tomamos una muestra al azar de diez piezas de la producción total de la empresa, y se registran dos piezas defectuosas, ¿cuál es la probabilidad de que procedan de la máquina 1?

Evidentemente, la probabilidad de elegir cada una de las máquinas es la misma, $p(M_1) = p(M_2) = 1/2$. La probabilidad de producir una pieza defectuosa en la máquina 1 es $p(d|M_1) = 1/150$, mientras que la de producirlo en la 2 es $p(d|M_2) = 1/1000$. Por otro lado, dado que una muestra de diez piezas puede verse como un suceso compuesto D formado por diez sucesos elementales, 2 piezas defectuosas y diez correctas ($D = ddccccccc$) estadísticamente independientes, podemos escribir que

$$p(D|M_1) = p^2(d|M_1)p^8(c|M_1) = \left(\frac{1}{150}\right)^2 \left(\frac{149}{150}\right)^8 = 4,2 \times 10^{-5}$$

$$p(D|M_2) = p^2(d|M_2)p^8(c|M_2) = \left(\frac{1}{1000}\right)^2 \left(\frac{999}{1000}\right)^8 = 9,9 \times 10^{-7}$$

Finalmente, aplicando el teorema de Bayes obtenemos

$$p(M_1|D) = \frac{p(D|M_1)p(M_1)}{p(D|M_1)p(M_1) + p(D|M_2)p(M_2)} = 0,977$$

Ejemplo 2.5

Se ha comprobado en un cierto estudio clínico que el 75 % de los fumadores (F) tiene una determinada forma de cáncer de pulmón, mientras que de los no fumadores (NF) únicamente un 5 % manifiesta síntomas de dicha enfermedad. Si el porcentaje de fumadores en la población es del 31 %, ¿cuál es la probabilidad de que un paciente con cáncer de pulmón sea fumador?

Solución:

$$C = \text{ "cáncer de pulmón"}$$

$$p(C|F) = 0,75 \quad ; \quad p(C|NF) = 0,05$$

Por el teorema de Bayes, tendremos

$$p(F|C) = \frac{p(C|F)p(F)}{p(C|F)p(F) + p(C|NF)p(NF)}$$

$$= \frac{0,75,0,31}{0,75,0,31 + 0,05,0,69} = 0,87$$

2.2.2. Entropía estadística. Principio de máxima entropía (PME).

Hasta el momento presente hemos venido construyendo las distribuciones de probabilidad a partir de datos suministrados por el muestreo exhaustivo de una determinada muestra de la población. Sin embargo, en la realidad normalmente trabajamos con muestras más o menos amplias, pero no con la población completa; o no disponemos de distribuciones de probabilidad con las que construir nuevas distribuciones, con lo cual hemos de hacer hipótesis sobre la asignación de probabilidades a sucesos. En muchas ocasiones, la información disponible procedente de las observaciones dista mucho de ser suficiente para determinar una distribución única, existiendo diferentes distribuciones capaces de reproducir los hechos conocidos. Surge entonces de manera inevitable la pregunta sobre cuál es la distribución de probabilidad más adecuada para el problema entre todas las que verifican las restricciones impuestas por la información conocida.

Esta cuestión tiene mucho que ver con la incertidumbre del proceso de asignación de probabilidades. Tenemos que buscar una medida que nos permita disminuir la incertidumbre de este proceso y conseguir la asignación óptima compatible con el nivel de información del que disponemos acerca del fenómeno en cuestión. Shannon (1948) demostró que una buena medida de la información no disponible en la descripción de un fenómeno estadístico en el que tenemos $\Omega = \{A_i\}_{i=1}^n$ sucesos posibles, y cuya distribución de probabilidad es $p_i = p(A_i)$, es la denominada *entropía* de la distribución de probabilidad⁶:

$$S(p_1 \dots p_n) = -k \sum_{i=1}^n p_i \ln p_i \quad k > 0, \quad (2.24)$$

donde la suma se extiende a todos los elementos del espacio muestral. S está definida positivamente puesto que $p_i \leq 1, \forall A_i$. Notemos también que, en ausencia de otros condicionantes, la información no disponible acerca del problema, S , alcanza su valor máximo cuando las diferentes probabilidades son iguales:

$$p_i = \frac{1}{n}, \forall A_i \in E \Rightarrow S = -k \sum_{i=1}^n \frac{1}{n} \ln \left(\frac{1}{n} \right) \Rightarrow S = k \ln n = S_{max} \quad (2.25)$$

i.e. la entropía es máxima cuando la incertidumbre es máxima y únicamente podemos suponer que se verifica el principio de igualdad de probabilidades a priori o de razón insuficiente. Como sabemos, esto corresponde a una situación de mínima información acerca de la distribución de probabilidad concreta.

El criterio de asignación de probabilidades lo proporciona el denominado principio de entropía máxima (PME) que establece que la distribución de probabilidad menos sesgada que podemos asignar a un fenómeno aleatorio es aquella que maximiza la entropía sometida a las restricciones de la información conocida (Jaynes, 1957). En otras palabras, la distribución de probabilidad

⁶Este funcional coincide con la expresión de Boltzmann y Gibbs de la Termodinámica y la Mecánica Estadística.

óptima es aquélla que maximiza nuestro desconocimiento acerca del fenómeno en cuestión.

Ejemplo 2.6

Asignación de probabilidad en un proceso electoral. Supongamos que ante un determinado proceso político queremos elaborar un modelo de probabilidad que permita el análisis del comportamiento electoral de la población. Consideremos únicamente dos grupos políticos (D, I). Se realiza una encuesta entre un grupo nutrido de votantes y se concluye lo siguiente:

La probabilidad de que la clase de mayor renta opte por la opción D es 0.95, i.e. el 95 % de los votantes de esta clase (a) optan por dicha opción electoral.

El porcentaje de clase baja que opta por la opción I es del 95 % (b).

Según datos que podemos obtener de la Administración, el 5 % de los ciudadanos con derecho a voto tiene una renta que nos permite clasificarlo en la clase de mayor renta, un 60 % en la clase media y un 35 % en la clase de menores ingresos.

No existe ningún porcentaje fiable de expectativa de voto entre los votantes de clase media (m), ya que depende, entre otras cuestiones, de su renta, factor que normalmente ocultan en las encuestas. En democracias avanzadas los votantes de este grupo suelen ser oscilante en su apoyo a una opción u otra.

Podemos considerar que los datos proporcionados para (a) y (b) son altamente fiables, y tenemos una importante incertidumbre en lo referente a (m) ¿Cuál es la mejor asignación de probabilidad?

Sol: El espacio muestral está formado por los elementos Da, Db, Dm, Ia, Ib e Im . Tomando como siempre un cuadrado de lado unidad, las áreas de los rectángulos correspondientes a las diferentes clases nos da las probabilidades de las mismas.

De los datos disponibles, podemos calcular la probabilidad de las siguientes clases:

$$p(Da) = 0,05 \times 0,95 = 0,0475$$

$$p(Ia) = 0,05 \times 0,05 = 0,0025$$

$$p(Bb) = 0,05 \times 0,35 = 0,0175$$

$$p(Ib) = 0,95 \times 0,35 = 0,3325$$

Evidentemente, y debido a la carencia de información razonable sobre la tendencia de voto de la clase media, tenemos que emitir una hipótesis sobre esta. Consideremos que una fracción α de la clase media que vota la opción D . Así, teniendo en cuenta la información conocida acerca de la normalización de la distribución de probabilidad, $\sum_i p_i = 1$, tenemos:

$$p(mD)(\alpha) = 0,6 \alpha$$

$$p(mI)(\alpha) = 0,6 (1 - \alpha)$$

Se trata de obtener ahora, con la información disponible, el parámetro α que optimiza la distribución de probabilidad, y, por tanto, que nos permite una mejor modelización del voto. Construimos el funcional entropía:

$$S = -k \sum_i p_i \ln p_i = S_0 - k [p(mD) \ln p(mD) + p(mI) \ln p(mI)] \quad (2.26)$$

donde $S_0 = -k[0,0475 \ln 0,0475 + 0,0025 \ln 0,0025 + \dots]$. Evidentemente, la entropía es una función del parámetro α :

$$S(\alpha) = S_0 - k \{0,6\alpha \ln(0,6\alpha) + 0,6(1 - \alpha) \ln[0,6(1 - \alpha)]\} \quad (2.27)$$

Maximizando la expresión anterior tenemos:

$$\frac{dS(\alpha)}{d\alpha} = -k[0,6 \ln(0,6\alpha) + 0,6 - 0,6 \ln[0,6(1 - \alpha)] - 0,6] = 0 \quad (2.28)$$

lo que nos conduce a:

$$\alpha = 1 - \alpha \Rightarrow \alpha = \frac{1}{2} \quad (2.29)$$

Luego, la asignación óptima de probabilidad en las condiciones informativas en las que estamos es:

$$\begin{aligned} p(mD) &= 0,3 \\ p(mI) &= 0,3 \end{aligned}$$

situación que corresponde al PRI o principio de igualdad de probabilidades a priori que, como sabemos, implica que tenemos un máximo de desinformación acerca de la situación concreta.

Esto mismo sucede en un sistema físico aislado en equilibrio, donde todos los microestados (sucesos) son igualmente probables (1^{er} postulado de la Mecánica Estadística) ya que, debido a su aislamiento, no tenemos información alguna sobre el sistema salvo la propia condición de normalización de la distribución de probabilidad.

En general, la implementación práctica del principio de entropía máxima de Jaynes implica la maximización de una función, la entropía, sometida a las restricciones asociadas a la información conocida, lo que supone la utilización del formalismo de multiplicadores indeterminados de Lagrange.

Ejemplo 2.7

Consideremos un fenómeno aleatorio cuyo espacio muestral es $\Omega = \{A_1 \dots A_n\}$ y denotemos por p_l la probabilidad del suceso l -ésimo. Supongamos que no tenemos información alguna acerca de la distribución de probabilidad que gobierna un determinado fenómeno salvo la propia condición de que debe encontrarse normalizada

$$\sum_{l=1}^n p_l = 1$$

Construyamos la función lagrangiana:

$$\begin{aligned} L[\{p_l\}] &= S(p_1 \dots p_n) - \alpha \sum_{l=1}^n p_l \\ &= -k \sum_{l=1}^n p_l \ln p_l - \alpha \sum_{l=1}^n p_l \end{aligned}$$

donde α es el multiplicador indeterminado. De acuerdo con el principio de entropía máxima de Jaynes, tendremos que $\delta L[\{p_l\}] = 0$, por lo que:

$$\begin{aligned} \delta L[\{p_l\}] &= \sum_{l=1}^n \frac{\partial L[\{p_l\}]}{\partial p_l} \delta p_l = \sum_{l=1}^n [-k(\ln p_l + 1) - \alpha] \delta p_l = \\ &= \sum_{l=1}^n [-k(\ln p_l + 1 + \alpha')] \delta p_l = 0 \end{aligned}$$

donde hemos introducido $\alpha' = \alpha/k$. Evidentemente, de la ecuación anterior se sigue que $(\ln p_l + 1 + \alpha') = 0$, o lo que es lo mismo, $p_l = \text{cte.} = 1/n$, de acuerdo con el principio de razón insuficiente.

Ejemplo 2.8

Supongamos ahora que por algún medio conocemos el valor medio de un determinada variable aleatoria E,

$$\bar{E} = \sum_{l=1}^n p_l E_l$$

además de la condición de normalización. En este caso la lagrangiana será:

$$L[\{p_l\}] = -k \sum_{l=1}^n p_l \ln p_l - \alpha' \sum_{l=1}^n p_l - \beta' \sum_{l=1}^n p_l E_l$$

donde α' y β' son ahora los multiplicadores indeterminados. De acuerdo con el principio de entropía máxima de Jaynes, tendremos nuevamente que $\delta L[\{p_l\}] = 0$, por lo que:

$$\begin{aligned} \delta L[\{p_l\}] &= \sum_{l=1}^n \frac{\partial L[\{p_l\}]}{\partial p_l} \delta p_l = \sum_{l=1}^n [-k \ln p_l + 1 - \alpha' - \beta' E_l] \delta p_l = \\ &= \sum_{l=1}^n [-k(\ln p_l + 1 + \alpha' + \beta' E_l)] \delta p_l = 0 \end{aligned}$$

donde hemos introducido $\alpha = \alpha'/k$, $\beta = \beta'/k$. Evidentemente, de la ecuación anterior se sigue que $(\ln p_l + 1 - \alpha - \beta E_l) = 0$, o lo que es lo mismo,

$$p_l = \frac{e^{-\beta E_l}}{Z}$$

$$Z = \sum_{l=1}^n e^{-\beta E_l}$$

donde hemos introducido la notación $Z^{-1} = \exp(-1 + \alpha)$ y usado la condición de normalización para calcular su expresión.⁷

2.3. Distribuciones de probabilidad

2.3.1. Variables aleatorias discretas y continuas

Sea Ω el espacio muestral de un determinado fenómeno aleatorio y sea X una función real definida sobre el espacio muestral que asocia a cada suceso un valor de la recta real. Se dice entonces que X es una *variable aleatoria unidimensional*:

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ A &\rightarrow X(A). \end{aligned} \tag{2.30}$$

Así pues, una variable aleatoria o estadística asigna un número a cada suceso de Ω , representando de esta manera de modo numérico una determinada característica de la población. Es obvio que para el mismo experimento aleatorio podrá haber diferentes asignaciones de la variable aleatoria. Por ejemplo, en el lanzamiento de una moneda ($\Omega = \{c, x\}$) podemos construir la variable aleatoria X de modo que $X(c) = 1$ y $X(x) = 0$. O bien podríamos hacer la asignación opuesta $X(c) = 0$ y $X(x) = 1$. Por ello, fundamental saber qué variable se considera en cada caso concreto.

Diremos que X es una *variable aleatoria discreta* si el número de valores que puede tomar es numerable (finito o infinito), esto es, se puede poner en relación biyectiva con un subconjunto de los números enteros. Un ejemplo de una variable aleatoria discreta es el número de veces que ocurre un suceso determinado (e.g. número de desintegraciones radiactivas en una muestra de materia), o bien el ejemplo anterior del lanzamiento de un dado o de una moneda.

⁷Es interesante destacar que, si E representa la energía y el índice l recorre el conjunto de los microestados de un sistema físico, esta distribución de probabilidad (debida a Gibbs, 1902) constituye un resultado central de la Mecánica Estadística, rama de la Física que, mediante la aplicación de métodos estadísticos, proporciona las propiedades macroscópicas (termodinámicas) de cuerpos en equilibrio a partir de las correspondientes propiedades microscópicas (mecánicas) de sus constituyentes.

Por contra, se dice que X es una *variable aleatoria continua* si su imagen está formada por uno o más intervalos de la recta real. Ejemplos de este tipo de variables podrían ser el peso o dimensiones de una persona, el tiempo de duración de un suceso, su frecuencia de ocurrencia, la energía de un sistema físico en Mecánica Clásica, etc. Notemos que no es posible conocer el valor *exacto* de una variable estadística de este tipo (asociada al resultado numérico de una medida), sino que medir su valor consiste en clasificarlo dentro de un intervalo determinado.

2.3.2. Variable aleatoria discreta. Función de probabilidad y función de distribución de probabilidad

Para una variable discreta cuyo espacio muestral es $\Omega = \{x_1, x_2, \dots, x_i, \dots\}$, aquella función que asigna a cada posible valor x_i de la variable aleatoria X su probabilidad se conoce como *función de probabilidad*⁸:

$$p : \Omega \rightarrow [0, 1] \quad (2.31)$$

$$x_i \rightarrow p(x_i) = p(X = x_i). \quad (2.32)$$

Obviamente la distribución debe encontrarse normalizada

$$\sum_{x_i \in \Omega} p(x_i) = 1. \quad (2.33)$$

Por otro lado, la función de distribución acumulativa de probabilidad (o simplemente *función de distribución*) $F(x)$, para una variable aleatoria discreta se define en un punto x_0 , $F(x_0)$, como la probabilidad de que la variable tome un valor menor o igual que x_0 :

$$F(x_0) = p(x \leq x_0) \quad (2.34)$$

Por convenio, $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ y $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$. Si ordenamos los valores de la variable aleatoria, x_1, \dots, x_n , de manera creciente $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$, entonces

$$\begin{aligned} F(x_1) &= p(x_1) \\ F(x_2) &= p(x_1) + p(x_2) \\ &\vdots \\ F(x_n) &= \sum_{i=1}^n p(x_i) = 1 \end{aligned} \quad (2.35)$$

Ejemplo 2.9

⁸ Notemos que la asignación de probabilidad a los diferentes valores de una variable aleatoria es equivalente a su atribución a los sucesos correspondientes del espacio muestral.

Variable uniforme: Consideremos la variable aleatoria

$$\begin{aligned} X : \Omega = \{A_1, A_2, \dots, A_N\} &\rightarrow \mathbb{R} \\ A_i &\rightarrow X(A_i) = i \end{aligned} \quad (2.36)$$

y la distribución de probabilidad

$$p(A_i) \equiv p(x_i) = \frac{1}{N} \quad (2.37)$$

que obviamente se ajusta a la definición y verifica la propiedad (2.33). En lo que respecta a la función de distribución de la variable aleatoria, es evidente que toma los valores

$$\begin{aligned} F(x_1) &= p(x_1) = \frac{1}{N} \\ F(x_2) &= p(x_1) + p(x_2) = \frac{2}{N} \\ &\vdots \\ F(x_n) &= \sum_{i=1}^n p(x_i) = \frac{N}{N} = 1 \end{aligned}$$

2.3.3. Variable aleatoria continua. Función de densidad de probabilidad y función de distribución de probabilidad.

Consideremos una variable que puede tomar valores de manera continua en un intervalo que tomaremos, por simplicidad, $\mathbb{R} = (-\infty, +\infty)$. Ejemplos de este tipo de variables puede ser la posición de una partícula en una dimensión que, evidentemente, no podemos precisar de manera absolutamente exacta. Como hemos mencionado en el capítulo anterior, lo que debemos hacer en estos casos es agrupar los valores de la variable aleatoria en clases, discretizando el problema, lo que nos permitirá construir un determinado histograma y su curva continua asociada. Haciendo particiones cada vez más y más finas (i.e. aumentando el número de observaciones) la línea quebrada correspondiente a la curva continua de frecuencias tendería convertirse en una curva suave, $f(x)$.

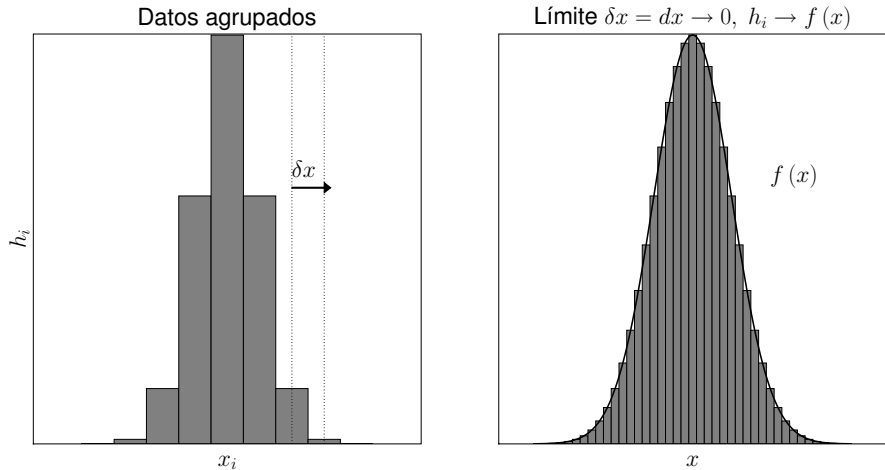
Como sabemos, el área de cada rectángulo del histograma es proporcional a la frecuencia de aparición de cada dato, por lo que podemos relacionarlo con la probabilidad de ocurrencia del suceso asociado al valor x_i de la variable aleatoria:

$$p(x_i) \sim h_i \delta x \quad (2.38)$$

donde h_i es la altura del rectángulo i -ésimo del histograma, cuya marca de clase es x_i . En el límite $\delta x \rightarrow 0$, la anchura de los intervalos del histograma

tiende a convertirse en una cantidad diferencial (dx) y la altura de cada intervalo (rectángulo) viene dada por el valor de una función $f(x)$ en el punto correspondiente a la marca de clase:

$$\begin{aligned} h_i &\rightarrow f(x_i) \\ \delta x &\rightarrow dx. \end{aligned} \quad (2.39)$$



Así, podemos afirmar que la probabilidad de obtener un valor de la magnitud x que se encuentre entre x_0 y $x_0 + dx$ será

$$dp(x_0) = f(x_0)dx, \quad (2.40)$$

que, lógicamente, se trata de una probabilidad diferencial. Así, $f(x_0)$ es una probabilidad por unidad de longitud

$$\frac{dp(x_0)}{dx} = f(x_0), \quad (2.41)$$

de donde procede el nombre de función densidad que recibe $f(x)$. En virtud de lo anterior, llamaremos función densidad de probabilidad a una función $f(x)$ que verifica las condiciones:

$$\begin{aligned} f(x) &\geq 0 \\ \int_{-\infty}^{+\infty} f(x)dx &= 1, \end{aligned} \quad (2.42)$$

y puede interpretarse como la curva límite que obtendríamos disminuyendo la anchura de las bases de un histograma de probabilidad. El conocimiento de la función densidad permite calcular cualquier probabilidad por integración de la ecuación

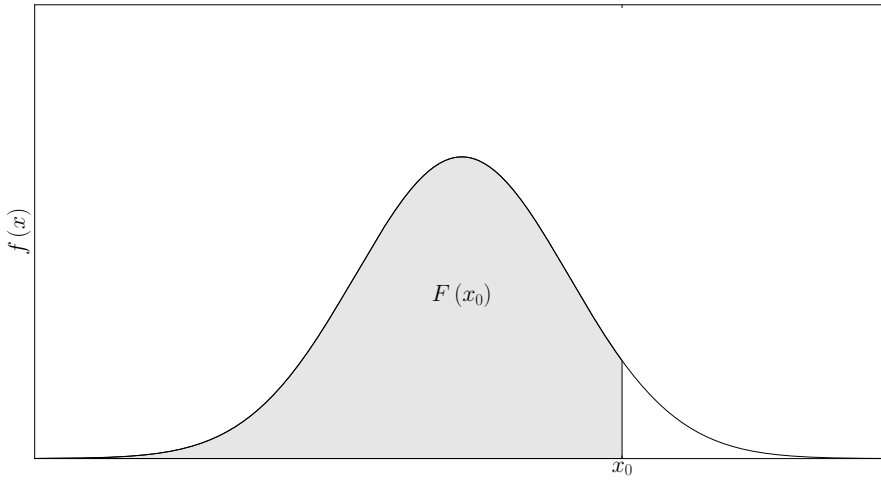
$$dp(x) = f(x)dx. \quad (2.43)$$

Así, la probabilidad de que la variable aleatoria x tome un valor comprendido en el intervalo $[a, b]$ será:

$$p(a \leq x \leq b) = \int_a^b dp(x) = \int_a^b f(x)dx. \quad (2.44)$$

La función de distribución acumulativa de probabilidad de una variable aleatoria continua se define, al igual que el caso discreto, de la forma:

$$F(x_0) = p(x \leq x_0) = \int_{-\infty}^{x_0} f(x)dx. \quad (2.45)$$

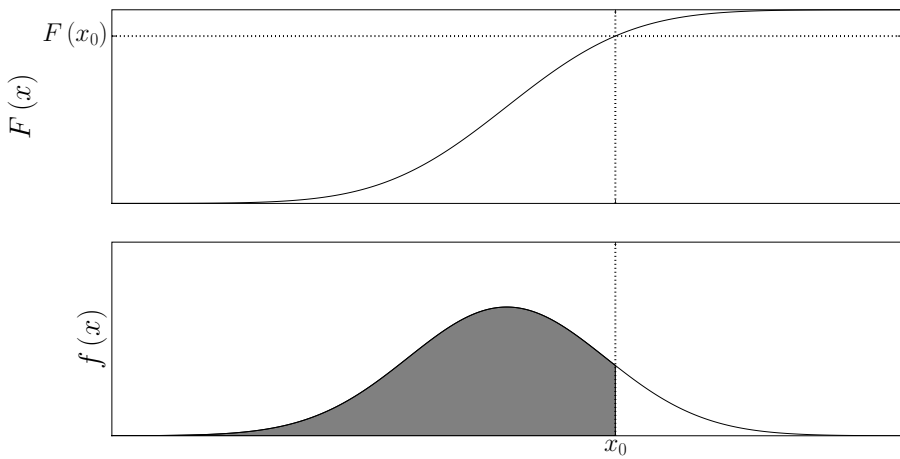


Equivalentemente, aplicando el teorema fundamental del cálculo integral, la función densidad en el caso continuo se calcula como derivada de la función de distribución:

$$f(x) = \frac{dF(x)}{dx} \quad (2.46)$$

La función de distribución de probabilidad de una variable aleatoria verifica tres propiedades derivadas de su definición:

- i) $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$
- ii) $F(+\infty) = \lim_{x \rightarrow \infty} F(x) = 1$
- iii) F monótona creciente $F(x_1) \geq F(x_2)$ si $x_1 \geq x_2$


Ejemplo 2.10

- Distribución uniforme: Se trata de una densidad de probabilidad cuyo valor es constante en un intervalo y nula fuera de él.

$$f(x) = \begin{cases} f_0 & a \leq x \leq b \\ 0 & \text{resto} \end{cases} \quad (2.47)$$

Evidentemente, para que $f(x)$ sea una auténtica función densidad ha de verificar las propiedades de estas:

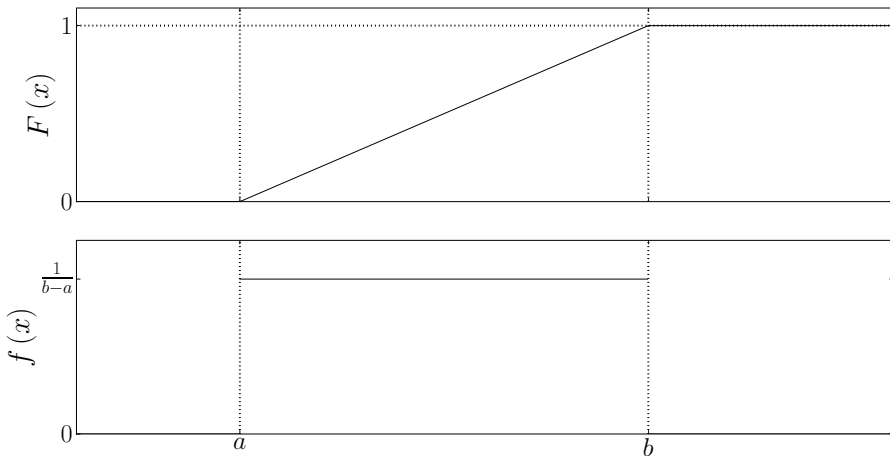
$$\begin{aligned} f(x) &\geq 0 \quad \forall x & f_0 &\geq 0 \\ \int_{-\infty}^x f(x) dx &= 1 \Rightarrow f_0 \int_a^b dx = f_0(b-a) = 1 \\ f_0 &= \frac{1}{(b-a)} \end{aligned}$$

La correspondiente función de distribución será:

$$F(x) = \int_{-\infty}^x f(u) du = \int_a^x \frac{1}{(b-a)} du = \frac{x-a}{b-a} \quad (2.48)$$

Observemos que esta función satisface las propiedades asignadas a las funciones de distribución de probabilidad:

- $F(-\infty) = 0$; $F(+\infty) = 1$
- F monótona creciente



- Distribución exponencial: Consideremos una variable estadística cuya función densidad de probabilidad sea:

$$f(x) = \begin{cases} 0 & x \in (-\infty, 0] \\ f_0 e^{-\beta x} & x \in (0, +\infty) \end{cases} \quad (2.49)$$

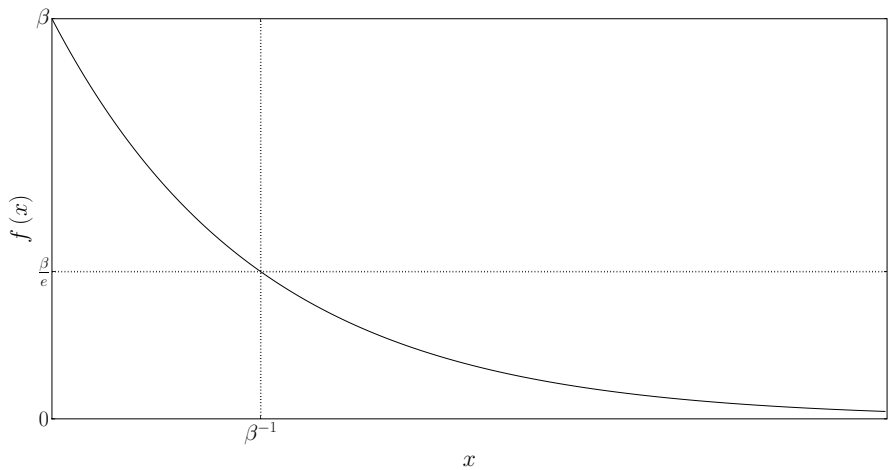
Normalizando la distribución anterior tenemos:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Rightarrow \int_0^{\infty} f_0 e^{-\beta x} dx = 1 \Rightarrow f_0 = \beta \quad (2.50)$$

La función distribución de probabilidad será (para $x \geq 0$), por lo tanto;

$$F(x) = \int_{-\infty}^x f(u) du = \int_0^x \beta e^{-\beta u} du = 1 - e^{-\beta x} \quad x \geq 0 \quad (2.51)$$

Siendo $F(x) = 0$ para $x < 0$. Como vemos de nuevo se cumplen las propiedades básicas para esta función de distribución.



2.3.4. Medidas características de una distribución de probabilidad.

De la misma forma que hemos definido las medidas características para una distribución de frecuencias muestrales podemos definir las para una distribución de probabilidad, tanto en el caso de que se trate de una variable aleatoria discreta como de una variable aleatoria continua.

Variable aleatoria discreta

Una vez que conocemos la función de probabilidad de una variable aleatoria, $p(x_i)$, podemos obtener cualquier medida asociada a ella.

1. *Medidas de centralización o de posición central.* Las principales medidas de centralización de la distribución de probabilidad son:

a) *Media:*

La media de una distribución de probabilidad de una variable aleatoria que toma k valores posibles es:

$$\mu = \sum_{i=1}^k p(x_i)x_i. \quad (2.52)$$

Cuando calculamos la media de una variable aleatoria X a partir de la distribución de probabilidad correspondiente a la población y no a partir de distribuciones de frecuencia, solemos denominar a dicha media valor esperado o esperanza matemática de X , $E\{x\} \equiv \mu$.

Esta forma de definir el valor esperado se puede extender a funciones arbitrarias de la variable aleatoria de forma que el valor esperado de una función cualquiera $H(X)$ de la variable aleatoria será:

$$E\{H(X)\} = \sum_{i \in E} p(x_i)H(x_i) \quad (2.53)$$

En el cálculo de valores esperados es muy importante la propiedad de linealidad. Sea X una variable aleatoria y $f(x)$, $h(x)$ dos funciones de esta variable aleatoria, se cumple que

$$E\{ag(x) + bh(x)\} = aE\{g(x)\} + bE\{h(x)\} ; \quad a, b \in \mathbb{R}. \quad (2.54)$$

La demostración de esta propiedad es trivial.

Es habitual hablar de modo indistinto de media de una distribución de probabilidad o de la media *poblacional*.

b) *Moda*:

Es el valor o valores de la variable aleatoria que maximiza(n) $p(x_i)$. Es decir aquellos valores $Md \in \Omega$ que cumplen

$$p(x_i) \leq p(Md) \quad \forall x_i \in \Omega. \quad (2.55)$$

c) *Mediana*:

Es aquel valor Me de la variable aleatoria tal que la probabilidad de obtener un valor superior o inferior a él es de $\frac{1}{2}$. Será, pues, el menor valor que verifique $F(x) \leq 0,5$.

2. *Medidas de dispersión*. Consideraremos únicamente las más importantes de todas ellas, que sin duda son la varianza y la desviación típica. La obtención de otras medidas introducidas con anterioridad en el caso de una distribución de probabilidad es inmediato. La varianza σ^2 de la distribución de probabilidad de una variable aleatoria discreta X se define de la forma:

$$\sigma^2 = \sum_{i=1}^k p(x_i)(x_i - \mu)^2 \quad (2.56)$$

La desviación típica se define como la raíz cuadrada positiva de la varianza:

$$\sigma = E \left\{ (x - E\{x\})^2 \right\} = \sqrt{\sum_{i=1}^k p(x_i)(x_i - \mu)^2}. \quad (2.57)$$

Usaremos de modo indistinto la expresión de desviación típica de una distribución de probabilidad o bien desviación típica *poblacional*.

3. Momentos de la distribución de la variable aleatoria.

El momento respecto al punto $x = c$ de orden j de una distribución de probabilidad de una variable aleatoria discreta está definido mediante

$$M_j(c) = \sum_{i=1}^k p(x_i) (x_i - c)^j \quad (2.58)$$

Notemos que la media de la variable coincide con el momento de primer orden respecto al origen ($c = 0$) y que la varianza es el momento de segundo orden central o respecto a la media ($c = \mu$) de la distribución de probabilidad.⁹ El resto de las medidas características de la variable pueden obtenerse a partir de los momentos sustituyendo las distribuciones de frecuencia por las de probabilidad en las expresiones que se introdujeron en el capítulo 1.

Variable aleatoria continua

Este caso puede tratarse de la misma forma que el anterior sin más que tener en cuenta ciertas reglas para pasar al continuo la variable aleatoria x_i . Este se realiza, de acuerdo con lo que hemos visto anteriormente, de manera que un valor de la variable aleatoria discreta x_i se corresponde ahora con un intervalo infinitesimal de la recta real en torno al valor x de la correspondiente variable aleatoria continua, $(x, x + dx)$. Consecuentemente, las sumas de las ecuaciones anteriores a todos los valores de la variable aleatoria se convertirán en integrales y la probabilidad de tener un determinado valor en la probabilidad diferencial de obtener un valor x de la variable aleatoria en un intervalo de anchura dx :

$$\begin{aligned} \sum_{i=1}^k &\longrightarrow \int_{-\infty}^{+\infty} \\ p(x_i) &\longrightarrow f(x)dx = dp(x) \end{aligned} \quad (2.60)$$

Mediante estas reglas es trivial generalizar las ecuaciones anteriores para el caso de una variable aleatoria continua.

⁹ A lo largo de este texto trataremos de establecer una distinción en la notación entre las medidas características de una muestra y las de la población. En general notaremos estas magnitudes de la forma siguiente:

	Muestra	Población
media	\bar{x}	μ
varianza	s^2	σ^2
momentos	$m_j(c)$	$M_j(c)$

(2.59)

1. *Medidas de centralización.*a) *Media:*

Modificando según lo anteriormente dicho la expresión del valor esperado de una variable discreta en (2.52) obtenemos¹⁰:

$$\mu \equiv E\{x\} = \int_{-\infty}^{+\infty} xf(x)dx \quad (2.61)$$

b) *Mediana:*

Es el valor que divide la probabilidad en dos mitades iguales. Por ello, para una variable continua, tenemos.

$$F(Me) = 0,5 = p(x \leq Me) \quad (2.62)$$

c) *Moda:*

La moda Md es el valor más probable de la variable aleatoria, por lo que, en caso de ser $f(x)$ diferenciable, puede calcularse mediante:

$$\left. \frac{df}{dx} \right|_{x=Md} = 0; \quad \left. \frac{d^n f}{dx^n} \right|_{x=Md} < 0, \quad (2.63)$$

siendo n un número par, el orden de la primera derivada no nula de f en la moda.

2. *Medidas de dispersión.*

Definimos la varianza σ^2 mediante el equivalente continuo de la ecuación (2.56):

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx \quad (2.64)$$

Lógicamente, la desviación típica es, como de costumbre, la raíz cuadrada positiva de la magnitud anterior.

3. *Momentos de la distribución de la variable aleatoria:*

El momento de orden j respecto al punto c de la distribución de probabilidad, $f(x)$, asociada a la variable aleatoria continua X será:

$$M_j(c) = \int_{-\infty}^{+\infty} (x - c)^j f(x) dx \quad (2.65)$$

Como de costumbre, cuando $c = 0$ hablaremos de momentos respecto al origen y de momentos centrales o respecto a la media cuando $c = \mu$:

$$M_j(\mu) = \int_{-\infty}^{+\infty} (x - \mu)^j f(x) dx \quad (2.66)$$

¹⁰Los límites de esta integral y las que siguen pueden restringirse a la imagen $[a, b]$ de la variable aleatoria.

Notemos que la media de la variable coincide con el momento de primer orden respecto al origen, y que la varianza es el momento de segundo orden.

Los momentos de una distribución de probabilidad pueden considerarse como valores esperados de ciertas funciones de la variable aleatoria. En efecto, los momentos de orden r respecto al punto c son

$$M_r(c) = E\{(x - c)^r\} = \sum_{i \in E} (x_i - c)^r p(x_i) \quad \text{discreto} \quad (2.67)$$

$$M_r(c) = E\{(x - c)^r\} = \int_{-\infty}^{+\infty} dx (x - c)^r f(x) \quad \text{continuo.} \quad (2.68)$$

Obviamente, los momentos respecto al origen ($c = 0$) son los valores esperados de la función

$$H(x) = x^r \quad r = 0, 1, 2, \dots, \quad (2.69)$$

que pueden ser escritos como

$$M_r(0) = E\{x^r\} \quad (2.70)$$

Del mismo modo se pueden definir los momentos centrales ($c = \mu$) de la distribución

$$\begin{aligned} M_r(\mu) &= E\{(x - \mu)^r\} \\ M_0(\mu) &= E\{(x - \mu)^0\} = E\{1\} = 1 \\ M_1(\mu) &= E\{(x - \mu)^1\} = E\{x\} - E\{\mu\} = \mu - \mu = 0 \\ M_2(\mu) &= E\{(x - \mu)^2\} = \sigma^2 \\ &\dots \end{aligned} \quad (2.71)$$

2.4. Función característica

Sea x una variable aleatoria continua¹¹ de función de densidad $f(x)$ cuyo espacio muestral coincide con toda la recta real. Se define la *función característica* de la distribución de la forma

$$\phi(k) = E\{e^{ikx}\} = \int_{-\infty}^{+\infty} f(x)e^{ikx} dx = \int_{-\infty}^{\infty} f(x) (\cos(kx) + i \operatorname{sen}(kx)), \quad (2.72)$$

que no es sino la transformada de Fourier inversa de la función de densidad de probabilidad. Es evidente que invirtiendo esta transformada de Fourier podemos obtener de manera unívoca la función densidad a partir de la función

¹¹Los resultados de la presente sección son directamente aplicables también al caso discreto, sin más que sustituir las integrales por sumas discretas y las funciones de densidad por las correspondientes funciones de probabilidad.

característica de la forma:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \phi(k) e^{-ikx} dx. \quad (2.73)$$

Entre las diferentes utilidades de la función característica, quizá la más importante es que su conocimiento nos permite obtener los momentos respecto al origen de cualquier orden de la distribución. Efectivamente, el momento de orden n respecto al origen de una distribución de probabilidad de función de densidad $f(x)$ y función característica $\phi(k)$, $E\{x^n\}$, está dado por¹²:

$$E\{x^n\} = \frac{1}{i^n} \left. \frac{d^n \phi(k)}{dk^n} \right|_{k=0} \quad (2.75)$$

lo que se puede demostrar de manera inmediata de la forma siguiente. Derivando la ecuación (2.72) obtenemos:

$$\begin{aligned} \frac{d^n \phi(k)}{dk^n} &= \frac{d^n}{dk^n} \int_{-\infty}^{+\infty} f(x) e^{ikx} dx = \int_{-\infty}^{+\infty} \frac{\partial^n}{\partial k^n} [f(x) e^{ikx}] dx = \\ &= \int_{-\infty}^{+\infty} i^n x^n f(x) e^{ikx} dx. \end{aligned} \quad (2.76)$$

Evaluando la expresión anterior en $k = 0$, obtenemos de manera trivial el resultado deseado:

$$\frac{d^n \phi(k=0)}{dk^n} = \int_{-\infty}^{+\infty} i^n x^n f(x) dx = i^n \langle x^n \rangle. \quad (2.77)$$

Por otro lado, el conocimiento de los diferentes momentos de la distribución de probabilidad nos permite la reconstrucción de la función característica:

$$\begin{aligned} \phi(k) &= E\{e^{ikx}\} = \int_{-\infty}^{+\infty} f(x) \sum_{n=0}^{\infty} \frac{(ikx)^n}{n!} dx \\ &= \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} E\{x^n\}. \end{aligned} \quad (2.78)$$

Lógicamente, la expresión anterior únicamente tiene sentido cuando los momentos de orden superior existen y la serie converge para el k de interés.

¹²Análogamente podríamos probar que el momento central de orden n de la distribución de probabilidad es:

$$(x - \mu)^n = \frac{1}{i^n} \left. \frac{d^n \phi_{\mu}(k)}{dk^n} \right|_{k=0} \quad (2.74)$$

donde $\phi_{\mu}(k) = \langle \exp[ik(x - \mu)] \rangle = e^{-ik\mu} \phi(k)$.

2.4.1. Función generatriz de momentos

Un concepto estrechamente relacionado con el de función característica es el de función generadora (o generatriz) de momentos, que se define como el valor medio o esperado $\exp(tx)$ ¹³:

$$\mathcal{M}_X(t) = E\{e^{tx}\} = \int_{-\infty}^{+\infty} dx e^{tx} f(x). \quad (2.79)$$

En el caso de que estemos ante una variable aleatoria discreta tendremos, obviamente,

$$\mathcal{M}_X(t) = \sum_{i \in \Omega} p(x_i) \cdot \exp(tx_i). \quad (2.80)$$

Al igual que se hizo con $\phi(k)$, es posible probar mediante estas definiciones que a partir de $\mathcal{M}_X(t)$ podemos obtener los diferentes momentos de la distribución de probabilidad de la forma:

$$M_r^{(\mu)} = E\{(X - \mu)^r\} = \left. \frac{d^r \mathcal{M}_{X-\mu}(t)}{dt^r} \right|_{t=0}. \quad (2.81)$$

Efectivamente, si existe $\mathcal{M}_X(t)$ para un cierto conjunto de valores $t \in [-c, c]$ entonces existen las derivadas de $\mathcal{M}_X(t)$ con respecto a t de todos los órdenes en torno a $t = 0$:

$$\begin{aligned} \left. \frac{d\mathcal{M}_X(t)}{dt} \right|_{t=0} &= \left. \frac{d}{dt} E\{\exp(tx)\} \right|_{t=0} = \\ &= E\left\{ \left. \frac{d}{dt} \exp(tx) \right|_{t=0} \right\} = E\{x \exp(tx)\} \Big|_{t=0} = \\ &= E\{x\} = M_1(0) \end{aligned}$$

$$\begin{aligned} \left. \frac{d^2 \mathcal{M}_X(t)}{dt^2} \right|_{t=0} &= \left. \frac{d^2}{dt^2} E\{\exp(tx)\} \right|_{t=0} = \\ &= E\left\{ \left. \frac{d}{dt} (x \exp(tx)) \right|_{t=0} \right\} = \\ &= E\{x^2 \exp(tx)\} \Big|_{t=0} = E\{x^2\} = M_2(0) \end{aligned}$$

Este proceso se puede justificar también mediante el desarrollo en serie de Taylor

$$\begin{aligned} \exp(tx) &= 1 + tx + \frac{1}{2!}t^2x^2 + \frac{1}{3!}t^3x^3 + \dots \\ E\{\exp(tx)\} &= E\{1\} + tE\{x\} + \frac{1}{2!}t^2E\{x^2\} + \dots \\ \frac{d}{dt^r} E\{\exp(tx)\} &= E\{x^r\} + tE\{x^{r+1}\} + \dots \\ \left. \frac{d}{dt^r} E\{\exp(tx)\} \right|_{t=0} &= E\{x^r\} = M_r(0). \end{aligned}$$

¹³ Por simplicidad, en lo que sigue usaremos esta función preferentemente, en lugar de la función característica, ya que, como demostraremos posteriormente, sus propiedades son muy similares y la función generadora permite trabajar en el espacio real.

Además, se pueden definir los momentos centrales a partir de la función generadora de momentos centrales

$$\mathcal{M}_{X-\mu}(t) = E\{\exp(t(X - \mu))\}, \quad (2.82)$$

que, igual que antes, verifica

$$M_r = E\{X - \mu)^r\} = \left. \frac{d^r}{dt^r} \mathcal{M}_{X-\mu}(t) \right|_{t=0}. \quad (2.83)$$

Por otro lado, aunque no lo probaremos, dada la función generadora de momentos de una distribución de probabilidad, la distribución de probabilidad queda perfectamente definida. En efecto, si dos variables aleatorias X_1, X_2 poseen dos funciones generadoras de momentos iguales

$$\mathcal{M}_{X_1}(t) = \mathcal{M}_{X_2}(t), \quad (2.84)$$

entonces sus funciones de densidad de probabilidad son iguales. Esto permite caracterizar de modo unívoco una distribución de probabilidad mediante la función generadora de momentos.

Sean X_1 y X_2 dos variables aleatorias independientes. Entonces, se verifica que la variable aleatoria

$$Y = X_1 + X_2 \quad (2.85)$$

tiene como función generadora de momentos

$$\mathcal{M}_Y(t) = \mathcal{M}_{X_1}(t)\mathcal{M}_{X_2}(t), \quad (2.86)$$

siendo $\mathcal{M}_{X_1}(t)$ y $\mathcal{M}_{X_2}(t)$ las funciones generadoras de momentos de las variables X_1 y X_2 respectivamente. En efecto, suponiendo que X_1 y X_2 sean variables aleatorias continuas con funciones de densidad $g(x_1)$ y $h(x_2)$, respectivamente:

$$\mathcal{M}_Y(t) = \int_{-\infty}^{+\infty} dy \exp[ty] f(y), \quad (2.87)$$

siendo $f(y)$ la densidad de probabilidad de y . Como $Y = X_1 + X_2$ y además X_1, X_2 son independientes

$$f(y) = f(x_1 + x_2) = \int_{-\infty}^{+\infty} dx_1 g(x_1) h(y - x_1) \quad (2.88)$$

Así, podemos escribir

$$\begin{aligned} \mathcal{M}_Y(t) &= \int_{-\infty}^{+\infty} dy \exp[ty] \int_{-\infty}^{+\infty} dx_1 g(x_1) h(y - x_1) = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dy dx_1 \exp[ty] g(x_1) h(y - x_1). \end{aligned}$$

Por otro lado,

$$\mathcal{M}_Y(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dy \exp[tx_1] \exp[ty - tx_1] g(x_1) h(y - x_1) \quad (2.89)$$

y haciendo el cambio $x_2 = y - x_1$ obtenemos

$$\begin{aligned} \mathcal{M}_Y(t) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 \exp[tx_1] \exp[tx_2]g(x_1)h(x_2) = \\ &= \int_{-\infty}^{+\infty} dx_1 \exp[tx_1]g(x_1) \int_{-\infty}^{+\infty} dx_2 \exp[tx_2]h(x_2) = \\ &= \mathcal{M}_{X_1}(t)\mathcal{M}_{X_2}(t). \end{aligned}$$

Esto nos permite averiguar cuál es la distribución de probabilidad cuando sumamos dos variables aleatorias independientes. Obviamente, una relación idéntica es cierta para las funciones características.

2.4.2. Cumulantes. Función generatriz de cumulantes.

En ocasiones resulta útil escribir la función característica de la distribución de probabilidad en términos de una expansión de cumulantes en lugar de expandirla directamente en los momentos de la distribución como se ha hecho anteriormente. Así, podemos desarrollar la función característica de la forma

$$\phi(k) = \exp \left[\sum_{n=0}^{\infty} \frac{(ik)^n}{n!} c_n(x) \right], \quad (2.90)$$

donde $c_n(x)$ es el cumulante de orden n de la distribución de probabilidad. Evidentemente, si expandimos la expresión anterior en serie de potencias e igualamos los coeficientes con los de la serie de la ecuación (2.78) obtenemos:

$$\begin{aligned} c_1(x) &= E \{x\} \\ c_2(x) &= E \{x^2\} - E \{x\}^2 = \sigma^2 \\ c_3(x) &= E \{x^3\} - 3E \{x^2\} E \{x\} + 2E \{x\}^3 \\ &\dots \end{aligned} \quad (2.91)$$

Es inmediato comprobar que la función generatriz de cumulantes es simplemente el logaritmo neperiano de la función característica:

$$\ln \phi(k) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} c_n(x) \Rightarrow c_n(x) = \frac{1}{i^n} \left. \frac{d^n \ln \phi(k)}{dk^n} \right|_{k=0}. \quad (2.92)$$

2.5. Transformación de variables aleatorias.

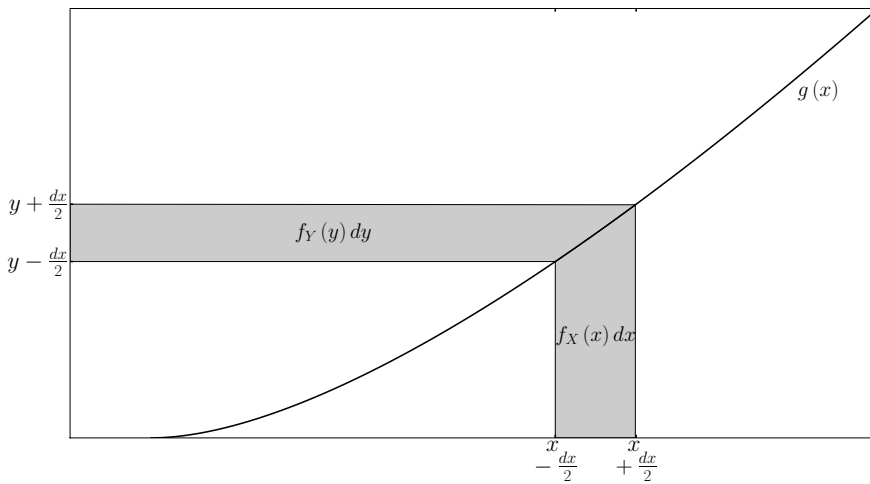
Sucede a menudo que conocemos, ya sea a partir de razonamientos teóricos o de resultados experimentales, la distribución de probabilidad de una variable aleatoria X y deseamos hallar la de otra variable, Y , que está relacionada con la primera mediante una función $y = g(x)$.

Si X es una variable aleatoria discreta que toma valores³ en un conjunto $\{x_i\}$, el espacio muestral de Y será $\{y_i = g(x_i)\}$, y es evidente que el suceso

“ Y toma el valor y_i ” es equivalente a “ X toma un valor x_j tal que $y_i = g(x_j)$ ”. En consecuencia, la probabilidad de y_j será igual a la suma de las probabilidades de esos x_j :

$$p(Y = y_i) = p_Y(y_i) = \sum_{j|g(x_j)=y_i} p(X = x_j) = \sum_{j|g(x_j)=y_i} p_X(x_j). \quad (2.93)$$

El razonamiento para variables continuas es conceptualmente el mismo, aplicando la receta estándar de cambio de sumas por integrales que se ha introducido en secciones anteriores. Para el caso más elemental, aquél en que la función $g(x)$ es biyectiva, se puede prescindir de la suma y aplicar el cambio $p \rightarrow f(x) dx$ en (2.93), lo cual lleva a un resultado de sencilla interpretación geométrica, una igualdad de áreas:



$$f_X(x) dx = f_Y(y) dy \Rightarrow f_Y(y) = \frac{f_X(x)}{\left| \frac{dg}{dx} \right|} \quad (2.94)$$

De forma más general y detallada, si X toma valores en un intervalo A de la recta real y viene caracterizada por una función de densidad de probabilidad $f_X(x)$, Y estará definida sobre la imagen $g(A)$ y el valor de su función de densidad para un punto y de esa imagen será la integral de $dp = f_X(x) dx$ sobre los valores de X tales que $y = g(x)$. Esta integración se expresa matemáticamente mediante el funcional δ de Dirac:

$$f_Y(y) = \int_A \delta[y - g(x)] f_X(x) dx \quad (2.95)$$

Supondremos, por simplicidad analítica, que para cada valor y de Y hay un número finito $n(y)$ de soluciones $\{x_j(y)\}_{j=1}^{n(y)}$ de la ecuación $y = g(x)$, que g es diferenciable y que su derivada en esos puntos no es nula. En tal caso, es posible aplicar una propiedad bien conocida del funcional δ para escribir:

$$f_Y(y) = \int_A \sum_{j=1}^{n(y)} \frac{\delta(x - x_j)}{\left| \frac{dg}{dx} \right|_{x=x_j}} f_X(x) dx, \quad (2.96)$$

con lo que llegamos a la ecuación equivalente a (2.93) para el caso continuo:

$$f_Y(y) = \sum_{\substack{g(x_j)=y \\ j=1}}^{n(y)} \frac{f_X(x_j)}{\left| \frac{dg}{dx} \right|_{x=x_j}}, \quad (2.97)$$

Conocida la función de densidad de probabilidad de Y , podemos emplearla para calcular sus momentos, si éstos existen. No obstante, también es posible realizar ese cálculo mediante la función de densidad de X , ya que, usando (2.95):

$$\begin{aligned} M_{Y,j} &= \int_{g(A)} f_Y(y) (y - c)^j dy = \int_{g(A)} \left\{ \int_A \delta[y - g(x)] f_X(x) dx \right\} (y - c)^j dy = \\ &= \int_A f_X(x) \left\{ \int_{g(A)} (y - c)^j \delta[y - g(x)] dy \right\} dx = \int_A [g(x) - c]^j f_X(x) dx. \end{aligned} \quad (2.98)$$

En particular, $E\{y\} = E\{g(x)\}$, como es de esperar en un formalismo consistente. Si $f_X(x)$ está lo suficientemente centrada alrededor de su valor esperado μ , y cuando $c \in \{0, \mu\}$, los valores de x que más contribuyen a las integrales en (2.98) son los próximos a μ . En consecuencia se puede llegar a aproximaciones útiles a estos momentos desarrollando g en serie de Taylor. Suponiendo que tenga derivada no nula en μ :

$$\begin{aligned} g(x) &\simeq g(\mu) + \left. \frac{dg}{dx} \right|_{x=\mu} (x - \mu) \\ \Rightarrow E\{y\} &= E\{g(x)\} \simeq E\{g(\mu)\} + \left. \frac{dg}{dx} \right|_{x=\mu} E\{x - \mu\} = g(\mu), \end{aligned} \quad (2.99)$$

aproximación a la media que, sustituida en (2.98), a su vez permite dar otra para los momentos centrales:

$$M_{Y,j} \simeq \int_A \left(\left. \frac{dg}{dx} \right|_{x=\mu} \right)^j (x - \mu)^j f_X(x) dx = \left(\left. \frac{dg}{dx} \right|_{x=\mu} \right)^j M_{X,j}(\mu) \quad (2.100)$$

En particular, cuando $j = 2$ tendremos una relación entre varianzas:

$$\sigma_Y^2 \simeq \left(\left. \frac{dg}{dx} \right|_{x=\mu} \right)^2 \sigma_X^2 \quad (2.101)$$

Ejemplo 2.11

Consideremos una explosión en la que de un proyectil salen fragmentos en direcciones al azar. Colocamos una pantalla a distancia unidad del punto de la explosión. Si llamamos x a la distancia del punto de impacto sobre la pantalla al punto de ésta más próximo al de la explosión, es evidente que está relacionada con el ángulo θ medido con respecto a la normal a la pantalla mediante:

$$x = g(\theta) = \tan \theta. \quad (2.102)$$

Usualmente, la distribución de probabilidad para el ángulo θ se asume como uniforme (proceso isótropo):

$$f(\theta) = \frac{1}{\pi}; \quad -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}. \quad (2.103)$$

Queremos calcular la distribución de probabilidad de x . Utilizamos para ello la relación general que acabamos de hallar, particularizada al caso $x = \tan \theta$, que tiene una única solución física para cada valor de $x \in [0, \infty)$:

$$f_X(x) = \frac{f_\theta(\theta = \arctan x)}{\left| \frac{d \tan \theta}{d\theta} \right|_{\theta = \arctan x}} = \frac{2}{\pi(1+x^2)}, \quad (2.104)$$

que se denomina distribución de Cauchy y, como se puede ver, está normalizada. Una interesante propiedad de esta distribución es que la varianza y los momentos de orden superior al segundo divergen.

En Física y otras disciplinas es de gran importancia un conjunto de métodos numéricos de integración conocidos como de Montecarlo, que se basan en el muestreo aleatorio del dominio de integración de la función de interés. Estos métodos requieren del uso de números aleatorios que se ajusten a una determinada distribución de probabilidad. Es habitual que los ordenadores proporcionen un generador de números aleatorios con una distribución uniforme entre 0 y 1. Se nos plantea, en consecuencia, el problema de la transformación de variables desde el punto de vista inverso: teniendo una variable $X \in [0, 1)$ con $f_X(x) = 1$, ¿qué $g(x)$ hay que escoger de modo que $y = g(x)$ venga descrita por la $f_Y(y)$ buscada? Para solucionarlo partiremos de nuevo¹⁴ de la ecuación (2.95). Integrando en y se halla la función de distribución de Y :

¹⁴Haremos el desarrollo para variables continuas por ser el caso físicamente más habitual. Estrictamente hablando, un número representado en un ordenador por un conjunto finito de bits puede tomar valores en un conjunto discreto, aunque no entraremos aquí en tales sutilezas, pues esos valores son lo suficientemente próximos para considerarlos continuos.

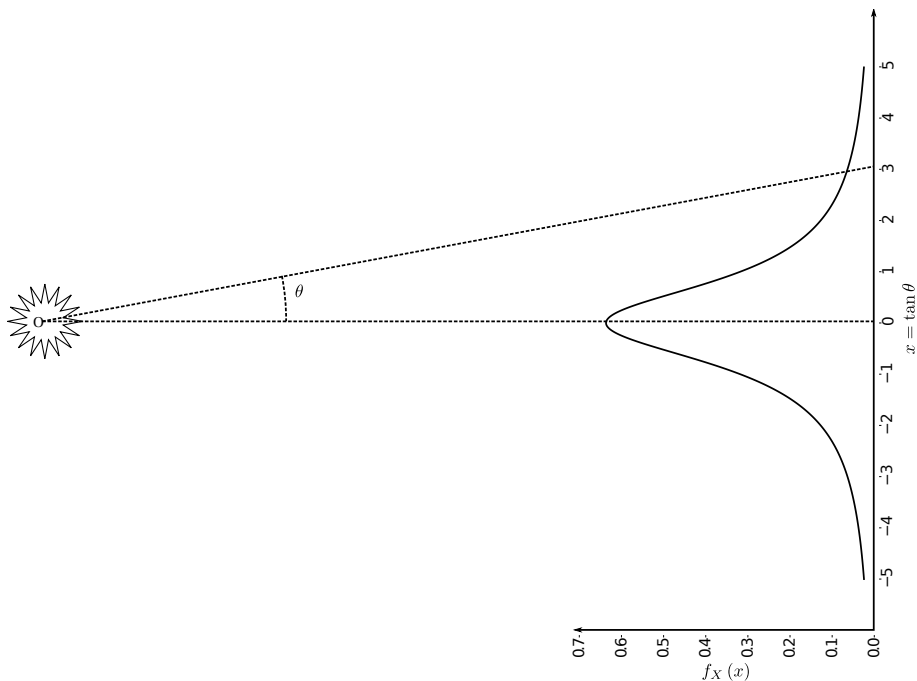


Figura 2.1: Función de densidad de probabilidad para el ejemplo de la explosión (distribución de Cauchy).

$$\begin{aligned}
 F_Y(y) &= \int_{\min g(A)}^y \int_A \delta[y' - g(x)] f_X(x) dx dy' \\
 &= \int_A f_X(x) \int_{\min g(A)}^y \delta[y' - g(x)] dy' dx.
 \end{aligned}$$

La integral interna vale 1 cuando $g(x) < y$ y 0 cuando $g(x) > y$, de modo que $F_Y(y)$ es simplemente la integral de $f_X(x) = 1$ sobre el subconjunto $B(y) \subset A$ que verifica $g(x) > y \forall x \in B(y)$. Para simplificar, buscaremos una función g continua y monótona (en consecuencia, biyectiva). En ese caso $B(y)$ es simplemente el intervalo delimitado por $\min A$ y por $x = g^{-1}(y)$ y se puede realizar la integración explícitamente para dar $F_Y(y) = g^{-1}(y)$, o lo que es lo mismo:

$$y = g(x) = F_Y^{-1}(x). \quad (2.105)$$

Desgraciadamente, el cálculo analítico de F_Y no es siempre posible (mucho menos su inversión) por lo que resulta de gran interés disponer de un camino

alternativo. Un método muy sencillo y general es el llamado algoritmo de aceptación-rechazo de von-Neumann, que se describirá a continuación. Si Y toma valores en $[a, b) \subset \mathbb{R}$:

- i) De la distribución uniforme sobre $[0, 1)$ se extraen dos números, x_0 y c .
- ii) Se calcula $y_0 = a + (b - a)x_0$, que será el candidato a número aleatorio.
- iii) Si $c \leq f_Y(y_0)$, el candidato se acepta. De lo contrario, se vuelve al paso (i) y se genera y_1 por el mismo procedimiento. Se repite el proceso hasta que se acepte un candidato.

Es posible ver, por aplicación de la regla de Laplace, que el proceso lleva a la distribución correcta. Además, su implementación computacional es muy sencilla, aún para distribuciones de probabilidad complicadas. Su principal defecto es el hecho de que, en general, es necesario generar más de un número aleatorio uniforme por cada número final, lo que unido a que cada iteración requiere una evaluación de la función de densidad, que puede ser computacionalmente costosa, degrada su rendimiento.

2.6. Algunas distribuciones de probabilidad relevantes

Realizamos en esta sección un estudio pormenorizado de las principales propiedades de algunas distribuciones de probabilidad que, por su utilidad tanto práctica como teórica, presentan un interés particular dentro de la teoría de probabilidades.

2.6.1. Distribución binomial

Consideremos un ensayo aleatorio que admite únicamente dos resultados excluyentes:

$$\begin{aligned} A \text{ (éxito)} &\longrightarrow p(A) = p \\ B \text{ (fracaso)} &\longrightarrow p(B) = 1 - p = q \end{aligned} \quad (2.106)$$

Un proceso de estas características se denomina prueba de Bernoulli, y la variable aleatoria discreta x asociada a este experimento que toma el valor 1 cuando ocurre el suceso A y 0 cuando ocurre el B recibe el nombre de variable de Bernoulli. Un sinnúmero de sucesos aleatorios reales son reducibles en último término a este esquema. Es inmediato verificar que los momentos de la distribución de Bernoulli dependen únicamente de p , en particular $E\{x\} = p$ y $s = \sqrt{p \cdot q}$. En efecto, el valor esperado de la variable anterior

$$E\{x\} = \sum_i p_i x_i = 1 \cdot p + 0 \cdot q = p \quad (2.107)$$

y

$$\begin{aligned} \sigma^2 = \sum_i (x - E\{x\})^2 \cdot p_i &= (1 - p)^2 p + p + (0 - p)^2 (1 - p) = \\ &= p(1 - p) = pq \end{aligned} \quad (2.108)$$

Consideremos ahora un experimento de Bernoulli consistente en la realización de n pruebas sucesivas e independientes de Bernoulli, cada una de las cuales puede arrojar dos resultados excluyentes A (éxito) y B (fracaso), y definamos la variable aleatoria discreta:

$$X = \text{"número de veces que ocurre el suceso } A \text{ en los } n \text{ ensayos"} \quad (2.109)$$

Evidentemente, el espacio muestral de la variable anterior es el conjunto discreto

$$\Omega = \{0, 1, 2, \dots, n\} \quad (2.110)$$

Para la obtención de la función de probabilidad asociada a la variable aleatoria anterior consideremos el suceso con r éxitos y $n - r$ fracasos, $A^r B^{n-r}$. La probabilidad de este elemento del espacio muestral es, por la hipótesis de independencia estadística entre pruebas sucesivas, $p^r(1-p)^{n-r}$. Obviamente, el orden en el que se hayan obtenido los elementos es irrelevante, pues únicamente importa su número. Así pues, son sucesos compuestos admisibles todos aquellos construidos a partir del suceso $A^r B^{n-r}$ permutando los éxitos y los fracasos. El total de permutaciones de los r éxitos en un conjunto de n elementos es

$$\frac{n!}{r!(n-r)!} = \binom{n}{r} \text{ combinaciones.} \quad (2.111)$$

Consecuentemente la probabilidad de obtener r éxitos A en n intentos será

$$p(x = r) = \binom{n}{r} p^r (1-p)^{n-r}, \quad (2.112)$$

distribución que se conoce como *distribución binomial*, y que posee, como veremos, una notable importancia en la teoría estadística. Verifiquemos que se trata de una función de probabilidad admisible y obtengamos sus primeros momentos.

a) Condición de normalización: Obviamente, para ser una función de probabilidad admisible, la distribución binomial en la Ec. (2.112) debe verificar los axiomas de Kolmogorov que introdujimos al inicio del presente capítulo. Es evidente que $p(x = r) \geq 0$, y que se verifica la relación de adición de probabilidad de sucesos aleatorios excluyentes, por lo que nos resta verificar que la distribución binomial está normalizada. Para ello debemos obtener el valor de la suma de las probabilidades de que la variable aleatoria tome cada uno de los valores del espacio muestral

$$\sum_{r=0}^n p(x = r) = \sum_{r=0}^n \binom{n}{r} p^r q^{n-r} \quad (2.113)$$

y comprobar si es igual a uno. Usando el teorema del binomio de Newton¹⁵

¹⁵Recordamos al lector que este resultado establece que

$$(a + b)^n = \sum_{r=0}^n \binom{n}{r} p^r q^{n-r} \quad (2.114)$$

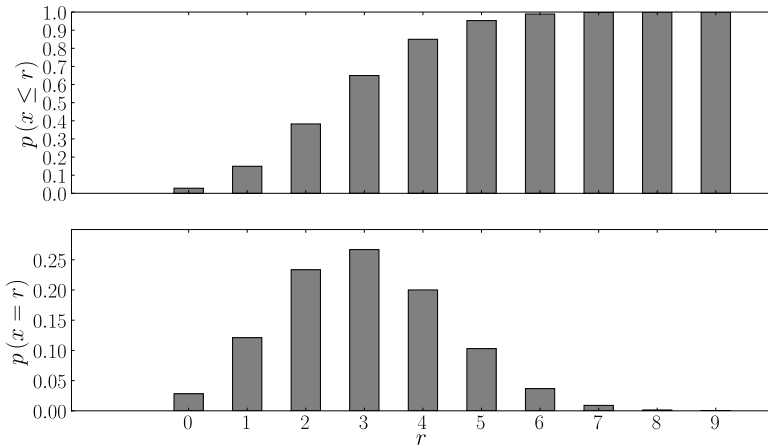


Figura 2.2: Distribución binomial para $n = 10$, $p = 0,3$. Obsérvese que es claramente asimétrica. Este efecto es menor cuanto mayor sea n .

tenemos

$$\sum_{r=0}^n p(x=r) = (p+q)^n = 1^n = 1, \quad (2.115)$$

lo que demuestra que la distribución binomial se encuentra normalizada y es por lo tanto una buena función de probabilidad.

- b) *Valor esperado*: Comprobemos que la esperanza matemática de la variable aleatoria binomial es $E\{x\} = np$. Usando la definición del valor esperado de la variable, obtenemos:

$$\begin{aligned} E\{x\} &= \sum_{k=0}^n kp(x=k) = \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k}. \end{aligned} \quad (2.116)$$

Haciendo simultáneamente $k \rightarrow k' = k - 1$ y $n \rightarrow n' = n - 1$, obtenemos de manera directa

$$E\{x\} = (np) \sum_{k'=0}^{n'} \frac{n'!}{k'!(n'-k')!} p^{k'} q^{n'-k'}, \quad (2.117)$$

y usando la condición de normalización, la suma anterior es igual a la unidad. Por lo tanto:

$$E\{x\} = np. \quad (2.118)$$

- c) *Varianza y desviación típica*: Veamos que la desviación típica de la distribución binomial es $\sigma^2 = npq$. En efecto, en la varianza de la distribución binomial se obtiene de la manera usual

$$\begin{aligned}
 \sigma^2 &= \sum_{k=0}^n (k - E\{x\})^2 p(x = k) = \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k q^{n-k} \\
 &= \sum_{k=0}^n [k^2 - 2npk + (np)^2] \binom{n}{k} p^k q^{n-k} = \\
 &= \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k} - 2np \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} \\
 &\quad + (np)^2 \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}.
 \end{aligned} \tag{2.119}$$

Usando la condición de normalización, $\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = 1$, y la media de la distribución binomial, $\sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np$, obtenemos

$$\begin{aligned}
 \sigma^2 &= \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k} - 2npE\{x\} + (np)^2 \\
 &= \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k} - (np)^2
 \end{aligned} \tag{2.120}$$

Finalmente, usando el hecho de que $k^2 = k(k-1) + k$, podemos reexpresar la suma del miembro de la derecha como:

$$A = \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^n k(k-1) \binom{n}{k} p^k q^{n-k} + \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} \tag{2.121}$$

que podemos escribir como

$$\begin{aligned}
 A &= \sum_{k=2}^n k(k-1) \binom{n}{k} p^k q^{n-k} + np = \\
 &= p^2 \sum_{k=2}^n \frac{n!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} + np
 \end{aligned} \tag{2.122}$$

Usando las transformaciones

$$\begin{aligned}
 k &\rightarrow k' = k - 2 \\
 n &\rightarrow n' = n - 2
 \end{aligned} \tag{2.123}$$

podemos hacer:

$$\begin{aligned} A &= n(n-1)p^2 \sum_{k'=2}^{n'} \frac{n!}{k'!(n'-k')!} p^{k'} q^{n'-k'} + np = \\ &= (np)^2 - np^2 + np \end{aligned}$$

Por lo tanto, sustituyendo el resultado anterior en la Ec. (2.120) la desviación típica de la variable binomial será

$$s = \sqrt{\sigma^2} = \sqrt{(np)^2 - np^2 + np - (np)^2} = \sqrt{np(1-p)} = \sqrt{npq} \quad (2.124)$$

Ley de los grandes números

Normalmente no conocemos a priori la probabilidad de un determinado evento, sino que la distribución ha sido determinada a partir de resultados experimentales correspondientes a distribuciones de frecuencia. Debemos tener un criterio para discernir cuándo estamos asignando una probabilidad de manera adecuada a un suceso. Aunque en el próximo capítulo analizaremos con más detalle el proceso de estimación de la probabilidad de un evento aleatorio a partir de su frecuencia observada en un experimento aleatorio, introduciremos ahora un tratamiento elemental de este problema en el contexto de la distribución binomial para analizar las implicaciones de la denominada ley de los grandes números. Ésta establece que, cuando el número de ensayos de un experimento aleatorio tiende a infinito, los valores de las frecuencias relativas de los diferentes sucesos del espacio muestral, $f(A)$, tienden a estabilizarse en torno a unos valores que son las probabilidades de dichos sucesos, $p(A)$. Esta ley se conoce también como ley fundamental del azar y es un resultado central dentro del formalismo estadístico. Tanto es así que se conoce como primer teorema fundamental de la probabilidad. Demostremos este resultado para el caso particular de la distribución binomial.

Consideremos un fenómeno aleatorio que sigue una distribución binomial. Si realizamos un experimento de Bernouilli, la frecuencia del evento de tipo A (éxito) en el experiencia será

$$f_A = \frac{1}{n} x_A \quad (2.125)$$

donde x_A es el número de veces, del total de n observaciones del experimento, en que se ha obtenido el suceso A . A diferencia de la probabilidad, la magnitud anterior es una variable aleatoria, ya que depende de la muestra analizada y de los sucesos observados. Es posible entonces determinar los momentos de su distribución de probabilidad, particularmente los dos primeros: la media y la varianza. La primera de ellas toma el valor

$$E\{f_A\} = E\left\{\frac{n_A}{n}\right\} = \frac{E\{x_A\}}{n}. \quad (2.126)$$

Recordando que el número de aciertos en un experimento de Bernouilli es una variable binomial, es obvio que $E\{x_A\} = np$ y por lo tanto:

$$E\{f_A\} = p \quad (2.127)$$

de tal manera que vemos que el promedio de todas las frecuencias relativas del suceso A que podamos obtener en una serie de experimentos de Bernoulli será justamente la probabilidad de dicho suceso. Análogamente podemos obtener la varianza de la variable aleatoria f_A :

$$\begin{aligned}\sigma^2(f_A) &= \sigma^2\left(\frac{x_A}{n}\right) = \frac{1}{n^2}\sigma^2(x_A) = \frac{1}{n^2}npq \Rightarrow \\ \Rightarrow \sigma^2(f_A) &= \frac{p(1-p)}{n}\end{aligned}\quad (2.128)$$

El producto $p(1-p)$ en la distribución anterior está acotado, por lo que podemos asegurar que la desviación estándar de la frecuencia de observación de un fenómeno determinado tiende a cero a medida que el número de observaciones crece. Así pues:

$$\sigma(f_A) \sim \frac{1}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} 0 \quad (2.129)$$

Frecuentemente, el propósito de una investigación experimental es la obtención de la probabilidad de un determinado suceso. De acuerdo con lo anterior, podemos tomar la frecuencia de ocurrencia del mismo como una estimación de la probabilidad del fenómeno y la incertidumbre de dicha asignación se irá reduciendo a medida que vayamos incrementando el número de observaciones, esto es, para muestras muy grandes las frecuencias se estabilizan en torno a un valor bien definido que asociamos a las probabilidades de los eventos¹⁶. La incertidumbre de esta aproximación es inversamente proporcional a la raíz cuadrada del número de experimentos individuales. Este tipo de incertidumbre que tiene su origen en el número finito de ensayos se denomina incertidumbre estadística, y tiene una enorme relevancia en experimentos de conteo (e.g. conteo de partículas nucleares, animales con determinadas características hereditarias, artículos defectuosos en experimentos de control, clientes que compran un determinado producto e un centro comercial, etc.).

Ejemplo 2.12

Se sabe que una fracción de $\frac{1}{200}$ de un determinado tipo de moscas desarrolla una propiedad A cuando se expone a rayos X. Se quiere determinar dicha fracción con una precisión del 1%. ¿Qué tamaño debe tener la muestra empleada?

De acuerdo con la fracción conocida, $p_A = 0,005$, $1 - p_A = 0,995 \simeq 1$. Queremos ajustar n de manera que $\sigma(f_A) = 0,001 f_A$ (incertidumbre relativa de la frecuencia). Luego:

$$\begin{aligned}\sigma^2(f_A) &= \left[\frac{0,001}{200}\right]^2 = \frac{1}{n}0,005(1 - 0,005) \Rightarrow \\ \Rightarrow n &= 1,99 \cdot 10^6 \sim 2,10^6\end{aligned}$$

¹⁶En el capítulo siguiente diremos que la frecuencia del suceso A es un estimador fiel y consistente de la probabilidad del suceso.

Se necesitará comprobar la ocurrencia de la mutación A en unos 2 millones de moscas para dar la fracción con un error menor o igual al 1 %.

2.6.2. Distribución geométrica o de Pascal

Íntimamente relacionada con la distribución binomial que acabamos de introducir está la denominada distribución geométrica, que describe el comportamiento estadístico de la denominada variable geométrica:

$$X = \text{"número de intentos hasta el primer fracaso"} \quad (2.130)$$

en un experimento de Bernouilli¹⁷. El espacio muestral de la variable aleatoria anterior es:

$$\Omega = \{1, 2, \dots\} \quad (2.131)$$

La probabilidad de que tengamos r éxitos es:

$$p(x = r) = p^r q \quad (A^{r-1}AB) \quad (2.132)$$

Esta distribución se conoce como distribución geométrica o de Pascal. Analizamos a continuación la normalización y primeros momentos de la distribución. En primer lugar, es inmediato verificar que la distribución está correctamente normalizada. En efecto, si consideramos la suma

$$\sum_{r=1}^{\infty} p(x = r) = \sum_{r=1}^{\infty} p^{r-1} q = q \sum_{r=1}^{\infty} p^{r-1} = q \frac{1}{1-p} = 1, \quad (2.133)$$

donde hemos usado la suma de una progresión geométrica, $\sum_{n=1}^{\infty} a^n = \frac{1}{1-a}$. Por otro lado, consideremos el valor esperado de la variable aleatoria de Pascal

$$E\{r\} = \sum_{r=1}^{\infty} r p^{r-1} q = \sum_{r=1}^{\infty} r p^{r-1} (1-p), \quad (2.134)$$

teniendo en cuenta que

$$\sum_{r=0}^{\infty} p^r q = 1 \quad \frac{d}{dp} \sum_{r=1}^{\infty} p^r (1-p) = 0, \quad (2.135)$$

por lo que

$$\sum_{r=0}^{\infty} r p^{r-1} (1-p) - \sum_{r=0}^{\infty} p^r = 0 \Rightarrow \sum_{r=0}^{\infty} r p^{r-1} q = \frac{1}{q}. \quad (2.136)$$

¹⁷En ocasiones se define como $X = \text{"número de éxitos antes del primer fracaso"}$. En ese caso el espacio muestral comienza en 0 y $p(x = r) = p^r q$. La media es una unidad inferior a la aquí descrita, y la varianza no se ve modificada.

Sustituyendo la expresión anterior en la expresión del valor esperado de la variable geométrica (2.134) resulta que

$$E\{r\} = \frac{1}{q} \quad (2.137)$$

Por idénticos procedimientos podemos obtener la varianza de la distribución:

$$\begin{aligned} \sigma^2 &= \sum_{r=1}^{\infty} \left(r - \frac{1}{1-p} \right)^2 p^{r-1} (1-p) \\ &= \sum_{r=1}^{\infty} \left[r^2 - \frac{2r}{1-p} + \frac{1}{(1-p)^2} \right] p^{r-1} (1-p) \\ &= \sum_{r=1}^{\infty} r^2 p^{r-1} (1-p) - 2 \underbrace{\sum_{r=1}^{\infty} r p^{r-1}}_{1/(1-p)^2} + \frac{1}{1-p} \underbrace{\sum_{r=1}^{\infty} p^{r-1}}_{1/1-p}. \end{aligned} \quad (2.138)$$

Ahora bien, considerando que

$$\sum_{r=1}^{\infty} r^2 p^{r-1} (1-p) \equiv \sum_{r=1}^{\infty} r(r-1) p^{r-1} (1-p) + \underbrace{\sum_{r=1}^{\infty} r p^{r-1} (1-p)}_{\frac{1}{1-p}}, \quad (2.139)$$

y teniendo en cuenta las relaciones

$$\begin{aligned} \sum_{r=0}^{\infty} r p^{r-1} &= \frac{1}{(1-p)^2} \\ \sum_{r=0}^{\infty} r(r-1) p^{r-2} &= \frac{2}{(1-p)^3} \\ \sum_{r=0}^{\infty} r(r-1) p^{r-1} &= \frac{2p}{(1-p)^3} \\ \Rightarrow \sum_{r=0}^{\infty} r(r-1) p^{r-1} (1-p) &= \frac{2p}{(1-p)^2} \end{aligned} \quad (2.140)$$

obtenemos la varianza de la distribución geométrica o de Pascal:

$$\sigma^2 = \frac{2p}{(1-p)^2} + \frac{1}{(1-p)} - \frac{1}{(1-p)^2} = \frac{2p + 1 - p - 1}{(1-p)^2} = \frac{p}{(1-p)^2} \quad (2.141)$$

2.6.3. Distribución de Poisson

Consideremos un experimento en el que observamos la aparición de sucesos puntuales sobre un soporte continuo (espacio o tiempo), como por ejemplo al analizar la ocurrencia de averías de máquinas en el tiempo, llegada de aviones a un aeropuerto, defectos en una plancha de metal, impactos de partículas en un blanco, etc. Las características principales de este tipo de procesos son:

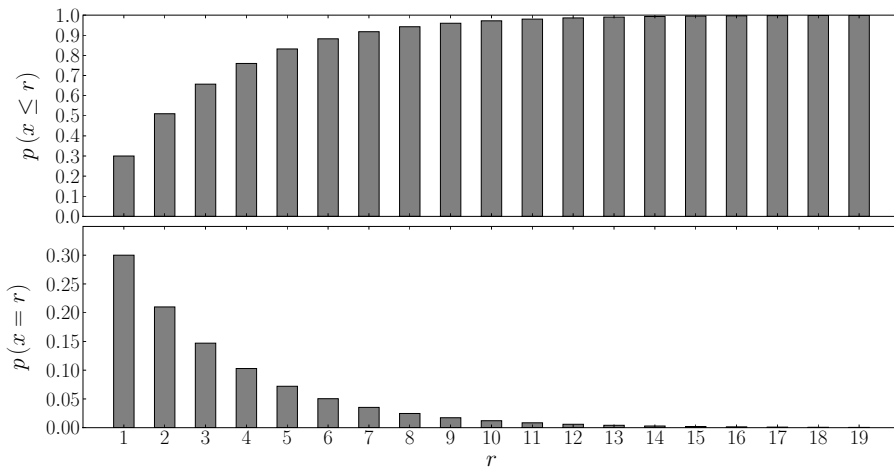


Figura 2.3: Distribución de Pascal para $p = 0,7$. La caída es más suave cuanto mayor sea p .

1. Son estables, i.e. producen a largo plazo un número medio de sucesos constante por unidad de observación.
2. Los sucesos aparecen aleatoriamente de manera independiente, i.e. el proceso no tiene memoria.

Estos procesos se denominan procesos de Poisson y, como veremos, son la generalización de los correspondientes procesos binomiales.

Definamos la variable aleatoria discreta (para un proceso de Poisson):

$$X = \text{”número de sucesos en un intervalo de longitud fija”} \tag{2.142}$$

Dividiendo el intervalo en subintervalos de anchura tendiendo a cero, podemos ver la realización del proceso de Poisson como un experimento de Bernouilli de un número muy grande de pruebas con una probabilidad muy baja de que tenga lugar un suceso en cada uno de los ensayos. En consecuencia, la distribución de Poisson aparece como límite de la distribución binomial si suponemos que el número de elementos observados en un experimento de Bernouilli es muy grande y su número medio constante, i.e.

$$\left. \begin{array}{l} n \rightarrow \infty \\ p \rightarrow 0 \end{array} \right\} p \cdot n = \lambda = \text{cte.} \tag{2.143}$$

En función de la variable λ anterior podemos reexpresar la distribución binomial de la forma:

$$p(x = r) = \binom{n}{r} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} \tag{2.144}$$

Tomando límites:

$$\begin{aligned} \lim_{n \rightarrow \infty} p(x=r) &= \frac{\lambda^r}{r!} \lim_{n \rightarrow \infty} \frac{n(n-1)\dots(n-r+1)}{\left(1 - \frac{\lambda}{n}\right)^r n^r} \left(1 - \frac{\lambda}{n}\right)^n = \\ &= \frac{\lambda^r}{r!} \lim_{n \rightarrow \infty} \frac{n(n-1)\dots(n-r+1)}{(n-\lambda)^r} \left(1 - \frac{\lambda}{n}\right)^n \end{aligned} \quad (2.145)$$

y usando que

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\frac{n}{n-\lambda}\right) \left(\frac{n-1}{n-\lambda}\right) \dots \left(\frac{n-r+1}{n-\lambda}\right) &= 1 \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda}, \end{aligned}$$

obtenemos finalmente la distribución de Poisson:

$$p(x=r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad r = 0, 1, 2, \dots \quad (2.146)$$

Ejemplo 2.13

En un estudio de tráfico se considera la variable aleatoria $X =$ “número de accidentes en 100 horas de conducción en un tramo de carretera”. Supongamos que el número medio de estos accidentes por unidad de tiempo en un determinado grupo de conductores es λ . Podemos convertir la variable X en una variable binomial considerando las 100 horas divididas en un número muy grande de intervalos temporales, en cada uno de los cuales la probabilidad de sufrir un accidente es muy pequeña. Tomemos, por ejemplo como intervalo de referencia el minuto. Tendremos entonces 6000 repeticiones de la observación, consistentes en verificar cada minuto si un accidente ha ocurrido o no en ese minuto. Obviamente el proceso de observación se ha convertido en un experimento aleatorio de Bernoulli con $n = 6000$ ensayos (lo que en la práctica equivale a que $n \rightarrow \infty$) y X en una variable binomial. La probabilidad de accidente en una prueba determinada será

$$E\{x\} = \lambda = np \Rightarrow p = \frac{\lambda}{n} \quad (2.147)$$

que, obviamente es muy pequeña ($p \rightarrow 0$).

Verifiquemos, al igual que en el caso de la distribución binomial, que la distribución de Poisson en (2.146) es una distribución de probabilidad bien definida y obtengamos sus principales momentos.

a) Normalización: Es inmediato verificar que $p(x=r) \geq 0$, por lo que, para comprobar que la distribución (2.146) verifica los axiomas de Kolmogorov,

debemos comprobar que la distribución de Poisson se encuentra normalizada. Sumemos a tal efecto $p(x = r)$ para todos los posibles valores de la variable:

$$\sum_{r=0}^{\infty} p(x = r) = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} e^{-\lambda} = e^{\lambda} e^{-\lambda} = 1 \quad (2.148)$$

b) *Valor esperado de la variable aleatoria de Poisson:* La media de la variable aleatoria x es:

$$\begin{aligned} E\{x\} &= \sum_{r=0}^{\infty} r p(x = r) = \sum_{r=1}^{\infty} r \frac{\lambda^r}{r!} e^{-\lambda} = \lambda \sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} e^{-\lambda} \\ &= \lambda e^{\lambda} e^{-\lambda} = \lambda \end{aligned} \quad (2.149)$$

c) *Varianza:* Usando la fórmula usual de la varianza obtenemos

$$\begin{aligned} \sigma^2(x) &= \sum_{r=0}^{\infty} p(x = r) (x - E\{x\})^2 = \sum_{r=0}^{\infty} (r - \lambda)^2 p(x = r) \\ &= \sum_{r=0}^{\infty} (r^2 - 2\lambda r + \lambda^2) p(x = r) \\ &= \sum_{r=0}^{\infty} r^2 p(x = r) - 2\lambda \sum_{r=0}^{\infty} r p(x = r) + \lambda^2 \sum_{r=0}^{\infty} p(x = r) = \\ &= \sum_{r=0}^{\infty} r^2 p(x = r) - \lambda^2. \end{aligned} \quad (2.150)$$

Usando que $r^2 = r(r-1) + r$ podemos escribir la suma de la ecuación anterior como

$$\begin{aligned} \sum_{r=0}^{\infty} r^2 p(x = r) &= \sum_{r=0}^{\infty} r^2 \frac{\lambda^r}{r!} e^{-\lambda} = \sum_{r=0}^{\infty} [r(r-1) + r] \frac{\lambda^r}{r!} e^{-\lambda} = \\ &= e^{-\lambda} \left[\sum_{r=2}^{\infty} \frac{\lambda^r}{(r-2)!} + \sum_{r=2}^{\infty} r \frac{\lambda^r}{r!} \right] = \\ &= e^{-\lambda} \left[\lambda^2 \sum_{r=2}^{\infty} \frac{\lambda^{r-2}}{(r-2)!} + \lambda \sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} \right] = \\ &= e^{-\lambda} (\lambda^2 + \lambda) e^{\lambda} = \lambda(\lambda + 1) \end{aligned} \quad (2.151)$$

Luego, la varianza buscada será

$$\sigma^2(x) = \lambda^2 + \lambda - \lambda^2 = \lambda \quad (2.152)$$

Ejemplo 2.14

En un gran aeropuerto, el número medio de operaciones de despegue o aterrizaje sigue una distribución de Poisson con media 120 operaciones por hora. Obténgase la probabilidad de:

- a) Ninguna operación en un minuto.
- b) Menos de cinco en un minuto.
- c) Menos de seis en cuatro minutos.

Sol:

$$p(x = r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad ; \quad \lambda = 2 \quad (\text{por minuto})$$

$$p(x = 0) = \frac{2^0}{0!} e^{-2} = \frac{1}{e^2} = 0,140$$

$$p(x \leq 4) = p(x = 0) + p(x = 1) + p(x = 2) + p(x = 3) + p(x = 4) =$$

$$= e^{-2} \left(1 + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!} \right) = 0,947$$

La media de operaciones en cuatro minutos será $\lambda' = 4\lambda = 8$ operaciones/minuto, por lo que las operaciones en intervalos de un mes siguen la distribución de Poisson:

$$p(x = r) = \frac{8^r}{r!} e^{-8} \quad (2.153)$$

Luego:

$$p(x \leq 5) = e^{-8} \left(1 + \frac{8^1}{1!} + \frac{8^2}{2!} + \frac{8^3}{3!} + \frac{8^4}{4!} + \frac{8^5}{5!} + \dots \right) = 0,191 \quad (2.154)$$

2.6.4. Distribución exponencial

La variable aleatoria exponencial resulta al considerar en un proceso de Poisson la variable aleatoria continua¹⁸,

$$t = \text{"tiempo transcurrido entre dos sucesos consecutivos"} \quad (2.155)$$

El espacio muestral asociado a la variable anterior es $\Omega = (0, \infty)$. La distribución de probabilidad de la variable anterior puede obtenerse fácilmente a partir de la distribución de Poisson para una variable $X = \text{"número de sucesos por unidad de tiempo"}$. En efecto, consideremos

$$P(t > t_0) = P(X = 0 \text{ en } (0, t_0)) = e^{-\lambda t_0}, \quad (2.156)$$

¹⁸También puede utilizarse para procesos de Poisson que ocurren en el espacio.

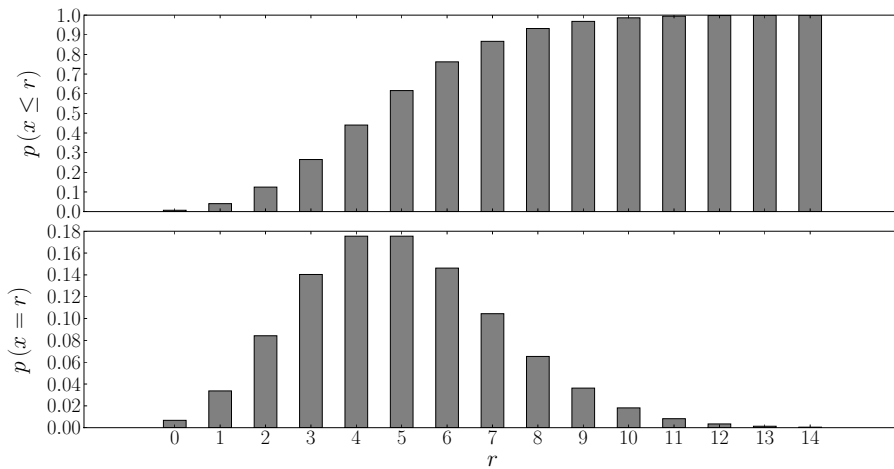


Figura 2.4: Distribución de Poisson para $\lambda = 5$.

donde λ es la media de la variable X , $\lambda = E\{x\}$, número de sucesos por unidad de tiempo. La función de distribución asociada es

$$F(t_0) = p(t \leq t_0) = 1 - p(t > t_0) = 1 - e^{-\lambda t_0}, \quad (2.157)$$

por lo que la función densidad asociada a la variable aleatoria t será

$$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t} \quad \lambda > 0 ; t > 0 \quad (2.158)$$

que se conoce como distribución exponencial. Si la distribución de Poisson es el equivalente de la distribución binomial para procesos en soporte continuo, la exponencial es el equivalente continuo de la distribución geométrica.

2.6.5. Distribución normal o gaussiana

Introducimos ahora una distribución de probabilidad especialmente importante, la distribución normal o gaussiana. Son realmente una infinidad los fenómenos que siguen esta distribución, hasta el punto de que cuando no disponemos de información acerca de la distribución de probabilidad que sigue un fenómeno, suponemos inicialmente que se somete a la distribución normal. Veremos posteriormente que esto es consecuencia del teorema de límite central y del carácter estable de la distribución gaussiana.

Una variable aleatoria X se dice normal o gaussiana de media μ y desviación típica σ ($X \in N(\mu, \sigma)$), si su función de densidad de probabilidad asociada es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]. \quad (2.159)$$

cuya representación gráfica para $\mu = 1$ y $\sigma = 0$ puede verse en la Fig. 2.5. La función de distribución asociada a la variable gaussiana anterior vendrá dada

por:

$$P(X \leq x) = F(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x dt \exp \left[-\frac{(t-\mu)^2}{2\sigma^2} \right] \quad (2.160)$$

La integral anterior no puede evaluarse en términos de funciones elementales, ya que la función $f(x) = \exp(-x^2)$ carece de función primitiva así expresable. Para su evaluación se define la función de error

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dn e^{-n^2}, \quad (2.161)$$

que se encuentra tabulada.

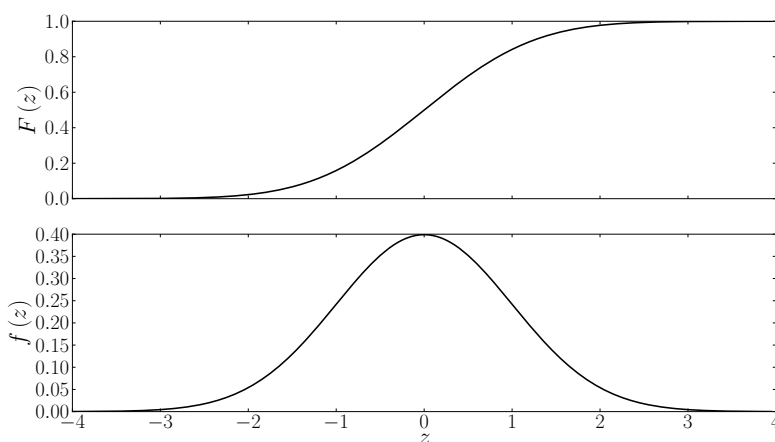


Figura 2.5: Distribución normal estándar.

Habitualmente se trabaja con la variable aleatoria normal estándar, Z , construida a partir de una variable gaussiana arbitraria $X \in N(\mu, \sigma)$ mediante la transformación

$$z = \frac{(x - \mu)}{\sigma}, \quad (2.162)$$

que produce una variable aleatoria de media nula y desviación típica unidad, $Z \in N(0, 1)$. En virtud de lo dicho para transformaciones generales de variables aleatorias, la función de densidad de probabilidad de la nueva variable transformada se obtendrá a partir de la anterior de la forma

$$f_z(z) = f_x(x) \left| \frac{dx}{dz} \right|, \quad (2.163)$$

por lo que la función de densidad de la variable aleatoria normal estándar será

$$f_z(z) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{z^2}{2} \right], \quad (2.164)$$

donde hemos usado que $\left|\frac{dx}{dz}\right| = \sigma$. Tendremos entonces la relación

$$P(X \leq x) = P\left[Z \leq \frac{x - \mu}{\sigma}\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{(x-\mu)}{\sigma}} dz \exp\left[-\frac{z^2}{2}\right], \quad (2.165)$$

que nos permite obtener la probabilidad de los diferentes sucesos aleatorios en términos de la función de error en (2.161) que es lo que se encuentra tabulado. De modo inverso, la distribución de una variable X siempre se puede obtener a partir de la distribución normal estándar de modo que

$$f_x(x) \equiv \frac{1}{\sigma} f_z\left(\frac{x - \mu}{\sigma}\right) \quad (2.166)$$

siendo

$$f_z(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]. \quad (2.167)$$

También se puede obtener la distribución de probabilidad acumulativa ya que

$$\begin{aligned} F_x(x; \mu; \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x dt \exp\left[-\frac{(t - \mu)^2}{2\sigma^2}\right] = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} \exp\left[-\frac{z^2}{2}\right] dz = F_z(z; 0; 1) \end{aligned} \quad (2.168)$$

De la misma forma:

$$\begin{aligned} P(a \leq X \leq b) &= \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\left[-\frac{(x - \mu)^2}{\sigma^2}\right] dx = \\ &= \frac{1}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \exp\left[-\frac{z^2}{2}\right] dz = P\left(\frac{a - \mu}{\sigma} \leq z_1 \leq \frac{b - \mu}{\sigma}\right) = \\ &= F_z\left(\frac{b - \mu}{\sigma}; 0, 1\right) - F_z\left(\frac{a - \mu}{\sigma}; 0, 1\right) \end{aligned} \quad (2.169)$$

Ejemplo 2.15

Un supermercado hace un estudio de producción de un determinado bien y obtiene que la demanda mensual es de 200 unidades y la desviación estándar es igual a 40 unidades. Si consideramos la demanda como normal, ¿cuál debe ser la fabricación mensual para que la probabilidad de que se agote no sea mayor de un 5%?

Supondremos que X = "número de unidades demandadas" sigue una distribución $N(200, 40)$, pretendemos encontrar el percentil del 95%, $x_{0,95}$, que verifica

$$P(X > x_{0,95}) = 0,05; \quad P(X \leq x_{0,95}) = 0,95 \quad (2.170)$$

Realizando el cambio a una distribución normal estándar, la variable $z = \frac{x-200}{40}$ tiene una distribución $N(0, 1)$, por lo tanto

$$P(X \leq x_{0,95}) = P(Z \leq z_{0,95}) = 0,95. \quad (2.171)$$

De las tablas obtenemos que

$$z_{0,95} \simeq 1,645 \quad (2.172)$$

por lo que

$$\begin{aligned} \frac{x_{0,95} - 200}{40} &= 1,645 \\ x_{0,95} &= 40 \times 1,645 + 200 = 266. \end{aligned}$$

Por tanto, fabricando 266 unidades a la semana, la probabilidad de que no se puedan atender las demandas es menor del 5%.

Función generatriz de momentos para la distribución normal estándar

Consideremos la función de densidad de probabilidad de una variable normal estándar, $Z \in N(0, 1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] \quad (2.173)$$

La función generatriz de momentos será

$$\begin{aligned} \mathcal{M}_z(t) &= E\{e^{zt}\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} e^{zt} dz = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}(z^2 - 2zt + t^2 - t^2)\right] dz = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}(z - t)^2\right] \exp\left[+\frac{t^2}{2}\right] dz = \\ &= \frac{1}{\sqrt{2\pi}} \exp\left[+\frac{t^2}{2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}n^2\right] dn = \exp\left[+\frac{t^2}{2}\right] \quad (2.174) \end{aligned}$$

que como vemos tiene la misma forma funcional que la distribución original¹⁹ Equivalentemente podríamos probar que la función característica de la distribución gaussiana estándar es $\phi(k) = e^{-k^2/2}$, sin más que hacer $t \rightarrow ik$.

¹⁹Es muy útil en Estadística la integral:

$$I_n = \int_{-\infty}^{+\infty} x^n \exp(-\alpha x^2) dx$$

Integrando por partes la distribución anterior es inmediato demostrar que verifica la relación de recurrencia:

$$I_{n+2} = \frac{n+1}{2\alpha} I_n$$

siendo

$$I_0 = \int_{-\infty}^{+\infty} \exp(-\alpha x^2) dx = \sqrt{\frac{\pi}{\alpha}}$$

En el caso de que la integral se extendiese únicamente al intervalo $[0, \infty)$, se incluiría un factor $1/2$ en la integral anterior.

Esta es una característica particularmente importante de esta distribución de probabilidad y se encuentra en la base de muchas de sus peculiares propiedades.

Como sabemos, a partir de la función generatriz de momentos podemos obtener los diferentes momentos de la distribución, ya que, como vimos,

$$E\{z^n\} = \left. \frac{d^n \mathcal{M}_z(t)}{dt^n} \right|_{t=0}. \quad (2.175)$$

En particular, la media de la variable aleatoria se obtiene a partir de la primera derivada de la función generatriz como

$$\mu = E\{z\} = \left. \frac{d\mathcal{M}_z(t)}{dt} \right|_{t=0} = \left[t \exp\left(+\frac{t^2}{2}\right) \right]_{t=0} = 0. \quad (2.176)$$

La segunda derivada proporciona el momento de segundo orden respecto al origen, que coincide en este caso con el segundo momento central al tratarse de una distribución de media nula:

$$\begin{aligned} \frac{d^2 \mathcal{M}_z(t)}{dt^2} &= \exp\left[+\frac{t^2}{2}\right] + t^2 \exp\left[+\frac{t^2}{2}\right] \\ \sigma^2 &= E\{z^2\} = \left. \frac{d^2 \mathcal{M}_z(t)}{dt^2} \right|_{t=0} = 1. \end{aligned} \quad (2.177)$$

De modo análogo, las siguientes derivadas de la función generatriz proporcionan los momentos de orden superior de la distribución:

$$\begin{aligned} \frac{d^3 \mathcal{M}_z(t)}{dt^3} &= t e^{t^2/2} + 2te^{t^2/2} + t^3 e^{t^2/2} \\ E\{z^3\} &= \left. \frac{d^3 \mathcal{M}_z(t)}{dt^3} \right|_{t=0} = 0 \\ \frac{d^4 \mathcal{M}_z(t)}{dt^4} &= e^{t^2/2} + t^2 e^{t^2/2} + 2e^{t^2/2} + 2t^2 e^{t^2/2} + 3t^2 e^{t^2/2} + t^4 e^{t^2/2} = \\ &= 3et^{2/2} + 6t^2 e^{t^2/2} + t^4 e^{t^2/2} \\ E\{z^4\} &= \left. \frac{d^4 \mathcal{M}_z(t)}{dt^4} \right|_{t=0} = 3 \end{aligned} \quad (2.178)$$

Es inmediato probar que, dada una distribución normal de media μ y de varianza σ^2

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}, \quad (2.179)$$

las funciones generatrices de momentos con respecto al origen y centrados son

$$\mathcal{M}_x(t) = e^{\mu t} \exp[\sigma^2 t^2/2], \quad (2.180a)$$

$$\mathcal{M}_{x-\mu}(t) = \exp[\sigma^2 t^2/2]. \quad (2.180b)$$

En ocasiones es también útil conocer el valor de la integral

$$I_0 = \int_{-\infty}^{+\infty} \exp(-\alpha x^2 + \beta x + \gamma) dx = e^{-\gamma} e^{-\frac{\beta^2}{4\alpha}} \sqrt{\frac{\pi}{\alpha}}$$

De nuevo se pueden obtener de aquí los momentos de orden arbitrario de la distribución, en este caso centrales.

Distribución normal como límite de la binomial

Es posible deducir la anterior distribución gaussiana a partir de la distribución binomial en el límite de n grande y p pequeña aunque finita, de tal manera que $\bar{r} = np$ se mantenga finito. Efectivamente, consideremos la distribución de probabilidad binomial para la variable de Bernoulli que se introdujo en secciones anteriores:

$$p(x = r) = \binom{n}{r} p^r q^{N-r}, \quad (2.181)$$

siendo $p + q = 1$ como es habitual. Tomemos el logaritmo neperiano de ambos miembros de la distribución anterior utilizando la aproximación de Stirling:²⁰

$$\ln p(x = r) \simeq n \ln n - r \ln r - (n - r) \ln (n - r) + r \ln p + (n - r) \ln q \quad (2.185)$$

Desarrollando la distribución anterior en serie de Taylor en torno al valor medio $\bar{r} = np$:

$$\ln p(x = r) = \ln p(x = \bar{r}) + \frac{\partial \ln p(x = \bar{r})}{\partial r} (r - \bar{r}) \quad (2.186)$$

$$+ \frac{1}{2!} \frac{\partial^2 \ln p(x = \bar{r})}{\partial r^2} (r - \bar{r})^2 + O(r - \bar{r})^3. \quad (2.187)$$

Derivando la distribución de probabilidad binomial en la ecuación (2.181) tenemos:

$$\begin{aligned} \frac{\partial \ln p(x = \bar{r})}{\partial r} &= -\ln \bar{r} + \ln [n(1-p)] + \ln p - \ln q = 0 \\ \frac{\partial^2 \ln p(x = \bar{r})}{\partial r^2} &= -\frac{1}{\bar{r}} - \frac{1}{n - \bar{r}} = -\frac{1}{npq}. \end{aligned} \quad (2.188)$$

Truncando el desarrollo en serie de Taylor en (2.187) a segundo orden en $(r - \bar{r})$ y usando los valores anteriores de las derivadas obtenemos:

$$\ln p(x = r) = \ln p(x = \bar{r}) - \frac{(r - \bar{r})^2}{2npq}, \quad (2.189)$$

²⁰La aproximación de Stirling establece que, para $N \rightarrow \infty$,

$$\ln N! \simeq N \ln N - N \quad (2.182)$$

Efectivamente, teniendo en cuenta que

$$\ln N! = \sum_{n=1}^N \ln n \quad (2.183)$$

y aproximando la suma por una integral a la manera convencional, tenemos:

$$\ln N! \simeq \int_1^N dn \ln n = [n \ln n - n]_1^N \simeq N \ln N - N \quad (2.184)$$

donde hemos despreciado la unidad frente a los términos en N .

de tal manera que:

$$p(x = r) = p(x = \bar{r}) \exp \left[-\frac{(r - \bar{r})^2}{2npq} \right]. \quad (2.190)$$

Teniendo en cuenta el valor de la varianza de la distribución binomial, $\sigma^2 = npq$, es inmediato obtener que:

$$p(x = r) = p(x = \bar{r}) \exp \left[-\frac{(r - \bar{r})^2}{2\sigma^2} \right]. \quad (2.191)$$

Tomando ahora el límite continuo para la variable aleatoria, $r \rightarrow x$, $p(x) \rightarrow f(x)$, y normalizando la distribución de probabilidad en la ecuación anterior tenemos:

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx &= p(x = \bar{x}) \int_{-\infty}^{+\infty} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma^2} \right] dx \\ &= p(x = \bar{x}) \sqrt{2\pi}\sigma = 1, \end{aligned} \quad (2.192)$$

donde hemos usado los valores conocidos de la integral gaussiana. Finalmente, sustituyendo la distribución en la ecuación (2.191) obtenemos la distribución gaussiana:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma^2} \right]. \quad (2.193)$$

Teorema del límite central

En muchas ocasiones los fenómenos aleatorios que se nos presentan son susceptibles de ser descompuestos en suma de contribuciones independientes más elementales. Para describirlos más formalmente, introducimos las siguientes definiciones:

1. *Variables aleatorias independientes e idénticamente distribuidas (iid)*: Un conjunto de variables aleatorias $\{x_i\}_{i=1}^N$ se dice independiente e idénticamente distribuido cuando las variables que lo forman tienen asociada la misma distribución de probabilidad y son independientes entre sí:

$$p(x_i | x_j) = p(x_i). \quad (2.194)$$

2. *Procesos aleatorios infinitamente divisibles*: Un proceso aleatorio x es infinitamente divisible cuando puede ser representado como una suma de n variables iid $\{x_i\}_{i=1}^n$, $\forall n \in \mathbb{Z}^+$:

$$x = \sum_{i=1}^n x_i. \quad (2.195)$$

Son muchas la situaciones prácticas en las que esto se produce, y como muestra basten los siguientes ejemplos.

Ejemplo 2.16

1. Tamaño de un objeto, que es la suma de las contribuciones de los tamaños de sus constituyentes fundamentales (átomos, moléculas, etc.), cuyo número puede ser macroscópicamente grande.
2. Desplazamiento de una partícula en un fluido, gobernado por las colisiones de la partícula con las moléculas del fluido que producen una probabilidad no nula de desplazarse en un sentido u otro en cada una de las tres direcciones del espacio (movimiento browniano). Tras n colisiones, su posición será

$$\vec{r} = \vec{r}_0 + \Delta\vec{r}; \quad \Delta\vec{r} = \sum_{i=1}^n \vec{r}_i. \quad (2.196)$$

Dados los tiempos característicos de los movimientos atómicos o moleculares en el seno de fluidos, es evidente que el número colisiones en intervalos de tiempo del orden de nuestro tiempo de observación (del orden de segundos) es muy grande ($n \rightarrow \infty$).

3. Variación del precio de las acciones de una compañía en el mercado financiero, que tras n intervalos de cotización será el resultado de lo que haya variado en cada uno de los intervalos por separado, δY_i

$$\Delta Y = \sum_{i=1}^n \delta Y_i. \quad (2.197)$$

A la hora de establecer cuál es la distribución de probabilidad de la variable $x = \sum_{i=1}^n x_i$, debemos tener en cuenta que, como hemos visto, la adición de variables aleatorias modifica en general la distribución de probabilidad, de manera que la distribución de probabilidad que sigue

$$x = x_1 + x_2 \quad (2.198)$$

es diferente a la que siguen tanto x_1 como x_2 , incluso aunque éstas sean la misma. Así, sería de esperar que aunque conociésemos la distribución de probabilidad de cada una de las contribuciones, x_i , por separado -algo que en la práctica raras veces sucederá- no podríamos afirmar nada sobre la distribución de probabilidad de x . Pues bien, el teorema del límite central que introducimos a continuación establece que todos los fenómenos que podamos describir en términos de una suma finita o infinita de variables aleatorias en la que ninguna de estas tenga distribuciones con varianza infinita, acaban sometiéndose a una

distribución de probabilidad de tipo gaussiano, distribución que actuaría por tanto como una especie de atractor en el espacio de las funciones de distribución. Este teorema es de tal importancia que en ocasiones recibe el nombre de *segundo teorema fundamental de la probabilidad*, siendo el primero la ley de los grandes números.

Teorema del límite central. Sean $\{X_i\}_{i=1}^n$ variables aleatorias iid con una distribución de probabilidad no especificada, cada una de ellas con media μ y varianza σ^2 finita. En estas condiciones, el promedio muestral

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.199)$$

es una variable aleatoria con media μ y varianza σ^2/n . Además, cuando $n \rightarrow \infty$, la distribución de \bar{x} tiende a ser una distribución normal $N(\mu, \sigma)$, o equivalentemente la variable

$$\bar{y} = \frac{(\bar{x} - \mu)}{\sigma/\sqrt{n}} \quad (2.200)$$

tiende a ser una distribución normal estándar $N(0, 1)$.

Demostración: Consideremos en primer lugar los momentos de la variable aleatoria \bar{x} . La media de esta variable aleatoria es

$$\begin{aligned} E\{\bar{x}\} &= \frac{1}{n}[E\{x_1\} + E\{x_2\} + \dots + E\{x_n\}] = \\ &= \frac{1}{n}[\mu + \mu + \dots + \mu] = \mu \end{aligned} \quad (2.201)$$

y su varianza se obtiene de la manera habitual como

$$\begin{aligned} E\{(\bar{x} - \mu)^2\} &= \frac{1}{n^2} E \left\{ \left(\sum_{i=1}^n x_i - n\mu \right)^2 \right\} = \\ &= E\{(x - \mu)^2\} = \frac{1}{n^2} E \left\{ \left[\sum_{i=1}^n (x_i - \mu) \right]^2 \right\} = \\ &= \frac{1}{n^2} \sum_{i=1}^n E \left\{ (x_i - \mu)^2 \right\} + 2 \sum_{i=1}^n \sum_{j=1}^n E \{ (x_i - \mu)(x_j - \mu) \} \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned} \quad (2.202)$$

donde hemos usado que $\text{cov}(x_i, x_j) = E \{ (x_i - \mu)(x_j - \mu) \} = 0$ por tratarse de variables aleatorias independientes.

Veamos finalmente que

$$\bar{y} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \rightarrow \infty} z \in N(0, 1), \quad (2.203)$$

para lo que comprobaremos que en el límite de un gran número de variables, la función generatriz de momentos (equivalentemente la función característica) de la variable aleatoria \bar{y} es la de la distribución gaussiana, i.e. que

$\lim_{n \rightarrow \infty} \mathcal{M}_{\bar{y}}(t) = \exp[t^2/2]$. En efecto, escribamos

$$\begin{aligned}\bar{y} &= \frac{\sqrt{n}\bar{x} - \mu}{\sigma} = \frac{\sqrt{n} \frac{1}{n} \sum_{i=1}^n x_i - \mu}{\sigma} \\ \bar{y} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)\end{aligned}\quad (2.204)$$

y denotemos por $z_i = (x_i - \mu)/\sigma$, variable aleatoria de media nula y varianza unidad $E\{z_i\} = 0$ y $E\{z_i^2\} = 1$. La función generatriz de momentos de la distribución de probabilidad asociada a la variable aleatoria \bar{y} será entonces

$$\begin{aligned}\mathcal{M}_{\bar{y}}(t) &= E\{\exp(t\bar{y})\} = E\left\{\exp\left(t \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i\right)\right\} \\ &= E\left\{\prod_{i=1}^n \exp\left(\frac{t}{\sqrt{n}} z_i\right)\right\}.\end{aligned}\quad (2.205)$$

Desarrollando en serie las exponenciales del producto anterior, tenemos

$$\exp\left[\frac{t}{\sqrt{n}} z_i\right] = 1 + \frac{t}{\sqrt{n}} z_i + \frac{t^2}{2} \left(\frac{z_i}{\sqrt{n}}\right)^2 + \frac{t^3}{3!} \left(\frac{z_i}{\sqrt{n}}\right)^3 + \dots, \quad (2.206)$$

por lo que

$$E\left\{\exp\left[\frac{t}{\sqrt{n}} z_i\right]\right\} = 1 + \frac{t}{\sqrt{n}} E\{z_i\} + \frac{t^2}{2n} E\{z_i^2\} + \frac{t^3}{3!n^{3/2}} E\{z_i^3\} + \dots \quad (2.207)$$

Dado que, como hemos visto, $E\{z_i\} = 0$ y $E\{z_i^2\} = 1$, la ecuación anterior se transforma en

$$E\left\{\exp\left[\frac{tz_i}{\sqrt{n}}\right]\right\} = 1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{3/2}} E\{z_i^3\} + \dots, \quad (2.208)$$

con lo cual, la función generatriz de momentos para \bar{y} será

$$\mathcal{M}_{\bar{y}}(t) = \prod_{i=1}^n \left[1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{3/2}} E\{z_i^3\} + \dots \right]. \quad (2.209)$$

Por hipótesis, $E\{z_1^3\} = E\{z_2^3\} = \dots = E\{z_i^3\}$ y así para el resto de los términos. En este caso,

$$\begin{aligned}\mathcal{M}_{\bar{y}}(t) &= \left[1 + \frac{1}{n} \underbrace{\left(\frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}} E\{z^3\} + \dots \right)}_a \right]^n \\ \mathcal{M}_{\bar{y}}(t) &= \left[1 + \frac{a}{n} \right]^n.\end{aligned}\quad (2.210)$$

Luego, finalmente llegamos a que la función generatriz de momentos de la variable \bar{y} será

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathcal{M}_{\bar{y}}(t) &= \lim_{n \rightarrow \infty} e^a = e^{\left[\frac{t^2}{2} + \frac{t^3}{3! \sqrt{n}} E\{z^3\} + \dots\right]} \\ &= \lim_{n \rightarrow \infty} e^{t^2/2}, \quad \text{q.e.d.}\end{aligned}\tag{2.211}$$

que coincide con la de la distribución gaussiana.

Este teorema puede generalizarse al caso en que las x_i , aun siendo independientes, no sigan la misma distribución de probabilidad, siempre que todas sus distribuciones tengan media y varianza finita y se cumplan ciertas condiciones (como el *criterio de Lindeberg*) que no discutiremos en detalle pero que a nivel intuitivo implican que ninguna variable hace contribuciones mucho mayores que las demás, i.e. todas las variables aleatorias en las que descomponemos el fenómeno global siguen distribuciones que tienen una escala característica (que se concreta en una varianza finita), de tal manera que ninguna de ellas realiza contribuciones extremas al conjunto.

El teorema del límite central no dice nada sobre cuán buena es la distribución gaussiana como aproximación a una suma finita de n variables, es decir, sobre la velocidad de convergencia al límite. Sin embargo, cuando la distribución de las x_i tiene, además de media y varianza, tercer momento central finito, el *teorema de Berry-Esséen* afirma que la distancia entre las funciones de distribución acumulativas exacta y gaussiana tiene una cota superior que decrece como $\frac{1}{\sqrt{n}}$.

Cuando las variables iid no cumplen la condición de tener los dos primeros momentos finitos, se puede generalizar el teorema introduciendo las distribuciones de Lévy, que trataremos más adelante en este capítulo.

2.6.6. Distribución log-normal

Como consecuencia del teorema del límite central, si una variable aleatoria y es el producto de n variables aleatorias independientes diferentes

$$y = x_1, x_2 \dots x_n\tag{2.212}$$

entonces su logaritmo podrá escribirse de la forma

$$\ln(y) = \ln(x_1) + \ln(x_2) + \dots + \ln(x_n),\tag{2.213}$$

por lo que se verificará que cuando $n \rightarrow \infty$, $\ln(y)$ tiende a tener un comportamiento gaussiano, es decir, la función de densidad de probabilidad asociada a $\ln(y)$ será de la forma $N(\mu, \sigma)$ siendo

$$\mu = \sum_i \ln(\mu_i).\tag{2.214}$$

Por lo tanto

$$\begin{aligned}f(z) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right] \\ z &= \ln(y) \quad y > 0.\end{aligned}\tag{2.215}$$

La función de densidad asociada a la variable aleatoria original, $g(y)$, podemos obtenerla a partir de la de la ecuación anterior mediante las reglas convencionales del cambio de variable

$$\begin{aligned} g(y) &= f(z) \left| \frac{dz}{dy} \right| \\ g(y) &= f[\ln y] \frac{1}{y}, \end{aligned} \quad (2.216)$$

por lo que la función de densidad de probabilidad de la variable y será:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(\ln y - \mu)^2}{2\sigma^2} \right] \frac{1}{y} \quad y > 0. \quad (2.217)$$

A esta nueva distribución de probabilidad se denomina distribución log-normal y, como la función logaritmo es monótona, los percentiles de la variable aleatoria y son los transformados a través del logaritmo de los percentiles de la variable z . Es evidente de la definición que la distribución log-normal sólo está bien definida para valores positivos de la variable aleatoria y , y que presenta un aspecto asimétrico (Fig. 2.6).

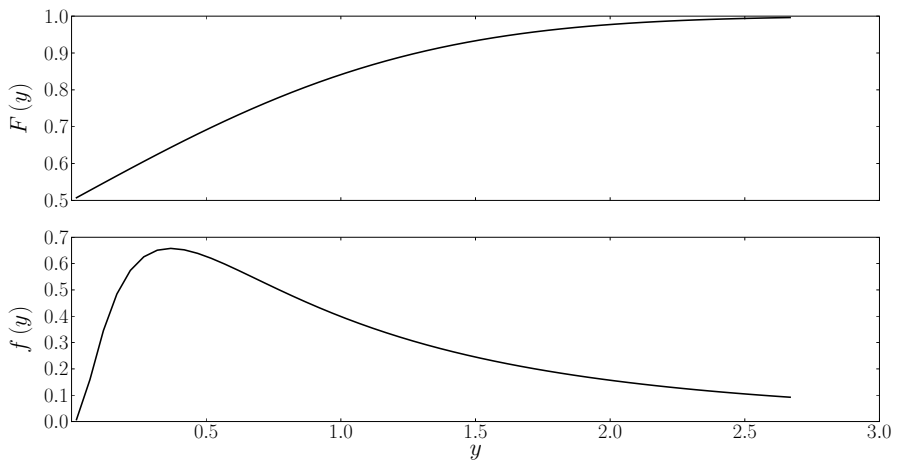


Figura 2.6: Distribución log-normal asociada a la normal estándar.

2.6.7. Distribución de Weibull

Uno de los problemas básicos sustanciales en la descripción de un buen número de fenómenos aleatorios es la tasa de fallo o de supervivencia, asociada a un proceso temporal. Este problema se presenta en un buen número de situaciones prácticas en las que debemos describir el tiempo de duración de algún elemento. En estas situaciones consideramos siempre una variable aleatoria continua t =" tiempo de duración del elemento", cuyo espacio muestral es

$\Omega = (0, \infty)$. Para la obtención de la distribución de probabilidad que gobierna esta variable, estudiemos la probabilidad de fallo entre t_0 y $t_0 + \Delta t$ condicionada a la supervivencia previa hasta t_0 :

$$P(t_0 \leq t \leq t_0 + \Delta t | t > t_0) = \frac{P(t_0 < t \leq t_0 + \Delta t)}{P(t > t_0)}. \quad (2.218)$$

Sean $f(t)$ y $F(t_0)$ la función densidad de probabilidad y la función de distribución de la variable t , respectivamente. Entonces

$$F(t_0) = \int_0^{t_0} f(t) dt ; \quad f(t) = \frac{dF(t)}{dt}, \quad (2.219)$$

y por tanto, para Δt pequeño,

$$\left. \begin{aligned} P(t_0 < t \leq t_0 + \Delta t) &= f(t_0) \Delta t \\ P(t > t_0) &= 1 - F(t_0) \end{aligned} \right\}. \quad (2.220)$$

Consecuentemente,

$$\frac{P(t_0 < t \leq t_0 + \Delta t | t > t_0)}{\Delta t} \simeq \frac{f(t_0)}{1 - F(t_0)}, \quad (2.221)$$

Observemos que esta es la probabilidad por unidad de tiempo de que se produzca el suceso en el tiempo t condicionado a que no se haya producido antes, $P(t_0 \leq t \leq t_0 + \Delta t | t > t_0)$. Luego podemos asignar una densidad de probabilidad de fallo condicionada a la supervivencia previa

$$\psi(t) = \frac{f(t)}{1 - F(t)}, \quad (2.222)$$

cuya función de distribución acumulativa será

$$\Psi(t) = \int_0^t \frac{f(s)}{1 - F(s)} ds = \int_0^t \psi(s) ds. \quad (2.223)$$

Dado que $dF(s)/ds = f(s)$, entonces

$$\Psi(t) = \int_0^t \frac{\frac{dF(s)}{ds}}{1 - F(s)} ds = - \{ \ln(1 - F(s)) \}_0^t, \quad (2.224)$$

o lo que es lo mismo,

$$\Psi(t) = -\ln [1 - F(t)] \Leftrightarrow F(t) = 1 - \exp[-\Psi(t)] \quad (2.225)$$

Derivando la ecuación anterior, obtendremos finalmente la función de densidad de probabilidad asociada a la variable aleatoria t buscada

$$f(t) = \psi(t) \exp[-\Psi(t)], \quad (2.226)$$

que, como vemos, depende de la forma que presente la tasa de fallo del problema concreto. Obviamente ésta debe ser una función creciente con el tiempo, ya

que la probabilidad de que el fallo se produzca debe aumentar con el paso del tiempo.

Los tres tipos de función más frecuentemente empleados en este tipo de problemas son los correspondientes a una tasa de fallo constante, una tasa de fallo potencialmente creciente y una tasa de fallo exponencialmente creciente. Los dos primeros casos se pueden representar por una tasa de fallo dada por una distribución potencial del tipo

$$\psi(t) = \beta t^{\alpha-1}, \quad (2.227)$$

a partir de la cual, usando la ecuación (2.226), se obtiene una densidad de probabilidad de fallo tras la supervivencia previa del sistema igual a

$$f(t) = \beta t^{\alpha-1} \exp \left[-\frac{\beta}{\alpha} t^\alpha \right], \quad (2.228)$$

distribución que se conoce como distribución de Weibull y de la cual la distribución exponencial es un caso particular ($\alpha = 1$) correspondiente a una tasa de fallo constante $\psi(t) = \beta$.

Finalmente, en el caso de que consideremos una tasa de fallo exponencialmente creciente con el tiempo, $\psi(t) = \beta e^{\alpha t}$, obtendremos una distribución para el tiempo de fallo condicionado a la supervivencia previa del sistema igual a

$$f(t) = \beta e^{\alpha t} \exp \left[-\frac{\beta}{\alpha} e^{\alpha t} + \frac{\beta}{\alpha} \right] \quad (2.229)$$

distribución denominada de Gompertz, que modela la duración de la vida humana por encima de los veinte años.

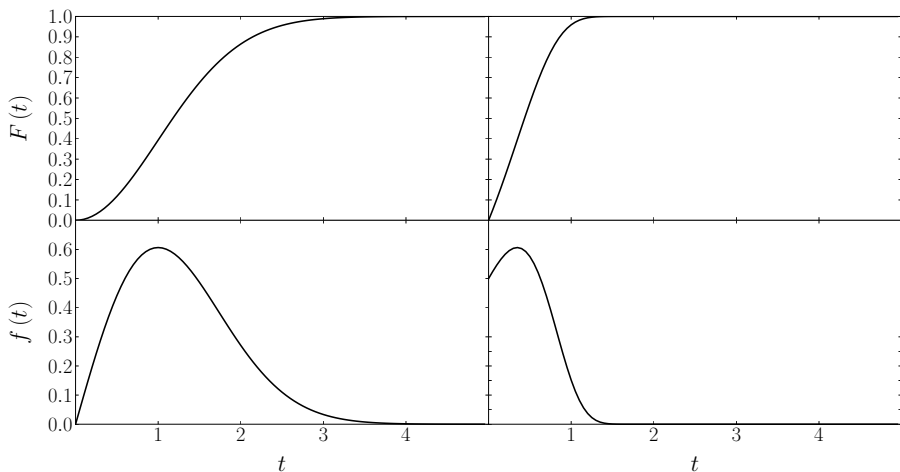


Figura 2.7: Distribuciones de Weibull (izquierda) y Gompertz para $\alpha = 2$ y $\beta = 1$.

2.6.8. Procesos aleatorios estables. Distribuciones de Lévy. Distribución de Pareto.

Como se ha mencionado, en general la distribución de probabilidad asociada a una suma de variables aleatorias es diferente de la función de distribución de probabilidad que las gobierna por separado. Sin embargo, existen determinados procesos aleatorios que muestran la propiedad de estabilidad, que consiste en la invariancia de la distribución de probabilidad frente a la adición de variables aleatorias.

Ejemplo 2.17

La distribución normal es estable en el sentido de que la distribución de probabilidad asociada a la suma de un conjunto de variables aleatorias gaussianas, sigue siendo una gaussiana. Efectivamente, sean x_1 y x_2 variables aleatorias gaussianas estadísticamente independientes, y supongamos, por simplicidad, que sus distribuciones de probabilidad asociadas son gaussianas normales estándar. Como ya hemos visto, en este caso la función característica de la variable aleatoria $x = x_1 + x_2$ es:

$$\phi(k) = \phi_1(k)\phi_2(k) = e^{-k^2} \quad (2.230)$$

donde hemos usado la función característica de la distribución gaussiana $\phi(k) = \exp(-k^2/2)$. Así pues, la función característica asociada a la suma de variables aleatorias es nuevamente la de una distribución gaussiana.

Para analizar estadísticamente este tipo de procesos debemos introducir la noción de *procesos aleatorios estables*. Como hemos mencionado anteriormente, la función característica asociada a la distribución de probabilidad conjunta es el producto de funciones características:

$$\phi(k) = \prod_{i=1}^N \phi_i(k), \quad (2.231)$$

o lo que es lo mismo:

$$\ln \phi(k) = \sum_{i=1}^N \ln \phi_i(k). \quad (2.232)$$

Teniendo en cuenta que la función generatriz de cumulantes es el logaritmo neperiano de la función característica, podemos escribir a partir de la ecuación anterior:

$$\begin{aligned} c_j(x) &= \frac{1}{i^n} \frac{d^j \ln \phi(k=0)}{dk^j} = \sum_{i=1}^N \frac{1}{i^n} \frac{d^j \ln \phi_i(k=0)}{dk^j} \\ &= \sum_{i=1}^N c_j(x_i) \end{aligned} \quad (2.233)$$

lo que constituye una demostración de la aditividad de los cumulantes, i.e. el cumulante de orden j asociado a la función de distribución de probabilidad de la variable aleatoria conjunta no es sino la suma de los cumulantes de orden j de cada una de las variables aleatorias.

En el caso de que tengamos un conjunto de N variables iid, las funciones de distribución de probabilidad y las funciones características de todas ellas coinciden, la función característica y los cumulantes resultantes verificarán:

$$\begin{aligned}\phi(k) &= [\phi_i(k)]^N \\ c_j(x) &= Nc_j(x_i)\end{aligned}\quad (2.234)$$

La clase de distribuciones de probabilidad que tienen la propiedad de que la suma de variables iid así distribuidas sigue la misma distribución se denomina *clase de distribuciones estables*, y los procesos aleatorios gobernados por estas distribuciones *procesos aleatorios estables*. Un ejemplo de esta clase es la distribución gaussiana, ya que si $\phi_i(k) = e^{-\frac{k^2\sigma^2}{2}}$, de acuerdo con (2.234) la función característica de la suma será $e^{-\frac{Nk^2\sigma^2}{2}}$, que claramente tiene la misma forma.

Lévy y Khintchine demostraron que la forma más general de la función característica de un proceso aleatorio estable es:

$$\ln \left[\sqrt{2\pi}\phi(k) \right] = \begin{cases} i\mu k - \gamma |k|^\alpha \left[1 - i\beta \frac{k}{|k|} \tan\left(\frac{\pi}{2}\alpha\right) \right] & \alpha \neq 1 \\ i\mu k - \gamma |k| \left[1 + i\beta \frac{k}{|k|} \frac{2}{\pi} \ln |k| \right] & \alpha = 1 \end{cases} \quad (2.235)$$

donde $0 < \alpha \leq 2$ es un factor de escala positivo, $\mu \in \mathbb{R}$ representa la media de la distribución, y $\beta \in [-1, 1]$ es un parámetro de asimetría. Hay que destacar que la expresión analítica de las distribuciones estables de Lévy²¹ se conoce únicamente para los casos siguientes:

- a) $\alpha = 2$, $\beta = 0$. En este caso la función característica es $\frac{1}{\sqrt{2\pi}}e^{i\mu k}e^{-\gamma k^2}$, que como ya sabemos (o se puede comprobar muy fácilmente invirtiendo la transformada de Fourier) se corresponde con una gaussiana de media μ y varianza 2γ .
- b) $\alpha = 1$, $\beta = 0$. La función característica de esta distribución será igual a $\frac{1}{\sqrt{2\pi}}e^{i\mu k}e^{-\gamma|k|}$, y su función de densidad de probabilidad el resultado de la integral:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\gamma|k|} e^{-ik(x-\mu)} dk, \quad (2.236)$$

que se resuelve inmediatamente partiendo en dos la región de integración en el cero, con el resultado:

²¹Conviene mencionar una diferencia entre nuestra nomenclatura y la de otros textos. Aquí se denominan distribuciones de Lévy todas las estables, mientras que otros autores aplican este nombre sólo a la de Lévy-Smirnov.

$$f(x) = \frac{1}{\pi\gamma} \frac{1}{1 + \left(\frac{x-\mu}{\gamma}\right)^2}. \quad (2.237)$$

Esta es la distribución lorentziana o de Cauchy, con la que ya nos hemos encontrado al hablar de transformación de variables. Su media es μ y no tiene una varianza finita (la correspondiente integral diverge). Se denomina distribución lorentziana estándar a la que tiene $\mu = 0$, $\gamma = 1$, y según la expresión anterior cualquier lorentziana se puede reducir a ésta, de manera análoga a lo hecho en detalle para la gaussiana, con el cambio $x \rightarrow \frac{x-\mu}{\gamma}$.

c) $\alpha = \frac{1}{2}$, $\beta = 1$. La distribución de probabilidad asociada es la de Lévy-Smirnoff:

$$f(x) = \frac{\gamma}{2\pi} \frac{1}{(x-\delta)^{\frac{3}{2}}} e^{-\frac{\gamma}{2(x-\delta)}} \text{ para } x \in (\delta, \infty), \quad (2.238)$$

aunque, por brevedad, no incluiremos aquí la demostración.

Aproximación asintótica de una distribución estable: colas paretianas

Dada la imposibilidad general de obtener una forma analítica para las funciones de distribución de Lévy, cobra una especial relevancia analizar el comportamiento de la función en determinados intervalos, especialmente en el régimen asintótico. En lo que sigue consideraremos exclusivamente procesos de media nula y simétricos, sin pérdida alguna de generalidad para los resultados que obtendremos. La parte real de la función de densidad asociada a una variable aleatoria distribuida de acuerdo con una distribución estable será, a partir de la ecuación (2.235):

$$\begin{aligned} \operatorname{Re} f(x) &= \operatorname{Re} \left[\frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi(k) e^{-ikx} dk \right] = \operatorname{Re} \left[\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\gamma|k|^\alpha} e^{-ikx} dk \right] \\ &= \frac{1}{\pi} \int_0^{+\infty} e^{-\gamma|k|^\alpha} \cos(kx) dk. \end{aligned} \quad (2.239)$$

Cuando $\gamma = 1$, una expansión en serie válida para $|x| \rightarrow \infty$ es

$$f(x) = -\frac{1}{\pi} \sum_{j=1}^n \frac{(-1)^j \Gamma(j\alpha + 1)}{j! |x|^{j\alpha+1}} \operatorname{sen} \left(\frac{j\pi\alpha}{2} \right) + R_n(|x|) \quad (2.240)$$

donde $\Gamma(x)$ es la función gamma de Euler y el residuo $R_n(|x|) \sim |x|^{-\alpha(m+1)-1}$. El término de mayor alcance en la expansión anterior es el correspondiente a

$j = 1$, por lo que asintóticamente $f(x)$ toma la forma:

$$f(|x|) \sim \frac{\Gamma(\alpha + 1)}{|x|^{\alpha+1}} \operatorname{sen}\left(\frac{\pi\alpha}{2}\right) \sim |x|^{-(\alpha+1)} \quad (2.241)$$

A este tipo de distribuciones potenciales también se las denomina de cola paretoiana o de Pareto. Como puede verse, las distribuciones estables de Lévy siguen una ley potencial en el límite asintótico de valores de la variable aleatoria, lo que tiene importantes consecuencias en el comportamiento de los momentos de la distribución, como veremos a continuación.

Tal como se ha dicho anteriormente, el momento de orden r de la función de distribución de probabilidad se define de la forma:

$$M_r(c) = E\{(x - c)^r\} = \int_{-\infty}^{+\infty} dx (x - c)^r f(x) \quad (2.242)$$

de tal manera que es evidente que la convergencia de la integral anterior (lo mismo podría aplicarse a la serie en el caso de variables aleatorias discretas) exige que la función densidad de probabilidad decaiga de manera más rápida que x^r , lo que no sucede, en general, para las distribuciones estables de Lévy con colas potenciales de Pareto, para las que:

- i) El valor medio existe únicamente si $\alpha > 1$.
- ii) La varianza existe si y sólo si $\alpha = 2$. La gaussiana es, en consecuencia, la única distribución estable con varianza finita.

Las distribuciones con varianzas infinitas son usadas especialmente para describir fenómenos sin escala característica, i.e. que presentan fluctuaciones estadísticas de variables aleatorias de todos los tamaños posibles (eventos extremos).

Las colas de las distribuciones estables de Lévy son mucho más gruesas que las de una gaussiana, por lo que resultan muy útiles para describir los fenómenos multiescala, en los que tanto valores muy grandes como muy pequeños de la variable aleatoria en cuestión pueden ser observados con una probabilidad finita. Esto sucede en infinidad de sistemas físicos, biológicos y sociales: fallos en un sistema, amplitud del ruido electrónico, amplitud de terremotos y catástrofes naturales, distribución de la renta entre los individuos de una determinada sociedad, tamaño de fondos de pensiones, variación diaria de los precios de activos financieros como acciones u opciones, tamaño de los clusters de magnetización en las inmediaciones del punto crítico de una transición de fase paramagnético-ferromagnético, etc. Por ejemplo, el parámetro de orden de un sistema físico en las inmediaciones de una transición de fase experimenta fluctuaciones en todas las escalas posibles, lo que dota al sistema de las propiedades de invariancia de escala y universalidad características (Le Bellac, 1992). Este comportamiento de sucesos aleatorios en los que se presentan de manera relativamente probable eventos extremos es característico de sistemas con un gran número de partes mutuamente interaccionantes, normalmente abiertos al entorno, que autoorganizan su estructura interna y su dinámica mostrando

propiedades macroscópicas novedosas (emergentes). Este tipo de sistemas son los denominados sistemas complejos, categoría que presenta un enorme interés multidisciplinar al abordar exitosamente propiedades de sistemas físicos (las citadas transiciones de fase), biológicos (evolución, redes biológicas, origen de la vida, epidemias...), Neurología (redes neuronales), Geología (placas tectónicas, terremotos...), Economía y Sociología (mercados financieros, agentes interaccionantes, aprendizaje, mercados políticos de voto...), etc. La mayoría de estos sistemas complejos presentan eventos extremos en tiempos muy inferiores a la escala temporal característica de su evolución posterior.

Notemos que las distribuciones con colas potenciales asignan a eventos muy distanciados de la media de la distribución una probabilidad mucho mayor que la asignada por la distribución gaussiana porque las potencias negativas de x decrecen más lentamente que las exponenciales. Los procesos estables de Lévy tienen todos ellos colas potenciales que provocan que no todos sus momentos estén bien definidos, con la excepción de la distribución gaussiana, única distribución estable con todos sus momentos finitos.

Ejemplo 2.18

Una máquina que produce hilo de cobre generara un fallo cada cierta longitud. La distribución de Pareto puede describir la longitud del hilo entre fallos:

$$f(x) = \begin{cases} 0 & x < x_0 \\ \frac{\alpha x_0^\alpha}{x^{\alpha+1}} & x \leq x_0 \end{cases} \quad (2.243)$$

Para esta distribución, los valores esperados son:

$$\begin{aligned} E\{x\} &= \frac{\alpha x_0}{\alpha - 1} \\ E\{x^2\} &= \frac{\alpha x_0^2}{\alpha - 2} \end{aligned}$$

$$\sigma^2 = E\{x^2\} - (E\{x\})^2 = \frac{\alpha x_0^2}{\alpha - 2} - \frac{\alpha x_0}{\alpha - 1} = \frac{\alpha x_0^2(\alpha - 1) - \alpha x_0(\alpha - 2)}{(\alpha - 2)(\alpha - 1)}. \quad (2.244)$$

$$\sigma^2 = \frac{\alpha x_0^2}{(\alpha - 1)^2(\alpha - 2)}. \quad (2.245)$$

Observemos que para que exista la media se ha de dar que $\alpha > 1$, y para que exista la varianza $\alpha > 2$.

Las distribuciones estables permiten generalizar el teorema del límite central a sumas de distribuciones de varianza infinita: sea $\{x_i\}_{i=1}^n$ un conjunto de variables aleatorias iid cuya función de densidad de probabilidad tiende asintóticamente a una cola de Pareto para un valor dado de α . Entonces, puede probarse que la distribución de la suma de estas variables tiende a una

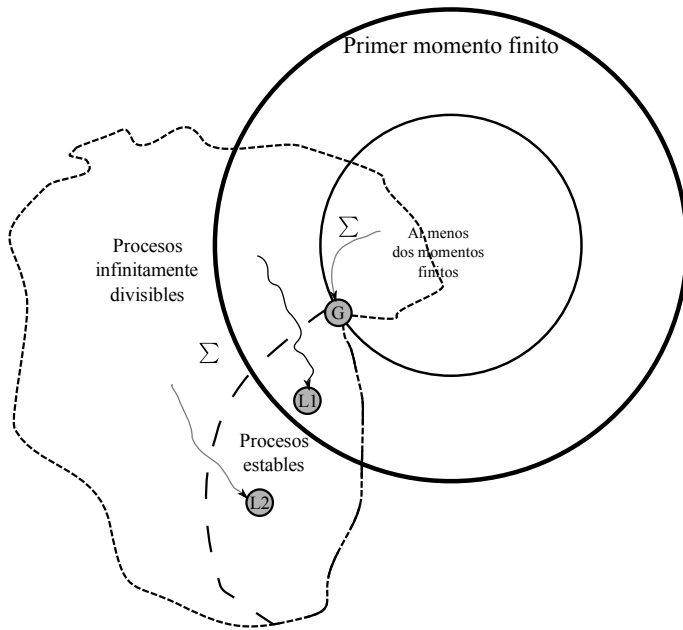


Figura 2.8: Distribuciones estables como límites de sumas de distribuciones de procesos iid. Para sumas de distribuciones con momentos finitos, el límite es la gaussiana G .

distribución de Lévy del mismo índice α . Existen, en consecuencia, infinitos atractores en el espacio de distribuciones para sumas de distribuciones con varianza infinita, mientras que la gaussiana presenta la propiedad de ser el único para distribuciones con varianza finita.

2.7. Distribuciones de probabilidad multidimensionales

En algunos casos, en cada suceso aleatorio se pueden medir conjuntamente varias magnitudes (o asociar varias características) de interés. Por ejemplo, en el estudio de una población podremos asociar a cada individuo un peso, edad y estatura. En esta situación, la variable aleatoria que surge de la medida o de asociar a cada característica cualitativa una cantidad numérica se denomina variable aleatoria vectorial o multidimensional.

A una variable aleatoria cuyo valor en cada caso es un conjunto de n cantidades numéricas se le asocia un vector de n componentes o de dimensión n . De modo que denotaremos esta variable aleatoria n -dimensional de la forma

$$X : \Omega \rightarrow \mathbb{R}^n \quad (2.246)$$

$$A \rightarrow X(A) = (x_1, x_2, \dots, x_n) \quad (2.247)$$

Una variable aleatoria vectorial $\vec{x} = (x_1, \dots, x_n)$ se denomina discreta cuando cada una de sus componentes es discreta y continua cuando las componentes son variables aleatorias continuas. En el caso de que tengamos componentes de ambos tipos diremos que \vec{x} es una variable vectorial aleatoria mixta.

Sea \vec{x} una variable aleatoria n -dimensional $\vec{x} = (x_1, x_2, \dots, x_n)$. Llamaremos *distribución de probabilidad conjunta* a la probabilidad de cada conjunto de valores (x_1, x_2, \dots, x_n) . Esto es, a la probabilidad de que, simultáneamente, la primera característica tome el valor x_1 , la segunda x_2 y así hasta la n -ésima característica, que debe tomar el valor x_n .

Ejemplo 2.19

Supongamos que lanzamos dos monedas diferentes (i.e. distinguibles) simultáneamente, podremos definir el vector bidimensional asociando a *cara* el valor 1 y a *cruz* el valor 0. De modo que el espacio muestral es²²

$$\Omega = \{(0, 0); (0, 1); (1, 0); (1, 1)\}. \quad (2.248)$$

Cuando se trate de una variable vectorial discreta, la distribución de probabilidad debe verificar:

$$\begin{aligned} p(\vec{x}_i) &= p(x_{1i}, x_{2i}) \\ 1 &\geq p(x_{1i}, x_{2i}) \geq 0 \quad \forall x_i \\ \sum_i p(\vec{x}_i) &= \sum_i p(x_{1i}, x_{2i}) = 1. \end{aligned} \quad (2.249)$$

En el caso de una variable continua las probabilidades vendrán determinadas por la función densidad de probabilidad conjunta $f(x_1, x_2)$, que verifica:

$$\begin{aligned} f(\vec{x}) = f(x_1, x_2) &\geq 0 \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx_1 dx_2 f(x_1, x_2) &= 1. \end{aligned} \quad (2.250)$$

Las probabilidades asociadas a intervalos de valores de la variable aleatoria se obtienen de la manera usual como:

$$P(a \leq x_1 \leq b; c \leq x_2 \leq d) = \int_a^b dx_1 \int_c^d dx_2 f(x_1, x_2). \quad (2.251)$$

²²Para simplificar la notación consideraremos a continuación el caso de una variable aleatoria vectorial de dos componentes. Es fácil extender las definiciones y resultados a los casos de variables multivariantes con más dimensiones.

Por otro lado, se denomina *distribución marginal de probabilidad asociada a la componente i -ésima* (x_i) de una variable aleatoria n -dimensional $\vec{x} = (x_1, \dots, x_n)$, a una distribución univariante de dicha componente $f(x_i)$ que proporciona la probabilidad de que la componente i -ésima de la variable tome el valor x_i con independencia del valor que tomen las demás variables. Así, para el caso de una variable discreta bidimensional:

$$\begin{aligned} p(x_1) &= \sum_i p(x_1, x_{2i}) \\ p(x_2) &= \sum_i p(x_{1i}, x_2), \end{aligned} \quad (2.252)$$

donde se entiende que $p(x_1, x_{2i})$ corresponde a las probabilidades $p(x_{1i}, x_{2i})$ que cumplen $x_{1i} = x_1$ (análogamente para $p(x_{1i}, x_2)$ se trata de las probabilidades de los sucesos $p(x_{1i}, x_{2i})$ donde $x_{2i} = x_2$). También podremos escribir esta probabilidad marginal en la forma

$$\begin{aligned} p(x_1) &= \sum_{\forall x_2} p(x_1, x_2) \\ p(x_2) &= \sum_{\forall x_1} p(x_1, x_2) \end{aligned} \quad (2.253)$$

En el caso de variables continuas, las funciones densidad de probabilidad marginales serán:

$$\begin{aligned} f(x_1) &= \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 \\ f(x_2) &= \int_{-\infty}^{+\infty} f(x_1, x_2) dx_1. \end{aligned} \quad (2.254)$$

Como ya se ha mencionado en este capítulo, la *probabilidad condicionada* de una variable x_1 cuando el valor de otra variable se supone fijo e igual a un determinado valor x_2 la denotaremos mediante $p(x_1|x_2)$. Esta probabilidad puede calcularse para cada valor de x_1 , obteniendo una distribución de probabilidad univariante de x_1 en los elementos de la población que tienen el valor fijado de x_2 . Esta distribución de probabilidad condicionada se obtiene normalizando la distribución de probabilidad conjunta, $p(x_1, x_2)$ adecuadamente, de manera que

$$\begin{aligned} p(x_2|x_1) &= \frac{p(x_1, x_2)}{p(x_1)} \\ \sum_{\forall x_2} p(x_1, x_2) &= p(x_1) \end{aligned} \quad (2.255)$$

Análogamente, la distribución de probabilidad condicionada de la variable x_1 condicionada a que la otra variable tome un valor fijo x_2 será

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)}. \quad (2.256)$$

En el marco de esta introducción de las distribuciones de probabilidad condicionadas, puede resultar útil al lector recordar el teorema de Bayes,

$$p(x_1|x_2) = \frac{p(x_2|x_1)p(x_1)}{\sum_{\forall x_1} p(x_2|x_1)p(x_1)} \quad (2.257)$$

que se obtiene de manera inmediata a partir de la definición de probabilidad condicionada haciendo:

$$\begin{aligned} p(x_1, x_2) &= p(x_2|x_1)p(x_1) = p(x_1|x_2)p(x_2) \\ p(x_2) &= \sum_{\forall x_1} p(x_1, x_2) = \sum_{\forall x_1} p(x_2|x_1)p(x_1). \end{aligned} \quad (2.258)$$

Para el caso de variables continuas, podemos reformular los conceptos anteriores en términos de la función densidad de probabilidad:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)}. \quad (2.259)$$

De modo análogo se pueden extender estos resultados a la variable aleatoria continua

$$\begin{aligned} f(x_1, x_2) &= f(x_2|x_1)f(x_1) \\ f(x_2) &= \int_{-\infty}^{+\infty} dx_1 f(x_1, x_2) = \int_{-\infty}^{+\infty} dx_1 f(x_2|x_1)f(x_1). \end{aligned} \quad (2.260)$$

De la misma forma, en términos de la función de densidad el teorema de Bayes se escribe de la forma:

$$f(x_1|x_2) = \frac{f(x_2|x_1)f(x_1)}{\int_{-\infty}^{+\infty} dx_1 f(x_2|x_1)f(x_1)}. \quad (2.261)$$

En el estudio de problemas multivariantes la cuestión de la independencia estadística de las variables aleatorias es de carácter fundamental. Se dirá que dos variables aleatorias, x_1 y x_2 , son estadísticamente independientes cuando el estudio de una de ellas no proporcione información alguna respecto de la otra. Esto quiere decir que la variable x_1 toma valores con independencia de aquellos que adopte la variable x_2 , i.e. x_1 tiene unos determinados valores cualquiera que sea el valor de x_2 . Matemáticamente se traduce en que

$$p(x_1|x_2) = p(x_1), \quad (2.262)$$

o bien, para el caso de variables aleatorias continuas,

$$f(x_1|x_2) = f(x_1). \quad (2.263)$$

Esto es, la variable x_1 no está condicionada en absoluto por la x_2 . Usando la definición de función densidad condicionada

$$\begin{aligned} f(x_1|x_2) &= \frac{f(x_1, x_2)}{f(x_1)} = f(x_2) \Leftrightarrow \\ \Leftrightarrow f(x_1, x_2) &= f(x_1)f(x_2), \end{aligned} \quad (2.264)$$

llegamos al importante resultado de que dos variables aleatorias son independientes si su distribución conjunta de probabilidad es igual al producto de sus distribuciones de probabilidad independientes (marginales). Generalizando el resultado anterior al caso de un conjunto $\{x_i\}_{i=1}^n$ de variables independientes, tendremos

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i) \quad (2.265)$$

2.7.1. Medidas características de una distribución de probabilidad multivariante

a) *Media de una variable aleatoria multidimensional*: definimos el vector de valores medios del vector aleatorio \vec{x} como el vector n -dimensional cuyas componentes son las medias de cada componente del vector \vec{x} :

$$\vec{\mu} = (E\{x_1\}, E\{x_2\}, \dots, E\{x_n\}) = (\mu_1, \mu_2, \dots, \mu_n), \quad (2.266)$$

lo cual también se puede escribir como

$$E\{\vec{x}\} = \vec{\mu} \quad (2.267)$$

siendo μ_i la media de la distribución marginal para x_i . Es posible también definir los valores medios para las sumas y productos de componentes de la variable aleatoria vectorial. En el caso de la suma, tendremos

$$E\left\{\sum_{i=1}^n x_i\right\} = \sum_{i=1}^n E\{x_i\} = \sum_{i=1}^n \mu_i \quad (2.268)$$

que es trivial de demostrar usando las definiciones de función densidad de probabilidad y de valor medio o valor esperado de una variable aleatoria. Si las variables x_1, \dots, x_n son independientes se verificará también

$$E\left\{\prod_{i=1}^n x_i\right\} = \prod_{i=1}^n E\{x_i\}. \quad (2.269)$$

En efecto, veamos que

$$\begin{aligned} E\left\{\prod_{i=1}^n x_i\right\} &= E\{x_1 \dots x_n\} = \int f(x_1 \dots x_n) x_1 \dots x_n dx_1 \dots dx_n = \\ &= \int f(x_1) \dots f(x_n) x_1 \dots x_n dx_1 \dots dx_n = \\ &= \int f(x_1) x_1 dx_1 \dots \int f(x_n) x_n dx_n = \\ &= E\{x_1\} \dots E\{x_n\}. \end{aligned} \quad (2.270)$$

b) *Covarianza*: tal como se dijo en el capítulo anterior, la covarianza es una medida de la relación lineal entre dos variables, y está definida por:

$$\text{cov}(x, y) = E\{(x - E\{x\})(y - E\{y\})\} = E\{xy\} - E\{x\}E\{y\} \quad (2.271)$$

Si los valores esperados de las variables son $E\{x\} = \mu$ y $E\{y\} = \nu$ entonces podremos escribir

$$\text{cov}(x, y) = E\{(x - \mu)(y - \nu)\} = E\{xy\} - \mu\nu. \quad (2.272)$$

Es inmediato comprobar que se verifica que si las variables son independientes su covarianza es nula. En efecto, si x e y son independientes:

$$E\{xy\} = E\{x\}E\{y\} \Rightarrow \text{cov}(x, y) = 0. \quad (2.273)$$

Lo contrario no es cierto, sin embargo, ya que las variables pueden estar relacionadas aunque no de manera lineal.

Ejemplo 2.20

Sea x una variable aleatoria $N(0, 1)$ y sea $y = x^2$. La covarianza será:

$$\text{cov}(x, y) = E\{xy\} - E\{x\}E\{y\} = (x - E\{x\})(y - E\{y\}) \quad (2.274)$$

donde

$$\begin{aligned} E\{x\} &= 0 \quad \text{variable normal estándar} \\ E\{y\} &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} x^2 dx = \frac{1}{\sqrt{2\pi}} \frac{1}{2\frac{1}{2}} \sqrt{\frac{\pi}{(1/2)}} \\ &\Rightarrow E\{y\} = 1. \end{aligned}$$

Luego

$$\text{cov}(x, y) = E\{x(y - 1)\} = E\{x^3 - x\} = 0 \quad (2.275)$$

ya que cualquier integral:

$$\int_{-\infty}^{+\infty} x^n \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = 0 \quad \forall n \text{ impar.} \quad (2.276)$$

Aunque $\text{cov}(x, y) = 0$, esto no significa, obviamente que sean independientes. Indica únicamente que no se relacionan de manera lineal.

Si hacemos una transformación de escala de la forma:

$$\begin{aligned} \tau &= ax + b \\ \omega &= cy + d \end{aligned}$$

obtenemos:

$$\begin{aligned}\operatorname{cov}(z, \omega) &= E\{z\omega\} - E\{z\}E\{\omega\} = \\ &= E\{(ax+b)(cy+d)\} - E\{(ax+b)\}E\{(cy+d)\} = \\ &= ac(E\{xy\} - E\{x\}E\{y\}) = ac \operatorname{cov}(x, y). \quad (2.277)\end{aligned}$$

Como vemos en la relación anterior, la covarianza es dependiente de la escala, i.e. de las unidades en que midamos las diferentes magnitudes. Una medida que nos evita este inconveniente puesto que es adimensional, es el denominado *coeficiente de correlación*, definido de la forma:

$$\rho(x, y) = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} \quad (2.278)$$

que verifica:

$$|\rho(x, y)| \leq 1 \quad (2.279)$$

$$y = ax + b \Rightarrow |\rho(x, y)| = 1; \operatorname{signo}[\rho(x, y)] = \operatorname{signo}(a) \quad (2.280)$$

$$x, y \text{ independientes} \Rightarrow \rho(x, y) = 0. \quad (2.281)$$

La covarianza interviene en la expresión de la varianza para sumas y diferencias de variables aleatorias:

$$\begin{aligned}\sigma^2(x \pm y) &= E\{(x \pm y - E\{x \pm y\})^2\} = \\ &= E\{(x - E\{x\})^2 + (y - E\{y\})^2 \pm 2(x - E\{x\})(y - E\{y\})\} = \\ &= \sigma_x^2 + \sigma_y^2 \pm 2\operatorname{cov}(x, y). \quad (2.282)\end{aligned}$$

c) *Matriz de varianzas y covarianzas*: denominaremos así a la matriz cuadrada de orden n :

$$\mathbf{M} = \begin{pmatrix} \sigma^2(x_1) & \operatorname{cov}(x_1, x_2) & \dots & \operatorname{cov}(x_1, x_n) \\ \vdots & \cdot & \dots & \vdots \\ \operatorname{cov}(x_n, x_1) & \cdot & \dots & \sigma^2(x_n) \end{pmatrix} \quad (2.283)$$

i.e. es una matriz que tiene como elementos diagonales la varianza de cada uno de los componentes del vector aleatorio, mientras que los elementos no diagonales son las diferentes covarianzas entre las observaciones. El término general podemos escribirlo de la forma:

$$M_{ij} = E\{x_i x_j\} - E\{x_i\}E\{x_j\}. \quad (2.284)$$

Ejemplo 2.21

Consideremos dos variables aleatorias x e y que podemos suponer como componentes de un vector aleatorio:

$$\vec{x} = (x, y). \quad (2.285)$$

Sea x una variable $N(0, 1)$ e $y = x^2$. Ya hemos visto que sus respectivos valores esperados son 0 y 1. También demostramos anteriormente que:

$$\text{cov}(x, y) = 0 \quad (2.286)$$

puesto que no existe relación lineal entre ambas variables. Luego:

$$M = \begin{pmatrix} \sigma^2(x) & 0 \\ 0 & \sigma^2(y) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2(y) \end{pmatrix} \quad (2.287)$$

$$\begin{aligned} \sigma^2(y) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (y - E\{y\})^2 dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (x^2 - 1)^2 dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x^4 - 2x^2 + 1) e^{-x^2/2} dx \end{aligned}$$

Usando la relación:

$$\begin{aligned} I_n &= \int_{-\infty}^{+\infty} x^n e^{-\alpha x^2} dx = \left(\frac{n+1}{2\alpha} \right)^{-1} I_n + 2 \\ I_0 &= \int_{-\infty}^{+\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} \end{aligned}$$

tenemos:

$$\begin{aligned} \sigma^2(y) &= \frac{1}{\sqrt{2\pi}} [I_4 - 2I_2 + I_0] = \\ &= \frac{1}{\sqrt{2\pi}} [3\sqrt{2\pi} - 2\sqrt{2\pi} + \sqrt{2\pi}] = 2 \end{aligned}$$

Luego la matriz de varianzas y covarianzas será:

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}. \quad (2.288)$$

Dos propiedades de la matriz de covarianzas que se derivan directamente de su definición son que es simétrica ($M_{ij} = M_{ji}$) y definida positiva (i.e. todos sus menores son positivos).

2.7.2. Distribuciones multinomial y normal n-dimensional

Distribución multinomial

Se trata de una generalización de la distribución binomial a un proceso multivariante. Supongamos un determinado proceso cuyos resultados pueden ser asignados a k clases distintas. Tomemos una muestra de n elementos y definamos k variables aleatorias:

$$x_i = n^o \text{ de elementos de la clase } i, \quad i = 1, \dots, k. \quad (2.289)$$

Entonces,

$$\vec{x} = (x_1 \dots x_l) \quad (2.290)$$

es un vector aleatorio k -dimensional (aunque sólo $k-1$ de las componentes son independientes, ya que la variable k -ésima está definida una vez que conocemos los valores de las $k-1$ restantes). La función de probabilidad se obtiene calculando la probabilidad de observar n_1 elementos de la clase 1, \dots , n_i de la clase, etc, en cualquier orden:

$$P(x_1 = n_1; \dots; x_k = n_k) = \frac{n!}{n_1! \dots n_k!} \prod_{j=1}^k p_j^{n_j} \quad (2.291)$$

donde hemos supuesto que las observaciones son independientes. La distribución anterior se conoce como *distribución multinomial* y tiene las siguientes propiedades:

- a) *Distribuciones marginales y condicionadas*: es posible ver que si se suma la ecuación (2.291) en todos los valores de (n_2, \dots, n_k) compatibles con la ligadura $n = \sum_{i=2}^k n_i = n - n_1$ y se aplica la fórmula del binomio de Newton se obtiene la siguiente distribución marginal para x_1 :

$$P(x_1 = n_1) = \frac{n!}{n_1! (n - n_1)!} p_1^{n_1} (1 - p_1)^{n_1}. \quad (2.292)$$

Concluimos pues que las distribuciones marginales para la variable i -ésima de una multinomial son todas binomiales, de parámetros p_i , medias $E\{x_i\} = np_i$ y varianzas $\sigma^2(x_i) = np_i(1 - p_i)$.

Aún más sencillo resulta ver que si se fija el valor de una o varias de las variables en (2.291) la distribución de las restantes sigue siendo multinomial. En consecuencia, todas las distribuciones condicionadas son multinomiales.

- b) *Covarianzas*: según acabamos de ver, la distribución de probabilidad conjunta de x_i y x_j (con $j \neq i$) es una multinomial de dos variables:

$$P(x_i = n_i, x_j = n_j) = \frac{n!}{n_i! n_j! (n - n_i - n_j)!} p_i^{n_i} p_j^{n_j} (1 - p_i - p_j)^{n - n_i - n_j}. \quad (2.293)$$

La covarianza de ambas variables puede calcularse como $\text{cov}(x_i, x_j) = E\{x_i x_j\} - E\{x_i\} E\{x_j\} = E\{x_i x_j\} - n^2 p_i p_j$, con

$$E\{x_i x_j\} = \sum_{n_i=0}^n \sum_{n_j=0}^{n-n_i} n_i n_j \frac{n!}{n_i! n_j! (n-n_i-n_j)!} p_i^{n_i} p_j^{n_j} (1-p_i-p_j)^{n-n_i-n_j}. \quad (2.294)$$

Esta suma se puede resolver de manera análoga a la usada al calcular la varianza de la distribución binomial en la ecuación (2.120) y sucesivas, dando como resultado $E\{x_i x_j\} = n(n-1)p_i p_j$. En consecuencia,

$$\text{cov}(x_i, x_j) = -n p_i p_j. \quad (2.295)$$

Esto implica que la matriz de varianzas y covarianzas asociada a una distribución multinomial es siempre singular. Por ejemplo, para $k=2$:

$$\begin{aligned} \text{cov}(x_1, x_2) &= -n p_1 p_2 = \text{cov}(x_2, x_1) \\ \sigma^2(x_1) &= n p_1 p_2; \quad \sigma^2(x_2) = n p_2 p_1 \end{aligned}$$

y en consecuencia

$$M = n p_1 p_2 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}; \quad \det M = 0. \quad (2.296)$$

Distribución normal n-dimensional

Un vector aleatorio \vec{x} sigue una distribución normal n -dimensional si su función densidad es:

$$f(\vec{x}) = \frac{1}{\sqrt{\det M} (2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu})^T M^{-1} (\vec{x} - \vec{\mu}) \right\}, \quad (2.297)$$

donde $\vec{\mu}$ es el vector de medias y M la matriz de covarianzas.

Ejemplo 2.22

Consideramos el caso de un vector aleatorio bidimensional $\vec{x} = (x, x^2)$. Hemos visto anteriormente que si x es una variable $N(0, 1)$, tenemos una matriz de covarianzas:

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \Rightarrow \det M = 2 \quad (2.298)$$

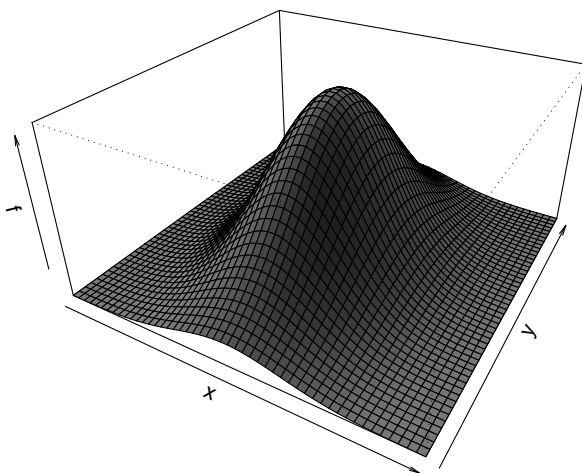
Luego:

$$\begin{aligned} (x \ x^2) \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} x \\ x^2 \end{pmatrix} &= (x \ y) \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} x & y \\ \frac{x}{2} & \frac{y}{4} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2} x^2 + \frac{y^2}{4} \end{aligned}$$

Por tanto, la distribución normal bidimensional asociada a \vec{x} será:

$$f(\vec{x}) = f(x, y) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(x^2 + \frac{y^2}{4}\right)} \quad (2.299)$$

Para una variable bidimensional la distribución tiene la forma de campana que se muestra en la figura siguiente para una matriz de covarianzas diagonal en la base canónica²³.



Al cortar la campana bidimensional por unos planos perpendiculares al xy se obtienen distribuciones normales. Por tanto las distribuciones marginales y condicionadas son también normales, como se puede comprobar analíticamente de forma muy sencilla.

²³Si no fuese así, por ser una matriz real, simétrica y definida positiva, la matriz de covarianzas siempre puede diagonalizarse, obteniendo los ejes principales de la gaussiana.

2.8. Cuestiones y problemas

- 2.1 Se tira un dado. Calcular la probabilidad de obtener un tres condicionada por haber obtenido un número impar.
- 2.2 Un feriante nos propone el siguiente juego: por una apuesta de 2 euros, podemos ganar 15 euros si conseguimos que nos salgan dos caras y un número par al lanzar dos monedas y un dado al aire. Presumiendo la honestidad del feriante (dado y monedas no trucadas), calcular nuestros ingresos esperados en n apuestas.
- 2.3 A la vista de los resultados de la cuestión anterior, calcúlese el valor esperado o medio de los ingresos en n apuestas.
- 2.4 Una fábrica tiene tres máquinas independientes que producen cierto tipo de pieza. La máquina 1 produce el 50 % de las piezas totales con un 2 % de piezas defectuosas, la máquina 2 produce el 40 % de las piezas con un 1 % defectuosas y la 3 produce el 10 % con un 0,5 % de piezas defectuosas. ¿Qué proporción de las piezas de la fábrica presentan defectos tras el proceso de producción? Si se selecciona al azar una pieza producida en la cadena de producción, ¿cuál es la probabilidad de que haya sido producida por la máquina 2 si no presenta defectos? Si la misma pieza resulta defectuosa, ¿cuál es entonces la probabilidad de que proceda de la máquina 2? ¿Cuál es la probabilidad total de piezas defectuosas en el proceso productivo?
- 2.5 Sabemos que se produce una forma incipiente de cáncer en tres de cada mil españoles. Como forma de detección precoz se ha desarrollado un protocolo de diagnóstico cuya fiabilidad es la siguiente: entre los pacientes sanos solamente un 5 % presenta una reacción positiva (falsa alarma). De los pacientes con cáncer incipiente sólo un 2 % presentan una reacción negativa (alarma fallida). Los pacientes que hayan presentado una reacción positiva serán hospitalizados para cirugía exploratoria. ¿Qué proporción de pacientes de los que se cree que tienen cáncer realmente lo tendrán?
- 2.6 El problema siguiente, denominado el problema de los ases, fue formulado en 1936 y se atribuye al matemático inglés J. H. C. Whitehead. Fue remitido a Marilyn von Savant por el maestro de los puzzles matemáticos Martin Gardner, señalando que era uno de sus favoritos.
- Si en una mano de bridge (se juega con 13 cartas) tenemos un as, ¿cuál es la probabilidad de que tengamos un segundo as en la misma mano?
 - Si el as de la mano es el as de corazones, ¿cuál es la probabilidad de que la mano tenga otro as?
 - ¿Son idénticas estas probabilidades?
- 2.7 Sea X una variable aleatoria. ¿Cuándo se verificará la igualdad de los valores esperados $E\{x^2\} = (E\{x\})^2$?

X	0	1	2	3	4	5	6
$p(X = x_i)$	0,47	0,30	0,10	0,06	0,04	0,02	0,01

2.8 La variable X = "número de hijos por familia en una cierta ciudad" tiene por distribución de probabilidad la siguiente tabla:

Hállese:

- Media o esperanza matemática. ¿Qué significado tiene este número?
 - Varianza y desviación típica.
 - Momento central de tercer orden.
 - Suponiendo que el Ayuntamiento de la ciudad paga 1.200 euros por hijo y que $Y = 1,200X$, ¿qué representa Y ? ¿Cuál es su distribución de probabilidad? ¿Cuál es el la cantidad que debe reservar en sus presupuestos la corporación municipal para política de incentivos a la natalidad? e) Media, varianza y desviación típica de Y .
- 2.9 Un fotomultiplicador de estado sólido está formado por cuatro contadores Geiger. Cada uno de los contadores cubre un cuarto de la superficie total del detector. Si dos o más fotones llegan simultáneamente a un contador, éste produce la misma señal s que si le llega un solo fotón. Las señales de los diferentes contadores se suman (i.e. si solo un contador se dispara la señal es s mientras que si dos detectan señal, ésta es de valor $2s$). Si al detector llegan dos fotones simultáneamente (y se reparten al azar entre los cuatro contadores), encuentra la distribución de probabilidad para la señal del detector (i.e. probabilidad de obtener s y $2s$). Calcula la señal media del detector y su varianza. Calcula el cociente de la desviación típica sobre el valor medio.
- 2.10 Consideremos un sistema físico formado por dos partículas cada una de las cuales puede encontrarse en dos niveles de energías 0 y ϵ . Consideremos que las partículas son idénticas, por lo que no podremos distinguir entre ambas mediante ningún procedimiento físico.
- Calcular el número de configuraciones (microestados) del sistema, i.e. el número de formas de colocar las dos partículas indistinguibles en los dos niveles de energía, así como la energía de cada microestado.
 - Denotemos mediante una etiqueta $l=1,2,\dots$, cada una de las configuraciones del apartado anterior y sea E_l la energía de cada una de ellas. Podemos considerar que cada una de ellas es un posible resultado de una variable aleatoria discreta y que el conjunto total de configuraciones obtenidas en el ejercicio anterior forman el espacio muestral o colectividad estadística. Supongamos que la función de probabilidad asociada a la variable aleatoria estuviese dada por

$$P_l = \frac{1}{Z} e^{-\beta E_l},$$

donde Z es una constante independiente del microestado. La distribución de probabilidad anterior se denomina distribución canónica o de Gibbs y nos da la probabilidad de que el sistema adopte un microestado determinado cuando se encuentra en equilibrio térmico con un termostato a la temperatura $T = 1/k_B\beta$, siendo k_B la constante de Boltzmann. Calcúlese el valor que debe tomar Z , denominada función de partición, para que P_i sea una función de probabilidad, el valor medio de la energía (energía interna) del sistema formado por dos partículas y el coeficiente de variación de Pearson de dicha magnitud, que mide sus fluctuaciones con respecto a la media.

- 2.11 Consideremos una variable cuya distribución de probabilidad es constante entre 0 y 1, esto es

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

¿cuál es el valor esperado y la varianza de x ? Calcúlese $E\{x^2\}$. Calcúlese el valor esperado de x^3 en esta distribución de probabilidad.

- 2.12 Se observó que que la variable $X =$ " número de centímetros a los que un dardo queda del centro al ser tirado por una persona" seguía una ley de probabilidad de la forma

$$f(x) = \begin{cases} k & 0 < x < 10 \\ 0 & \text{resto} \end{cases}$$

Se pide:

- Hallar k para que $f(x)$ sea función de densidad de probabilidad. Representar dicha función.
 - Hallar la función de distribución y representarla.
 - Media, varianza y desviación típica.
 - $p(X \leq 1)$.
 - Probabilidad de acertar en la diana.
- 2.13 Sea x una variable aleatoria con una función de densidad de probabilidad dada por:

$$f(x) = \begin{cases} k(x+a) & -a \leq x \leq 0 \\ k(a-x) & 0 \leq x \leq a \\ 0 & \text{en otro caso} \end{cases}$$

- Calcular k para que esta densidad de probabilidad esté correctamente definida.
- Calcular la función de distribución de probabilidad acumulativa.
- Calcular $E\{x\}$, $E\{x^2\}$, así como σ^2 .
- Si dos variables aleatorias x e y siguen esta distribución de probabilidad, calcular el valor esperado de $x+y$ y su varianza.

- 2.14 Un gran número de fenómenos en ciencias (Física, Química, Biología, Medicina, Economía, etc.) tienen asociada una variable aleatoria cuya ley de probabilidad es de la forma

$$f(x) = \begin{cases} ke^{-kx} & 0 < x < \infty \quad k > 0 \\ 0 & \text{resto} \end{cases} .$$

Se pide:

- ¿Puede tomar k cualquier valor?
 - Para $k = 0,1$, representar la función, encontrar la función de distribución y su gráfica. En este caso, calcular
 - $p(X > 10)$ y $p(50 \leq X \leq 100)$.
- 2.15 En un determinado experimento de determinación de la polidispersidad de tamaño de partículas coloidales, se ha determinado que la función densidad de probabilidad de que una partícula tenga un diámetro comprendido entre σ y $\sigma + d\sigma$ es la que se muestra en la figura adjunta, que puede aproximarse por una parábola de la forma:

$$f(\sigma) = a + b\frac{\sigma}{\bar{\sigma}} + c\left(\frac{\sigma}{\bar{\sigma}}\right)^2$$

donde $\bar{\sigma}$ es la media del diámetro de una partícula.

- Demostrar, usando los datos de la figura y cualquier otra propiedad de las distribuciones de probabilidad, que

$$f(\sigma) = \frac{3\sigma}{2\bar{\sigma}^2} \left(1 - \frac{\sigma}{2\bar{\sigma}}\right)$$

- Comprobar que la media de la distribución anterior es $\bar{\sigma}$.
- Calcúlese la dispersión de la distribución anterior y demuéstrese que la desviación estándar está dada por $s_\sigma = \bar{\sigma}/\sqrt{5}$.
- ¿Qué fracción de las partículas de la dispersión tiene diámetros comprendidos en el intervalo $(\bar{\sigma} - s_\sigma, \bar{\sigma} + s_\sigma)$? Compárese con la distribución normal.

Calcular la probabilidad de que al lanzar seis dados de póker obtengamos:

- exactamente un as
- al menos un as
- exactamente dos ases.

En el juego de la ruleta rusa en el tambor del arma hay únicamente una bala y los cinco compartimentos restantes permanecen vacíos. Se pide:

- Calcular la probabilidad de estar vivo después de pulsar N veces el gatillo.
- Calcular la probabilidad de sucumbir en el N -ésimo intento.

La probabilidad de éxito en un experimento de Bernoulli es p . Calcular probabilidad de que, en el experimento que ocupa el k -ésimo lugar, ocurra un éxito por l -ésima vez ($0 < l = k = n$).

En un negocio arriesgado la petrolera Aceito S.A. ha comprometido todos sus fondos (ya muy mermados por arriesgadas inversiones y corruptos administradores) para financiar una serie de 12 perforaciones. Cada perforación en esta región tiene una probabilidad de un 20% de producir petróleo con éxito, independientemente de las demás perforaciones. Para evitar la bancarrota, tres o más perforaciones deben producir petróleo. ¿Cuál es la probabilidad de lo anterior?

Modelo de Laplace de la incertidumbre experimental: En 1783, Laplace formuló un modelo para explicar el origen de los incertidumbres de observación. Sea x_0 el valor exacto de una determinada magnitud experimental. Supongamos que el proceso de medida está afectado por un número n de causas independientes, cada una de las cuales produce una desviación de la medida correcta de magnitud ε . Esta desviación se produce en una dirección o en otra ($x_0 \pm \varepsilon$) de manera equiprobable. La incertidumbre experimental total sería, según Laplace, el resultado de la acción combinada (suma) de las n causas independientes. Calcular la probabilidad de que se produzca una incertidumbre de medida Δx cuando actúan n causas.

Considérese un experimento aleatorio de Bernoulli. Obténgase el espacio muestral y la distribución de probabilidad de la variable aleatoria

$$X = \text{"número de intentos antes del primer fracaso"}$$

Calcúlese asimismo la esperanza matemática y la varianza de la distribución obtenida y compárese con la distribución geométrica o de Pascal.

- 2.16 Considérese un experimento aleatorio que consiste en retirar sin reemplazamiento bolas de una urna en las que existen m bolas blancas y $N - m$ bolas negras. Demuéstrese que la probabilidad de que en n intentos se hayan obtenido r bolas blancas es

$$p(x = r) = \frac{\binom{m}{r} \binom{N-m}{n-r}}{\binom{N}{n}}$$

Demuéstrese que la distribución anterior, denominada distribución hipergeométrica, está normalizada y obténganse sus dos primeros momentos característicos. Obsérvese que la distribución hipergeométrica es el equivalente a la binomial cuando el muestreo se hace sin reemplazamiento. Pruébese que la distribución anterior recupera la binomial cuando $N \rightarrow \infty$.

- 2.17 Supongamos que disponemos de dos cajas de cigarros habanos, cada una de ellas con n cigarros inicialmente, y que cada vez que necesitamos un habano extraemos uno de una de las cajas al azar. Calcular la probabilidad de que cuando en una de las cajas se hayan terminado los cigarros en la otra queden exactamente r cigarros.

2.18 *Los cinco problemas de Huygens.*

En 1654, Antoine Gombaud, el caballero de Méré, miembro de la aristocracia francesa, clase social muy aficionada entonces a los juegos de azar; y Damien Mitton, rico burgués de la corte, escritor, autor junto con Méré del concepto de "hombre honesto" del XVII y modelo de "libertino" para Pascal en sus *Pensées* (Pensamientos, 1670), plantean a Pascal el denominado *problème des partis* (problema de los puntos), que en esencia consiste en lo siguiente: dos jugadores A y B pactan jugar una serie de juegos justos hasta que uno de ellos haya ganado un número concreto, N . Si el juego debe interrumpirse súbitamente cuando A ha ganado N_1 y B N_2 , ¿cómo deben dividirse las apuestas? Este problema ya había sido considerado por Cardán y por Pacioli y Tartaglia aproximadamente en la misma época. Esta consulta se considera el desencadenante de la correspondencia entre Pascal y Fermat, siete cartas intercambiadas entre los meses de julio y octubre de 1654, que hoy día son consideradas como el origen de la moderna teoría de probabilidades. Además del antedicho problema de los puntos, consideraron también el problema de los dados, que analiza la cuestión de cuántas veces deben lanzarse dos dados antes de que se espere que salga el seis doble.

En 1655, Christiaan Huygens visita París, para informar a los matemáticos de esta ciudad de su descubrimiento de la primera luna de Saturno. En este viaje, Huygens conoció a través de Pierre Carcavy el trabajo sobre probabilidades contenido en la correspondencia entre Pascal y Fermat. De vuelta a Holanda, en 1657 Huygens publicó un pequeño trabajo titulado *De Ratiociniis in Ludo Aleae* (De la razón en los juegos de azar) sobre cálculo de probabilidades, el primer trabajo impreso sobre el tema, probablemente inspirado por la correspondencia de Pascal y Fermat, y basado eminentemente en la solución de problemas. Contiene 14 problemas resueltos y cinco para resolver por el lector, en parte debidos a Pascal y Fermat, y que serán considerados como una referencia ineludible por matemáticos posteriores como Jacob y Nicholas Bernoulli, de Moivre o Montmort. Estos problemas son los siguientes:

Problema 1

A y B juegan uno contra el otro, con dos dados, bajo la condición de que A gana si obtiene 6 puntos, y B gana si obtiene siete puntos. Le corresponde el primer tiro a A, los dos siguientes a B, los otros dos siguientes a A, y así sucesivamente. ¿Cuál es la probabilidad de que gane A sobre la de B?

Nota: Este problema fue propuesto por Fermat en carta a Huygens en junio de 1656, y resuelto por Huygens en carta a Carcavi el 6 de julio de 1656.

Problema 2

Tres jugadores A, B y C, con doce fichas cada uno, de las cuales cuatro son blancas y ocho negras, juegan con la condición de que gana el primer jugador que obtiene, extrayendo sin mirar, una ficha blanca. A extrae primero, luego B, luego C, luego A de nuevo, y así sucesivamente hasta que

se extraiga por primera vez la ficha blanca. ¿Cuál es la proporción de las probabilidades de ganar de cada jugador con respecto a los otros?

Nota: Problema resuelto por Huygens en 1665.

Problema 3

El jugador A apuesta al jugador B que de un mazo de cuarenta cartas, que contiene diez de cada color, extraerá cuatro siendo todas ellas de colores distintos. ¿Cuál es la ratio de las probabilidades de victoria de A sobre las de B?

Nota: Problema propuesto por Fermat a Huygens en junio de 1656, quien lo responde sin prueba en una carta a Carcavi de 6 de julio de 1656.

Problema 4

Dos jugadores tienen doce fichas cada uno, cuatro blancas y ocho negras. A apuesta a B que, escogiendo siete fichas sin mirar, obtendrá tres blancas. ¿Cuál es la probabilidad de que gane A sobre la de que gane B? Huygens también considera el caso en que se apuesta a escoger tres o más fichas blancas. Obténgase la solución en este último caso.

Nota: Huygens resuelve este problema en 1665.

Aunque no forma parte del problema original de Huygens, obténgase la solución tanto en el caso de extracción sin reemplazamiento de la bola como con reemplazamiento.

Problema 5

Los mismos jugadores A y B tienen doce fichas cada uno y juegan con tres dados. La condición es ahora que si se obtienen 11 puntos en un lanzamiento, A entrega una ficha a B, mientras que si se obtienen catorce, B entrega una a A, ganando el jugador que obtiene primero todas las fichas. Obténgase la fracción de las probabilidades de victoria de A y B.

Nota: Este problema es el conocido como la ruina del jugador, y fue planteado por Pascal a Fermat. Posteriormente, a través de Carcavi, se le plantea a Huygens en carta de 28 de septiembre de 1656 en la que se contienen las soluciones de Pascal y de Fermat. Por su parte, Huygens incluye su solución en una carta a Carcavi de 12 de octubre de 1656 y la demostración en una nota de 1676.

Problema adicional

En el caso de que tiremos dos dados podemos obtener un nueve de dos formas diferentes, $9=5+4=6+3$, el mismo número que de obtener un diez, $10=6+4=5+5$. Si se lanzan tres dados, el número de formas de obtener un nueve y de obtener un diez también es el mismo, seis maneras diferentes. Analícese por qué al lanzar dos dados es más frecuente obtener un nueve y al lanzar tres es más frecuente el diez.

Nota: Este problema se contiene en el *Liber de ludo aleae* (Libro de los juegos de azar), de Gerolamo Cardano, el primero publicado sobre probabilidades con técnicas eficaces, escrito durante la década de 1560 y publicado póstumamente en 1663.

- 2.19 Una partícula que ejecuta saltos independientes entre sí de longitud l en los nodos de una red unidimensional con probabilidades idéjemplo básico del denominado paseo al azar simple (*simple random walk*). Calcular la probabilidad de que, luego de $2n$ saltos, la partícula se encuentre a una distancia $x = nl$ del origen de coordenadas situado en el punto de partida. ¿Qué sucede cuando $n \rightarrow \infty$? La generalización a tres dimensiones de este modelo se denomina movimiento browniano. ¿Qué habría que modificar para es
- 2.20 En un *call center* se reciben una media de 2 llamadas por minuto. Calcular la probabilidad de que se reciban:
- Ninguna llamada en un minuto.
 - Menos de 5 llamadas por minuto.
 - Ninguna llamada en 10 minutos.
 - Menos de 5 llamadas en 10 minutos.
- ¿Cuál es el número más probable de llamadas durante un intervalo de 10 minutos? ¿Y en media hora?
- 2.21 Un fotomultiplicador de estado sólido está formado por un único contador Geiger cuya superficie es 1 mm^2 y se somete a una radiación pulsada cuya intensidad media es de 5 fotones/ mm^2 en el plano donde se encuentra el detector. Considérese que la señal s del detector es nula si no llega ningún fotón y $s = 1$ si llega un fotón o más de uno. ¿Cuál es la probabilidad de que en un detector incidan
- más de 5 fotones?,
 - cinco fotones?,
 - ningún fotón?
- 2.22 Una lámina fina de oro es bombardeada mediante un haz de neutrones de manera que podemos suponer que la probabilidad de que un neutrón impacte en un punto cualquiera de la lámina es uniforme. Si el número medio de colisiones por núcleo es de 3,
- ¿qué fracción de núcleos no sufre ningún impacto?
 - ¿Qué proporción del total sufre 3 impactos? ¿Y más de dos impactos?
- 2.23 Las colisiones de una partícula en el seno de un gas son sucesos puntuales que tienen lugar en un soporte continuo, el tiempo. Por ello, el número de colisiones de una partícula en el interior del gas es una variable aleatoria de Poisson. Teniendo en cuenta esto, obténgase la función densidad de

probabilidad de la variable t ="tiempo entre colisiones sucesivas de una partícula" y obténgase su valor esperado en términos de la frecuencia de colisión, λ (número de colisiones por unidad de tiempo). Considérese un gas formado por partículas de masa m encerradas en un recipiente de volumen V a la temperatura T . Sabiendo que según la teoría cinética de los gases, la velocidad media de las partículas en estas condiciones es

$$\bar{v} = \sqrt{\frac{8k_B T}{\pi m}},$$

donde k_B es la constante de Boltzmann ($k_B = 1,38 \times 10^{-23}$ J/K), calcúlese el recorrido libre medio de una partícula (espacio recorrido en promedio entre dos colisiones consecutivas). Como aplicación numérica tómesese, $m = 3,32 \times 10^{-27}$ kg (correspondiente a la masa de la molécula de hidrógeno) y $T = 298,15$ K.

(Nota: El orden de magnitud típico de la frecuencia de colisión, λ , para hidrógeno molecular a densidades moderadas es de 10^{11} colisiones/s).

- 2.24 La probabilidad de que un núcleo radiactivo con vida media τ que existe a $t = 0$ se desintegre durante el intervalo de tiempo $(t, t + \delta t)$ es:

$$p(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

Sabiendo que en una fuente radiactiva que contiene n núcleos la probabilidad de observar k desintegraciones en un determinado temporal sigue una distribución de Poisson de media λ , determínese λ en función de τ .

- 2.25 Consideremos un líquido en el interior de un capilar aproximadamente unidimensional. Debido a las colisiones aleatorias con las moléculas vecinas, la posición de una partícula en el interior del líquido es una variable aleatoria que está descrita por la distribución gaussiana con varianza dependiente del tiempo:

$$f(x, t) = A e^{-x^2/4Dt} \quad ; \quad x \in (-\infty, \infty)$$

donde D es el denominado coeficiente de difusión. Calcúlese la constante de normalización y demuéstrese que la curtosis de la distribución de probabilidad es igual a 3 en todo instante de tiempo.

- 2.26 Un líquido se expone a un haz de rayos X incidente que provoca la ionización de los átomos del material, los electrones emitidos colisionan con las moléculas que los rodean hasta detenerse. Si la densidad de probabilidad de que un electrón se detenga a una distancia x del punto en el que fue emitido viene descrita por la función:

$$\rho(x) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad x \in (0, +\infty)$$

donde σ depende de la naturaleza del líquido y de la temperatura, obtener la distancia media que recorren los electrones, denominada longitud de termalización.

- 2.27 En un procedimiento industrial de control de calidad se desechan las piezas cuyo tamaño difiere en al menos un 5% del tamaño medio (5,0 mm). Sabiendo que el tamaño de las piezas sigue una distribución gaussiana de desviación típica 0,2 mm, calcula la proporción de piezas rechazadas.
- 2.28 Se ha comprobado experimentalmente que el tiempo necesario para acabar un examen de una determinada carrera se distribuye normalmente con una media de 240 minutos con una desviación típica de unos 30 minutos.
- ¿Qué proporción de alumnos terminará su examen antes de media hora?
 - ¿Qué proporción lo hará en cuatro horas?
 - ¿Cuándo debe darse por terminado el examen para permitir que el 95% de los alumnos haya terminado?

Capítulo 3

Métodos estadísticos

En diversos momentos de esta obra nos hemos referido al hecho de que el problema fundamental de la Estadística es el de proporcionar métodos mediante los cuales podamos hacer afirmaciones o predicciones acerca del comportamiento de los elementos de una población a partir de la observación de subconjuntos de la misma denominados muestras. Éste es el objetivo de la denominada inferencia estadística, y sus principales métodos son objeto de estudio en el presente capítulo. En particular, estudiaremos los siguientes temas:

1. *Estimación puntual y por intervalos*, donde analizaremos las técnicas de estimación de parámetros de distribuciones de probabilidad a partir de los obtenidos en las distribuciones de frecuencia experimentales o empíricas. Un caso paradigmático de este problema es aquel en el que consideramos una población en la que tenemos definida una distribución de probabilidad asociada a una cierta variable aleatoria X :

$$\begin{array}{lll} \text{prob}(X \in (x, x + dx)) & = f(x) \cdot dx & \text{caso continuo} \\ \text{prob}(X = x_i) & = p_i & \text{caso discreto} \end{array}$$

$f(x)$ ó p_i deberán estar adecuadamente normalizadas tal y como vimos en el capítulo anterior:

$$\int_a^b f(x)dx = 1 \quad \sum_{i=1}^n p_i = 1 \quad (3.1)$$

En general estas distribuciones de probabilidad, ya sean discretas o continuas dependerán de ciertos parámetros (λ para la distribución de Poisson, por ejemplo) que serán inicialmente desconocidos. La información acerca de los parámetros provendrá de la distribución experimental de frecuencias obtenida de muestras finitas, lo que nos obligará a adoptar una serie de procedimientos para inducir el valor de los parámetros poblacionales a partir de los de las muestras analizadas. En particular, serán objeto de tratamiento detallado la estimación puntual de la media y la varianza de poblaciones madre normales y la construcción de intervalos de confianza para estos parámetros.

2. *Test de hipótesis estadísticas*, mediante el cual podremos estimar el nivel de aceptabilidad estadística de una afirmación acerca del comportamiento de una población conocida la información suministrada por una muestra. Un supuesto especialmente relevante de test de hipótesis estadísticas es el test de bondad de los ajustes de datos experimentales a leyes teóricas. Este caso particular de test de hipótesis ocupará un lugar destacado en la exposición contenida en esta obra.
3. *Método de máxima verosimilitud*, que introduciremos como una técnica de especial interés para la determinación de parámetros, ya que proporciona un procedimiento de evaluación de los parámetros que maximizan la probabilidad de obtención de los datos de una muestra determinada supuesta una cierta distribución de probabilidad para el fenómeno. Asimismo, la obtención de los mejores ajustes de datos experimentales a curvas teóricas mediante el método de máxima verosimilitud permite resaltar la auténtica naturaleza de los procesos de ajuste como métodos de estimación de los parámetros óptimos para maximizar la verosimilitud de la muestra experimental una vez supuesta la pertenencia de ésta a una población madre gobernada por una distribución de probabilidad elegida de antemano.

3.1. Estimación de parámetros

Consideremos la población (normalmente infinita) de todas las posibles medidas de una magnitud de interés y sea X la variable aleatoria objeto de estudio. Sea $f(x)$ (o p_i en el caso de una variable aleatoria discreta) la densidad de probabilidad de la población. Los parámetros que describen esta distribución de probabilidad (la media y la desviación típica, por ejemplo, en una distribución normal) los denominaremos parámetros de población¹

$\mu \equiv$ media de la población

$\sigma \equiv$ desviación típica de la población

etc...

Como ha quedado dicho en la introducción al presente capítulo, el problema de la estimación de parámetros corresponde a la obtención de estos parámetros a partir de información procedente de conjuntos de datos obtenidos mediante observación de elementos de la población. Así, las distribuciones de probabilidad se obtienen a partir de muestras de datos. Este problema se conoce como problema de estimación de parámetros.

¹Denotaremos los parámetros de la población con letras del alfabeto griego (μ , σ) y sus correspondientes para las muestras con letras del alfabeto latino (\bar{x} , s).

Ejemplo 3.1

Consideremos la desintegración radiactiva de un núcleo, sabemos que la probabilidad $p(n)$ de obtener n sucesos en un cierto intervalo de tiempo Δt puede parametrizarse mediante la distribución de Poisson

$$p(n) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

El parámetro λ , que coincide con la media de la distribución, es directamente proporcional a la actividad de la muestra. Podemos obtenerlo mediante la medida del número de desintegraciones que se producen en intervalos de la anchura especificada.

Así pues, para resolver el problema de la estimación de parámetros hemos de tomar muestras de datos que nos suministren información acerca del fenómeno en cuestión. Si tomamos l muestras de tamaño n de la magnitud de interés, obtendremos un conjunto de datos

$$\begin{array}{ll} 1^a \text{ muestra} & (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}) \\ 2^a \text{ muestra} & (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}) \\ \vdots & \\ \text{muestra } l & (x_1^{(l)}, x_2^{(l)}, \dots, x_n^{(l)}). \end{array} \quad (3.2)$$

Podemos considerar que cada muestra corresponde a una variable aleatoria n -dimensional $\vec{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)})$, y que esta variable aleatoria seguirá una cierta distribución de probabilidad multivariante

$$g(\vec{x}) = g(x_1, x_2, \dots, x_n). \quad (3.3)$$

Lógicamente, estas muestras de datos han de ser obtenidas mediante alguna técnica de muestreo que no condicione el resultado final. Es de crucial importancia que las muestras sean una imagen lo más neutra posible de la población (i.e. que no incorporemos sesgo alguno mediante nuestra selección de los elementos que la integran).

Ejemplo 3.2

Si tratamos de estudiar el comportamiento del cuerpo electoral de un determinado país, no podemos, en el caso de poblaciones no homogéneas, tratar de establecer valores para los parámetros de la distribución de probabilidad de la población (la totalidad del censo electoral) a partir de los datos obtenidos mediante encuestas entre 2000 habitantes de una única provincia (sesgo territorial) o de un mismo nivel de renta (sesgo socioeconómico).

Para poder garantizar que nuestro estudio se realiza en condiciones de muestreo aceptables vamos a considerar el caso más simple de poblaciones homogéneas, que se da a menudo en muchas medidas físicas. Consideraremos que se cumplen los siguiente requisitos del denominado *muestreo aleatorio simple*²:

1. Todas las medidas son independientes

$$g(\vec{x}) = g_1(x_1) g_2(x_2) \cdots g_n(x_n). \quad (3.4)$$

2. Las distribuciones de probabilidad individuales de cada medida son idénticas a la distribución de probabilidad de la población

$$g_1(x_1) = g_2(x_2) = \cdots = g_n(x_n) = f(x). \quad (3.5)$$

Estas condiciones extremadamente simplificadoras no son siempre aplicables con seguridad a una muestra estadística, ya sea porque no tenemos garantía absoluta de la independencia estadística de las medidas o porque no estamos seguros de que la distribución de probabilidad $f(x)$ describa el comportamiento estadístico de la propia población. Por otra parte, recordando la definición de probabilidad del capítulo anterior, esperamos que la distribución de probabilidad de la muestra tienda a la distribución de probabilidad de la población cuando $n \rightarrow \infty$.

Una función de los valores de las variables estadísticas que componen la muestra de datos se suele denominar *estadístico*. Un ejemplo de estadístico es la media muestral

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) \quad (3.6)$$

definido por la media aritmética de los valores de la variable estadística en la muestra. Notemos que los estadísticos son funciones de variables aleatorias y por tanto ellos mismos son variables aleatorias.

Denominaremos *estimador* S de un parámetro λ de la población a un estadístico o función de la muestra

$$S = S(x_1, x_2, \dots, x_n) \quad (3.7)$$

que nos permita obtener o estimar el parámetro λ de la población que deseamos conocer. Introduzcamos a continuación algunas características de los estimadores de particular importancia.

i) Fidelidad:

se dice que un estimador es *fiel* si para cualquier tamaño n de la muestra se cumple

$$E \{S(x_1, x_2, \dots, x_n)\} = \lambda. \quad (3.8)$$

²Otras técnicas de muestreo como el por conglomerados, el estratificado o el polietápico no serán objeto de estudio en este manual introductorio, aunque pueden encontrarse excelentes tratamientos en libros específicos.

ii) *Consistencia:*

un estimador es llamado *consistente* si su poder de estimación aumenta en precisión con el tamaño de la muestra

$$\lim_{n \rightarrow \infty} \sigma(S) = 0. \quad (3.9)$$

iii) *Eficiencia:*

una medida de la eficiencia relativa de dos estimadores fieles es el cociente de sus varianzas, es decir

$$\eta = \frac{E\{(S_1 - \lambda)^2\}}{E\{(S_2 - \lambda)^2\}} = \frac{\sigma^2(S_1)}{\sigma^2(S_2)}. \quad (3.10)$$

En algunos casos es posible establecer cuál es el estimador con la varianza más pequeña y en este caso el estimador ha de ser preferido frente a los otros.

3.1.1. Estimación puntual de la media poblacional

Además de ser fundamental para la construcción de la distribución de probabilidad de un fenómeno, la estimación de la media de una población a partir de muestras de la misma presenta un gran interés práctico. Ejemplos de esto son la determinación del peso medio de un producto en un proceso industrial, o la media de una magnitud física en un proceso de medición.

Consideremos el caso de muestreo aleatorio simple de una población continua con una distribución de probabilidad $f(x)$ y supongamos que mediante muestreo aleatorio simple se han obtenido los datos $\{x_i\}_{i=1}^n$. En principio es lógico pensar que una buena estimación puntual de la media poblacional la podremos obtener mediante la media muestral³, definida como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + \cdots + x_n). \quad (3.11)$$

Analicemos las características de este estadístico como estimador de la media poblacional. Veamos que éste es un estimador de la media de población μ fiel, consistente y el más eficiente de todos los estimadores lineales fieles de la media poblacional.

a) *Fidelidad.*

Para analizar la fidelidad de \bar{x} como estimador de la media poblacional, consideremos la esperanza matemática del estimador media muestral

³Sin embargo, la media muestral no es el único estimador que podemos construir a partir de una combinación lineal de los datos. Por ejemplo, es fácil demostrar que

$$M = \sum_{i=1}^n a_i x_i, \quad a_i \in \mathbb{R}$$

es también un estimador fiel de la media de población, siempre que $\sum_{i=1}^n a_i = 1$. Volveremos sobre este punto al estudiar la eficiencia relativa del estimador media muestral con respecto a otros estimadores lineales de la media poblacional.

$$E\{\bar{x}\} = \frac{1}{n} \sum_{i=1}^n E\{x_i\} = \frac{1}{n} (E\{x_1\} + E\{x_2\} + \cdots + E\{x_n\}). \quad (3.12)$$

En virtud de la hipótesis de muestreo aleatorio simple cada x_i tiene una distribución de probabilidad idéntica a la de la población de modo que $E\{x_1\} \equiv E\{x_2\} \equiv \cdots \equiv E\{x_n\} = \mu$, lo que conlleva

$$E\{\bar{x}\} = \frac{1}{n} n\mu = \mu. \quad (3.13)$$

Por tanto, la media muestral es un estimador fiel de la media de la población. Con este resultado, tendremos que dada una cierta distribución en la variable x con media μ , \bar{x} tiene una distribución cuya media también es μ , aunque, obviamente, la distribución de probabilidad de \bar{x} y de x no coinciden. Volveremos sobre este punto al final de la sección.

b) Consistencia.

La siguiente pregunta a la que nos interesa contestar es: ¿cuál es el grado de aproximación con el que la variable aleatoria \bar{x} se distribuye en torno a la media de la población μ ? Para dar respuesta a esta cuestión calcularemos la varianza de \bar{x} :

$$\sigma^2(\bar{x}) = E\left\{(\bar{x} - E\{\bar{x}\})^2\right\} = E\left\{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right\}. \quad (3.14)$$

Puesto que los datos x_i, x_j con $i \neq j$ son independientes estadísticamente, su covarianza es nula y todos los términos mixtos cumplen $E\{(x_i - \mu)(x_j - \mu)\} = 0, i \neq j$. De este modo,

$$\begin{aligned} \sigma^2(\bar{x}) &= \frac{1}{n^2} \left[\sum_{i=1}^n E\{(x_i - \mu)^2\} \right. \\ &\quad \left. + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E\{(x_{i-1} - \mu)(x_i - \mu)\} \right] = \\ &= \frac{1}{n^2} \sum_{i=1}^n E\{(x_i - \mu)^2\} \end{aligned} \quad (3.15)$$

por lo que, bajo las hipótesis de muestreo aleatorio simple, se cumple que

$$\sigma^2(\bar{x}) = \frac{1}{n} \sigma^2(x) \quad (3.16)$$

siendo σ^2 la varianza de la población de la variable estadística X . Esta expresión puede ser escrita como:

$$\sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{n}}. \quad (3.17)$$

De la ecuación anterior se sigue inmediatamente que,

$$\lim_{n \rightarrow \infty} \sigma(\bar{x}) = 0, \quad (3.18)$$

por lo que la media muestral \bar{x} es también un estimador consistente de la media de la población⁴.

c) *Eficiencia.*

Como ha quedado dicho, la eficiencia de un estimador únicamente puede analizarse de manera relativa a otro estimador del mismo parámetro. A continuación estudiaremos la eficiencia de la media muestral como estimador de la media población relativa a la familia de estimadores lineales del mismo parámetro. Consideremos en primer lugar el estimador lineal fiel más general de la media poblacional, que será una combinación lineal de la forma

$$M(a_1, \dots, a_n) = \sum_{i=1}^n a_i x_i. \quad (3.19)$$

El valor esperado de este estadístico es

$$E\{M\} = \sum_{i=1}^n a_i E\{x_i\} = \mu \sum_{i=1}^n a_i \quad (3.20)$$

Por tanto, como ya habíamos mencionado (§ nota al pie 3), el estimador M es un estimador fiel de la media poblacional sí y sólo sí $\sum_{i=1}^n a_i = 1$. Calculemos su varianza,

$$\begin{aligned} \sigma_M^2 &= E\{(M - \mu)^2\} = E\left\{\left(\sum_{i=1}^n a_i x_i - \mu\right)^2\right\} \\ &= E\left\{\left(\sum_{i=1}^n a_i x_i - \mu \sum_{i=1}^n a_i\right)^2\right\} \\ &= E\left\{\left[\sum_{i=1}^n a_i (x_i - \mu)\right]^2\right\}. \end{aligned} \quad (3.21)$$

Dado que la esperanza matemática de los términos cruzados será nula por independencia estadística de los datos procedentes del muestreo aleatorio

⁴Es importante notar que los resultados reflejados en esta sección, tales como $\sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{n}}$, son válidos para cualquier distribución de probabilidad.

simple, podemos reexpresar la ecuación anterior de la forma

$$\begin{aligned}\sigma_M^2 &= E \left\{ \sum_{i=1}^n a_i^2 (x_i - \mu)^2 \right\} \\ &= \sigma^2 \sum_{i=1}^n a_i^2,\end{aligned}\tag{3.22}$$

donde, de nuevo, hemos usado que $E \left\{ (x_i - \mu)^2 \right\} = \sigma^2$. Debemos calcular ahora los parámetros a_1, \dots, a_n que hacen que el estimador $M(a_1, \dots, a_n)$ tenga varianza mínima, i.e. sea el estimador más eficiente de todos los lineales. Para ello debemos resolver el problema de extremos condicionados de la función $\sigma_M^2(a_1, \dots, a_n)$ de la ecuación anterior sometida a la restricción $\sum_{i=1}^n a_i = 1$. Construyamos a tal efecto la lagrangiana⁵

$$L(a_1, \dots, a_n) = \sum_{i=1}^n a_i^2 + \alpha \sum_{i=1}^n a_i, \quad \alpha \in \mathbb{R},\tag{3.23}$$

y maximicemos esta última función. Los extremos de L vendrán dados por

$$\vec{\nabla} L(a_1, \dots, a_n) = 0 \Leftrightarrow \frac{\partial L}{\partial a_j} = 0 \quad \forall j = 1, \dots, n.\tag{3.24}$$

En nuestro caso,

$$\frac{\partial L}{\partial a_j} = 2a_j + \alpha = 0 \Leftrightarrow a_j = -\frac{\alpha}{2} = a \quad \forall j = 1, \dots, n.\tag{3.25}$$

Es inmediato demostrar que el extremo así obtenido es un mínimo de L y por tanto de la varianza de los estimadores tipo $M(a_1, \dots, a_n)$. Luego, el estimador lineal que tiene una mayor eficiencia en la estimación de la media poblacional será aquél cuyos coeficientes sean todos ellos idénticos

$$M = a \sum_{i=1}^n x_i\tag{3.26}$$

y, dado que su suma ha de ser igual a 1:

$$a_i = \frac{1}{n} \quad \forall i = 1, \dots, n.\tag{3.27}$$

De la combinación de las dos ecuaciones anteriores se sigue que el estimador lineal más eficiente de la media poblacional es

$$M = \frac{1}{n} \sum_{i=1}^n x_i \equiv \bar{x},\tag{3.28}$$

que corresponde a la media muestral.

⁵El mismo problema se puede resolver de forma más elemental si la ligadura se incorpora a la función escribiendo $\sigma_M^2 = \sigma^2 \sum_{i=1}^n a_i^2 + \sigma^2 \left(1 - \sum_{i=1}^{n-1} a_i\right)^2$. Basta con ver que todas las derivadas de esta expresión con respecto a las a_i tienen la misma forma para deducir que todos estos coeficientes son iguales entre sí y, por normalización, a n^{-1} .

Finalmente, por lo que respecta a la distribución de probabilidad del estadístico media muestral, \bar{x} , debemos notar que si la distribución de la población madre $f(x)$ es una distribución normal, entonces \bar{x} , que es suma de variables normales, tiene también una distribución normal. De las ecuaciones (3.13) y (3.17) se sigue que, en este caso, \bar{x} es una variable con distribución normal de media μ y varianza σ^2/n , $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, en pleno acuerdo con el teorema del límite central. Aún en el caso de que la distribución de la población, $f(x)$, no sea gaussiana, si tiene momentos finitos, entonces en virtud del teorema del límite central, \bar{x} tenderá a tener una distribución gaussiana cuando n sea suficientemente grande. Volveremos sobre este punto con mayor detalle una vez analizado el estimador de la varianza poblacional, lo que haremos en la sección siguiente.

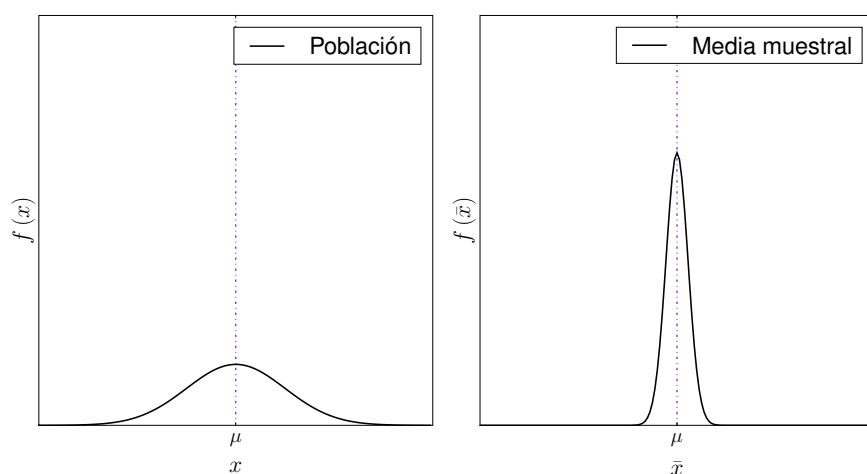


Figura 3.1: Distribución de probabilidad de cada dato x de una muestra y de la media muestral \bar{x} de 20 datos obtenida de una población normal estándar.

3.1.2. Estimación puntual de la varianza poblacional

La estimación de la varianza de una población a partir de datos de una muestra también presenta un gran interés en múltiples aplicaciones prácticas, como la cuantificación de las fluctuaciones en magnitudes físicas tales como la energía o la imanación, o la determinación de la incertidumbre de una medida física, que se tratará específicamente en el próximo capítulo.

Al igual que hemos hecho en la sección anterior con la estimación puntual de la media poblacional a partir de la media muestral, parece lógico utilizar la varianza muestral

$$s'^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.29)$$

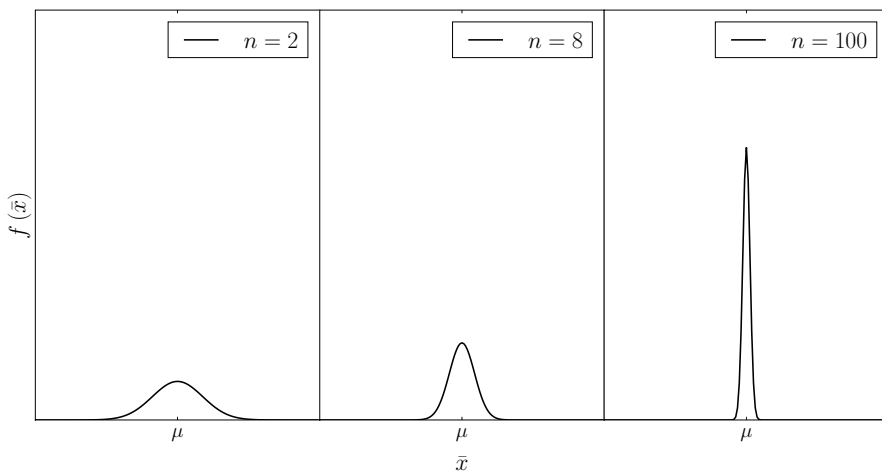


Figura 3.2: Evolución de la densidad de probabilidad de la media de una muestra obtenida de una población normal estándar al aumentar el número de elementos n de la muestra. Obsérvese la disminución de la varianza de \bar{x} cuando aumenta n .

como estimador de su homóloga poblacional, σ^2 . Veamos si éste es un estimador fiel de la varianza poblacional. Calculemos la esperanza matemática de esta variable aleatoria:

$$\begin{aligned}
 E\{s'^2\} &= E\left\{\frac{1}{n}\sum_{i=1}^n(x_i - \bar{x})^2\right\} = \frac{1}{n}E\left\{\sum_{i=1}^n(x_i - \mu + \mu - \bar{x})^2\right\} \\
 &= \frac{1}{n}E\left\{\sum_{i=1}^n[(x_i - \mu)^2 + (\mu - \bar{x})^2 + 2(x_i - \mu)(\mu - \bar{x})]\right\} \\
 &= \frac{1}{n}E\left\{\sum_{i=1}^n(x_i - \mu)^2 + \sum_{i=1}^n(\mu - \bar{x})^2 + 2(\mu - \bar{x})\sum_{i=1}^n(x_i - \mu)\right\}. \quad (3.30)
 \end{aligned}$$

Usando ahora que

$$\sum_{i=1}^n(x_i - \mu) = \sum_{i=1}^n x_i - n\mu = n\bar{x} - n\mu = n(\bar{x} - \mu) \quad (3.31)$$

el miembro de la derecha de la Ec. (3.30) puede reescribirse de la forma:

$$\begin{aligned}
 E\{s'^2\} &= \frac{1}{n} E \left\{ \sum_{i=1}^n (x_i - \mu)^2 + \sum_{i=1}^n (\mu - \bar{x})^2 - 2(\mu - \bar{x})n(\bar{x} - \mu) \right\} \\
 &= \frac{1}{n} E \left\{ \sum_{i=1}^n (x_i - \mu)^2 + n(\mu - \bar{x})^2 - 2n(\mu - \bar{x})^2 \right\} \\
 &= \frac{1}{n} E \left\{ \sum_{i=1}^n (x_i - \mu)^2 - n(\mu - \bar{x})^2 \right\}, \tag{3.32}
 \end{aligned}$$

con lo cual obtenemos que

$$E\{s'^2\} = \frac{1}{n} E \left\{ \sum_{i=1}^n (x_i - \mu)^2 \right\} - E\{(\mu - \bar{x})^2\}. \tag{3.33}$$

Ahora bien por la linealidad del operador E ,

$$\begin{aligned}
 E\{s'^2\} &= \frac{1}{n} \sum_{i=1}^n E\{(x_i - \mu)^2\} - E\{(\mu - \bar{x})^2\} \\
 &= \frac{1}{n} n\sigma^2 - \sigma^2(\bar{x}) = \frac{1}{n} n\sigma^2 - \frac{\sigma^2}{n} \\
 &= \frac{n-1}{n} \sigma^2 \tag{3.34}
 \end{aligned}$$

donde hemos usado que $E\{(\mu - \bar{x})^2\} = \sigma^2(\bar{x}) = \sigma^2/n$ y $E\{(x_i - \mu)^2\} = \sigma^2$. Dado que la esperanza matemática del estimador s'^2 no es igual al parámetro estimado, σ^2 , sino que difiere de él en un factor multiplicativo que depende del tamaño de la muestra, debemos concluir que s' no es un estimador fiel de la varianza de la población. En cambio, si construimos el estimador

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \tag{3.35}$$

tendremos que:

$$E\{s^2\} = \frac{n}{n-1} E\{s'^2\} = \sigma^2, \tag{3.36}$$

con lo que s^2 sí es un estimador fiel de la varianza poblacional. La aparición del factor $n-1$ en el denominador del estimador s^2 puede parecer extraña a primera vista. Sin embargo, en el caso $n=1$, tendríamos $\bar{x} = x_1$ y $s'^2 = 0$, lo cual no es en absoluto correcto. El problema fundamental es que en la construcción del estimador de la varianza poblacional hemos de usar \bar{x} en lugar de μ , por lo que el número de diferencias $(x_i - \bar{x})$ independientes es una unidad menor al número de datos. Así pues, el número efectivo de grados de libertad es $n-1$

al calcular la varianza muestral s^2 , correspondiente al número de desviaciones cuadráticas independientes en la expresión del estimador.

Ejemplo 3.3

Consideremos que hemos medido la longitud de un objeto siete veces obteniendo la siguiente muestra de datos:

$$L(\text{mm}) \ 10,5, \ 10,9, \ 9,2, \ 9,8, \ 9,0, \ 10,4, \ 10,7$$

Calculemos la media y la varianza muestral de estas medidas. La media muestral es, como siempre,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{7}(10,5 + 10,9 + 9,2 + 9,8 + 9,0 + 10,4 + 10,7) = 10,07 \text{ mm.}$$

Sin embargo, para disponer de una estimación fiel de la varianza poblacional a partir de los datos muestrales hemos de usar el valor obtenido anteriormente para \bar{x} , por lo que el número de grados de libertad es únicamente 6. Así, el estimador de la varianza poblacional será

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) = \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) = \\ &= \frac{1}{6}(0,43^2 + 0,83^2 + 0,87^2 + 0,27^2 + 1,07^2 + 0,33^2 + 0,63^2) = \\ &= 0,56 \text{ mm}^2, \end{aligned}$$

con lo que $s = 0,75$ mm será a su vez una estimación fiel de la desviación típica de la población. De nuevo cabe recordar que el resultado obtenido para s^2 garantiza que esta variable se distribuye en torno a un valor medio igual a la varianza de la población.

Dejamos para el lector (en los ejercicios) la demostración de que el estimador s^2 es también un estimador consistente de la varianza poblacional dado que

$$\sigma^2(s^2) = \sigma^4 \left(\frac{2}{n-1} + \frac{(g-3)}{n} \right) = 0 ; \quad g = \frac{E\{(x-\mu)^4\}}{\sigma^4} \quad (3.37)$$

y por tanto,

$$\lim_{n \rightarrow \infty} \sigma^2(s^2) = \lim_{n \rightarrow \infty} \sigma^4 \left(\frac{2}{n-1} + \frac{(g-3)}{n} \right) = 0 \quad (3.38)$$

En muchos casos la distribución de la varianza muestral tiene un comportamiento asimétrico. Su forma dependerá de la forma de la distribución de probabilidad de la población $f(x)$ y de n , el número de medidas de la muestra.

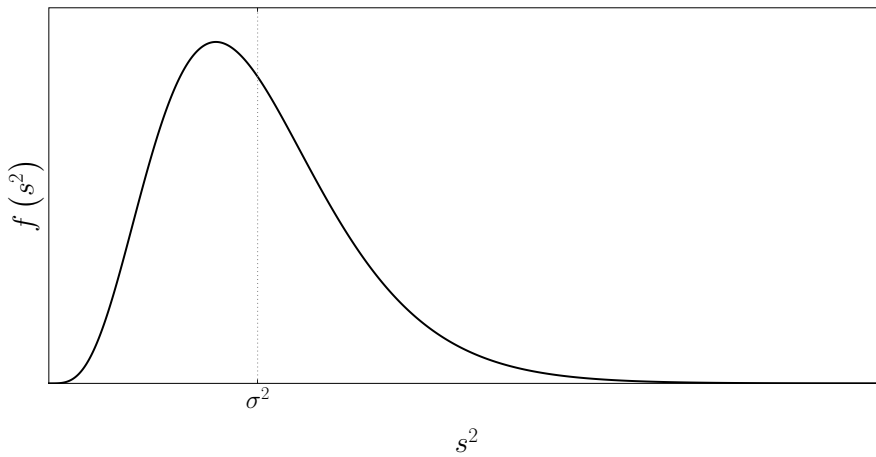


Figura 3.3: Distribución de la varianza de una muestra de 11 elementos obtenida de una población madre normal.

Analicemos finalmente la distribución de probabilidad de la variable s^2 . Ya vimos que en el caso de tener una distribución de la población normal $N(\mu, \sigma)$, entonces la variable lineal \bar{x} es normal con distribución $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Esto no es así para el estimador s^2 , dada la no linealidad de su expresión. Consideramos el estimador:

$$\frac{(n-1)}{\sigma^2} s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma^2}, \quad (3.39)$$

y realicemos el cambio de variable $\{x_i\}_{i=1}^n \rightarrow \{y_i\}_{i=1}^n$ definidas de la forma:

$$\begin{aligned} y_1 &= \frac{1}{\sqrt{1 \cdot 2}}(x_1 - x_2) \\ y_2 &= \frac{1}{\sqrt{2 \cdot 3}}(x_1 + x_2 - 2x_3) \\ y_3 &= \frac{1}{\sqrt{3 \cdot 4}}(x_1 + x_2 + x_3 - 3x_4) \\ &\vdots \\ y_{n-1} &= \frac{1}{\sqrt{(n-1)n}}(x_1 + x_2 + \dots + x_{n-1} - (n-1)x_n) \\ y_n &= \frac{1}{\sqrt{n}}(x_1 + x_2 + \dots + x_n) = \sqrt{n} \bar{x}. \end{aligned} \quad (3.40)$$

Es posible demostrar matemáticamente que (ejercicio 2):

$$\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2, \quad (3.41)$$

de modo que

$$\begin{aligned}
 (n-1)s^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 = \\
 &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n y_i^2 - y_n^2 = \sum_{i=1}^{n-1} y_i^2.
 \end{aligned} \tag{3.42}$$

Dada la linealidad de la transformación (3.40), si las variables x_i son gaussianas, las nuevas variables y_i tienen también una distribución normal con media nula, i.e. todas, salvo y_n , se distribuyen de acuerdo a $N(0, \sigma)$. Con el objetivo de normalizar la expresión dividimos las variables transformadas por su desviación típica, obteniendo las variables $z_i = y_i/\sigma$ que tienen una distribución de probabilidad normal estándar, $N(0, 1)$. Luego, el estadístico en la Ec. (3.39) será la suma de variables normales estándar al cuadrado:

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n-1} \frac{y_i^2}{\sigma^2}. \tag{3.43}$$

Para obtener la distribución de probabilidad del estadístico anterior, consideremos una muestra de n variables que siguen una distribución normal estándar con media nula y desviación típica unidad, z_1, z_2, \dots, z_n , y definamos una nueva variable a través de la suma de cuadrados de las variables aleatorias anteriores⁶

$$\chi^2 = z_1^2 + z_2^2 + \dots + z_n^2. \tag{3.44}$$

Esta nueva variable sigue una distribución de probabilidad de χ^2 de Pearson con n grados de libertad. Veamos cuál es la forma de esta distribución de probabilidad. Para ello consideremos el caso elemental de una única variable aleatoria $x^2 = z^2$ donde, al ser z una variable aleatoria $N(0, 1)$,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \tag{3.45}$$

La función de distribución acumulativa para χ^2 será

$$\begin{aligned}
 F(\chi^2) &= P(z^2 < \chi^2) = P\left(-\sqrt{\chi^2} < z < \sqrt{\chi^2}\right) \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{\chi^2}}^{\sqrt{\chi^2}} e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{\chi^2}} e^{-z^2/2} dz.
 \end{aligned} \tag{3.46}$$

Si queremos expresar esta distribución de probabilidad en función de χ^2 más que de su raíz cuadrada hagamos el cambio de variable,

⁶Nota: χ^2 es una variable y no el cuadrado de una variable. El exponente denota su origen como suma de cuadrados de variables distribuidas de acuerdo con una distribución normal.

$$\begin{aligned}
 u &= z^2 \\
 du &= 2z dz \\
 dz &= \frac{1}{2} u^{-1/2} du
 \end{aligned}
 \tag{3.47}$$

De esta manera, la función de distribución acumulativa de la Ec. (3.46) puede reexpresarse de la forma

$$F(\chi^2) = \frac{1}{\sqrt{2\pi}} \int_0^{\chi^2} du u^{-1/2} e^{-u/2}. \tag{3.48}$$

Aplicando el teorema fundamental del cálculo integral a la ecuación anterior la densidad de probabilidad para χ^2 es

$$f(\chi^2) = \frac{1}{\sqrt{2\pi}} (\chi^2)^{-1/2} e^{-\chi^2/2}. \tag{3.49}$$

Esta distribución de probabilidad es la distribución de χ^2 para 1 sólo grado de libertad. Si consideramos la distribución de n grados de libertad $\chi^2 = z_1^2 + z_2^2 + \dots + z_n^2$

$$f(\chi^2) = \frac{1}{\Gamma(\lambda) 2^\lambda} (\chi^2)^{\lambda-1} e^{-\chi^2/2}. \tag{3.50}$$

siendo $\lambda = \frac{n}{2}$, n el número de grados de libertad y

$$\Gamma(\lambda) = \int_0^\infty dt e^{-t} t^{\lambda-1}.$$

la función gamma de Euler ⁷.

La distribución de probabilidad de la variable aleatoria χ^2 posee propiedades muy interesantes. En particular, destacaremos:

1. La distribución de probabilidad de la suma de dos variables que se comporten como χ^2 con n_1 y n_2 grados de libertad respectivamente es una variable que se comporta como una χ^2 de $n = n_1 + n_2$ grados de libertad.
2. El valor esperado de una distribución de χ^2 es n .

$$E\{\chi^2\} = n. \tag{3.51}$$

3. La varianza de una distribución de χ^2 es $2n$.

$$E\{(\chi^2 - n)^2\} = 2n. \tag{3.52}$$

⁷Es interesante resaltar que la distribución de χ^2 es un caso particular de la conocida como distribución Gamma

$$f(x; k\theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}; x > 0, k, \theta > 0$$

para $k = n/2$ y $\theta = 2$.

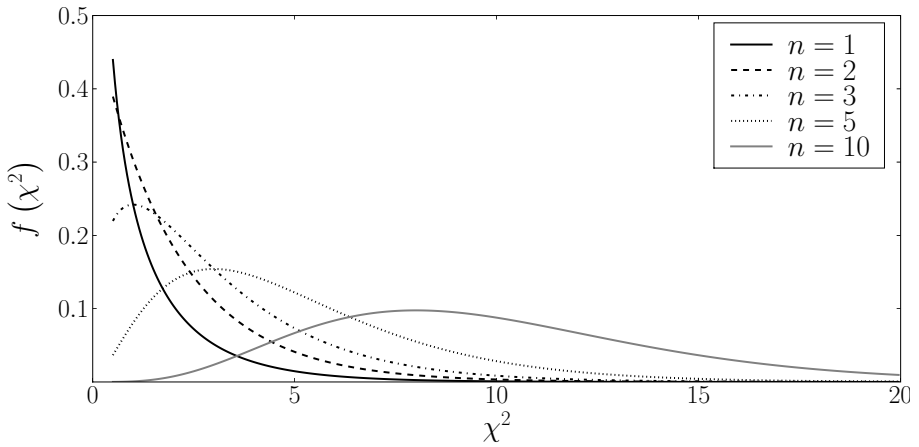


Figura 3.4: Distribución χ^2 para diferentes números de grados de libertad.

En la Fig. 3.4 vemos, además, que la distribución de χ^2 es asimétrica.

Como consecuencia de lo anteriormente expuesto, si las variables $\{x_i\}_{i=1}^n$ siguen una distribución normal $N(\mu, \sigma)$ entonces la variable

$$\chi^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{\sigma^2} \quad (3.53)$$

se comporta como una χ^2 de Pearson con n grados de libertad. Por tanto, la variable aleatoria originariamente considerada para el estudio de la distribución de probabilidad del estadístico s^2 en la Ec. (3.39),

$$\frac{(n-1)s^2}{\sigma^2} = \chi_{n-1}^2, \quad (3.54)$$

se comporta como una χ^2 de Pearson con $n-1$ grados de libertad. La razón de que sean $n-1$ se debe a que no todos los valores $(x_i - \bar{x})^2$ son independientes al estar ligados por

$$\sum_{i=1}^n (x_i - \bar{x}) = 0. \quad (3.55)$$

Cuando $n > 30$ la densidad de probabilidad de χ^2 se simetriza y se pueden usar fórmulas aproximadas para χ^2 basadas en aproximaciones gaussianas.

Ejemplo 3.4

Una fábrica de cojinetes tiene un control de calidad para las bolas de acero que garantiza que $\sigma^2 = 0,001 \text{ mm}^2$. ¿Cuál es la probabilidad de que en una muestra de 21 bolas la varianza sea mayor o igual a $0,0016 \text{ mm}^2$?

Queremos calcular $P(s^2 \geq 0,0016)$ el número de grados de libertad es $n = 21 - 1 = 20$. Sabemos que la variable aleatoria

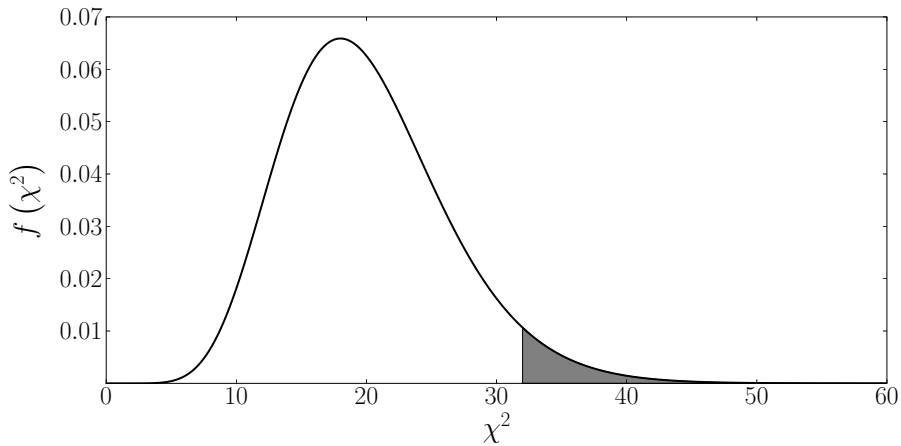
$$\frac{(n-1)s^2}{\sigma^2} = \chi_{n-1}^2$$

sigue una distribución χ^2 de Pearson con $n - 1$ grados de libertad. En el caso concreto considerado esto implica que

$$\chi_{20}^2 = \frac{20 \cdot 0,0016}{0,001} = 32$$

Por tanto, que la varianza muestral sea mayor que 0,0016 es equivalente a que χ_{20}^2 sea mayor que 32:

$$P(s^2 \geq 0,0016) = P(\chi_{20}^2 \geq 32)$$



$$P(\chi_{20}^2 \geq 32) = \int_{32}^{\infty} dn f(n).$$

Los valores de la χ^2 están tabulados (véase el apéndice) y en este caso

$$P(\chi_{20}^2 \geq 32) \simeq 0,045$$

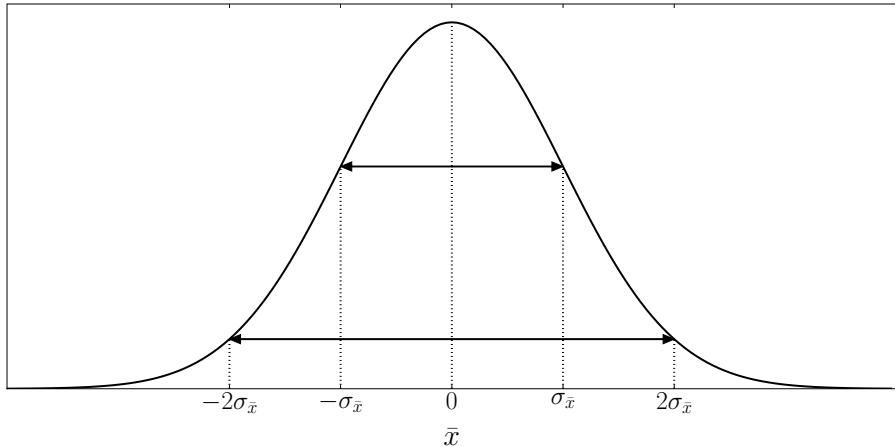
Por tanto la probabilidad de que $s^2 \geq 0,0016$ es un 4,5 %.

3.1.3. Distribución de \bar{x} para poblaciones madre normales

Como ya hemos mencionado, si la variable aleatoria X tiene una distribución $N(\mu, \sigma)$, entonces \bar{x} tiene una distribución $N(\mu, \frac{\sigma}{\sqrt{n}})$.

Ejemplo 3.5

La medida del peso de un huevo de gallina producido en Granjas S.A. tiene una distribución normal con media $\mu = 50$ g y varianza $\sigma^2 = 80$ g². Estudiar los intervalos para los que la probabilidad de obtener un cierto valor de \bar{x} con una muestra de 20 huevos sean 68,3 %, 95,4 % y 99,7 %.



Dado que la distribución de la variable X es $N(50, \sqrt{80})$, la distribución de \bar{x} para muestras de 20 huevos será normal, de media igual a la media de la población madre, 50 g, y varianza

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n} = \frac{80}{20} = 4 \text{ g}^2$$

$$\sigma(\bar{x}) = 2 \text{ g}$$

Así pues, de acuerdo con las propiedades de la distribución gaussiana estudiadas en el capítulo anterior, el 68,3 % de las medias muestrales caerán en el intervalo

$$\mu \pm \sigma^2(\bar{x}) = 50 \pm 1 \times 2 \text{ g} \quad [48, 52] \text{ g.}$$

De la misma forma el 95,4 % caerán en el intervalo

$$\mu \pm 2\sigma^2(\bar{x}) = 50 \pm 2 \times 2 \text{ g} \quad [46, 54] \text{ g,}$$

y el 99,7 % caerán en el intervalo

$$\mu \pm 3\sigma^2(\bar{x}) = 50 \pm 3 \times 2 \text{ g} \quad [44, 56] \text{ g.}$$

En el caso de una distribución madre que no sea normal pero que tenga varianza finita se considera que si $n \geq 30$ entonces la distribución de $z = \frac{x-\mu}{\sigma}$ es muy próxima a una distribución normal $N(0, 1)$, valor convencionalmente aceptado para considerar que tenemos un número suficiente de variables aleatorias para aplicar el teorema del límite central.

En el ejemplo anterior, cuando estimamos los intervalos para \bar{x} con un determinado valor de probabilidad consideramos conocida la σ de la población. En muchos casos σ no es conocida separadamente y ha de ser estimada a partir de la propia muestra, usando el valor del estimador varianza muestral, s . El estadístico que debemos considerar entonces es

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} = t. \quad (3.56)$$

Fijémonos que estamos utilizando en la construcción de este estadístico que un estimador fiel para $\sigma(\bar{x})$ es

$$\sigma(\bar{x}) \simeq \frac{s}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}. \quad (3.57)$$

Analícemos ahora la distribución de probabilidad que sigue la variable aleatoria t , que no será una distribución normal dado que se ha sustituido la varianza poblacional por su estimador en la expresión de t . Para ello reexpresemos la Ec. (3.56) de la forma:

$$\begin{aligned} t &= \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\bar{x} - \mu}{\frac{s}{\sigma\sqrt{n}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{s} \\ &= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{\sigma^2}}} = \frac{\bar{x} - \mu}{\sqrt{\frac{(n-1)s^2}{(n-1)\sigma^2}}}. \end{aligned} \quad (3.58)$$

Después de esta reordenación de términos, podemos considerar que el numerador $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ corresponde a una variable normal estándar $z \in N(0, 1)$ (si X es una variable gaussiana $N(\mu, \sigma)$). El denominador es esencialmente una variable $\chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma^2}$ de Pearson con $n-1$ grados de libertad. Por lo tanto podemos reescribir la variable aleatoria t de la forma:

$$t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}}. \quad (3.59)$$

Sea z una variable normal distribuida según $N(0, 1)$ y sea una variable χ_n^2 distribuida según una distribución de x^2 de Pearson con n grados de libertad, la variable

$$t_n = \frac{z}{\sqrt{\frac{\chi_n^2}{n}}}, \quad (3.60)$$

se dice que sigue una distribución t de Student⁸ con n grados de libertad, cuya densidad de probabilidad es

$$f(t) = \frac{\Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}n)} \frac{1}{\sqrt{\pi}\sqrt{n}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)}. \quad (3.61)$$

⁸Seudónimo de William Sealy Gosset Jr., matemático estadístico.

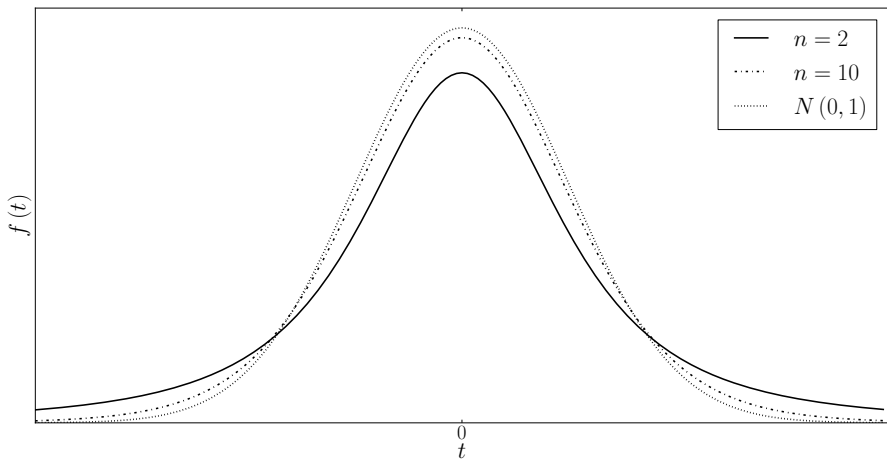


Figura 3.5: Distribución t de Student para $n = 2$ y $n = 10$ en comparación con una normal estándar.

La distribución de la t de Student es una distribución de probabilidad centrada en cero y simétrica. Cuando n es pequeña esta distribución de probabilidad es más ancha que la distribución normal (tiene una menor curtosis, es decir, es leptocúrtica). Obviamente, como un caso particular del teorema del límite central, la distribución $f(t)$ tiende a una distribución normal $N(0, 1)$ cuando el número de grados de libertad (i.e. el número de elementos de la muestra) es suficientemente grande, $n \rightarrow \infty$.

Entonces, usando la definición anterior, la distribución de probabilidad del estadístico t en la ecuación (3.56) es una t de Student con $n - 1$ grados de libertad.

3.2. Intervalos de confianza

Normalmente, cuando tratamos de estimar un parámetro no sólo nos interesa obtener un valor puntual sino un intervalo que nos informe acerca de la incertidumbre de tal estimación. El propio estimador ha de considerarse una variable estadística, por lo que es necesario completar la información procedente de la estimación puntual con un intervalo en el que esperamos encontrar el verdadero valor del parámetro con un determinado nivel de probabilidad o de confianza. Aunque en principio esto se aplica a todos los estimadores de parámetros que podamos considerar, en esta obra introductoria analizaremos únicamente los intervalos de confianza correspondientes a la media poblacional estimada a partir de la media muestral y los intervalos de confianza para la estimación de la varianza poblacional a partir de la muestral en poblaciones madre normales. Comencemos estudiando los supuestos de estimación de intervalos de confianza para la media poblacional.

3.2.1. Intervalos de confianza para la media poblacional

El objetivo es ahora la obtención de un intervalo en el cual, con una probabilidad predeterminada, se encuentre la media poblacional μ , una vez conocida la media de una muestra de dicha población, \bar{x} .

Asignemos una probabilidad $1 - \alpha$ al hecho de que el parámetro μ esté en el intervalo (a, b)

$$P(a \leq \mu \leq b) = 1 - \alpha.$$

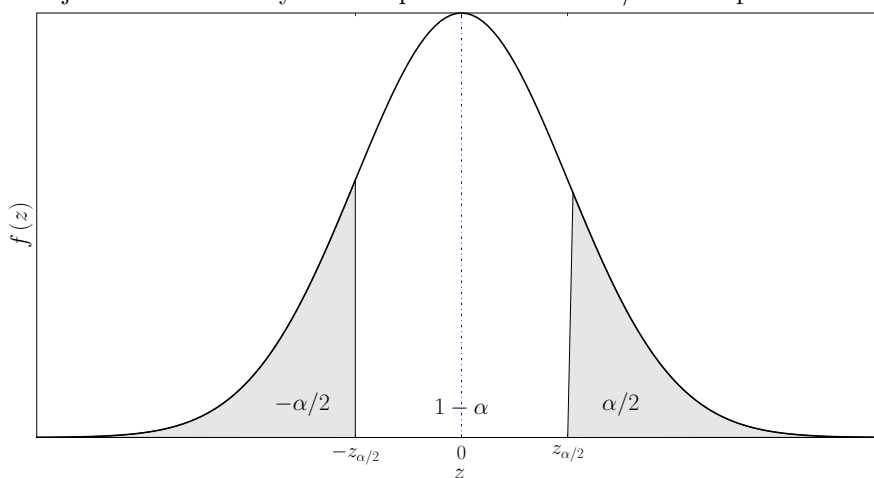
A $1 - \alpha$ lo conocemos como el nivel de confianza y definirá el tamaño del intervalo. Para construir este intervalo de confianza habremos de apoyarnos en \bar{x} . La probabilidad de error que admitimos en la definición de nuestro intervalo es α y en muchas ocasiones se denota en términos de tanto por ciento.

$$\begin{aligned} 0,90 &= 1 - \alpha && \text{nivel de confianza del 90 \%} \\ 0,10 &= \alpha && \text{probabilidad de error del 10 \%} \end{aligned}$$

Los valores más usados para α son 0,10, 0,05 y 0,01 que corresponden a intervalos de confianza del 90 %, 95 % y 99 % respectivamente. En general, el tamaño del intervalo de confianza crece con $1 - \alpha$ y depende también del número de datos de la muestra, (cuanto menor sea n , mayor es el intervalo). Analicemos a continuación diversos casos particulares de obtención de intervalos de confianza para la media poblacional en poblaciones madre normales (μ) y para los parámetros de las distribuciones binomial (p) y de Poisson (λ).

A) Intervalo de confianza para la media μ de una población madre normal con σ conocida

Consideremos el cálculo del intervalo en el que, con un nivel de confianza del $100(1 - \alpha)\%$, i.e. con una probabilidad $(1 - \alpha)$, encontraremos la media poblacional. Éste coincidirá con el intervalo en el que podemos encontrar la media muestral con una probabilidad $1 - \alpha$, limitado por los valores de los percentiles de la distribución de probabilidad de este estadístico que dejan a su derecha y a su izquierda un valor $\alpha/2$ de la probabilidad:



Dado que consideramos que la población madre de la muestra de datos es normal, la variable aleatoria

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

se distribuye de acuerdo a una distribución normal estándar $N(0, 1)$ y nos permite estimar la media poblacional. El intervalo que buscamos lo podremos obtener a partir de los percentiles de la distribución normal estándar que dejan a su izquierda y derecha un área igual a $\alpha/2$. Sabemos que, en el $100(1 - \alpha)\%$ de los casos,

$$-z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \quad (3.62)$$

de lo que se deduce que el intervalo de confianza para la media poblacional buscado es

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}. \quad (3.63)$$

Ejemplo 3.6

Consideremos que en una muestra de 25 datos se tiene una media $\bar{x} = 127$, y supongamos que conocemos que la población madre de la muestra tiene una desviación típica $\sigma = 5,4$. Calcular un intervalo de confianza del 95% para la media de la población madre de la muestra. En este caso,

$$\alpha = 0,05 \quad z_{\alpha/2} = 1,96$$

y el intervalo de confianza buscado será por tanto

$$\begin{aligned} 127 - (1,96) \frac{5,4}{\sqrt{25}} &\leq \mu \leq 127 + (1,96) \frac{5,4}{\sqrt{25}} \\ 124,9 &\leq \mu \leq 129,1. \end{aligned}$$

También podemos estimar el tamaño que ha de tener una muestra determinada con el objetivo de que un intervalo de tamaño dado contenga la media poblacional a un nivel de confianza aceptable. Para ello debemos considerar:

1. el nivel de confianza $(1 - \alpha)$ que deseamos.
2. La diferencia máxima que admitiremos entre el valor estimado y el valor verdadero, D .

Si consideramos la estimación de la media poblacional mediante

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}},$$

la distancia entre las medias poblacional y muestral estará acotada de la forma:

$$-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}. \quad (3.64)$$

Por lo tanto, la diferencia máxima admitida a este nivel de confianza $1 - \alpha$ será

$$D = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (3.65)$$

Despejando n en la expresión anterior obtenemos trivialmente:

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{D^2} \quad (3.66)$$

B) Intervalo de confianza para la media μ de una población madre normal con σ desconocida

En este caso debemos estimar la varianza (equivalentemente la desviación típica) de la población mediante la desviación típica muestral, y considerar el parámetro

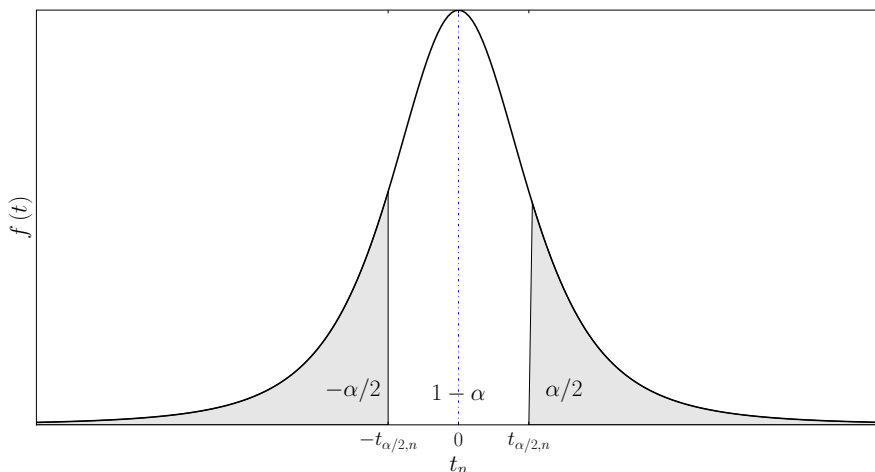
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}},$$

que, como hemos visto con anterioridad, se comporta como una distribución t de Student con $n - 1$ grados de libertad. De modo que si nuestro nivel de confianza es $100(1 - \alpha)\%$ debemos obtener los percentiles de la distribución t de Student que dejan a su izquierda y a su derecha un área igual a $\alpha/2$. Denotaremos por $t_{\alpha/2, n}$ al valor de la abscisa de la distribución t de Student con n grados de libertad que verifica esta propiedad. Entonces en el $100(1 - \alpha)\%$ de los casos se cumplirá que

$$-t_{\frac{\alpha}{2}, n-1} \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq t_{\frac{\alpha}{2}, n-1} \quad (3.67)$$

o, equivalentemente,

$$\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}.$$



Ejemplo 3.7

Un equipo de baloncesto obtiene las siguientes puntuaciones en 10 partidos: 69, 81, 67, 80, 71, 70, 78, 68, 57 y 59.

Si consideramos que los tantos del equipo siguen una distribución de Gauss, calcular el intervalo de confianza para su media con un nivel de confianza del 99%.

En este ejemplo se observa que la media y desviación típica muestrales son, respectivamente

$$\begin{aligned}\bar{x} &= 70 & s &= 8,10 \\ s(\bar{x}) &= \frac{s}{\sqrt{n}} = 2,56\end{aligned}$$

En este caso el número de grados de libertad es 9 y $\alpha = 0,01$, por lo que

$$t_{\frac{\alpha}{2}, n} = t_{0,005, 9} = 3,25$$

Luego, el intervalo de confianza buscado es:

$$\begin{aligned}70 - 3,25 \left(\frac{8,10}{3,16} \right) &\leq \mu \leq 70 + 3,25 \left(\frac{8,10}{3,16} \right) \\ 61,67 &\leq \mu \leq 78,33\end{aligned}$$

Vemos que se trata de un intervalo bastante amplio. Si queremos reducirlo manteniendo el nivel de confianza debemos aumentar n . Supongamos que aumentásemos el número de grados de libertad, n , tomando los datos de 30 encuentros, y que se mantuviese en este caso el valor de la media muestral \bar{x} y de la varianza s . En este caso el percentil asociado al intervalo de confianza del 99% es

$$t_{0,005, 20} = 2,756$$

y, consecuentemente, el intervalo de confianza buscado será

$$\begin{aligned}70 - 2,76 \left(\frac{8,1}{\sqrt{30}} \right) &\leq \mu \leq 70 + 2,76 \left(\frac{8,1}{\sqrt{30}} \right) \\ 65,92 &\leq \mu \leq 74,08\end{aligned}$$

Si por el contrario preferiésemos mantener el tamaño de la muestra y aumentar nuestro riesgo al 10 %, entonces

$$t_{0,05,9} = 1,83$$

con un intervalo de confianza del 90 %

$$\begin{aligned} 70 - 1,83 \left(\frac{8,10}{3,16} \right) &\leq \mu \leq 70 + 1,83 \left(\frac{8,10}{3,16} \right) \\ 65,3 &\leq \mu \leq 74,7 \end{aligned}$$

Observemos que, manteniendo el número de grados de libertad, el intervalo se hace mayor a medida que aumentamos el nivel de confianza. El centro de cada intervalo corresponde a la mejor estimación puntual de μ .

Como hemos visto, la obtención de intervalos de confianza para la media poblacional difiere notablemente dependiendo de si conocemos o no la varianza poblacional. En el primer caso podemos considerar que la distancia normalizada entre la media muestral y la media poblacional $(\bar{x} - \mu) / (\sigma / \sqrt{n})$, es una variable aleatoria gaussiana siempre que x lo sea. Sin embargo en el segundo, la necesidad de estimar la varianza poblacional a partir de su homóloga muestral, obliga a utilizar distancias entre medias que siguen una distribución t de Student. Esta diferencia se va haciendo más pequeña a medida que aumentamos el tamaño de la muestra, ya que en este caso, la distribución t de Student, tiende a la gaussiana. Normalmente si $n > 30$, entonces podremos considerar que el parámetro

$$\frac{\bar{x} - \mu}{s / \sqrt{n}}$$

presenta un comportamiento aproximadamente gaussiano con distribución normal estándar. Así pues, en este caso el intervalo de confianza de la media poblacional puede obtenerse como

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \quad (3.68)$$

siendo $z_{\frac{\alpha}{2}}$ el percentil correspondiente a la distribución normal que deja a su derecha $\alpha/2$ de la probabilidad. En algunos casos, por simplicidad, se suele adoptar un nivel de confianza del 68,3 % que corresponde a $z_{\alpha/2} = 1$, con lo cual en este caso, el intervalo de confianza buscado será:

$$\bar{x} - \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}} \leq \mu \leq \bar{x} + \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}. \quad (3.69)$$

Como veremos en el capítulo siguiente, este caso es particularmente importante en el tratamiento de las incertidumbres de medida de magnitudes físicas.

C) Intervalo de confianza para el parámetro p de una distribución binomial

Consideremos una variable aleatoria binomial X , y recordemos que si realizamos n pruebas de Bernoulli sucesivas e independientes la probabilidad de tener k éxitos de probabilidad p es

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (3.70)$$

Denotando $q = 1 - p$ suele escribirse

$$P(X = k) = \binom{n}{k} p^k q^{n-k}.$$

Como vimos en el capítulo anterior, esta distribución verifica $\mu = np$ y $\sigma^2 = npq$. Consideremos ahora el siguiente estimador para el parámetro p

$$\hat{p} = \frac{\text{número de éxitos en } n \text{ ensayos}}{n} = \frac{x}{n}. \quad (3.71)$$

La esperanza matemática y la varianza de este estimador son, respectivamente,

$$\begin{aligned} E\{\hat{p}\} &= \frac{1}{n} E\{x\} = \frac{1}{n}(np) = p \\ \sigma^2(\hat{p}) &= E\left\{\left(\frac{x}{n} - p\right)^2\right\} = \frac{1}{n^2} E\{(x - np)^2\} = \frac{npq}{n^2} = \frac{pq}{n}, \end{aligned} \quad (3.72)$$

por lo que vemos que \hat{p} es un estimador fiel y consistente del parámetro p . Además, si tenemos ya información segura de que la distribución de esta variable es binomial, en lugar de usar la expresión convencional de la varianza muestral en la Ec. (3.35), podemos realizar nuestra estimación de la varianza poblacional como

$$\sigma(\hat{p}) \simeq \sqrt{\frac{\frac{x}{n} \left(1 - \frac{x}{n}\right)}{n}}. \quad (3.73)$$

Dado que la distribución binomial tiene todos sus momentos finitos, para muestras suficientemente grandes, el parámetro

$$\frac{\hat{p} - p}{\sigma(\hat{p})} = \frac{\hat{p} - p}{\sqrt{\frac{\frac{x}{n} \left(1 - \frac{x}{n}\right)}{n}}}$$

tendrá a ser $N(0, 1)$, de modo que para un nivel de confianza $(1 - \alpha)$ el correspondiente intervalo de confianza⁹ para p es

$$\frac{x}{n} - z_{\alpha/2} \sqrt{\frac{\frac{x}{n} \left(1 - \frac{x}{n}\right)}{n}} \leq p \leq \frac{x}{n} + z_{\alpha/2} \sqrt{\frac{\frac{x}{n} \left(1 - \frac{x}{n}\right)}{n}}. \quad (3.74)$$

⁹Obsérvese que el intervalo de confianza para el parámetro p puede convertirse en un intervalo de confianza para el valor esperado $p \cdot n$.

Ejemplo 3.8

En una muestra de 144 familias se hallan 48 familias con 2 o más televisores, hallar un intervalo del 95% de confianza para la proporción de las familias con 2 o más televisores.

$$\hat{p} = \frac{x}{n} = \frac{48}{144} = \frac{1}{3}$$

$$\sigma(\hat{p}) \simeq \sqrt{\frac{\frac{x}{n}(1-\frac{x}{n})}{n}} = \sqrt{\frac{\frac{1}{3}(\frac{2}{3})}{144}} = 0,039.$$

Para el 95% de nivel de confianza, $\alpha = 0,05$ y $z_{\alpha/2} = 1,96$

$$\frac{1}{3} - 1,96(0,039) \leq p \leq \frac{1}{3} + 1,96(0,039)$$

$$0,26 \leq p \leq 0,41.$$

Así pues, para un nivel de confianza del 95% la proporción de familias con 2 o más televisores está entre el 26,0% y el 41,0%.

D) Intervalo de confianza para el parámetro λ de una distribución de Poisson

Consideremos una variable aleatoria de Poisson X con media λ . Como vimos, su distribución de probabilidad viene dada por

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

que tiene media $\mu = \lambda$ y varianza $\sigma^2 = \lambda$. Si tomamos una muestra genérica de la población formada por los elementos (sucesos) $\{x_1, x_2, \dots, x_n\}$, un estimador para λ es la media muestral

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i.$$

que, como sabemos, tiene una varianza $\sigma^2(\hat{\lambda}) = \frac{\sigma^2}{n} = \frac{\lambda}{n}$. Cuando n es suficientemente grande¹⁰, $\hat{\lambda}$ sigue una distribución normal $N\left(\lambda, \sqrt{\frac{\lambda}{n}}\right)$, por lo que el parámetro

$$z = \frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{n}}}$$

se comportará como una variable normal estándar $N(0, 1)$, siempre que se verifique que $\hat{\lambda}$ sea también grande. Notemos además que para la estimación de

¹⁰En el caso de contajes pequeños ha de usarse la propia distribución de Poisson.

λ en la desviación típica de la distribución usamos su estimador $\hat{\lambda}$. Así pues, a un nivel de confianza $1 - \alpha$, el intervalo de confianza del parámetro λ de una distribución de Poisson para contajes grandes es:

$$\hat{\lambda} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}} \leq \lambda \leq \hat{\lambda} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}}. \quad (3.75)$$

Ejemplo 3.9

Se realizan con un detector Geiger 100 medidas de 1 minuto de duración del número de desintegraciones de una muestra radiactiva y se obtiene un total de 640.000 desintegraciones durante las 100 medidas. Con un nivel de confianza del 95 % establecer un intervalo para la actividad de la muestra.

En este caso,

$$\hat{\lambda} = \frac{640,000}{100} = 6,400 \text{ cuentas/minuto}$$

$$\sigma(\hat{\lambda}) \simeq \sqrt{\frac{\hat{\lambda}}{100}} = \frac{80}{10} = 8 \text{ cuentas/minuto.}$$

Como consideramos un nivel de confianza $1 - \alpha = 0,95$, entonces $\alpha = 0,05$ y $z_{0,025} = 1,96$ (suponiendo que el fenómeno es ya aproximadamente gaussiano dado lo elevado del número de cuentas). Por tanto el nivel de confianza buscado es

$$6400 - 1,96 \times 8 \leq \lambda \leq 6400 + 1,96 \times 8$$

$$6384 \leq \lambda \leq 6416.$$

3.2.2. Intervalo de confianza para la varianza σ^2 de una distribución normal

Hemos visto anteriormente que la variable aleatoria $(n - 1)s^2/\sigma^2$ sigue una distribución χ^2 de Pearson con $n - 1$ grados de libertad. Si denotamos por $\chi_{\alpha/2, n}^2$ el valor de la abscisa de una distribución de χ^2 con n grados de libertad que deja a su derecha un área $\alpha/2$, se cumplirá que:

$$P \left[\chi_{1-\frac{\alpha}{2}; n-1}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{\frac{\alpha}{2}; n-1}^2 \right] = 1 - \alpha, \quad (3.76)$$

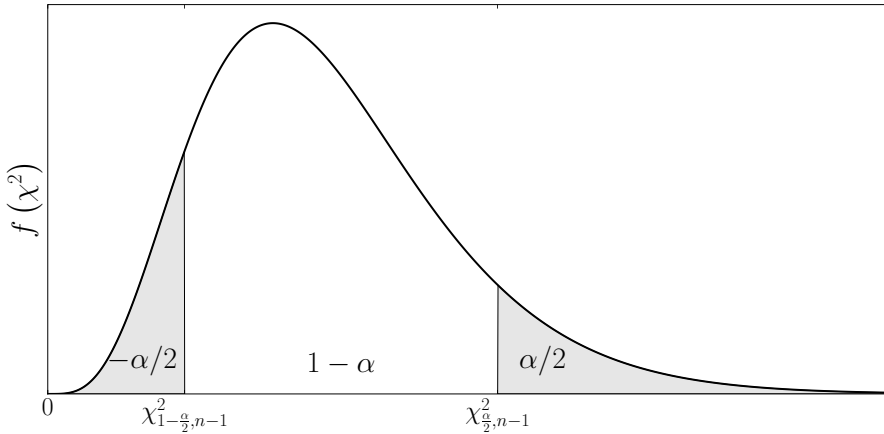
o lo que es lo mismo,

$$P \left[\sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}; n-1}^2} \right] = P \left[\chi_{1-\frac{\alpha}{2}; n-1}^2 \leq \frac{(n-1)s^2}{\sigma^2} \right] = 1 - \frac{\alpha}{2}$$

$$P \left[\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}; n-1}^2} \leq \sigma^2 \right] = P \left[\frac{(n-1)s^2}{\sigma^2} \leq \chi_{\frac{\alpha}{2}; n-1}^2 \right] = \frac{\alpha}{2} \quad (3.77)$$

Por tanto, los valores a y b que hacen que la probabilidad de que la varianza poblacional se encuentre en el intervalo (a, b) sea $1-\alpha$ (nivel de confianza buscado) serán¹¹

$$P \left[\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}; n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}; n-1}^2} \right] = 1 - \alpha \quad (3.78)$$



Ejemplo 3.10

Si consideramos de nuevo los tanteos de los partidos de baloncesto, obtendremos que $(n = 10)$

$$s^2 = \frac{1}{9}(590) = 65,6$$

Si deseamos un intervalo al 95% de nivel de confianza para la varianza σ^2 , entonces

$$\left. \begin{array}{l} \chi_{0,025;9}^2 = 19,02 \\ \chi_{0,975;9}^2 = 2,70 \end{array} \right\}$$

por lo que el intervalo de confianza buscado es:

$$\begin{aligned} \frac{9,65,6}{19,2} &\leq \sigma^2 \leq \frac{9,65,6}{2,7} \\ 31,1 &\leq \sigma^2 \leq 218,5. \end{aligned}$$

¹¹Debe tenerse en cuenta que la distribución de la variable aleatoria χ^2 es asimétrica, por lo que en los cocientes aparecen percentiles diferentes.

3.3. Test de hipótesis estadísticas

Una hipótesis estadística es una afirmación que se realiza sobre una o más características de una población. Una hipótesis estadística puede ser, por ejemplo, la que establece que la probabilidad de obtener cara al lanzar una moneda sea $1/2$, una ley física como la que establece que la intensidad en un circuito eléctrico es directamente proporcional a la diferencia de potencial aplicada sobre el mismo, siendo el coeficiente de proporcionalidad la resistencia equivalente del circuito; que un conjunto de puntos experimentales se ajusta a una determinada curva matemática o cualquier otra afirmación respecto al valor de un parámetro de la población. La emisión de estas afirmaciones acerca del comportamiento general de una población y las condiciones para su verificación son el objetivo último de la Estadística matemática. Obviamente, estas hipótesis deben ser verificadas o contrastadas mediante la observación empírica o experimental del fenómeno en un número más o menos amplio de elementos de la población. Una hipótesis estadística estaría completamente contrastada si pudiéramos comprobar su verificación sobre todos los elementos de una población, pero, ya sea porque la población es infinita o porque en la práctica no es accesible el conjunto de elementos de la misma, en general el proceso de verificación estadística de la hipótesis se realiza mediante una muestra de la población.

Ejemplo 3.11

Un ingeniero encargado del control de calidad de una fábrica ha de verificar la proporción p de artículos defectuosos producidos. Para ello adopta el valor crítico del 5% y realiza las hipótesis excluyentes y exhaustivas

$$p > 0,05 \quad (a)$$

$$p \leq 0,05 \quad (b)$$

A continuación toma una muestra de tamaño n donde se encuentran x artículos defectuosos. El estimador de la proporción poblacional de elementos defectuosos será

$$\hat{p} = \frac{x}{n},$$

y según los resultados muestrales habrá de decidir entre la hipótesis (a) o la hipótesis (b).

Distinguiremos en general entre hipótesis simples y compuestas. Por ejemplo, en el caso anterior si el ingeniero hace la hipótesis de que $p = 0,1$ ésta es una hipótesis simple. Se suele considerar como hipótesis compuesta aquella que afirma un recorrido o intervalo para el(los) parámetro(s) a estudio.

$$H(p = 0,1) \quad \text{simple}$$

$$H(p \geq 0,1) \quad \text{compuesta}$$

Como habíamos expresado antes, el test de hipótesis estadísticas se presenta mediante la contraposición de dos hipótesis, la hipótesis nula y la hipótesis alternativa. Por ejemplo, si se ha determinado mediante un cierto experimento que en una muestra un determinado parámetro μ toma el valor μ_0 y emitimos la hipótesis $\mu = \mu_0$ para la población, esta sería la hipótesis nula, ya que implicaría no esperar nada nuevo en otras mediciones, y $\mu \neq \mu_0$ sería la hipótesis alternativa:

$$\begin{aligned} H_0; \mu = \mu_0 &\rightarrow \text{hipótesis nula} \\ H_a; \mu \neq \mu_0 &\rightarrow \text{hipótesis alternativa.} \end{aligned} \quad (3.79)$$

También podrían usarse como alternativa $\mu > \mu_0$ o $\mu < \mu_0$, por ejemplo. Debe siempre intentarse que aquel valor que, según el criterio del investigador, corresponda al valor verdadero esté reflejado en la hipótesis nula H_0 , y que la hipótesis alternativa contenga el resto de los valores.

Ejemplo 3.12

Supongamos que mediante un experimento gravimétrico se ha podido determinar, a partir de una muestra de datos, que la aceleración de la gravedad en un cierto punto de la superficie terrestre es $g=9,810 \text{ m s}^{-2}$. Entonces, la hipótesis de que $g=9,810 \text{ m s}^{-2}$ para la población de todas las posibles medidas será la hipótesis nula.

El proceso de test de una hipótesis estadística se compone de los siguientes pasos:

1. Definición de la hipótesis nula H_0 y la hipótesis alternativa.
2. Definición de una medida de discrepancia entre los datos muestrales y la hipótesis H_0 . Se usará para ello una función de los datos de la muestra (por ejemplo la media muestral \bar{x}).
3. Definición de la región de aceptación de H_0 , constituida por el conjunto de valores que consideraremos aceptable si H_0 es cierta. Para ello es necesario fijar el nivel de significación α , que es la probabilidad de rechazar H_0 siendo correcta. Se suele denominar error de tipo I a aquel que se comete cuando se rechaza H_0 siendo correcta. Su probabilidad se denomina α . Se denomina error de tipo II a aquel en que se acepta H_0 siendo falsa, y su probabilidad se denota por β .
4. Cálculo de la función de decisión (estadístico de contraste) y resolución acerca de la aceptación o rechazo de H_0 .

Analizaremos a continuación los procesos de test de hipótesis estadísticas para la media y la varianza de poblaciones normales, tests de hipótesis acerca de distribuciones de probabilidad, para finalizar con el test χ^2 de bondad de un

ajuste de datos experimentales a una curva matemática expresión de alguna ley teórica.

3.3.1. Test de la media de una población normal

Consideremos una población supuesta gaussiana, sobre cuya media μ realizamos la siguiente afirmación:

$$\begin{aligned} H_0 & : \mu = \mu_0 \\ H_a & : \mu \neq \mu_0 \end{aligned} \quad (3.80)$$

Consideremos como estadístico de contraste la variable

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (3.81)$$

que mide la distancia normalizada entre los datos muestrales y la media poblacional (hipótesis estadística) y, según se ha visto con anterioridad, sigue una t -Student con $n - 1$ grados de libertad¹². La región crítica, es decir, el valor de t para el cual rechazamos la hipótesis a un nivel de confianza $1 - \alpha$, H_0 será

$$c = \{t : |t| > t_{\frac{\alpha}{2}; n-1}\}, \quad (3.82)$$

y la región de aceptación de la hipótesis será

$$\Delta = \{t : |t| \leq t_{\frac{\alpha}{2}; n-1}\}. \quad (3.83)$$

Así, aceptaremos la hipótesis estadística H_0 si

$$\frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} \leq t_{\frac{\alpha}{2}; n-1},$$

y rechazaremos H_0 si

$$\frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} > t_{\frac{\alpha}{2}; n-1}.$$

Observemos que esto está muy relacionado con los intervalos de confianza. Un intervalo de confianza puede considerarse simplemente como el conjunto de hipótesis aceptables para ese nivel de confianza.

El contraste de hipótesis analizado anteriormente es de tipo bilateral al estar interesados en rechazar los valores en los que el estadístico de contraste es mayor o menor que unos determinados valores. El test de hipótesis sería unilateral p.ej. si nos interesara rechazar únicamente aquellos casos en que t fuese mayor o igual que un cierto valor. En el caso del contraste unilateral consideraremos como región crítica en la cual rechazamos la hipótesis

$$c = \{t : t > t_{\alpha; n-1}\}, \quad (3.84)$$

¹²Como ya sabemos, en el caso de muestras grandes $n > 30$ puede sustituirse $t_{\frac{\alpha}{2}; n-1}$ por el percentil $z_{\frac{\alpha}{2}}$ de una distribución gaussiana.

y la región de aceptación será

$$\Delta = \{t : t \leq t_{\alpha;n-1}\} \quad (3.85)$$

De esta manera se acepta H_0 si

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} \leq t_{\alpha;n-1},$$

y se rechaza H_0 si

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{\alpha;n-1}.$$

Ejemplo 3.13

Un laboratorio farmacéutico afirma que un calmante es efectivo contra el dolor de cabeza en 14 min. Con el fin de comprobar estadísticamente esta afirmación se administra la sustancia a 18 personas con este problema y se obtiene

$$\bar{x} = 19 \text{ min} \quad s = 7 \text{ min}$$

Comprobar la validez de la afirmación a un nivel de significación del 5%.

En este caso,

Hipótesis nula, $H_0 : \mu = 14 \text{ min}$

Hipótesis alternativa $H_a : \mu \neq 14 \text{ min}$.

El estadístico de contraste para el test de la media es

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

Dado que en el supuesto que nos ocupa

$$\mu_0 = 14 \text{ min}; \quad \bar{x} = 19 \text{ min}; \quad s = 7 \text{ min}; \quad n = 18,$$

el valor del estadístico de contraste será:

$$t = \frac{19 - 14}{7/\sqrt{18}} = 3,03.$$

El nivel de significación es $\alpha = 0,05$ por lo que el percentil a considerar para el test bilateral es aquel valor que deja una área de $\alpha/2 = 0,025$ a la derecha para la distribución de probabilidad del estadístico de contraste, la t de Student en el presente caso. Este valor es $t_{0,025;17} = 2,11$, y dado que $t > t_{0,025;17}$ se rechaza la hipótesis, lo que indica que, con una probabilidad superior al 95% el tiempo de actuación del medicamento es superior a 14 min.

3.3.2. Contraste de la varianza de una población normal

Consideremos una población normal y supongamos que emitimos la hipótesis de que la varianza σ^2 toma un cierto valor σ_0^2 a un nivel de confianza $1 - \alpha$ o bien un nivel de significación α

$$\begin{array}{ll} \text{Hipótesis nula} & H_0 : \sigma^2 = \sigma_0^2 \\ \text{Hipótesis alternativa} & H_a : \sigma^2 \neq \sigma_0^2. \end{array} \quad (3.86)$$

Dado que, como se ha visto en secciones anteriores de esta obra, la distribución de probabilidad asociada a la varianza muestral es la χ^2 de Pearson, una elección obvia para el estadístico de contraste será

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (3.87)$$

que mide la discrepancia entre el valor de la varianza muestral, s^2 , y el supuesto para la población, σ_0^2 . Definimos la región crítica donde rechazaremos la hipótesis.

$$c = \left\{ \chi^2 : \chi^2 < \chi_{1-\frac{\alpha}{2}; n-1}^2 \quad \chi^2 > \chi_{\frac{\alpha}{2}; n-1}^2 \right\}. \quad (3.88)$$

La región de aceptación de la hipótesis estadística será

$$A = \left\{ \chi^2 : \chi_{1-\frac{\alpha}{2}; n-1}^2 \leq \chi^2 \leq \chi_{\frac{\alpha}{2}; n-1}^2 \right\}, \quad (3.89)$$

por lo que aceptaremos la hipótesis H_0 si

$$\chi_{1-\frac{\alpha}{2}; n-1}^2 \leq \frac{(n-1)s^2}{\sigma_0^2} \leq \chi_{\frac{\alpha}{2}; n-1}^2, \quad (3.90)$$

y la rechazaremos si

$$\chi_{1-\frac{\alpha}{2}; n-1}^2 > \frac{(n-1)s^2}{\sigma_0^2},$$

ó

$$\chi_{\frac{\alpha}{2}; n-1}^2 < \frac{(n-1)s^2}{\sigma_0^2}.$$

También podemos realizar en este caso un test unilateral si nos interesa estudiar que σ^2 sea mayor o menor que un cierto valor dado. Consideremos para ello alternativamente

$$\begin{array}{ll} H_0 & : \sigma^2 \leq \sigma_0^2 \\ H_1 & : \sigma^2 > \sigma_0^2 \end{array} \quad (3.91)$$

Entonces las regiones crítica y de aceptación son, respectivamente,

$$\begin{array}{ll} c & = \left\{ \chi^2 : \chi^2 > \chi_{\alpha; n-1}^2 \right\} \\ A & = \left\{ \chi^2 : \chi^2 \leq \chi_{\alpha; n-1}^2 \right\}. \end{array} \quad (3.92)$$

Ejemplo 3.14

Una empresa de bloques de hormigón armado dice que su material tiene una desviación típica menor o igual que $7,75 \text{ kg cm}^{-2}$ en términos de resistencia respecto al valor medio de 220 kg cm^{-2} . Se toma una muestra de 9 elementos, obteniendo como resultado 203, 229, 215, 220, 223, 233, 208, 228 y 209. Contrastar la hipótesis de que esta muestra provenga de una población con $\sigma_0 \leq 7,75 \text{ kg cm}^{-2}$ al 5% de nivel de significación.

Siguiendo el protocolo de contraste de hipótesis, estableceremos en primer lugar las hipótesis nula y alternativa para el supuesto que nos ocupa:

$$\begin{aligned} H_0 &: \sigma \leq 7,75 \text{ kg cm}^{-2} \\ H_a &: \sigma > 7,75 \text{ kg cm}^{-2} \end{aligned}$$

Teniendo en cuenta que la media y la desviación típica muestrales son

$$\begin{aligned} \bar{x} &= 218,67 \text{ kg cm}^{-2} \\ s &= \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2} = 10,52 \text{ kg cm}^{-2} \end{aligned}$$

el estadístico de contraste χ_{n-1}^2 toma el valor

$$\chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{8(10,52)^2}{(7,75)^2} = 14,74$$

El percentil de la distribución χ^2 de Pearson con 8 grados de libertad que deja a su derecha un área de 0.05 es

$$\chi_{0,05;8}^2 = 15,51.$$

y como $14,74 < 15,51$ aceptamos la hipótesis de que es cierta H_0 . Es interesante notar que a un nivel de significación del 10% hubiéramos rechazado la hipótesis dado que $\chi_{0,10;8}^2 = 13,9$.

3.3.3. Test de χ^2 de una distribución de probabilidad

En muchas situaciones debemos analizar la adecuación de un fenómeno aleatorio a una determinada distribución de probabilidad teórica. Cuando suponemos una distribución de probabilidad para un fenómeno concreto estamos emitiendo una hipótesis estadística que tendrá que ser debidamente verificada mediante las técnicas antes expuestas de test de hipótesis para poder ser utilizada con seguridad. Así, se realizará un experimento aleatorio del fenómeno en cuestión, obteniéndose una serie de frecuencias de sucesos discretos o agrupados en clases que se histogramarán de acuerdo con lo expuesto en el capítulo primero de esta obra. El test consistirá en comparar la frecuencia absoluta observada para los sucesos de la clase i -ésima, O_i , con la frecuencia esperada si los sucesos estuviesen gobernados por la distribución de probabilidad supuesta, E_i .

Consideremos una muestra de n datos de la forma $\{x_1, x_2, \dots, x_n\}$. Agruparemos los datos en clases para histogramarlos de modo que el número de clases k sea $k \geq 5$. Normalmente intentaremos que haya al menos tres datos de cada clase y que éstos se distribuyan de un modo homogéneo. A partir de la propia muestra haremos una estimación p_i que corresponde a la probabilidad de obtener un dato en la clase i -ésima

$$E_i = np_i,$$

que sería la frecuencia esperada de la clase i -ésima si el fenómeno se adecuase a la distribución de probabilidad supuesta. Por otro lado, si consideramos el intervalo correspondiente a la clase i -ésima, I_i , la variable aleatoria sólo tiene dos posibilidades: tomar valores dentro o fuera de él. Por ello, la hipótesis estadística que vamos a manejar es

$$H_0 : \left. \begin{array}{l} E_1 = E\{O_1\} \\ E_2 = E\{O_2\} \\ \vdots \\ E_k = E\{O_k\} \end{array} \right\}, \quad (3.93)$$

donde O_i es el número de veces que la variable aleatoria en cuestión toma valores en I_i , i.e. la frecuencia absoluta de la clase i -ésima observada en el experimento. Como de costumbre, $E\{O_i\}$ representa la esperanza matemática de la variable aleatoria O_i . Considerando nuevamente que la variable aleatoria únicamente puede o no tomar valores en una determinada clase, esperamos que la variable aleatoria O_i obedezca una distribución binomial con esperanza $E_i = np_i$ y desviación típica $\sqrt{np_i q_i}$. En el caso de que p_i sea pequeño, podremos considerar que su distribución es de Poisson y $q_i \approx 1$.

Construyamos ahora el estadístico de contraste

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}, \quad (3.94)$$

que seguirá una χ^2 de Pearson con un número determinado de grados de libertad dado que podemos suponer que la variable

$$z = \frac{O_i - E_i}{\sqrt{E_i}} = \frac{O_i - np_i}{\sqrt{np_i}} \quad (3.95)$$

es $N(0, 1)$ cuando $E_i \leq 5$. En cuanto al número de grados de libertad de la distribución que gobierna el estadístico de contraste de la Ec. (3.94):

1. Si el modelo había predicho la probabilidad p_i de cada clase de modo independiente de la muestra ($E_i = np_i$) entonces la χ^2 tiene $k - 1$ grados de libertad dado que existe una ligadura:

$$\sum_{i=1}^k O_i = n. \quad (3.96)$$

2. Si el modelo se ha ajustado usando información de la muestra (el valor esperado, por ejemplo) y se han estimado r parámetros a partir de dicha muestra, entonces χ^2 sigue una distribución χ^2 de Pearson con $k - r - 1$ grados de libertad.

Finalmente, a un nivel de confianza $1 - \alpha$, rechazaremos el modelo si se cumple

$$\chi^2 \geq \chi_{\alpha; k-r-1}^2. \quad (3.97)$$

En general si el modelo es rechazado convendrá estudiar la razón, ya que es suficiente una discrepancia importante, por ejemplo en una sola clase, para surtir este efecto.

Ejemplo 3.15

Durante una epidemia de gripe, en una ciudad se hace un muestreo contando el número de infectados en cien grupos conectados (p.ej. familias) de cinco personas, con los resultados:

x_i : número de afectados	0	1	2	3	4	5
O_i : frecuencia	33	41	19	4	3	0

A fin de saber si la epidemia se halla en su máximo o todavía progresando, se pide determinar al 90% de nivel de confianza si esta distribución es compatible con la hipótesis de una distribución uniforme.

Aunque estrictamente hablando el proceso estaría gobernado por una distribución binomial, dado que la media muestral es aproximadamente 1, significativamente menor que $n = 5$, podemos aproximarla por una distribución de Poisson. Por ello, nuestra hipótesis estadística acerca de la distribución de probabilidad será que si la gripe estuviese distribuida homogéneamente entre la población, los casos seguirían una distribución de Poisson de media

$$\hat{\lambda} = \frac{\sum_{i=0}^5 x_i O_i}{\sum_{i=0}^5 O_i} = 1,03.$$

Por lo tanto, testaremos si los datos siguen una distribución de probabilidad

$$P_i = \frac{(1,03)^i}{i!} e^{-1,03},$$

que proporciona unas frecuencias esperadas

$$\begin{aligned} E_i &= p_i \cdot 100 \\ E_0 &= 35,70 \\ E_1 &= 36,77 \\ E_2 &= 18,94 \\ E_3 &= 6,50 \\ E_4 &= 1,67 \\ E_5 &= 0,34 \end{aligned}$$

Por tanto, el estadístico de contraste toma el valor

$$\begin{aligned}\chi^2 &= \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} = \frac{(33 - 35,70)^2}{35,70} + \dots \\ &+ \dots + \frac{(0 - 0,34)^2}{0,34} = 3,05.\end{aligned}$$

Para realizar el test de la hipótesis acerca de la distribución de probabilidad del número de infectados por grupo humano, debemos calcular el percentil de la distribución χ^2 de Pearson con 4 grados de libertad que deja a su derecha un área de 0,05:

$$\chi_{0,05;4}^2 = 9,49.$$

Consecuentemente, a un nivel de confianza del 95 % debemos aceptar la hipótesis de que el número medio de infectados es una variable aleatoria que sigue una distribución de Poisson y , por tanto, con el mismo nivel de confianza, que de los datos empíricos obtenidos se sigue que la epidemia se halla uniformemente distribuida por la ciudad.

3.3.4. Test de χ^2 de la bondad de un ajuste

Se ha repetido en diversos lugares de esta obra que el objetivo de la actividad científica es inducir leyes de validez general que regulen las relaciones entre diferentes magnitudes a partir de observaciones empíricas o experimentales de determinados fenómenos. Por ello, buena parte de los resultados de los experimentos consiste en conjuntos de puntos en el plano correspondientes a observaciones simultáneas de dos variables, (x_i, y_i) , y , a partir de estos conjuntos deben inducirse leyes teóricas que regulen la dependencia de una variable en la otra. Estas leyes se traducirán en expresiones matemáticas de la forma $y = f(x)$. La asignación de una ley de este tipo al conjunto de los datos experimentales reviste, desde el punto de vista estadístico, el carácter de hipótesis estadística, que debe ser debidamente verificada de acuerdo con los procedimientos antes expuestos.

Consideremos que hemos ajustado un conjunto de datos $\{(y_i, x_i)\}_{i=1}^n$ a una cierta ley física $y = f(x)$, i.e. hemos emitido la hipótesis de que la variable estadística correspondiente a la magnitud física y se relaciona con la variable aleatoria correspondiente a la magnitud física x a través de la ley $y = f(x)$. Esperamos que, si cada y_i se distribuye de modo gaussiano en torno a su valor medio ¹³ \bar{y}_i , la variable

$$z_i = \frac{y_i - \bar{y}_i}{\sigma_i} \tag{3.98}$$

¹³Esto implica que si repitiésemos en varias ocasiones la medición de la magnitud y_i con x_i constante, obtendríamos un conjunto de datos que se distribuiría de manera normal o gaussiana en torno al valor medio, \bar{y}_i . Es razonable adoptar la hipótesis de distribución gaussiana dado que la mayoría de los fenómenos considerados en la práctica se adecúan a las hipótesis del teorema del límite central.

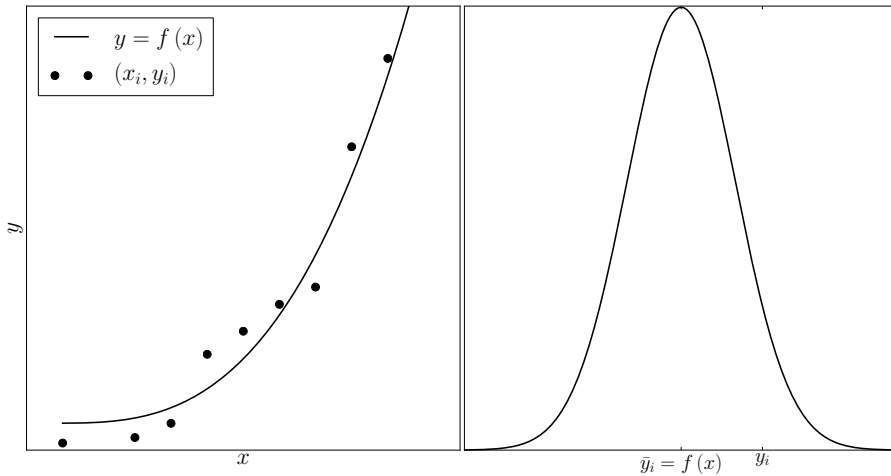


Figura 3.6: Izquierda: datos experimentales hipotéticos y curva de ajuste. Derecha: distribución normal que se supone centrada en la curva de ajuste para un y_i dado.

sea $z_i \in N(0, 1)$. Por tanto, esperaríamos que

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \bar{y}_i)^2}{\sigma_i^2} \quad (3.99)$$

responda a una distribución χ^2 de Pearson. Consideraremos que, a primer orden, los valores medios de la variable y coinciden con las predicciones de la ley teórica, razón por la cual $\bar{y}_i \simeq f(x_i)$, y consiguientemente

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - f(x_i))^2}{\sigma_i^2}, \quad (3.100)$$

que nos sirve como estadístico de contraste para el test de hipótesis. Las incertidumbres o desviaciones típicas de cada y_i , σ_i , serán en general diferentes. y podremos obtener una estimación de σ_i^2 como la varianza muestral obtenida mediante la repetición de su medida en varias ocasiones¹⁴,

$$\begin{aligned} \sigma_i^2 &\approx s^2(y_i) \\ \chi^2 &= \sum_{i=1}^n \frac{[y_i - f(x_i)]^2}{s^2(y_i)} \end{aligned} \quad (3.101)$$

El número de grados de libertad de la distribución χ^2 que gobierna el parámetro anterior será $n - r$ siendo r el número de parámetros que hayamos ajustado usando nuestra muestra de datos.

¹⁴En este valor de la incertidumbre no se incluye la incertidumbre intrínseca asociada al procedimiento experimental (aparato de medida, condiciones, etc.). Para un tratamiento detallado de la obtención de la incertidumbre de magnitudes experimentales remitimos al lector al capítulo siguiente del presente libro.

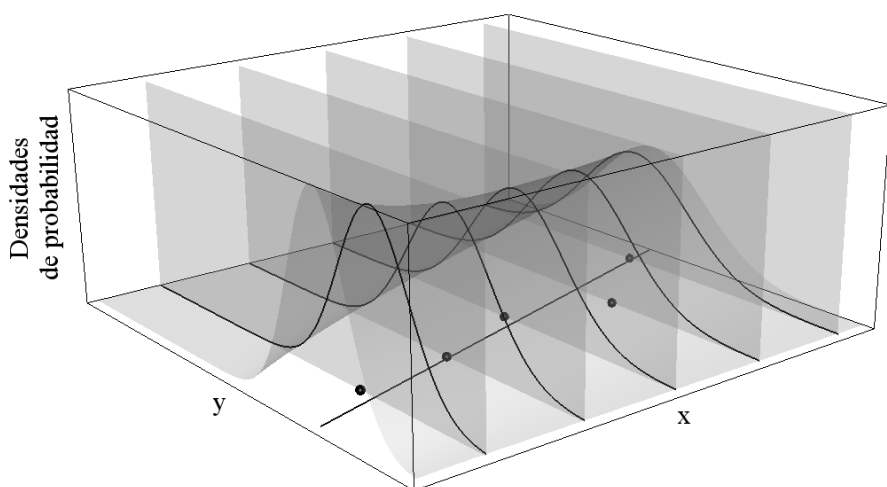


Figura 3.7: Visión conjunta de las distintas distribuciones de los y_i para un ajuste. En este caso se ajusta a una recta y $\sigma(y_i)$ crece logarítmicamente con y_i .

Para un nivel de confianza $1 - \alpha$, rechazaremos la hipótesis en que se basa el ajuste si

$$\chi^2 \geq \chi_{\alpha; n-r}^2 \quad (3.102)$$

Podría ocurrir que un ajuste pasara la prueba de χ^2 y sin embargo todos los puntos cumplieran $y_i > f(x_i)$ o manifestasen otro sesgo, con lo cual habría errores sistemáticos. Es siempre necesario realizar una inspección de los datos y su gráfica además de realizar un test de χ^2 para la calidad del ajuste.

Ejemplo 3.16

Se han tomado los siguientes datos experimentales de dos magnitudes x e y , con vistas a estudiar su correlación:

x_i	1	2	3	4	5
y_i	2,8	3,6	4,8	6,6	8,9

y todos los datos tienen $s(y) = 0,2$. Mediante un cierto método de ajuste un alumno afirma que un buen ajuste a los datos experimentales viene dado por

$$y = f(x) = 1,52x + 0,78,$$

donde se han ajustado la pendiente y la ordenada en el origen de la recta mediante estos datos. Realizar un test de χ^2 para probar a un nivel de confianza del 99% si este ajuste es válido.

De acuerdo con la ley teórica obtenida

x_i	1	2	3	4	5
$\tilde{y}_i = f(x_i)$	2,3	3,82	5,34	6,86	8,38

por tanto, el estadístico de contraste del ajuste será:

$$\chi^2 = \frac{(2,3 - 2,8)^2}{(0,2)^2} + \frac{(3,82 - 3,6)^2}{(0,2)^2} + \frac{(5,34 - 4,8)^2}{(0,2)^2} + \frac{(6,86 - 6,6)^2}{(0,2)^2} + \frac{(8,9 - 8,38)^2}{(0,2)^2} = 23,2.$$

Dado que hemos ajustado dos parámetros a partir de los datos experimentales, el número de grados de libertad de la distribución χ^2 de Pearson que gobierna el comportamiento del estadístico anterior será de $n-r=5-2=3$. El percentil de esta distribución que deja a su derecha un área de 0,01 (correspondiente a un nivel de significación del 10% o a un nivel de confianza del 90%) es:

$$\chi_{0,01;3}^2 = 11,3.$$

Como $\chi^2 > \chi_{0,01;3}^2$ rechazamos este ajuste al 99% de nivel de confianza, concluyendo que a este nivel de confianza: i) el ajuste no es lineal, o, ii) los parámetros determinados mediante el procedimiento empleado por el alumno no son los correctos.

3.3.5. Test de hipótesis para dos muestras

En ocasiones es necesario realizar test de hipótesis que afectan a más de una muestra, como por ejemplo el test de la hipótesis de que las medias de múltiples poblaciones normalmente distribuidas y con la misma varianza son iguales. Este es el ejemplo más característico del denominado test F de Fisher, una prueba en la que se analiza si el estadístico usado sigue la distribución F de Fisher si la hipótesis nula es cierta.

Consideremos dos muestras independientes, $\{x_i\}_{i=1}^{n_1}$ y $\{y_i\}_{i=1}^{n_2}$, procedentes de poblaciones madre normales de medias μ_1 y μ_2 respectivamente, y suponemos que queremos comparar sus medias y sus varianzas. En el primer caso buscamos un test estadístico para la hipótesis $\mu_1 = \mu_2$. En virtud del teorema del límite central, las medias muestrales están normalmente distribuidas, y por ello su diferencia $\delta = \bar{x} - \bar{y}$ también lo estará, con varianza

$$\sigma^2(\delta) = \sigma^2(\bar{x}) + \sigma^2(\bar{y}) = \frac{\sigma^2(x)}{n_1} + \frac{\sigma^2(y)}{n_2}, \quad (3.103)$$

estimada a través de las varianzas muestrales

$$s_\delta^2 = s_{\bar{x}}^2 + s_{\bar{y}}^2 = \frac{1}{n_1(n_1 - 1)} \sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \frac{1}{n_2(n_2 - 1)} \sum_{i=1}^{n_2} (y_i - \bar{y})^2, \quad (3.104)$$

En el caso de que la hipótesis de igualdad de las medias sea correcta, tendremos que $E\{\delta\} = 0$, por lo que el cociente

$$z = \frac{\delta}{\sigma(\delta)} \quad (3.105)$$

será una variable aleatoria normal estándar. Sin embargo, la desviación típica en el denominador de la expresión anterior únicamente puede estimarse a partir de las desviaciones típicas muestrales, por lo que hemos de construir el estimador

$$t = \frac{\delta}{s_\delta}, \quad (3.106)$$

que puede demostrarse que sigue una distribución t -Student con $n_1 + n_2 - 2$ grados de libertad. La igualdad de medias puede testarse ahora mediante el denominado test de diferencia de Student, que consiste en comparar el valor del estadístico anterior en un caso concreto con el percentil correspondiente al nivel de confianza deseado de la distribución t -Student. La región de aceptación de la hipótesis de igualdad de medias a un nivel de confianza $1 - \alpha$ sería, por lo tanto:

$$|t| = \frac{|\delta|}{\sigma(\delta)} \leq t_{\frac{\alpha}{2}; n_1 + n_2 - 2}. \quad (3.107)$$

Por otra parte, dadas dos muestras obtenidas de una población madre normal y de medias iguales, podríamos estar interesados en la realización de un test de igualdad de sus varianzas. Para ello consideremos las muestras introducidas al principio de la presente sección y construyamos el estadístico

$$F = \frac{s_x^2}{s_y^2}, \quad (3.108)$$

que puede reexpresarse de la forma

$$F = \frac{\frac{n_1 - 1}{n_1} \frac{s_x^2}{\sigma_x^2}}{\frac{n_2 - 1}{n_2} \frac{s_y^2}{\sigma_y^2}}, \quad (3.109)$$

variable que está formada por un cociente de variables χ^2 , pues, como hemos visto:

$$\chi_{n_1 - 1}^2 = \frac{(n_1 - 1) s_x^2}{\sigma_x^2}. \quad (3.110)$$

Es evidente que en el caso de que la hipótesis de igualdad de las varianzas de ambas muestras sea cierta tendremos

$$F = \frac{(n_2 - 1) \chi_{n_1 - 1}^2}{(n_1 - 1) \chi_{n_2 - 1}^2}. \quad (3.111)$$

Esta variable sigue la distribución de probabilidad $F(n_1, n_2)$ de Fisher o de Fisher-Snedecor¹⁵, y se trata de una distribución asimétrica con valores diferentes de cero únicamente en el semieje real positivo. Realizando un test F

¹⁵La variable aleatoria

$$F(n_1, n_2) = \frac{\chi_{n_1}^2/n_1}{\chi_{n_2}^2/n_2},$$

donde $\chi_{n_i}^2$ es una variable χ^2 de Pearson con n_i grados de libertad, sigue distribución F de

bilateral para la igualdad de varianzas a un nivel de confianza $1 - \alpha$ tenemos que la región de aceptación es

$$\Delta = \left\{ F(n_1, n_2)_{1-\frac{\alpha}{2}} \leq F \leq F(n_1, n_2)_{\frac{\alpha}{2}} \right\}, \quad (3.112)$$

donde, como de costumbre, utilizamos una notación según la cual

$$P \left[F(n_1, n_2) > F(n_1, n_2)_{1-\frac{\alpha}{2}} \right] = 1 - \frac{\alpha}{2}. \quad (3.113)$$

3.4. Método de máxima verosimilitud

Para completar el tratamiento de los métodos de determinación de parámetros vamos a introducir en la presente sección el método de máxima verosimilitud, que nos permitirá contemplar el problema de la estimación de parámetros y los ajustes de datos experimentales desde una perspectiva diferente a la utilizada hasta el momento. Supongamos que tenemos información para considerar que un determinado fenómeno se encuentra gobernado por una cierta distribución de probabilidad que depende de una serie de parámetros en principio desconocidos, $p(a_1, a_2, \dots, a_n)$. Para la determinación de estos parámetros debemos realizar un experimento para la obtención de datos que nos permitan calcularlos. El método de máxima verosimilitud establece que, una vez conocidos los datos, los parámetros óptimos para la distribución de probabilidad son aquellos que maximizan la probabilidad de obtención de ese conjunto de resultados experimentales. Este método es especialmente útil y sencillo en el caso de distribuciones normales, que es el caso que vamos analizar con detalle a continuación.

Supongamos que esperamos que nuestros datos, $\{x_i\}_{i=1}^n$, sigan una cierta distribución de probabilidad que está parametrizada por una serie de parámetros (a_1, a_2, \dots, a_n) . Denotemos por

$$p_i \equiv p(x_i; a_1, a_2, \dots, a_n) \quad (3.114)$$

la probabilidad de haber obtenido un cierto resultado x_i . Si se verifica la independencia estadística de los resultados sucesivos del experimento aleatorio, la probabilidad (*a posteriori*) de haber obtenido el conjunto $\{x_i\}_{i=1}^n$ será

$$P \equiv p_1 \cdot p_2 \cdot p_3 \dots p_n = \prod_{i=1}^n p_i. \quad (3.115)$$

Fisher o de Fisher-Snedecor, cuya función de densidad es:

$$f(x) = \frac{1}{B(n_1/2, n_2/2)} \left(\frac{n_1 x}{n_1 x + n_2} \right)^{n_1/2} \left(1 - \frac{n_1 x}{n_1 x + n_2} \right)^{n_2/2} \frac{1}{x}.$$

Los números n_1 y n_2 son enteros positivos, $x \in (0, \infty)$ y B representa la función beta:

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt.$$

Esta probabilidad *a posteriori* de haber obtenido la muestra se denomina función de verosimilitud

$$L(a_1, a_2, \dots, a_n) = \prod_{i=1}^n p_i \quad (3.116)$$

y es una función de los parámetros de la distribución de probabilidad supuesta a priori para el fenómeno. De acuerdo con el método de máxima verosimilitud, los valores óptimos para los parámetros a_i de la distribución son aquellos que maximizan la probabilidad de haber obtenido nuestros datos de la población¹⁶, i.e. que maximizan la función de verosimilitud. Esto es, escogeremos los parámetros a_i que hacen nuestros datos más verosímiles.

Veamos a continuación la aplicación del método de máxima verosimilitud para la obtención de estimadores de la media poblacional y de parámetros de ajuste.

3.4.1. Estimador de la media poblacional

Volvamos ahora desde este nuevo punto de vista a analizar los estimadores de parámetros de una población a través de una muestra. Si tomamos una muestra de n datos $\{x_i\}_{i=1}^n$ que sabemos que siguen una distribución gaussiana de varianza σ^2 , la probabilidad de obtener el dato i -ésimo es

$$p_i(\mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}\right], \quad (3.117)$$

Considerando la probabilidad de obtener todos los datos de la muestra construimos la función de verosimilitud, que dependerá en este caso únicamente de la media poblacional:

$$L(\mu) = \prod_{i=1}^n p_i = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}\right]. \quad (3.118)$$

Maximizar la probabilidad (verosimilitud) de nuestro conjunto de datos para obtener la media poblacional μ que hace más probable la obtención de los datos de la muestra es equivalente en este caso a minimizar el exponente

$$\chi^2(\mu) = +\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}, \quad (3.119)$$

lo que implica

$$\frac{d\chi^2}{d\mu} = \frac{1}{2} \sum_{i=1}^n \frac{d}{d\mu} \frac{(x_i - \mu)^2}{\sigma^2} = -\sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} = 0$$

¹⁶Ya que la búsqueda de máximos no está afectada por transformaciones monótonas, en ocasiones conviene convertir el producto anterior en una suma mediante el uso del logaritmo,

$$L' = \sum_{i=1}^n \ln p_i.$$

Consecuentemente el valor óptimo del parámetro (media poblacional) será

$$\sum_{i=1}^n (x_i - \mu') = 0 \Rightarrow \mu' = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.120)$$

que se corresponde con la media muestral. La media muestral se muestra así como la mejor asignación posible para la media poblacional, la que maximiza la verosimilitud de la muestra de datos. De esta manera, el método de máxima verosimilitud permite la obtención de los estimadores de parámetros de la población.

Como vimos en el primer capítulo y detallaremos en el siguiente, podemos ahora calcular la incertidumbre¹⁷ de la media μ mediante el procedimiento usual de propagación de incertidumbres estadísticas:

$$\sigma_\mu^2 = \sum_{i=1}^n \left(\frac{\partial \mu}{\partial x_i} \right)^2 \sigma_i^2. \quad (3.121)$$

Ahora bien, dado que de la Ec. (3.120) se sigue que

$$\frac{\partial \mu}{\partial x_i} = \frac{1}{n},$$

tenemos:

$$\begin{aligned} \sigma_\mu^2 &= \sum_{i=1}^n \frac{1}{n^2} \sigma_i^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = \frac{n\sigma^2}{n^2} \\ \sigma_\mu^2 &= \frac{\sigma^2}{n}, \end{aligned} \quad (3.122)$$

de acuerdo con lo que ya conocíamos de antemano por haberlo introducido con anterioridad en esta obra.

Hasta aquí hemos usado que todos los resultados experimentales están afectados por la misma incertidumbre, i.e. tienen el mismo peso estadístico en el conjunto de los datos, algo que no sucede en la mayoría de los supuestos experimentales. Consideremos ahora que se han realizado una serie de ensayos que arrojan un conjunto de resultados con diferentes incertidumbres estadísticas $\{x_i \pm \sigma_i\}_{i=1}^n$. La probabilidad de que cada uno de ellos provenga de una distribución gaussiana con media μ será ahora

$$p_i = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma_i^2} \right]. \quad (3.123)$$

y la probabilidad de haber obtenido la serie concreta de medidas experimentales (función de verosimilitud)

$$L(\mu) = \prod_{i=1}^n P_i = \left(\prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_i^2} \right]. \quad (3.124)$$

¹⁷Los motivos por los que nos referimos a la desviación típica quedarán claros en el siguiente capítulo.

Nuevamente el valor más verosímil de μ será aquél que maximiza $L(\mu)$, esto es aquél que minimiza

$$\chi^2(\mu) = \frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_i^2}, \quad (3.125)$$

lo que sucede para

$$0 = \frac{d\chi^2}{d\mu} = \frac{1}{2} \sum_{i=1}^n \frac{d}{d\mu} \frac{(x_i - \mu)^2}{\sigma_i^2} = -\frac{1}{2} \sum_{i=1}^n \frac{2(x_i - \mu)}{\sigma_i^2} = 0$$

De esta ecuación se sigue inmediatamente que

$$\begin{aligned} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} &= \sum_{i=1}^n \frac{\mu}{\sigma_i^2} \\ \mu &= \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \end{aligned} \quad (3.126)$$

de modo que cada dato x_i es pesado o ponderado con un factor inversamente proporcional a su varianza. Esta media de datos con diferentes incertidumbres estadísticas se conoce como media ponderada o pesada.

La incertidumbre de la media pesada μ la calcularemos nuevamente a través de la técnica de propagación de incertidumbres estadísticas. Así

$$\sigma_\mu^2 = \sum_{i=1}^n \left(\frac{\partial \mu}{\partial x_i} \right)^2 \sigma_i^2, \quad (3.127)$$

y usando que

$$\frac{\partial \mu}{\partial x_i} = \frac{1}{\sum_{j=1}^n 1/\sigma_j^2} \frac{1}{\sigma_i^2}, \quad (3.128)$$

obtenemos

$$\sigma_\mu^2 = \sum_{i=1}^n \frac{\left(\frac{1}{\sigma_i^2} \right)^2 \sigma_i^2}{\left(\sum_{j=1}^n \frac{1}{\sigma_j^2} \right)^2} = \frac{1}{\left(\sum_{j=1}^n \frac{1}{\sigma_j^2} \right)^2} \sum_{i=1}^n \frac{1}{\sigma_i^2}. \quad (3.129)$$

Por tanto, la incertidumbre estadística de la media ponderada será

$$\sigma_\mu^2 = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \quad (3.130)$$

Ejemplo 3.17

Un estudiante realiza una serie de medidas para obtener el voltaje en un experimento primero realiza 40 medidas obteniendo $\bar{x} = 1,022$ V con una varianza $s_1 = 0,01$ V. Después de revisar el experimento advierte que puede mejorar el montaje y reducir las incertidumbres de medida realizando 10 medidas más, en las que obtiene $\bar{x}_2 = 1,018$ V con $s_2 = 0,004$ V. Obtener el resultado combinado de ambas series de medidas.

Los valores de las incertidumbres de las medias muestrales de ambas series son

$$\begin{aligned}\sigma_{\bar{x}_1} &= \frac{s_1}{\sqrt{n_1}} = \frac{0,01}{\sqrt{40}} \text{ V} \\ \sigma_{\bar{x}_2} &= \frac{s_2}{\sqrt{n_2}} = \frac{0,004}{\sqrt{10}} \text{ V}\end{aligned}$$

por lo que podemos considerar que disponemos de una muestra de dos datos de la media muestral cada uno de ellos con una varianza diferente y resumen de una serie de medidas. La asignación óptima para la media poblacional será entonces la media ponderada de ambas observaciones:

$$\mu \simeq \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} = \frac{\frac{40(1,022)}{(0,01)^2} + \frac{10(1,018)}{(0,004)^2}}{\frac{40}{(0,01)^2} + \frac{10}{(0,004)^2}} = 1,0196 \text{ V.}$$

La incertidumbre de la media ponderada será

$$\sigma_\mu \simeq \left(\frac{40}{(0,01)^2} + \frac{10}{(0,004)^2} \right)^{1/2} = 0,0010 \text{ V.}$$

El resultado para el voltaje será

$$\mu = 1,0196 \pm 0,0010 \text{ V.}$$

Obsérvese que la precisión del resultado final es mayor que la de cada serie por separado, ya que

$$\sigma_{\bar{x}_1} = 0,0016 \text{ V} \quad \sigma_{\bar{x}_2} = 0,0013 \text{ V}$$

3.4.2. Ajuste por el método de mínimos cuadrados

En muchos casos de interés, existe una correlación entre dos magnitudes medibles x e y . Como hemos visto al introducir esta sección 3.4, la curva teórica que plasma esta relación, $y = f(x)$, no es sino una hipótesis estadística acerca de la dependencia de las medias de una magnitud y en las medias de la variable independiente x y dependerá en general de un conjunto de parámetros. A partir de este supuesto, y mediante el método de máxima verosimilitud, podemos obtener los parámetros óptimos de la función de ajuste, que serán aquellos que maximicen la probabilidad de haber obtenido la muestra experimental de datos. Como veremos a continuación, en el caso de que consideremos poblaciones madre normales (i.e. cuando podamos considerar que la probabilidad de obtener

un determinado valor de la variable aleatoria y se distribuye en torno a un valor medio de acuerdo con una distribución de tipo gaussiano), maximizar la verosimilitud de la muestra será equivalente a calcular la función que se encuentra a la mínima distancia vertical de los puntos de la muestra, de donde procede la denominación de ajuste por mínimos cuadrados.

Estudiemos mediante este enfoque diferentes casos de regresión de muestras de datos.

A) Regresión lineal

Consideremos un conjunto de datos de una variable aleatoria bidimensional $\{(x_i, y_i)\}_{i=1}^n$, y supongamos que podemos considerar que existe una relación lineal $y = a + bx$ entre ambas variables aleatorias. Nuestro interés estriba en calcular la recta de regresión o de mejor ajuste entre x e y mediante la determinación de los parámetros a y b .

Para ello, consideramos que la probabilidad de obtener un cierto valor y_i tiene una distribución normal de media \bar{y}_i ¹⁸

$$p_i = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[\frac{(y_i - \bar{y}_i)^2}{\sigma_i^2} \right] \right\}, \quad (3.131)$$

donde hemos tenido en cuenta que, en general, las incertidumbres asociadas a cada punto de ajuste $\sigma_i^2 = \sigma_{y_i}^2 = \sigma^2(y_i)$ son diferentes¹⁹. La regresión en la que se tiene en cuenta el peso estadístico de cada dato (dado por el inverso de su varianza $1/\sigma_i^2$) se conoce como regresión ponderada o pesada, como ya hemos dicho en secciones anteriores. Por otro lado, como vimos al estudiar el test χ^2 de un ajuste, es razonable suponer que las medias de las distribuciones de los diferentes puntos experimentales y_i se sitúan sobre la recta de regresión,

$$\bar{y}_i = a + bx_i, \quad (3.132)$$

razón por la cual podemos reescribir la probabilidad de obtención del dato i -ésimo como

$$p_i(a, b) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ \frac{1}{2} \left[\frac{(b_i - a - bx_i)^2}{\sigma_i^2} \right] \right\}. \quad (3.133)$$

La probabilidad conjunta de haber obtenido la serie de puntos que componen la muestra $\{(x_i, y_i)\}_{i=1}^n$ vendrá dada entonces por la función de verosimilitud

$$\begin{aligned} L(a, b) &= \prod_{i=1}^n p_i(a, b) \\ &= \left(\prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - a - bx_i)^2}{\sigma_i^2} \right\}. \end{aligned} \quad (3.134)$$

¹⁸No es cierto que la aproximación gaussiana sea siempre correcta, aunque en un gran número de casos prácticos sea muy adecuada.

¹⁹Observemos que las abscisas de los puntos experimentales son tratadas como si no fuesen variables aleatorias. Esto es así debido a que, en la práctica, las abscisas suelen corresponderse con variables controladas desde el exterior, por lo que, generalmente, podemos despreciar su incertidumbre frente a la de las ordenadas, $b^2 \sigma_{x_i}^2 \ll \sigma_{y_i}^2$. En un apartado posterior se explicará cómo proceder cuando no sea éste el caso.

La aplicación del método de máxima verosimilitud nos conduce a que los parámetros a y b que permiten reproducir las medias óptimas de las ordenadas experimentales, y_i , son aquellos que maximizan $L(a, b)$, lo que equivale a minimizar el exponente de la ecuación anterior. Por lo tanto, si definimos la siguiente función de los parámetros a y b

$$\chi^2(a, b) = \sum_{i=1}^n \frac{(y_i - a - bx_i)^2}{\sigma_i^2}, \quad (3.135)$$

que sigue una distribución χ^2 de Pearson, podemos obtener los parámetros a y b óptimos a partir del sistema de ecuaciones:

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= 0 = -2 \sum_{i=1}^n \frac{1}{\sigma_i^2} (y_i - a - bx_i) \\ \frac{\partial \chi^2}{\partial b} &= 0 = -2 \sum_{i=1}^n \frac{x_i}{\sigma_i^2} (y_i - a - bx_i), \end{aligned} \quad (3.136)$$

o, equivalentemente,

$$\begin{aligned} a \sum_{i=1}^n \frac{1}{\sigma_i^2} + b \sum_{i=1}^n \frac{x_i}{\sigma_i^2} &= \sum_{i=1}^n \frac{y_i}{\sigma_i^2} \\ a \sum_{i=1}^n \frac{x_i}{\sigma_i^2} + b \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} &= \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2}. \end{aligned} \quad (3.137)$$

La solución de este sistema de ecuaciones lineales proporciona los parámetros a y b que maximizan la probabilidad de obtención de la muestra de datos $\{(x_i, y_i)\}_{i=1}^n$ suponiendo que los valores de las y_i se distribuyan normalmente en torno a las medias determinadas por la recta de regresión. Resolviendo el sistema mediante la regla de Kramer obtenemos:

$$a = \frac{1}{\Delta} \begin{vmatrix} \sum_{i=1}^n \frac{y_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \end{vmatrix}, \quad b = \frac{1}{\Delta} \begin{vmatrix} \sum_{i=1}^n \frac{1}{\sigma_i^2} & \sum_{i=1}^n \frac{y_i}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} \end{vmatrix}, \quad (3.138)$$

siendo Δ el determinante de la matriz del sistema de ecuaciones en (3.137):

$$\Delta = \begin{vmatrix} \sum_{i=1}^n \frac{1}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \end{vmatrix}. \quad (3.139)$$

Podemos entonces obtener la forma explícita de los parámetros de mejor ajuste en la regresión lineal ponderada como

$$a = \frac{\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \sum_{i=1}^n \frac{b_i}{\sigma_i^2} - \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \sum_{i=1}^n \frac{x_i b_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2} \quad (3.140)$$

y

$$b = \frac{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} - \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \sum_{i=1}^n \frac{b_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2}. \quad (3.141)$$

Este método de ajuste lineal ponderado es más tedioso que el caso simplificado en el que la incertidumbre de todos los datos experimentales es la misma, $\sigma_i = \sigma$, que analizaremos posteriormente como caso de particular importancia. Sin embargo, resulta necesario si tenemos la certeza de la inhomogeneidad de las incertidumbres experimentales.

Para calcular las incertidumbres en a y en b seguiremos considerando que la incertidumbre en la variable independiente es despreciable frente a la de la dependiente, $b^2 \sigma_{x_i}^2 \ll \sigma_{y_i}^2$. Con esta aproximación tenemos, a partir de la fórmula de propagación de incertidumbres

$$\sigma_a^2 = \sum_{j=1}^n \sigma_j^2 \left(\frac{\partial a}{\partial y_j} \right)^2 \quad \sigma_b^2 = \sum_{j=1}^n \sigma_j^2 \left(\frac{\partial b}{\partial y_j} \right)^2. \quad (3.142)$$

Las derivadas parciales de los parámetros de ajuste son inmediatas a partir de sus expresiones en (3.140) y (3.141) respectivamente:

$$\begin{aligned} \frac{\partial a}{\partial y_j} &= \frac{1}{\Delta} \left[\frac{1}{\sigma_j^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \frac{x_j}{\sigma_j^2} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right] \\ \frac{\partial b}{\partial y_j} &= \frac{1}{\Delta} \left[\frac{x_j}{\sigma_j^2} \sum_{i=1}^n \frac{1}{\sigma_i^2} - \frac{1}{\sigma_j^2} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right]. \end{aligned} \quad (3.143)$$

Sustituyendo las expresiones de la ecuación anterior en (3.142) obtenemos

$$\begin{aligned} \sigma_a^2 &= \sum_{j=1}^n \frac{\sigma_j^2}{\Delta^2} \left(\frac{1}{\sigma_j^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \frac{x_j}{\sigma_j^2} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 = \\ &= \frac{1}{\Delta^2} \left[\sum_{j=1}^n \frac{1}{\sigma_j^2} \left(\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \right)^2 - 2 \sum_{j=1}^n \frac{x_j}{\sigma_j^2} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} + \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 \right] \\ &= \frac{1}{\Delta^2} \left[\sum_{j=1}^n \frac{1}{\sigma_j^2} \left(\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \right)^2 - 2 \left(\sum_{j=1}^n \frac{x_j}{\sigma_j^2} \right)^2 \sum_{i=1}^n \frac{x_i}{\sigma_i^2} + \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 \right] \\ &= \frac{1}{\Delta^2} \left[\sum_{j=1}^n \frac{1}{\sigma_j^2} \left(\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \right)^2 - \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 \right] = \\ &= \frac{1}{\Delta^2} \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} \left[\sum_{i=1}^n \frac{1}{\sigma_j^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 \right]. \end{aligned} \quad (3.144)$$

Teniendo en cuenta que el determinante del sistema de ecuaciones es el expresado en la Ec. (3.139), es decir

$$\Delta = \sum_{j=1}^n \frac{1}{\sigma_j^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2, \quad (3.145)$$

obtenemos para la incertidumbre del término independiente de la recta de mejor ajuste lineal ponderado:

$$\sigma_a^2 = \frac{1}{\Delta} \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2}. \quad (3.146)$$

De modo análogo es posible calcular la varianza de la pendiente de la recta de mejor ajuste, b , sustituyendo la derivada correspondiente en la Ec. (3.143) en (3.142):

$$\begin{aligned} \sigma_b^2 &= \sum_{j=1}^n \frac{\sigma_j^2}{\Delta^2} \left(\frac{x_j}{\sigma_j^2} \sum_{i=1}^n \frac{1}{\sigma_i^2} - \frac{1}{\sigma_j^2} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 = \frac{1}{\Delta^2} \left[\sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^2 \right. \\ &\quad \left. - 2 \sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{j=1}^n \frac{x_j}{\sigma_j^2} \sum_{i=1}^n \frac{x_i}{\sigma_i^2} + \sum_{j=1}^n \frac{1}{\sigma_j^2} \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 \right], \end{aligned} \quad (3.147)$$

de donde

$$\begin{aligned} \sigma_b^2 &= \frac{1}{\Delta^2} \left[\sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^2 - \sum_{j=1}^n \frac{1}{\sigma_j^2} \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 \right] = \\ &= \frac{1}{\Delta^2} \sum_{j=1}^n \frac{1}{\sigma_j^2} \left[\sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} \sum_{i=1}^n \frac{1}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 \right] = \\ &= \frac{1}{\Delta} \sum_{j=1}^n \frac{1}{\sigma_j^2}. \end{aligned} \quad (3.148)$$

Ejemplo 3.18

Un alumno estudia una fuente radiactiva y quiere investigar si la radiactividad registrada a una distancia r de la fuente decae con el inverso del cuadrado de la distancia, o, equivalentemente, muestra un comportamiento lineal en $1/r^2$. Para ello toma medidas con un contador Geiger durante un intervalo de tiempo fijo colocando la fuente entre 20 y 100 cm de distancia. Sus datos son

d_i (cm)	20	25	30	35	40	45	50	60	75	100
$x_i = d_i^{-2}$ (m^{-2})	25,00	16,00	11,11	8,16	6,25	4,94	4,00	2,78	1,78	1,00
c_i	901	652	443	339	283	281	240	220	180	154

Queremos ajustar

$$c_i \simeq a + bx_i = a + b/d_i^2.$$

Considerando que las posiciones d_i se miden con una incertidumbre pequeña (magnitud controlada desde el exterior) comparada con la incertidumbre del número de cuentas (magnitud medida), la contribución más importante a nuestras incertidumbres son las fluctuaciones estadísticas del número de cuentas, c_i , que siguen según sabemos una distribución de Poisson. De todos modos si la media de esta distribución es $\lambda \simeq 10$, podemos considerar que la distribución de Poisson es aproximadamente igual a una distribución de Gauss con $\sigma \simeq \sqrt{\lambda}$. En este caso se cumple esta condición dado lo elevado del número de cuentas registrado, por lo que podremos realizar un ajuste por mínimos cuadrados. Observemos que los valores observados tienen incertidumbres que difieren notablemente entre sí

$$\sigma_i \simeq \sqrt{c_i} \quad ; \quad \frac{1}{\sigma_i^2} = \frac{1}{c_i},$$

por lo que será necesario realizar un ajuste lineal ponderado. El determinante del sistema de la regresión lineal ponderada será en este caso

$$\begin{aligned} \Delta &= \sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 = \\ &= \sum_{i=1}^n \frac{1}{y_i} \sum_{i=1}^n \frac{x_i^2}{y_i} - \left(\sum_{i=1}^n \frac{x_i}{y_i} \right)^2 = \\ &= 0,03571,912 - (0,1868)^2 = 0,0334 \end{aligned}$$

y los parámetros de la recta de mejor ajuste

$$\begin{aligned} a &= \frac{1}{\Delta} \left[\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \sum_{i=1}^n \frac{y_i}{\sigma_i^2} - \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} \right] = \\ &= \frac{1}{\Delta} \left[\sum_{i=1}^n \frac{x_i^2}{y_i} \sum_{i=1}^n 1 - \sum_{i=1}^n \frac{x_i}{y_i} \sum_{i=1}^n x_i \right] \\ a &= \frac{1}{\Delta} [101,912 - 0,1868 \cdot 81,0] = 119,5 \end{aligned}$$

y

$$\begin{aligned} b &= \frac{1}{\Delta} \left[\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \sum_{i=1}^n \frac{b_i}{\sigma_i^2} \right] = \\ &= \frac{1}{\Delta} \left[\sum_{i=1}^n \frac{1}{y_i} \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{x_i}{y_i} \sum_{i=1}^n 1 \right] = \\ &= \frac{1}{\Delta} [0,0357 \cdot 81,0 - 0,1868 \cdot 10] = 30,7 \text{ m}^2. \end{aligned}$$

Finalmente, las incertidumbres de dichos parámetros son

$$\begin{aligned} \sigma_a^2 &= \frac{1}{\Delta} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} = \frac{1}{\Delta} \sum_{i=1}^n \frac{x_i^2}{y_i} = \frac{1,912}{0,0334} = 57,3 \\ \sigma_a &\simeq 7,6 \end{aligned}$$

y

$$\sigma_b^2 = \frac{1}{\Delta} \sum_{i=1}^n \frac{1}{\sigma_i^2} = \frac{1}{\Delta} \sum_{i=1}^n \frac{1}{y_i} = \frac{0,0357}{0,0334} = 1,07 \text{ m}^2$$

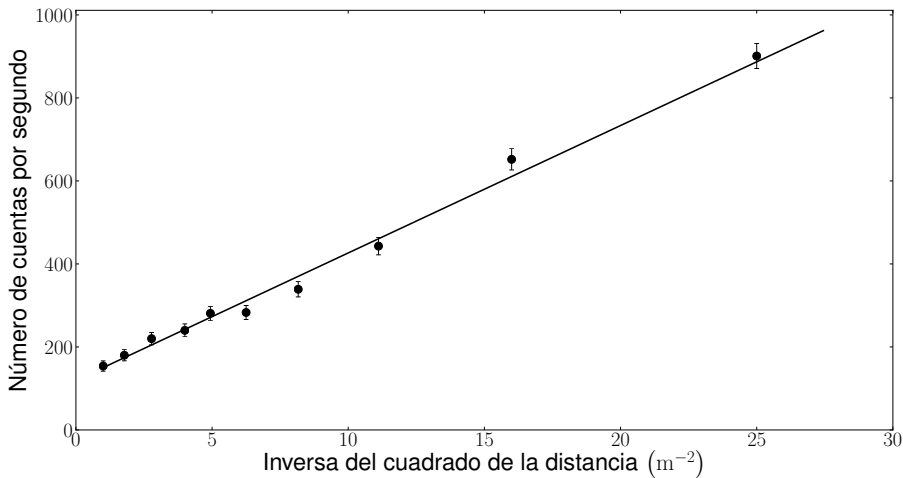
$$\sigma_b \simeq 1,1 \text{ m}^2,$$

con lo cual, podemos afirmar que $c = a + b/r^2$ con

$$a = 119,5 \pm 7,6$$

$$b = 30,7 \pm 1,1 \text{ m}^2.$$

No obstante, en el presente caso si calculamos el χ^2 de este ajuste obtendremos una probabilidad del 20%, lo que debe considerarse como un indicio de que el ajuste no es determinante.



En el caso de que todos los datos experimentales estén afectados por la misma incertidumbre, $\sigma_i = \sigma$, la regresión, denominada entonces regresión lineal simple, se vuelve considerablemente más sencilla. Los diferentes parámetros de ajuste pueden obtenerse a partir de los correspondientes resultados del ajuste lineal ponderado sin más que considerar que $\sigma_i = \sigma$. El sistema de ecuaciones al que conduce la maximización de la función de verosimilitud puede obtenerse a partir del sistema en la Ec. (3.139) resulta:

$$\sum_{i=1}^n y_i = a \sum_{i=1}^n 1 + b \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2, \quad (3.149)$$

recuperando el sistema que introdujimos por métodos más sencillos y menos generales en el capítulo 1 de la presente obra. La solución de este sistema es, como ya vimos entonces,

$$\begin{aligned}
 a &= \frac{1}{\Delta} \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^2 \end{vmatrix} = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \\
 b &= \frac{1}{\Delta} \begin{vmatrix} n & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i y_i \end{vmatrix} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad (3.150)
 \end{aligned}$$

siendo Δ el determinante del sistema

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix} = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2. \quad (3.151)$$

Finalmente, operando de manera similar a lo hecho en el caso del ajuste lineal ponderado, y suponiendo nuevamente que $b\sigma_{x_i} \ll \sigma_{y_i}$, podemos obtener de manera inmediata las incertidumbres de los parámetros de mejor ajuste:

$$\sigma_a^2 = \frac{\sigma^2}{\Delta} \sum_{i=1}^n x_i^2; \quad \sigma_b^2 = n \frac{\sigma^2}{\Delta}. \quad (3.152)$$

En general, si somos capaces de estimar σ a partir de las características de nuestras medidas, podremos introducirla en esta expresión. Otra forma de operar consiste en asumir que la media para cada variable y_i se estima mediante el valor arrojado por la recta de mejor ajuste obtenida mediante la técnica de regresión, $a + bx_i$, y obtener la incertidumbre de y a partir de la varianza muestral. Si tenemos en cuenta que, como hemos usado los datos para calcular dos parámetros (a y b), tenemos ahora únicamente $n - 2$ grados de libertad,

$$\sigma^2 \simeq s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - a - bx_i)^2. \quad (3.153)$$

Consecuentemente, podremos escribir las incertidumbres de los parámetros en (3.152) como

$$\begin{aligned}
 \sigma_a^2 &\simeq \frac{1}{\Delta} \sum_{i=1}^n x_i^2 \left[\frac{1}{n-2} \sum_{i=1}^n (y_i - a - bx_i)^2 \right] \\
 \sigma_b^2 &\simeq \frac{n}{\Delta} \left[\frac{1}{n-2} \sum_{i=1}^n (y_i - a - bx_i)^2 \right]. \quad (3.154)
 \end{aligned}$$

Tengamos en cuenta que σ_a y σ_b en estas expresiones miden las incertidumbres provocadas por la dispersión de los datos respecto a la recta de ajuste. Otros errores e incertidumbres en las medidas no estarán reflejados en σ_a ni en σ_b

y pueden constituir contribuciones importantes. Imaginemos, por ejemplo, un voltímetro mal calibrado para el que la magnitud de salida será $y_i = V_i$, siendo $V_i = 0,5 \bar{V}_i$ donde \bar{V}_i es el valor real de la magnitud. Este error de calibración provocará que la pendiente de un ajuste x_i versus V_i se desvíe del valor real un factor 2, aunque la dispersión de los datos sea mínima.

Ejemplo 3.19

Un estudiante mide la diferencia de potencial entre un punto de un cable de Ni-Ag y el borne de una batería con los siguientes resultados:

$x_i(\text{cm})$	10,0	20,0	30,0	40,0	50,0	60,0	70,0	80,0	90,0
$V_i(\text{V})$	0,37	0,58	0,83	1,15	1,36	1,62	1,90	2,18	2,45

Teniendo en cuenta que dispone de un voltímetro analógico, considera que todos los valores de V_i están afectados por la misma incertidumbre $\sigma = 0,05$ V y que los x_i se miden con una incertidumbre de 1 mm que despreciaremos. Se pide un ajuste lineal entre x_i e V_i .

Los parámetros de la recta de mejor ajuste se obtienen como solución del sistema de ecuaciones lineales (3.149), cuyos coeficientes valen en este supuesto:

$$\begin{aligned} \sum_{i=1}^9 x_i &= 450,00 & \sum_{i=1}^9 x_i^2 &= 28500 \\ \sum_{i=1}^9 x_i v_i &= 779,30 & \sum_{i=1}^9 v_i &= 12,44 \end{aligned}$$

Por su parte, el determinante de la matriz del sistema será

$$\Delta = N \sum_{i=1}^9 x_i^2 - \left(\sum_{i=1}^9 x_i \right)^2 = 9 \times 28500 - (450)^2 = 54000$$

lo que nos conduce a los siguientes valores de los parámetros de la recta de mejor ajuste:

$$\begin{aligned} a &= \frac{1}{\Delta} \left(\sum_{i=1}^9 x_i^2 \sum_{i=1}^9 v_i - \sum_{i=1}^9 x_i \sum_{i=1}^9 x_i v_i \right) = \frac{(28500 \cdot 12,44 - 450,0 \cdot 779,3)}{54000} = \\ &= 0,0714 \text{ V} \end{aligned}$$

y

$$\begin{aligned} b &= \frac{1}{\Delta} \left(N \sum_{i=1}^9 x_i v_i - \sum_{i=1}^9 x_i \sum_{i=1}^9 v_i \right) = \frac{9 \cdot 779,3 - 450,0 \cdot 12,44}{54000} = \\ &= 0,0262 \frac{\text{V}}{\text{cm}}. \end{aligned}$$

Por otro lado, las incertidumbres de estos parámetros serán

$$\sigma_a^2 = \sigma_v^2 \frac{1}{\Delta} \sum_{i=1}^9 x_i^2 = (0,05)^2 \frac{28500}{54000} = 0,001319 \text{ V}^2$$

$$\sigma_a = 0,036 \text{ V}$$

$$\sigma_b^2 = N\sigma_v^2 \frac{1}{\Delta} = 9(0,05)^2 \frac{1}{54000} = 0,417 \cdot 10^{-6} \left(\frac{\text{V}}{\text{cm}} \right)^2$$

$$\sigma_b = 0,00065 \frac{\text{V}}{\text{cm}}$$

de donde obtenemos finalmente que de los datos experimentales se deduce que $V = a + bx$, siendo

$$a = 0,071 \pm 0,036 \text{ V}$$

$$b = 0,02620 \pm 0,00030 \text{ V cm}^{-1}$$

Si en lugar de considerar σ_y estimado independientemente del ajuste, considerásemos una estimación directa a partir de la muestra y del propio ajuste, obtendríamos

$$(\sigma'_y)^2 \simeq \frac{a}{N-2} \sum_{i=1}^9 (y_i - a - bx_i)^2$$

por lo que teniendo en cuenta la tabla de valores

x_i (cm)	10,0	20,0	30,0	40,0	50,0	60,0	70,0	80,0	90,0
y_i (V)	0,37	0,58	0,83	1,15	1,36	1,62	1,90	2,18	2,45
$a + bx_i$ (V)	0,33	0,60	0,86	1,12	1,38	1,64	1,91	2,17	2,43

obtendríamos

$$\sum_{i=1}^9 (y_i - a - bx_i)^2 = 4,88 \cdot 10^{-3}$$

$$(\sigma'_y)^2 = \frac{1}{9-2} 4,88 \cdot 10^{-4}$$

$$\sigma'_y \simeq 0,03 \text{ V.}$$

Con este valor de σ'_y se pueden volver a calcular σ'_a y σ'_b :

$$\sigma_a'^2 = (\sigma'_y)^2 \frac{1}{\Delta} \sum_{i=1}^9 x_i^2 = 3,6 \cdot 10^{-4}$$

$$\sigma_b' = (\sigma'_y)^2 \frac{N}{\Delta} = 1,16 \cdot 10^{-7}$$

$$\sigma_a' = 0,019 \text{ V} \quad \sigma_b' = 0,00034 \text{ V cm}^{-1}.$$

Otro aspecto que podemos considerar del ajuste es el test de χ^2 asociado a éste. Recordemos que para realizar este test de la hipótesis estadística en la que consiste el ajuste $y = a + bx$ debemos construir el estimador

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - a - bx_i)^2}{\sigma_i^2}$$

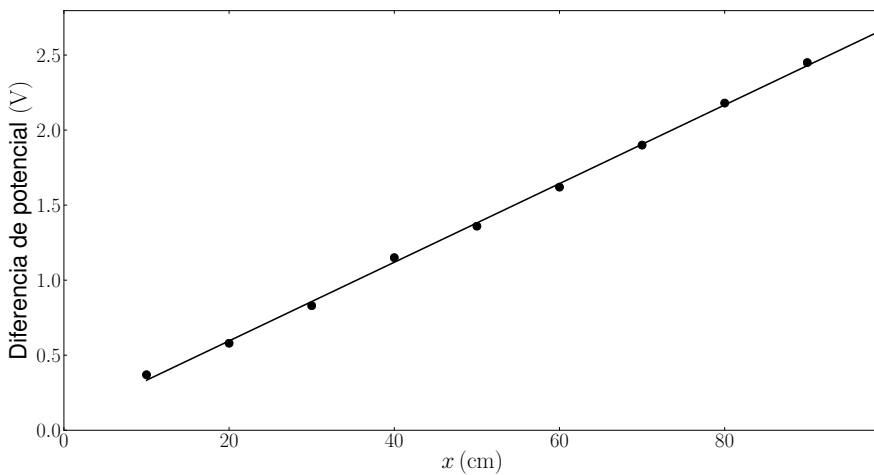
que, para variables y_i distribuidas normalmente, sigue una distribución χ^2 de Pearson. En nuestro caso, este estimador toma el valor

$$\begin{aligned}\chi_{9-2}^2 &= \sum_{i=1}^9 \frac{(y_i - a - bx_i)^2}{\sigma_i^2} = \frac{1}{\sigma^2} \sum_{i=1}^9 (y_i - a - bx_i)^2 = \\ &= 1,95\end{aligned}$$

Aceptaremos la hipótesis $y = a + bx$ implícita en el ajuste si, a un determinado nivel de confianza, el percentil correspondiente de la distribución χ^2 de Pearson es mayor que este valor. Alternativamente, si tenemos en cuenta que la probabilidad de que la variable con $7=9-2$ grados de libertad tome un valor mayor que 1,95 es

$$P(\chi_7^2 \geq 1,95) \simeq 96 \%,$$

asignaremos una probabilidad al ajuste del 96 %. Dado el valor tan pequeño de χ_7^2 podemos concluir, con un alto nivel de probabilidad, que estamos ante un buen ajuste de los datos experimentales.



B) Linealización. Ajuste a una exponencial

Como ya vimos en el ejemplo 3.18, mediante las transformaciones oportunas de los datos, en algunos casos es posible realizar ajustes a funciones no lineales mediante métodos de linealización. Supongamos que las variables x e y cuya correlación estadística estudiamos se encuentran teóricamente relacionadas por una relación $y = f(x)$. Para poder realizar un ajuste lineal de la variable y en términos de la variable x , debemos conseguir una dependencia lineal de alguna función de y de otra de x ,

$$g(y) = a + bh(x) \quad (3.155)$$

Este proceso se denomina linealización y es frecuente su utilización para aprovechar la simplicidad y precisión de las regresiones lineales mediante la técnica de regresión de mínimos cuadrados para ajustes entre variables que siguen relaciones en principio no lineales.

Supongamos, por ejemplo, que consideramos la función exponencial

$$y = ae^{-bx}, \quad (3.156)$$

siendo a y b parámetros desconocidos que han de ser ajustados a partir de los resultados experimentales. En el caso presente, la linealización consiste en tomar logaritmos de la ecuación anterior, de modo que

$$\ln y = \ln a - bx. \quad (3.157)$$

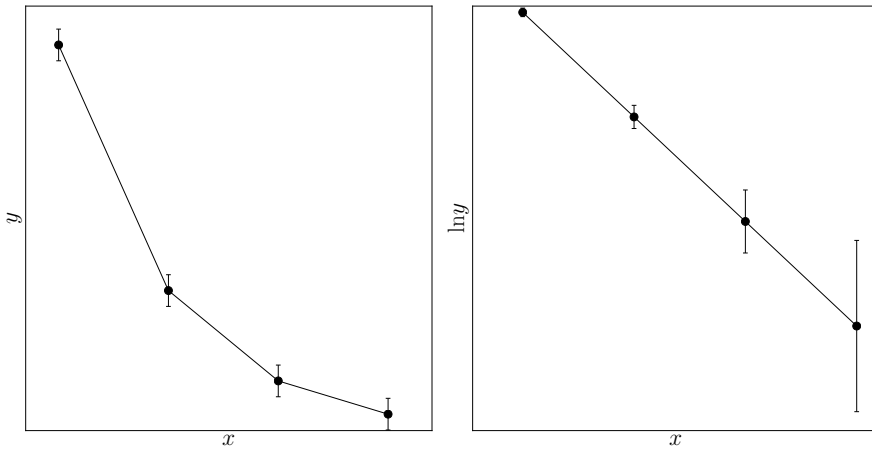
Para considerar el método de mínimos cuadrados como apropiado, debemos asumir que la variable $\ln y$ sigue una distribución aproximadamente gaussiana. Suponiendo que esto se satisface, para obtener los parámetros de la recta de mejor ajuste en la expresión anterior deberemos minimizar

$$\chi^2(a, b) = \sum_{i=1}^n \frac{(\ln y_i - \ln a + bx_i)^2}{(\sigma'_i)^2} \quad (3.158)$$

donde σ'_i ya no es la incertidumbre original que teníamos en los datos y_i sino la incertidumbre de la nueva variable estadística $\ln y$:

$$\sigma'_i = \frac{d \ln y_i}{dy_i} \sigma_i = \frac{\sigma_i}{y_i}.$$

Vemos entonces que si los datos de y_i tenían incertidumbres homogéneas, $\sigma_i = \sigma$, los datos de $\ln y_i$ no van a tener incertidumbres de tipo homogéneo sino que serán tanto mayores cuanto menor sea y_i . Ignorar este efecto nos llevaría a sobrevalorar los puntos donde y_i es pequeño en el ajuste.



Ejemplo 3.20

Consideremos un ejemplo donde un estudiante mide la respuesta de un circuito RC en función del tiempo, de modo que se espera

$$V(t) = V_0 e^{-\frac{t}{RC}}.$$

Se pide que verifique esta ley y obtenga el valor de RC a partir de los datos experimentales obtenidos.

$x_i = t_i$ (s)	1,0	2,0	4,0	6,0	8,0	10,0	20,0	40,0
V_i (V)	9,1	8,2	6,7	5,2	4,5	3,6	1,4	0,1
$y_i = \ln(V_i/1 \text{ V})$	2,21	2,10	1,90	1,65	1,50	1,3	0,3	-2

Consideraremos que $\sigma_i \simeq \sigma = 0,1 \text{ V}$ para la medida del voltaje. Por tanto, las incertidumbres de los diferentes datos experimentales serán:

$$\sigma'_i = \frac{\sigma}{V_i}$$

Utilizando las relaciones convencionales del ajuste lineal ponderado para los datos anteriores, tenemos:

$$\begin{aligned} \Delta &= \sum_{i=1}^n \frac{1}{\sigma'^2_i} \sum_{i=1}^n \frac{x_i^2}{\sigma'^2_i} - \left(\sum_{i=1}^n \frac{x_i}{\sigma'^2_i} \right)^2 = \\ &= \sum_{i=1}^n \frac{V_i^2}{\sigma^2} \sum_{i=1}^n \frac{V_i^2 x_i^2}{\sigma^2} - \left(\sum_{i=1}^n \frac{x_i V_i^2}{\sigma^2} \right)^2 \\ \sum_{i=1}^8 \frac{V_i^2}{\sigma^2} &= 2,572 \cdot 10^4 & \sum_{i=1}^8 \frac{x_i^2 V_i^2}{\sigma^2} &= 5,42 \cdot 10^5 \text{ s}^2 \\ \sum_{i=1}^8 \frac{x_i V_i^2}{\sigma^2} &= 8,9 \cdot 10^4 \text{ s}^2 & \Delta &= 6,026 \cdot 10^9 \text{ s}^2 \end{aligned}$$

$$a = \frac{1}{\Delta} \left[\sum_{i=1}^n \frac{x_i^2 V_i^2}{\sigma^2} \sum_{i=1}^n \frac{y_i V_i^2}{\sigma^2} - \sum_{i=1}^n \frac{x_i V_i^2}{\sigma^2} \sum_{i=1}^n \frac{x_i y_i V_i^2}{\sigma^2} \right]$$

$$b = \frac{1}{\Delta} \left[\sum_{i=1}^n \frac{y_i^2}{\sigma^2} \sum_{i=1}^n \frac{x_i V_i^2 y_i}{\sigma^2} - \sum_{i=1}^n \frac{x_i V_i^2 y_i}{\sigma^2} - \sum_{i=1}^n \frac{x_i V_i^2}{\sigma^2} \sum_{i=1}^n \frac{y_i V_i^2}{\sigma^2} \right]$$

$$\left. \begin{array}{l} a = 2,31 \quad b = -0,103 \text{ s}^{-1} \\ \sigma_a = 0,01 \quad \sigma_b = 0,002 \text{ s}^{-1} \end{array} \right\} \Rightarrow \chi^2 = 7,54.$$

Como vemos en la Fig. 3.8, el valor correcto del parámetro de ajuste está dentro de las barras de incertidumbre. Si realizásemos este ajuste sin tener en cuenta la inhomogeneidad de las incertidumbres de los puntos experimentales, entonces tendríamos

$$\Delta = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2$$

$$a = \frac{a}{\Delta} \left[\sum_{i=1}^n V_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i \right] \quad \begin{array}{l} a = 2,49 \\ b = -0,131 \text{ s}^{-1} \\ \sigma_a = 0,05 \\ \sigma_b = 0,003 \text{ s}^{-1} \end{array}$$

$$b = \frac{1}{\Delta} \left[n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n b_i \right]$$

$$\chi^2 = 370,8.$$

Ambos resultados se comparan en la figura 3.8.

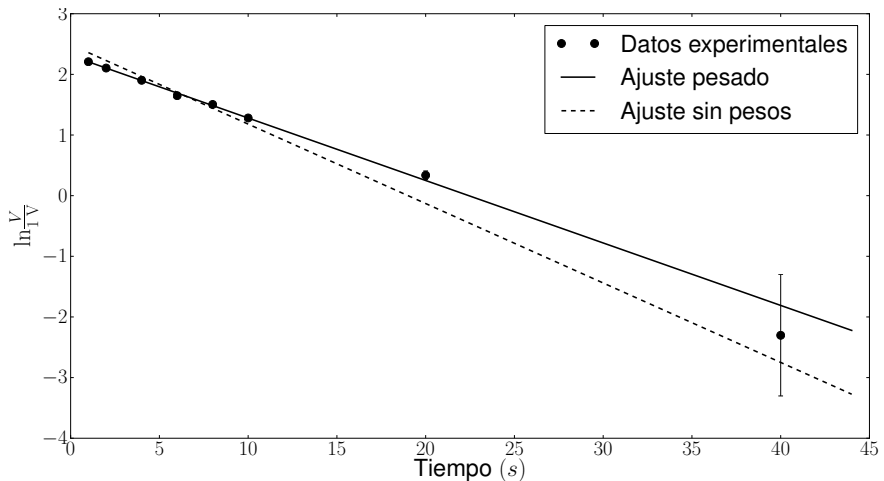


Figura 3.8: Comparación entre los ajustes pesados y sin pesos del circuito RC. Puede observarse fácilmente que el pesado es mejor al no sobrevalorar el dato a tiempos largos, que tiene gran incertidumbre.

C) Ajustes lineales generalizados y ajustes no lineales

En ocasiones, la relación no lineal que existe entre las variables aleatorias analizadas es de un tipo que no permite la linealización, por lo que ha de realizarse un ajuste directamente no lineal. En primer lugar, supongamos que tenemos que ajustar un conjunto de datos para la x e y a la supuesta relación de tipo polinómico

$$y = a + bx + cx^2. \quad (3.159)$$

De nuevo, siempre que las variables sigan distribuciones normales, la maximización de la función de verosimilitud equivale a la minimización de

$$\chi^2 = \sum_{i=1}^n \left(\frac{y_i - \bar{y}_i}{\sigma_i} \right)^2 = \sum_{i=1}^n \left(\frac{y_i - a - bx_i - cx_i^2}{\sigma_i} \right)^2, \quad (3.160)$$

donde, una vez más, consideramos que la función de ajuste pasa por las medias de los valores de la variable dependiente, $\bar{y}_i = a + bx_i + cx_i^2$. El ajuste más verosímil corresponde al valor de χ^2 mínimo

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= 0 \Rightarrow \sum_{i=1}^n \frac{(y_i - a - bx_i - cx_i^2)}{\sigma_i^2} = 0 \\ \frac{\partial \chi^2}{\partial b} &= 0 \Rightarrow \sum_{i=1}^n \frac{(y_i - a - bx_i - cx_i^2)}{\sigma_i^2} x_i = 0 \\ \frac{\partial \chi^2}{\partial c} &= 0 \Rightarrow \sum_{i=1}^n \frac{(y_i - a - bx_i - cx_i^2)}{\sigma_i^2} x_i^2 = 0, \end{aligned} \quad (3.161)$$

que conducen al sistema de ecuaciones lineales:

$$\begin{aligned} \sum_{i=1}^n \frac{y_i}{\sigma_i^2} &= a \sum_{i=1}^n \frac{1}{\sigma_i^2} + b \sum_{i=1}^n \frac{x_i}{\sigma_i^2} + c \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} &= a \sum_{i=1}^n \frac{x_i}{\sigma_i^2} + b \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} + c \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i^2 y_i}{\sigma_i^2} &= a \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} + b \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} + c \sum_{i=1}^n \frac{x_i^4}{\sigma_i^2}, \end{aligned} \quad (3.162)$$

cuyas soluciones son los parámetros de mejor ajuste:

$$a = \frac{1}{\Delta} \begin{vmatrix} \sum_{i=1}^n \frac{y_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i^2 y_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^4}{\sigma_i^2} \end{vmatrix} \quad (3.163)$$

$$b = \frac{1}{\Delta} \begin{vmatrix} \sum_{i=1}^n \frac{1}{\sigma_i^2} & \sum_{i=1}^n \frac{y_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2 y_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^4}{\sigma_i^2} \end{vmatrix} \quad (3.164)$$

$$c = \frac{1}{\Delta} \begin{vmatrix} \sum_{i=1}^n \frac{1}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{y_i}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2 y_i}{\sigma_i^2} \end{vmatrix} \quad (3.165)$$

$$\Delta = \begin{vmatrix} \sum_{i=1}^n \frac{1}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^4}{\sigma_i^2} \end{vmatrix}. \quad (3.166)$$

Hay que recordar siempre que si queremos ajustar un polinomio de grado 2 tendremos que tener más de 3 puntos como mínimo. En general, si queremos ajustar r parámetros debemos disponer de un número de puntos experimentales mayor o igual que $r + 1$.

Vamos a generalizar esta idea de ajustes a funciones arbitrarias y obtener las incertidumbres de los parámetros de ajuste. Supongamos que disponemos de una colección de n datos experimentales de las variables aleatorias x e y , $\{(x_i, y_i)\}_{i=1}^n$ y que la relación teórica entre dichas variables es una función general de m parámetros ($m < n$)

$$y = f(x; a_1, a_2, \dots, a_m). \quad (3.167)$$

Supondremos como de costumbre que los datos x_i son muy precisos en comparación con los y_i , los cuales son la principal fuente de incertidumbre estadística en el ajuste. Suponiendo que los datos y_i siguen una distribución gaussiana, la maximización de la verosimilitud de la muestra de datos se alcanza para los parámetros de la función de ajuste que minimizan la función χ^2 asociada

$$\chi^2 = \sum_{i=1}^n \frac{[y_i - f(x_i; a_1, a_2, \dots, a_m)]^2}{\sigma_i^2}. \quad (3.168)$$

En el caso de una función genérica $f(x; a_1 \dots a_m)$, no es posible minimizar esta función analíticamente, pero con frecuencia el problema se puede afrontar numéricamente, evaluando la función $\chi^2 = \chi^2(a_1, \dots, a_m)$ y si es posible sus derivadas parciales, y usando algoritmos para buscar sus mínimos locales.

No obstante, el problema admite una solución analítica si la función f tiene una expresión lineal de la forma

$$f(x; a_1, a_2, \dots, a_m) = \sum_{i=1}^m a_i f_i(x) \quad (3.169)$$

donde las $f_i(x)$ pueden ser cualquier función (exponencial, logaritmo, trigonométrica, etc), pero los parámetros a_k sólo deben aparecer multiplicando a estas funciones (y no en sus argumentos). Se habla en este caso de *ajustes lineales generalizados*. La función χ^2 a minimizar es

$$\chi^2(a_1, a_2, \dots, a_m) = \sum_{i=1}^n \frac{[y_i - \sum_{j=1}^m a_j f_j(x_i)]^2}{\sigma_i^2} \quad (3.170)$$

y las condiciones de mínimo para la función χ^2 serán

$$\begin{aligned} \frac{\partial \chi^2}{\partial a_1} &= -2 \sum_{i=1}^n \frac{[y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \dots - a_n f_n(x_i)]}{\sigma_i^2} f_1(x_i) = 0 \\ \frac{\partial \chi^2}{\partial a_2} &= -2 \sum_{i=1}^n \frac{[y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \dots - a_n f_n(x_i)]}{\sigma_i^2} f_2(x_i) = 0 \\ &\vdots \\ \frac{\partial \chi^2}{\partial a_n} &= -2 \sum_{i=1}^n \frac{[y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \dots - a_n f_n(x_i)]}{\sigma_i^2} f_n(x_i) = 0. \end{aligned} \tag{3.171}$$

Equivalentemente, el mínimo de la función χ^2 se alcanza para los parámetros $a_k; k = 1, \dots, m$, que verifican el sistema de n ecuaciones lineales y m incógnitas:

$$\sum_{i=1}^n \frac{[y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \dots - a_n f_n(x_i)]}{\sigma_i^2} f_k(x_i) = 0 \quad k = 1, 2, \dots, n, \tag{3.172}$$

que en su forma desarrollada puede escribirse como

$$\left\{ \begin{aligned} \sum_{i=1}^n y_i \frac{f_1(x_i)}{\sigma_i^2} &= a_1 \sum_{i=1}^n \frac{f_1^2(x_i)}{\sigma_i^2} + \dots + a_n \sum_{i=1}^n \frac{f_1(x_i) f_n(x_i)}{\sigma_i^2} \\ \sum_{i=1}^n y_i \frac{f_2(x_i)}{\sigma_i^2} &= a_1 \sum_{i=1}^n \frac{f_1(x_i) f_2(x_i)}{\sigma_i^2} \dots + a_n \sum_{i=1}^n \frac{f_2(x_i) f_n(x_i)}{\sigma_i^2} \\ &\vdots \\ \sum_{i=1}^n y_i \frac{f_n(x_i)}{\sigma_i^2} &= a_1 \sum_{i=1}^n \frac{f_1(x_i) f_n(x_i)}{\sigma_i^2} + \dots + a_n \sum_{i=1}^n \frac{f_n^2(x_i)}{\sigma_i^2}. \end{aligned} \right. \tag{3.173}$$

$\underbrace{\hspace{10em}}_Y \qquad \underbrace{\hspace{10em}}_{HA}$

La expresión matricial del sistema anterior es $Y = HA$, donde

$$\begin{aligned} H &= \begin{pmatrix} \sum_{i=1}^n \frac{f_1^2(x_i)}{\sigma_i^2} & \dots & \sum_{i=1}^n \frac{f_1(x_i) f_n(x_i)}{\sigma_i^2} \\ \vdots & & \vdots \\ \sum_{i=1}^n \frac{f_1(x_i) f_n(x_i)}{\sigma_i^2} & \dots & \sum_{i=1}^n \frac{f_n^2(x_i)}{\sigma_i^2} \end{pmatrix} \\ Y &\equiv \begin{pmatrix} \sum_{i=1}^n \frac{y_i f_1(x_i)}{\sigma_i^2} \\ \vdots \\ \sum_{i=1}^n \frac{y_i f_n(x_i)}{\sigma_i^2} \end{pmatrix} \quad A \equiv \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \end{aligned} \tag{3.174}$$

o lo que es lo mismo

$$\left\{ \begin{aligned} Y_k &= \sum_{i=1}^n \frac{y_i f_k(x_i)}{\sigma_i^2} \\ H_{kj} &= \sum_{i=1}^n \frac{f_k(x_i) f_j(x_i)}{\sigma_i^2} \\ A_j &= a_j \end{aligned} \right\} \quad Y_k = \sum_{j=1}^m H_{kj} a_j. \tag{3.175}$$

El cálculo de los parámetros de mejor ajuste implica pues la solución del problema matricial $Y = HA$ que, obviamente, viene dada por $A = H^{-1}Y$, aunque a efectos prácticos la inversión explícita de la matriz H no es, en general, un buen método de solución, porque da lugar a pérdida de precisión e inestabilidades numéricas. Es preferible utilizar otros métodos, como la descomposición en valores singulares, que se hallan implementados en numerosos paquetes de *software* (e.g. LAPACK, Scipy, R, Matlab...).

Consideremos ahora las incertidumbres de los parámetros a_1, a_2, \dots, a_m . Si obtuviéramos una serie de datos, los y_i variarían dando lugar a diferentes valores de a_1, a_2, \dots, a_m . Supongamos que los valores medios de los parámetros son $\alpha_1, \alpha_2, \dots, \alpha_m$. Pretendemos calcular la covarianza de los parámetros a_k y a_q , que corresponderá como sabemos a

$$\text{cov}(a_k, a_q) = E \{ (a_k - \alpha_k)(a_q - \alpha_q) \}, \quad (3.176)$$

con E representando, como de costumbre, el valor esperado. Tomaremos la aproximación de primer orden al valor esperado de a_k , α_k , que viene dada en términos de la matriz del sistema de ajuste H por

$$\alpha_k = \sum_{j=1}^m H_{kj}^{-1} \bar{Y}_j, \quad (3.177)$$

siendo \bar{Y}_j el valor medio que obtendríamos para esta componente del vector columna Y si repitiéramos muchas veces la toma de datos. Así pues,

$$\begin{aligned} a_k - \alpha_k &= \sum_{j=1}^m (H^{-1})_{kj} (Y_j - \bar{Y}_j) = \\ &= \sum_{j=1}^m (H^{-1})_{kj} \sum_{i=1}^n \frac{f_j(x_i)}{\sigma_i^2} (y_i - \bar{y}_i) \\ &= \sum_{j=1}^m \sum_{i=1}^n (H^{-1})_{kj} \frac{f_j(x_i)}{\sigma_i^2} (y_i - \bar{y}_i). \end{aligned} \quad (3.178)$$

Análogamente,

$$a_q - \alpha_q = \sum_{p=1}^m \sum_{l=1}^n (H^{-1})_{qp} \frac{f_p(x_l)}{\sigma_l^2} (y_l - \bar{y}_l). \quad (3.179)$$

Consecuentemente, la covarianza que buscamos entre los parámetros la podremos escribir en términos de la matriz H de la regresión como:

$$\begin{aligned} E [(a_k - \alpha_k)(a_q - \alpha_q)] &= \\ &= E \left\{ \sum_{p=1}^m \sum_{l=1}^n \sum_{j=1}^m \sum_{i=1}^n (H^{-1})_{qp} (H^{-1})_{kj} \frac{f_j(x_i)}{\sigma_i^2} \frac{f_p(x_l)}{\sigma_l^2} (y_i - \bar{y}_i)(y_l - \bar{y}_l) \right\} \\ &= \sum_{p=1}^m \sum_{l=1}^n \sum_{j=1}^m \sum_{i=1}^n (H^{-1})_{qp} (H^{-1})_{kj} \frac{f_j(x_i)}{\sigma_i^2} \frac{f_p(x_l)}{\sigma_l^2} E \{ (y_i - \bar{y}_i)(y_l - \bar{y}_l) \} \end{aligned} \quad (3.180)$$

Considerando las medidas y_i estadísticamente independientes, entonces

$$E \{(y_i - \bar{y}_i)(y_l - \bar{y}_l)\} = \sigma_i^2 \delta_{il} \begin{cases} 0 & i \neq l \\ \sigma_i^2 & i = l \end{cases} \quad (3.181)$$

por lo cual la Ec. (3.180) se puede reescribir de la forma:

$$\begin{aligned} E \{(a_k - \alpha_k)(a_q - \alpha_q)\} &= \\ &= \sum_{p=1}^m \sum_{l=1}^n \sum_{j=1}^m \sum_{i=1}^n (H^{-1})_{kj} \frac{f_j(x_i)}{\sigma_i^2} \frac{f_p(x_l)}{\sigma_l^2} \sigma_i^2 \delta_{il} \\ &= \sum_{p=1}^m \sum_{j=1}^m (H^{-1})_{qp} (H^{-1})_{kj} \underbrace{\sum_{i=1}^n \frac{f_j(x_i) f_p(x_i)}{\sigma_i^2}}_{H_{jp}} \end{aligned} \quad (3.182)$$

con lo cual,

$$E \{(a_k - \alpha_k)(a_q - \alpha_q)\} = \sum_{p=1}^m \sum_{j=1}^m (H^{-1})_{qp} (H^{-1})_{kj} H_{jp} \quad (3.183)$$

Ahora bien, teniendo en cuenta que $\sum_{j=1}^m (H^{-1})_{kj} H_{jp} = \delta_{kp}$ tendremos:

$$E \{(a_k - \alpha_k)(a_q - \alpha_q)\} = \sum_{p=1}^m (H^{-1})_{qp} \delta_{kp} = (H^{-1})_{qk}, \quad (3.184)$$

con lo que (3.183) se convierte en:

$$\left. \begin{aligned} \sigma^2(a_k) &= (H^{-1})_{kk} \\ \text{cov}(a_k, a_q) &= (H^{-1})_{kq}. \end{aligned} \right\} \quad (3.185)$$

Ejemplo 3.21

Consideremos un ajuste lineal convencional $y = a + bx$. Como ha quedado dicho, la matriz del sistema es

$$H = \begin{pmatrix} \sum_{i=1}^n \frac{1}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \end{pmatrix}$$

por lo que su inversa será:

$$H^{-1} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2} \begin{pmatrix} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{-x_i}{\sigma_i^2} \\ \sum_{i=1}^n \frac{-x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{1}{\sigma_i^2} \end{pmatrix}.$$

De sus elementos extraemos:

$$\begin{aligned}\sigma^2(a) &= \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2}\right)^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \\ \sigma^2(b) &= \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2}\right)^2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \\ \text{cov}(a, b) &= \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2}\right)^2} \sum_{i=1}^n \frac{-x_i}{\sigma_i^2}.\end{aligned}$$

Obsérvese que esta covarianza está relacionada con el valor medio ponderado de todas las abscisas. Así, para un conjunto de puntos con incertidumbres iguales se obtendrá $\text{cov}(a, b) = 0$ si conseguimos $\sum_{i=1}^n x_i = 0$. En caso contrario, la covarianza tendrá signo opuesto a esta suma.

Es importante señalar que el uso de la regresión lineal para predecir el valor de y para algún valor de x no medido directamente implica un cálculo de incertidumbre adecuado. Supongamos que existe un cierto valor de la abscisa x_0 para el que queremos estimar y_0 mediante la recta de ajuste: $y_0 = a + bx_0$. La incertidumbre asociada a y_0 será:

$$\sigma^2(y_0) = \sigma^2(a) + x_0^2 \sigma^2(b) + 2x_0 \text{cov}(a, b). \quad (3.186)$$

Ajuste a una recta con incertidumbres en ambos ejes

Cuando no se cumple la condición $b^2 \sigma_{x_i}^2 \ll \sigma_{y_i}^2$, el desarrollo de los apartados anteriores es claramente inadecuado. En particular, no es posible considerar que las x_i son valores perfectamente definidos, sino que es necesario trabajar con variables aleatorias bidimensionales (x_i, y_i) con sus correspondientes distribuciones de probabilidad. Si existe una relación lineal entre las variables x e y , y de manera análoga a lo que hemos venido haciendo, supondremos que las medias de estas distribuciones se sitúan sobre la recta de ajuste, es decir, que $\bar{y}_i = a + b\bar{x}_i$. Se estudiará el caso frecuente en que estas distribuciones son gaussianas bidimensionales. Según lo visto en el capítulo dos, esto significa que la probabilidad de obtener un valor (x_i, y_i) es

$$p_i = \frac{1}{\sqrt{2\pi} \det M} \exp \left\{ -\frac{1}{2} \begin{pmatrix} x_i - \bar{x}_i \\ y_i - \bar{y}_i \end{pmatrix} M^{-1} (x_i - \bar{x}_i, y_i - \bar{y}_i) \right\}, \quad (3.187)$$

donde M es la matriz de covarianzas de la variable aleatoria:

$$M = \begin{pmatrix} \sigma_{x_i}^2 & \text{cov}(x_i, y_i) \\ \text{cov}(x_i, y_i) & \sigma_{y_i}^2 \end{pmatrix}$$

$$M^{-1} = \frac{1}{\sigma_{x_i}^2 \sigma_{y_i}^2 - [\text{cov}(x_i, y_i)]^2} \begin{pmatrix} \sigma_{y_i}^2 & -\text{cov}(x_i, y_i) \\ -\text{cov}(x_i, y_i) & \sigma_{x_i}^2 \end{pmatrix}. \quad (3.188)$$

El método de máxima verosimilitud aplicado a este caso permite afirmar, mediante un desarrollo completamente análogo al de apartados anteriores, que los valores de a y b que maximizan la probabilidad de obtener el conjunto de datos experimentales es el que minimiza

$$\chi^2(a, b) = \sum_i \frac{(y_i - a - bx_i)^2}{\sigma_{y_i}^2 + b^2 \sigma_{x_i}^2 - 2b \text{cov}(x_i, y_i)}. \quad (3.189)$$

Puesto que $\sigma^2 \{y_i - a - bx_i\} = \sigma_{y_i}^2 + b^2 \sigma_{x_i}^2 - 2b \text{cov}(x_i, y_i)$, es evidente que esta fórmula es análoga a la de un ajuste ponderado, pero considerando el peso estadístico de $y_i - a - bx_i$ en lugar de solamente el de y_i .

Cada uno de los sumandos de la expresión (3.189) es el cociente de un cuadrado y una varianza. Salvo el caso trivial y no físico en que los puntos estén perfectamente determinados o se hallen exactamente sobre la recta, ambos son estrictamente positivos. Esto conduce a que $\chi^2(a, b)$ sea continua y diferenciable en todo \mathbb{R}^2 , de manera que sus extremos se pueden caracterizar por derivación:

$$0 = \frac{\partial \chi^2(a, b)}{\partial a} = -2 \sum_i W_i (y_i - a - bx_i) \quad (3.190)$$

donde, para abreviar la notación, se ha definido $W_i = [\sigma^2 \{y_i - a - bx_i\}]^{-1}$. Si mediante una doble barra indicamos los promedios ponderados por W_i ($\bar{f} = \frac{\sum_i W_i f_i}{\sum_i W_i}$), podemos reescribir lo anterior como:

$$a = \bar{y} - b\bar{x}. \quad (3.191)$$

Como $\chi^2(a, b)$ es una función cuadrática de a con coeficiente de grado dos positivo, esta ecuación arroja un mínimo absoluto en a para un valor de b dado. La otra derivada parcial a anular es:

$$0 = \frac{\partial \chi^2(a, b)}{\partial b} = -2 \sum_i W_i x_i (y_i - a - bx_i) - 2 \sum_i W_i^2 (y_i - a - bx_i)^2 [b\sigma_{x_i}^2 - \text{cov}(x_i, y_i)]. \quad (3.192)$$

Para establecer la existencia de un mínimo, estudiemos el comportamiento de estas funciones en los extremos de la recta real. Cuando $b \rightarrow \pm\infty$,

$$\begin{aligned}
\bar{x} &\rightarrow \frac{\sum_i \frac{x_i}{\sigma_{x_i}^2}}{\frac{1}{\sigma_{x_i}^2}} \\
a &\rightarrow -b\bar{x} \\
\chi^2(\bar{y} - b\bar{x}, b) &\rightarrow \sum_i \frac{x_i + \bar{x}}{\sigma_{x_i}^2} \\
\left. \frac{\partial \chi(a, b)}{\partial b} \right|_{a=\bar{y}-b\bar{x}} &\rightarrow -\frac{2}{|b|} \bar{x} \sum_i \frac{\bar{x} + x_i}{\sigma_{x_i}^2}.
\end{aligned} \tag{3.193}$$

La función $\chi^2(b) = \chi^2(\bar{y} - b\bar{x}, b)$ tiene asíntotas horizontales con el mismo valor en los dos extremos de la recta real, y su derivada tiene el mismo signo al aproximarse a ambos. Como además es continua, y no es igual a una constante, tiene al menos un máximo y un mínimo. Como cociente de polinomios, posee un número finito de extremos, y en consecuencia existe un mínimo absoluto. Obsérvese, no obstante, que la ecuación no lineal (3.192) no admite solución explícita, por lo que se deben emplear métodos numéricos para calcularla de modo aproximado. La mayor parte de estos métodos requieren una estimación inicial de la posición del mínimo. En este caso, y puesto que es de esperar que la inclusión de la incertidumbre de y y de la correlación no cambie drásticamente el ajuste, en general el valor de b obtenido mediante los procedimientos de apartados anteriores será un punto de partida razonable para la minimización.

Un método con buenas propiedades de convergencia es el desarrollado para este problema por York et al. Para llegar a él, observemos que la ecuación (3.192) se puede expresar, mediante manipulaciones inmediatas, como:

$$0 = \sum_i W_i \beta_i (V_i - bU_i) \tag{3.194}$$

si se hacen las siguientes definiciones:

$$\begin{aligned}
U_i &= x_i - \bar{x} \\
V_i &= y_i - \bar{y} \\
\beta_i &= W_i [bV_i\sigma_{x_i}^2 + U_i\sigma_{y_i}^2 - (V_i + bU_i) \text{cov}(x_i, y_i)].
\end{aligned} \tag{3.195}$$

Esto permite diseñar un método iterativo en que

$$b_{k+1} = \left. \frac{\sum_i \beta_i W_i V_i}{\sum_i \beta_i W_i U_i} \right|_{b=b_k}, \tag{3.196}$$

iteración cuyos puntos fijos son, obviamente, aquellos valores de b para los que la derivada de $\chi^2(b)$ se anula. Puede probarse que los únicos puntos fijos estables son los mínimos, por lo que si la iteración converge lo hará siempre a un mínimo. El criterio de convergencia más habitual es que $|b_{k+1} - b_k| / |b_k|$ sea menor de un cierto umbral δ durante un número n_δ de iteraciones. En el

caso de la estimación de b , es suficiente con que δ sea del orden de la milésima parte de $\frac{\sigma_b}{|b_k|}$, ya que un cambio menor no afectaría a las cifras significativas de b . Aunque en principio para asegurarse de hallar el mínimo absoluto habría que buscarlos todos y compararlos, en la práctica se ha comprobado en la literatura científica que si se escoge $b_0 = 0$ se suele llegar al resultado correcto en pocas iteraciones. Obsérvese que en este caso b_1 es la pendiente del ajuste ponderado con incertidumbres sólo en las ordenadas. Cabe mencionar también que en algunos casos pueden existir varios ajustes que arrojen valores iguales o muy próximos de χ^2 .

Para evaluar las incertidumbres de los parámetros no hay más que recurrir a la fórmula estándar de propagación. Comenzando por b :

$$\sigma_b^2 = \sum_i \left\{ \left(\frac{\partial b}{\partial x_i} \right)^2 \sigma_{x_i}^2 + \left(\frac{\partial b}{\partial y_i} \right)^2 \sigma_{y_i}^2 + 2\text{cov}(x_i, y_i) \left(\frac{\partial b}{\partial x_i} \right) \left(\frac{\partial b}{\partial y_i} \right) \right\}. \quad (3.197)$$

Ahora bien, ha de tenerse en cuenta que las derivadas parciales han de ser evaluadas en los valores medios (\bar{x}_i, \bar{y}_i) . Por definición, este punto se halla sobre la recta de ajuste. Se trata, por tanto, de hallar el punto de la recta que maximiza la probabilidad expresada en la ecuación (3.187), es decir, el \bar{x}_i que hace menos negativo el exponente si $\bar{y}_i = a + b\bar{x}_i$. Como M no depende de \bar{x}_i , la parte del exponente a minimizar es:

$$\sigma_{y_i}^2 (x_i - \bar{x}_i)^2 + \sigma_{x_i}^2 (y_i - a - bx_i)^2 - 2\text{cov}(x_i - \bar{x}_i)(y_i - a - b\bar{x}_i). \quad (3.198)$$

Si desarrollamos los cuadrados, inmediatamente identificaremos W_i^{-1} como el coeficiente de \bar{x}_i^2 , de modo que hay un mínimo y ningún máximo. La localización de este punto se obtiene fácilmente por derivación:

$$\begin{aligned} \bar{x}_i &= W_i [\sigma_{y_i}^2 x_i + (y_i - a) b \sigma_{x_i}^2 - \text{cov}(x_i, y_i)(y_i - a + bx_i)] = \\ &= W_i \{ \sigma_{y_i}^2 2 + (U_i + \bar{x}) + b \sigma_{x_i}^2 (V_i + b\bar{x}) - \text{cov}(x_i, y_i) [V_i + b(U_i + 2\bar{x})] \} = \\ &= W_i \{ U_i \sigma_{y_i}^2 + b V_i \sigma_{x_i}^2 - (b U_i + V_i) \text{cov}(x_i, y_i) + \\ &\quad + \bar{x} [\sigma_{y_i}^2 + b^2 \sigma_{x_i}^2 - 2b \text{cov}(x_i, y_i)] \}. \end{aligned} \quad (3.199)$$

A la luz de las definiciones de esta sección, esta última forma es simplemente:

$$\bar{x}_i = \bar{x} + \beta_i. \quad (3.200)$$

La forma de las derivadas de b se obtiene, de manera explícita, de la derivación de la ecuación implícita (3.196). Por brevedad expositiva no incluiremos aquí el desarrollo completo, que da como resultado:

$$\sigma_b^2 = \frac{\sum_i W_i}{\left(\sum_i W_i \bar{x}_i^2 \right) \left(\sum_i W_i \right) - \left(\sum_i W_i \bar{x}_i \right)^2} \quad (3.201)$$

El caso de la incertidumbre de a es mucho más sencillo, ya que las derivadas relevantes pueden relacionarse con las de b usando que $a = \bar{y} - b\bar{x}$. Sin más que sustituir en la ecuación equivalente a (3.197) para a y agrupar términos, se llega a:

$$\sigma_a^2 = \frac{1}{\sum_i W_i} + \bar{x}^2 \sigma_b^2, \quad (3.202)$$

lo cual completa la obtención de los parámetros y sus incertidumbres. Es de resaltar que las ordenadas y las abscisas juegan papeles equivalentes en todas ellas (aunque en algún caso no sea inmediatamente obvio por la forma escogida) en contraste a lo que ocurría en apartados anteriores.

Es ilustrativo mencionar algún caso particular especialmente simple y que se pueda dar en la práctica. El más obvio es el de $b^2 \sigma_{x_i}^2 \ll \sigma_{y_i}^2$ (o viceversa) y sin correlación. Como es esperable, permite recuperar las fórmulas del ajuste ponderado ordinario, y se propone como ejercicio para el lector al final del capítulo. Un caso algo más general y muy habitual es aquél en el que no se puede despreciar ninguna de las incertidumbres, pero sí la correlación (ya sea porque se ha comprobado empíricamente que es pequeña o porque se tienen razones para pensar que las magnitudes de influencia sobre ordenadas y abscisas están desconectadas entre sí). En tal caso se habla de una regresión ortogonal. Se puede comprobar fácilmente que esta simplificación no es suficiente para convertir en explícita la ecuación para b . Para ello y para poder comentar analíticamente el efecto de las incertidumbres sobre los resultados, haremos la asunción adicional $\sigma_{x_i} = D\sigma_{y_i}$, siendo D una constante real positiva. Con esto las expresiones anteriores se convierten en

$$\begin{aligned} W_i &= [(1 + b^2 D^2) \sigma_{y_i}]^{-1} \\ \bar{x} &= \frac{\sum_i \frac{x_i}{\sigma_{y_i}^2}}{\sum_i \frac{1}{\sigma_{y_i}^2}} \\ \bar{y} &= \frac{\sum_i \frac{y_i}{\sigma_{y_i}^2}}{\sum_i \frac{1}{\sigma_{y_i}^2}} \\ \beta_i &= \frac{U_i + D^2 V_i}{1 + b^2 D^2} \\ b &= \frac{\sum_i \frac{1}{\sigma_{y_i}^2} (U_i + D^2 V_i) V_i}{\sum_i \frac{1}{\sigma_{y_i}^2} (U_i + D^2 V_i) U_i} \\ \frac{b}{b|_{D=0}} &= \frac{1 + D^2 b|_{D \rightarrow \infty}}{1 + D^2 b|_{D=0}}. \end{aligned} \quad (3.203)$$

El centro de masas ponderado del ajuste (\bar{x}, \bar{y}) es independiente del parámetro D , que representa la contribución relativa de la incertidumbre en las x . En b , sin embargo, se observa que al aumentar D (partiendo de cero, que representa

el ajuste ponderado antes estudiado) aparece una contribución a la pendiente de las desviaciones sobre el eje de ordenadas de los datos con respecto a los promedios ajustados. La derivada de la última ecuación con respecto a D es:

$$\frac{d}{dD} \left(\frac{b}{b|_{b=0}} \right) = 2D \frac{b|_{D \rightarrow \infty} - b|_{D=0}}{(1 + D^2 b|_{D=0})^2}. \quad (3.204)$$

El signo de esta derivada es igual al de $b|_{D \rightarrow \infty} - b|_{D=0}$, que se puede expandir a:

$$\begin{aligned} b|_{D \rightarrow \infty} - b|_{D=0} &= \frac{\sum_i \frac{V_i^2}{\sigma_{y_i}^2}}{\sum_i \frac{U_i V_i}{\sigma_{y_i}^2}} - \frac{\sum_i \frac{U_i V_i}{\sigma_{y_i}^2}}{\sum_i \frac{U_i^2}{\sigma_{y_i}^2}} = \\ &= \frac{\left(\sum_i \frac{V_i^2}{\sigma_{y_i}^2} \right) \left(\sum_i \frac{U_i^2}{\sigma_{y_i}^2} \right) - \left(\sum_i \frac{U_i V_i}{\sigma_{y_i}^2} \right)^2}{\left(\sum_i \frac{U_i V_i}{\sigma_{y_i}^2} \right) \left(\sum_i \frac{U_i^2}{\sigma_{y_i}^2} \right)}. \end{aligned} \quad (3.205)$$

El numerador es siempre positivo en virtud de la desigualdad de Cauchy-Schwartz. En consecuencia la diferencia tiene el mismo signo que $\sum_i \frac{U_i V_i}{\sigma_{y_i}^2}$, que a su vez es el signo de los límites $b|_{D \rightarrow \infty}$ y $b|_{D=0}$. De este desarrollo se extraen dos importantes conclusiones para este caso particular de incertidumbres proporcionales: que la inclusión de las incertidumbres en los dos ejes nunca cambia el signo de la pendiente del ajuste y que su valor absoluto, $|b|$, es una función monótonamente creciente de D , luego el ajuste con incertidumbres en un sólo eje siempre subestima la pendiente (en valor absoluto).

Cuando $D \ll b|_{D=0}^{-\frac{1}{2}}$, es decir, cuando la incertidumbre en las ordenadas no es aún muy relevante, la última de las ecuaciones en (3.203) se puede aproximar por

$$\frac{b}{b|_{D=0}} \simeq 1 + D^2 (b|_{D \rightarrow \infty} - b|_{D=0}) \quad (3.206)$$

donde se aprecia claramente la contribución del ajuste con incertidumbres sólo en las abscisas ($D \rightarrow \infty$) a la pendiente.

Puesto que \bar{x} e \bar{y} no dependen de D , el cambio en la ordenada en el origen obtenida en el ajuste es $a - a|_{D=0} = \bar{x} (b - b|_{D=0})$, que tiene el signo de $b\bar{x}$.

En cuanto a las incertidumbres de los parámetros, al dividirse el peso estadístico de cada dato entre $1 + b^2 D^2$, la incertidumbre de b dada por (3.201) se ve multiplicada por la raíz cuadrada de ese factor, por lo que en todo caso aumenta, como era de esperar.

3.5. Cuestiones y problemas

3.1 Dada una muestra con tres datos se construye el estadístico $\psi = x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3$. Estudiar si se trata de un estimador fiel y consistente de la media de la población.

3.2 Demuéstrese que $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ es un estimador fiel de la varianza poblacional y que además es un estimador consistente, dado que

$$\sigma_{(s^2)}^2 = \sigma^4 \left(\frac{2}{n-1} + \frac{(g-3)}{n} \right); \quad g = \frac{E\{(x-\mu)^4\}}{\sigma^4}$$

3.3 Una variable aleatoria verifica que la media de su población es nula ($\mu = 0$). Considerando una muestra de n datos de esta población, $\{x_i\}_{i=1}^n$, estudiar si el estimador

$$t = \frac{1}{n-1} \sum_{i=1}^n x_i^2$$

es un estimador fiel de la varianza de la población.

3.4 Dadas dos variables aleatorias x e y se toma una muestra formada por n pares de datos $[(x_1, y_1); (x_2, y_2); \dots; (x_n, y_n)]$. Los estimadores de las correspondientes medias poblacionales de x e y (μ y ν respectivamente) son

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Demostrar que la covarianza de \bar{x} e \bar{y} está relacionada con la covarianza poblacional $\text{cov}(x, y)$ mediante

$$\text{cov}(\bar{x}, \bar{y}) = \frac{\text{cov}(x, y)}{n}$$

Considérese que $E\{(x_i - \mu) \cdot (y_j - \nu)\} = 0 \quad i \neq j$

3.5 Dada una muestra de n datos procedente de una población madre normal, calcula la varianza del estadístico,

$$\lambda = \frac{(\bar{x} - \mu)}{s/\sqrt{n}}$$

donde consideramos conocida la media de la población μ .

3.6 Para una muestra de n pares de datos como la del ejercicio anterior, mostrar que un estimador fiel de la covarianza $\text{cov}(x, y)$ viene dado por

$$s(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

3.7 Considérese que las variables aleatorias independientes $\{x_1, x_2, \dots, x_n\}$ son normales con media μ y desviación típica σ . Demostrar que la variable

$$y_{n-1} = \frac{1}{\sqrt{n(n-1)}}(x_1 + x_2 + \dots + x_{n-1} - (n-1)x_n)$$

es normal con media nula y desviación típica σ .

Nota: Se verificará que $E\{(x_i - \mu) \cdot (x_j - \mu)\} = 0 \quad i \neq j$

3.8 Demuéstrese que las variables definidas en la transformación (3.40) verifican la relación $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$.

3.9 En un experimento se ha tomado una serie de 10 datos obteniéndose los resultados $\{18, 21, 23, 19, 20, 21, 20, 19, 20, 17\}$ si suponemos que la distribución madre es normal, dar un intervalo con un nivel de confianza del 95 % para la media y la varianza de la población.

3.10 Se realiza una encuesta a 100 personas antes de unas elecciones para averiguar la proporción de votantes que votarán al partido A. De esta muestra 36 manifiestan esta intención de voto. Calcular con un nivel de confianza del 99 % la proporción de votantes del partido A. ¿A cuántas personas habría que entrevistar para que con el mismo nivel de confianza la proporción se conociera con una incertidumbre de $\pm 2\%$?

3.11 Se mide la actividad de una fuente radiactiva mediante un contador de pozo, obteniéndose un total de 3224 cuentas en un minuto. Obtener con un nivel de confianza del 99 % un intervalo para el valor medio de la tasa de desintegraciones por minuto.

3.12 En un experimento se ha tomado una serie de 10 datos obteniéndose los resultados $\{18, 21, 23, 19, 20, 21, 20, 19, 20, 17\}$ si suponemos que la distribución madre es normal, dar un intervalo con un nivel de confianza del 95 % para la media y la varianza de la población.

3.13 Se mide la actividad de una fuente radiactiva mediante un contador de pozo, obteniéndose un total de 3224 cuentas en un minuto. Obtener con un nivel de confianza del 99 % un intervalo para el valor medio de la tasa de desintegraciones por minuto.

3.14 Para estudiar la intención de voto al partido PDX se entrevista a un conjunto de 1000 personas. De esta muestra 256 manifiestan su intención de voto favorable a este partido.

i) Encontrar un intervalo de confianza para la fracción p de votantes del partido PDX, con un nivel de confianza del 95 % (considérese una aproximación gaussiana).

ii) Si suponemos que la proporción de votantes no cambia significativamente al aumentar el tamaño de la muestra, ¿a cuántas personas habría que entrevistar para que la semianchura de este intervalo de confianza representase el 5 % del valor estimado de p ?

- 3.15 Un fabricante de pinturas afirma que con un bote de 1 litro se puede pintar una superficie de 20 m^2 . Se realiza un muestreo aleatorio simple y se obtienen los siguientes resultados para 6 casos:

Superficie (m^2)	18	15	21	16	19	22
-----------------------------	----	----	----	----	----	----

Considerando que la variable aleatoria “superficie” sigue una distribución madre normal, ¿qué podemos decir de la hipótesis del fabricante a un nivel de confianza del 90 %?

- 3.16 En una población total de 10000 hogares se realiza una encuesta de audiencia de un programa de televisión, a lo largo de distintos días de emisión, obteniéndose los resultados

5200	6100	4900	4600	4100	4500	3900	6300	4700	4200
------	------	------	------	------	------	------	------	------	------

Suponiendo que la distribución de probabilidad de audiencia es normal, hacer un test de hipótesis para la afirmación de que la audiencia media es del 50 % a un nivel de confianza del 90 %.

- 3.17 La dosis media que se desea impartir a un paciente en un tratamiento de radioterapia es de 2,00 Gy. Se realizan 10 medidas mediante una cámara de ionización con los siguientes resultados (en Gy):

1,98 1,93 1,98 1,89 1,92 1,95 1,94 1,96 1,93 1,92

Establecer si, al 99 % de nivel de confianza, la dosis media es distinta de la deseada.

- 3.18 Consideremos que en un experimento de Bernoulli se lanza una moneda y se obtienen 10 caras y 20 cruces. Si se asigna a este suceso una probabilidad binomial, usar el método de máxima verosimilitud para obtener el valor de la probabilidad de sacar cara. Obsérvese que esta forma de estimar el valor de la probabilidad coincide con la aproximación convencional (número de éxitos sobre el número total de sucesos).

- 3.19 En un experimento de determinación de la constante elástica de un resorte se han medido en tres ocasiones las elongaciones para diferentes masas de objetos suspendidos, empleando un metro con una precisión de 0.002 cm, obteniéndose los siguientes resultados:

m (g)	l_1 (cm)	l_2 (cm)	l_3 (cm)
1.0	0.993	1.011	0.979
2.0	1.954	1.944	1.978
3.0	2.918	2.949	3.011
4.0	4.143	4.157	3.987

- i) Obténganse los valores de la elongación correspondiente a cada masa junto con sus incertidumbres. Representense los anteriores valores en una gráfica $l(m)$ mostrando su barra de incertidumbre.
- ii) Realícese un ajuste lineal ponderado de los valores de la elongación frente a los de masa, y proporciónese un valor para la constante elástica del resorte problema y para su incertidumbre.

iii) La hipótesis de linealidad de la elongación frente a la masa, ¿es aceptable al 95 % de nivel de confianza? ¿Y a un 99 %?

3.20 En un determinado experimento de medición de la densidad de una disolución acuosa diluida a una temperatura de 298,15 K frente a la concentración de soluto (en moles/l), se obtuvieron los siguientes resultados:

c (mol l ⁻¹)	0,000	0,005	0,010	0,015	0,020
ρ (kg m ⁻³)	990,0	992,2	994,4	996,6	998,7

Representense los datos obtenidos y hágase una regresión lineal ponderada $\rho(c) = a + b c$ utilizando el método de mínimos cuadrados, teniendo en cuenta que la incertidumbre relativa del densímetro es del 1 % del valor del mensurando. Calcular explícitamente el coeficiente de correlación de las variables aleatorias c y ρ . Si la velocidad del sonido, u , en dicha disolución puede considerarse constante e igual a (1500 ± 100) ms⁻¹ en el rango de concentración analizado, obténgase la representación de la compresibilidad adiabática del fluido

$$K_s = \frac{1}{\rho u^2}$$

frente a la concentración así como la incertidumbre de cada punto.

3.21 Se quiere ajustar por el método de mínimos cuadrados un conjunto de datos $\{(x_1, y_1), \dots, (x_n, y_n)\}$ a una recta $y = a$. ¿Cuál es el valor de la constante a ? ¿Cuál sería la expresión de la constante en el caso de que cada uno de los datos y_i tuviese una incertidumbre σ_i ?

3.22 Compruébese que, en el caso de un ajuste a una recta $y = a + bx$, las incertidumbres de los parámetros a y b halladas según la ecuación (3.185) coinciden con las dadas por las expresiones que se obtuvieron previamente para la regresión lineal ponderada, ecuaciones (3.146) y (3.148).

3.23 Demuéstrese que las fórmulas desarrolladas para el ajuste a una recta con incertidumbres en ambos ejes se convierten en las del ajuste pesado ordinario cuando la incertidumbre en las abscisas y las covarianzas $\text{cov}(x_i, y_i)$ son despreciables.

Capítulo 4

Expresión de la incertidumbre de la medida

4.1. Introducción. Error e incertidumbre

La necesidad de comunicar los resultados de una medida de forma que puedan ser comprendidos por todos los miembros de la comunidad científica, industrial o comercial, ha llevado a intentar establecer reglas que normalicen la expresión de los valores arrojados por el proceso de medida. Estas reglas o recomendaciones se refieren tanto a las unidades de las magnitudes medidas, como a la forma en que la incertidumbre inherente al proceso de medida debe reflejarse junto al resultado numérico de ésta. La correcta expresión de la incertidumbre de la medida tiene como objetivo garantizar la intercomparación de los datos y establecer el grado de precisión con el que un dispositivo permite determinar el valor de una magnitud.

En el presente capítulo introduciremos la nomenclatura en español asociada a los conceptos de error e incertidumbre. La presente nomenclatura se basa en la “Guía para la expresión de la incertidumbre de la medida” editada en el año 2000 por el Centro Español de Metrología. Esta Guía es la traducción de la recomendación CI-1981 del Comité Internacional de Pesas y Medidas y está apoyada por la mayoría de las organizaciones científicas de carácter internacional (Comité Internacional de Pesas y Medidas, Comisión Electrotécnica Internacional, International Organization for Standardization, Organización Internacional de Metrología Legal, Unión Internacional de Química Pura y Aplicada, y Unión Internacional de Física Pura y Aplicada).

La realización de una medida consiste en comparar una cantidad de una cierta magnitud con otra de su misma clase que se adopta como patrón. El proceso de la medida involucra a tres elementos básicos que son el *mensurando* (lo que se mide), el *instrumento* de medida (aquello con lo que se mide) y el *operador* (el que mide). Alrededor de este proceso pueden existir factores que afecten al resultado de la medida, pero cuyo control *a priori* no sea posible. Este conjunto de factores se conocen como *magnitudes de influencia* y pueden incluir, por ejemplo, la temperatura ambiente, la humedad relativa del aire, la

presión atmosférica absoluta, errores de calibración instrumental, defectos del operador al realizar la medida, etc.

Las *magnitudes de influencia* son aquellas que, no siendo el objeto de la medida, afectan al resultado de ésta de un modo determinante.

Como se puede deducir de lo expuesto anteriormente, la medición física es realmente un experimento aleatorio y resulta en la práctica imposible conocer el valor exacto del mensurando, por lo que, en lugar de arrojar un valor numérico como resultado de una medida, la expresaremos escribiendo un intervalo dentro del cual se encontrará, con mayor o menor probabilidad, el valor verdadero del mensurando. A la semiamplitud de dicho intervalo la denominaremos *incertidumbre de la medida*. El resultado de una medición se expresará en la forma:

$$x = (\bar{x} \pm u_x) \text{ unidades,} \quad (4.1)$$

siendo \bar{x} su valor convencionalmente aceptado, obtenido como la media muestral de un conjunto de observaciones, y u_x su incertidumbre. Debemos resaltar, para conectar con los temas anteriores, que esta expresión no da un intervalo de confianza, salvo en el caso en que la incertidumbre es la combinada.

Todo resultado de una medida debe expresarse con su correspondiente incertidumbre. La incertidumbre es un parámetro que caracteriza el proceso de medida de una magnitud, y corresponde a la dispersión de los valores que podrían asignarse a ésta con un cierto grado de verosimilitud. Cuanto menor sea el valor de la incertidumbre de una medida mayor será la calidad de ésta. La incertidumbre se calcula normalmente *a posteriori* del proceso de medida teniendo en cuenta todas las magnitudes de influencia que participan en él, y debe ser diferenciada de algunos conceptos relacionados.

Así, en primer lugar, debe distinguirse claramente entre incertidumbre y tolerancia. La tolerancia nos viene impuesta *a priori* como un intervalo prefijado dentro del cual deben encontrarse los valores de una magnitud para ésta sea aceptada como válida. La tolerancia se deriva normalmente de un reglamento o instrucción de carácter técnico que regula los valores aceptables de una magnitud. Evidentemente, para que una medida pueda tener un valor discriminatorio frente a un cierto intervalo de tolerancia (de semiamplitud T), la incertidumbre asociada a ésta no debe ser demasiado grande, ya que en otro caso el resultado de la medida es de dudoso valor. Si U es la incertidumbre de la medida, se suele adoptar en metrología la regla práctica $3 \leq \frac{T}{2U} \leq 10$.

Por otro lado, la exactitud de una medida es el grado de aproximación entre el resultado de una medición o serie de medidas y el valor del mensurando aceptado como referencia. Podemos decir que la exactitud como cualidad de una medida tiene dos componentes fundamentales: la veracidad y la precisión.

Veracidad o justeza: es el grado de coincidencia existente entre el valor medio obtenido en una serie de medidas y el valor del mensurando convencionalmente aceptado como correcto.

Precisión o fidelidad: es el grado de coincidencia entre los resultados independientes de una serie de medidas obtenidos en condiciones estipuladas.

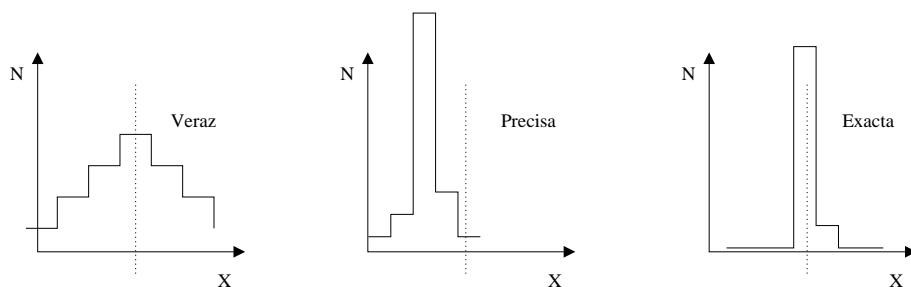


Figura 4.1: Concepto de exactitud, veracidad y precisión

Es también importante destacar la diferencia existente entre los conceptos de *error* e *incertidumbre*. Las medidas son siempre realizadas mediante instrumentos y métodos imperfectos. Si en la medida hubiera algún defecto pero éste pudiera ser *perfectamente* conocido, entonces sería posible corregirlo y la medida quedaría libre de error. En la práctica, la calibración del instrumental de uso en el laboratorio permitiría subsanar las desviaciones de carácter *sistemático* de tales instrumentos.

Por tanto, en el proceso de medida podremos hacer ciertas correcciones a nuestros instrumentos. Estas correcciones necesitan a su vez ser medidas y tal vez corregidas. De esta manera vemos que el proceso de la medida podría considerarse un bucle infinito. Debemos en general buscar un compromiso para establecer una *corrección residual* y limitar el número de iteraciones del proceso. Este planteamiento implica la hipótesis de que los instrumentos de medida son razonablemente buenos y la *corrección residual* puede acotarse.

El *error* designa adecuadamente la desviación sistemática de los resultados de una medida respecto al valor convencionalmente correcto del mensurando.

Por su parte, la *incertidumbre* hace referencia al grado de posible variación en los resultados del proceso de medida que subsiste tras haber aplicado las correcciones pertinentes. Es decir, la incertidumbre es una cota superior estimada de esta corrección residual, consecuencia de la imperfección de nuestra medida corregida.

4.2. Unidades y expresión de la medida

Las unidades de medida surgen históricamente como consecuencia de los intercambios comerciales entre los pueblos. Al incrementarse tales transacciones comerciales aparece la necesidad de armonizar y establecer las unidades de medida utilizadas. Desde el siglo XVIII se produjeron en Europa distintas iniciativas cuyo objetivo era unificar las diferentes unidades vigentes en cada Estado. En el año 1791 la Asamblea Nacional de Francia decide adoptar el llamado Sistema Métrico Decimal, que tenía como unidad básica de longitud el metro (definido como 10^{-7} la longitud del cuadrante del meridiano terrestre). Este sistema fue adoptado progresivamente por otros países. Por ejemplo,

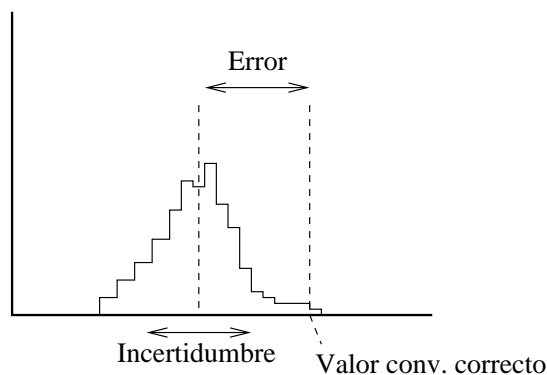


Figura 4.2: Concepto de error e incertidumbre

España lo hizo en su ley de Pesas y Medidas de 1849. Posteriormente surgió el sistema CGS de medida, basado en tres unidades de la mecánica: el centímetro, el gramo y el segundo. Este sistema de unidades tuvo una gran relevancia en el desarrollo de la experimentación en Física.

Al sistema CGS le seguiría el sistema de medida MKS, tras la firma en París del Tratado de la Convención del Metro, cuyas unidades básicas eran el metro, el kilogramo y el segundo. Durante los años siguientes se fueron incorporando unidades de medida eléctricas, apareciendo en 1939 el sistema MKSA con la incorporación del amperé como unidad básica. Finalmente la Conferencia General de Pesas y Medidas (C.G.P.M.) celebrada en 1960 adoptó el denominado Sistema Internacional de Unidades y recomendó su uso en todos los ámbitos de carácter científico o técnico.

4.2.1. Sistema Internacional de Unidades

Dentro del Sistema Internacional de Unidades (en adelante SI) distinguiremos dos tipos de unidades:

- Unidades básicas o fundamentales
- Unidades derivadas

Aunque esta división supone una cierta arbitrariedad, la decisión de la C.G.P.M. siguió el criterio de designar las siete unidades básicas como aquellas mejor definidas y que se consideran independientes desde el punto de vista dimensional. Los avances científicos y técnicos, así como la disponibilidad de instrumentos de una mayor precisión, han conducido a que las definiciones (a enero de 2000) de las unidades básicas se realicen en base a experimentos físicos (con excepción del kilogramo). La segunda categoría de unidades, denominadas derivadas, se obtiene a través de relaciones algebraicas de las unidades fundamentales del SI. El conjunto de ambas clases de unidades forman un sistema coherente según los criterios que se indican a continuación:

1. Las unidades básicas se definen (a excepción del kilogramo) en función de constantes físicas.
2. Cada magnitud se expresa en términos de una única unidad, obtenida por multiplicación o división de las unidades base y de las unidades derivadas adimensionales.
3. Los múltiplos y submúltiplos se obtienen mediante la multiplicación por una potencia exacta de diez.
4. Las unidades derivadas se expresan en términos de las unidades básicas mediante relaciones algebraicas que no conllevan factores numéricos.

Las siete unidades básicas son las siguientes:

Magnitud	Nombre	Símbolo
Longitud	metro	m
Masa	kilogramo	kg
Tiempo	segundo	s
Intensidad de corriente eléctrica	ampére	A
Temperatura termodinámica	kelvin	K
Intensidad luminosa	candela	cd
Cantidad de sustancia	mol	mol

Las unidades derivadas, como ya hemos mencionado, se forman a partir de las unidades básicas mediante relaciones algebraicas. Algunas de estas unidades reciben nombres y símbolos específicos que pueden usarse para expresar unidades derivadas de un modo más simple que si las escribiéramos directamente en función de las unidades fundamentales. Estas unidades derivadas pueden clasificarse en tres grupos principales que enumeramos a continuación con algunos ejemplos.

1. Unidades derivadas a partir de las unidades básicas:

Superficie	metro cuadrado	m²
Volumen	metro cúbico	m³
Velocidad	metro por segundo	m/s
Masa volúmica	kilogramo por metro cúbico	kg/m³

2. Unidades derivadas con nombre específico:

Frecuencia	hertz	Hz
Fuerza	newton	N
Presión	pascal	Pa
Capacidad eléctrica	farad	F

3. Unidades derivadas a partir de las que poseen nombre específico:

Viscosidad dinámica	pascal segundo	Pa s
Entropía	joule por kelvin	J/K
Campo eléctrico	volt por metro	V/m
Momento	newton metro	N m

Finalmente existe un grupo de unidades no incluidas en el Sistema Internacional pero reconocidas por la C.G.P.M. como consecuencia de la necesidad de los usuarios de expresar sus medidas en otras unidades diferentes. Dentro de estas unidades estableceremos cuatro grupos fundamentales.

i) Unidades utilizadas con el SI.

Se trata de unidades cuya aplicación se admite complementariamente a las definidas en el SI como resultado de su amplia difusión.

minuto	min	1 min = 60 s
hora	h	1 h = 60 min = 3 600 s
día	d	1 d = 24 h = 86 400 s
grado de arco	°	1 ° = ($\pi/180$) rad
minuto de arco	'	1 ' = (1/60)°
segundo de arco	"	1 " = (1/60)'
litro	l	1 l = 10⁻³ m³
tonelada	t	1 t = 10³ kg

ii) Unidades cuyo valor se obtiene experimentalmente.

Debido a su definición práctica se desconoce su valor exacto.

unidad de masa atómica	u	1 u \approx 1,660 540 2 \times 10⁻²⁷ kg
electronvolt	eV	1 eV \approx 1,602 177 33 \times 10⁻¹⁹ J
unidad astronómica	ua	1 ua \approx 1,495 978 706 91 \times 10¹¹ m

iii) Unidades admitidas temporalmente.

Se admite su uso en base a la difusión práctica de éstas. Sin embargo, a medio plazo habrán de ser sustituidas.

milla marina		1 milla marina = 1 852 m
nudo		1 nudo = 1 milla marina por hora
ångström	Å	1 Å = 1 \times 10⁻¹⁰ m
área	a	1 a = 100 m²
hectárea	ha	1 ha = 10 000 m²
bar	bar	1 bar = 1 \times 10⁵ Pa
barn	b	1 b = 1 \times 10⁻²⁸ m²

iv) Unidades desaconsejadas.

quilate métrico		1 quilate métrico = 2 \times 10⁻⁴ kg
tex	tex	1 tex = 1 \times 10⁻⁶ kg m⁻¹
atmósfera normal	atm	1 atm = 101 325 Pa

Los símbolos de las unidades del SI se expresarán en caracteres romanos, excepto Ω , y con letra minúscula, excepto aquellos derivados de nombres propios. Por su parte, el nombre del símbolo se escribirá siempre en minúscula. Por tanto, escribiremos pascal para el nombre del símbolo de la presión pero usaremos Pa como símbolo de la unidad; o bien escribiremos joule para nombrar la unidad de energía cuyo símbolo es J, mientras que el símbolo del metro, por ejemplo, es m.

4.2.2. Múltiplos y submúltiplos

El Sistema Internacional contempla un conjunto de múltiplos y submúltiplos en potencias enteras de diez que permiten expresar cantidades de las magnitudes en un enorme rango de valores.

Factor	Prefijo	Símbolo	Factor	Prefijo	Símbolo
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	centi	c
10^{18}	exa	E	10^{-3}	mili	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^9	giga	G	10^{-12}	pico	p
10^6	mega	M	10^{-15}	femto	f
10^3	kilo	k	10^{-18}	atto	a
10^2	hecto	h	10^{-21}	zepto	z
10^1	deca	da	10^{-24}	yocto	y

4.2.3. Reglas prácticas de escritura

En base a la normativa ISO 31-0 vamos a resumir algunas de las reglas convencionalmente admitidas como convenientes a la hora de expresar el resultado de una medida.

- El signo de separación decimal deberá ser una coma¹.
- Deben usarse las unidades del Sistema Internacional que acabamos de introducir.
- No es recomendable mezclar símbolos y nombres de las unidades.
kilómetros por hora ó km/h *pero no* km por hora ó kilómetros/hora

- Los dígitos deberán escribirse en grupos de tres cifras separados por un pequeño espacio pero no por un punto o coma. Esta regla no se sigue a veces para números con cuatro cifras a uno de los lados de la coma decimal en ese caso es deseable al menos la uniformidad de nomenclatura.

8012,1 ó 8 012,1 *es preferible a* 8.012,1
 43 279,168 29 *pero no* 43.279,16829
 8012,5947 ó 8 012,594 7 *pero no* 8 012,5947 ó 8012,594 7

- Los resultados numéricos se escribirán separados por un espacio de las unidades correspondientes. Las abreviaturas tales como sec en lugar de s o segundo, o bien cc en lugar de cm^3 o centímetro cúbico son incorrectas.

10,21 cm^3 *pero no* 10,21cc
 35 m \pm 23 m ó (35 \pm 23) m *pero no* 35 \pm 23 m
 35 m \times 2 m *pero no* 35 \times 2 m

¹La recomendación ISO admite el uso del punto como separador de las cifras decimales cuyo uso está extendido en los países anglosajones. Se recomienda también la notación 0,27 en lugar de ,27 en el caso de no haber cifras enteras no nulas en el resultado.

- f) No se pone punto a continuación del símbolo de las unidades ni se añades al plural.
- la longitud es 10 m *pero no* la longitud es 10 ms
siendo 10 C la carga eléctrica *pero no* siendo 10 C. la carga eléctrica
- g) Para multiplicar dos símbolos se pone un punto a media altura o se deja un espacio en blanco. Por tanto, el producto de metro por segundo (ó metro segundo) se escribirá m s ó m·s.
- h) Para dividir dos símbolos se utiliza una barra oblicua, una barra horizontal o potencias negativas. Esto significa que metro dividido por segundo (ó metro por segundo) se expresa como m/s, ó $\frac{m}{s}$ ó $m s^{-1}$.

4.3. Cifras significativas y redondeo

Supongamos que un estudiante realiza una medida de la longitud de una vara de madera con una precisión entorno al milímetro. Nuestro estudiante puede arrojar un resultado numérico diferente dependiendo de cuales sean las unidades de longitud escogidas, por ejemplo:

208 000 μm 208 mm 20,8 cm
0,208 m 2,08 dm 0,000 208 km

o incluso

$2,08 \times 10^5 \mu\text{m}$ $2,08 \times 10^2 \text{ mm}$ $2,08 \times 10^1 \text{ cm}$
 $2,08 \times 10^{-1} \text{ m}$ $2,08 \times 10^0 \text{ dm}$ $2,08 \times 10^{-4} \text{ km}$

Las expresiones de esta medida difieren en el número total de dígitos que hemos empleado y en la posición de la coma decimal. Sin embargo, representan la misma cantidad con igual incertidumbre (aproximadamente del orden de 1 mm). Lo que deducimos de ello es que la precisión que nos indican las cifras de esta medida no está dada por el número total de dígitos de la expresión ni tan siquiera por el número de éstos a la derecha de la coma decimal.

En el segundo grupo de cifras hemos usado notación científica y en ellos es particularmente evidente que sólo los dígitos “208” nos dan una información relevante de la medida. ¿Cómo expresa una cifra el grado de precisión de la medida? Tomemos nuestro ejemplo. Es obvio que los ceros a la izquierda del “2” no tienen valor e igualmente ocurre con los ceros a la derecha del “8”. Al conjunto de dígitos comprendidos entre el 2 y el 8 (incluidos) los denominaremos *cifras significativas*. En esta medida cuyo valor numérico es 208 mm entendemos que la incertidumbre en esta cifra es de algunas unidades, digamos del orden del 1%, esto es, que el dígito que tal vez sea incierto es el “8”. Siguiendo este mismo argumento, si en lugar de tres dígitos significativos nuestra medida tuviera cuatro, su precisión relativa sería mayor. En general, llamaremos cifra significativa a aquella que aporta información no ambigua ni superflua acerca de una determinada medida experimental. A continuación detallaremos las reglas prácticas para conocer el número de cifras significativas en cada caso.

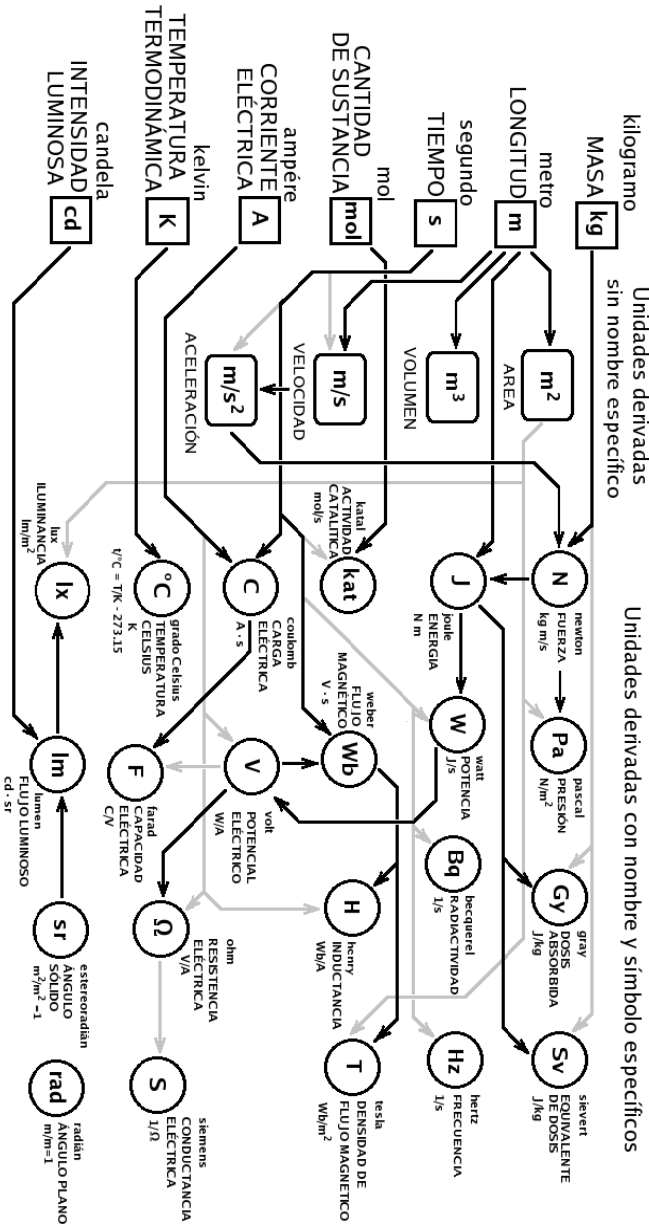


Figura 4.3: Esquema del Sistema Internacional de unidades. Las flechas negras indican producto y las flechas en tono gris indican división. Por ejemplo, $1 \text{ C} = 1 \text{ A} \cdot 1 \text{ s}$, por lo que la casilla del coulomb recibe flecha negras de las dos fundamentales de las que deriva.

4.3.1. Reglas para contar el número de cifras significativas

La precisión de un resultado experimental está implícita en el número de dígitos con los que se expresa la medida, aunque hemos de añadir también el valor de la incertidumbre de la medida para identificar las cifras realmente significativas en un caso concreto. Las reglas básicas para el conteo del número de cifras significativas son:

1. La primera cifra no nula empezando por la izquierda es la cifra más significativa.
2. Si no hay coma decimal, la primera cifra no nula empezando por la derecha es la cifra menos significativa.
3. Si (y sólo si) hay una coma decimal, la primera cifra por la derecha es la cifra menos significativa incluso aunque sea un 0.
4. Todas las cifras situadas entre la más y menos significativa (incluidas éstas) se cuentan como cifras significativas.

En el caso de que la expresión numérica haya sido dada en notación científica, aplicaremos estas reglas a la mantisa, ignorando el exponente. Por ejemplo, los números siguientes poseen todos ellos cuatro cifras significativas

$$1\ 234 ; \quad 12,34 ; \quad 0,001\ 234 ; \quad 123\ 400$$

Debemos tener aquí sumo cuidado en cómo usamos la coma decimal, ya que debido a nuestra definición de cifras significativas su inclusión puede modificar sustancialmente el número de éstas. Consideremos los siguientes números:

Expresión	Número de cifras significativas	Precisión
210 s	2	decenas de segundos
210, s	3	segundos
210,0 s	4	décimas de segundo
210,00 s	5	centésimas de segundo.

Por tanto, en las expresiones numéricas anteriores estamos arrojando medidas completamente distintas. Cuando escribimos 210, entendemos que el valor *verdadero* del mensurando tal vez sea 202 s ó 216 s. Mientras que cuando escribimos 210,0 estamos afirmando que probablemente el valor *verdadero* se encuentra (por ejemplo) entre 209,7 y 210,4. Es importante recordar que en las manipulaciones aritméticas realizadas con calculadoras se ignora la segunda regla que hemos dado y se considera que en el número 210 el cero es también una cifra significativa. Esto puede ser muy grave cuando escribimos un resultado tal como 210 000 s, ya que (tal como está escrito) la precisión de la medida son decenas de miles de segundos (!) (ninguno de los ceros a la derecha del 1 son significativos). ¿Cómo podría escribirse una cifra así expresando que nuestra

precisión es, digamos, de cientos de segundos? La mejor solución consiste en expresarlo en notación científica y escribir $2,100 \times 10^5$ s donde dejamos claro que nuestra precisión está en el rango de los centenares de segundos.

Cuando expresamos un resultado experimental el criterio indicado por la GUM² indica que se debe expresar la incertidumbre experimental con dos cifras significativas. La cifra menos significativa de nuestra medida corresponderá al dígito menos significativo de la incertidumbre. Así, por ejemplo, si medimos una longitud de valor medio 1,979 m y cuya incertidumbre es 0,12 m, expresaremos el resultado como $L=(1,98 \pm 0,12)$ m y la cifra menos significativa de nuestra medida sería el último 8, que corresponde a una precisión de cm. Por consiguiente la incertidumbre definirá la precisión con la que expresaremos el resultado numérico de la medida.

4.3.2. Reglas de redondeo

Al expresar nuestro resultado experimental en el párrafo anterior hemos *redondeado* la cifra 1,979 a 1,98 con el objetivo de no incluir un número excesivo de cifras significativas en el resultado final de nuestro proceso de medida. Cuando eliminamos dígitos no significativos de un número, el dígito menos significativo que conservamos debe ser redondeado para no deteriorar la precisión. Para truncar un número dado a un valor con menos cifras significativas, consideramos las cifras a eliminar como una fracción decimal (ponemos una coma decimal a la izquierda de la primera cifra que se elimina). Por ejemplo, si queremos quedarnos con 3 cifras significativas del número 1,39872, la fracción a eliminar respecto a la cifra menos significativa será 0,872. Entonces consideraremos los siguientes casos:

1. Si la fracción remanente es mayor que 0,5 aumentaremos en una unidad la cifra menos significativa.
2. Si esta fracción es menor que 0,5 dejamos la cifra menos significativa sin modificar.
3. Si esta fracción es igual a 0,5 entonces aumentamos en una unidad la cifra menos significativa sólo si ésta es impar.

La razón para la última regla consiste en evitar tendencias sistemáticas en el redondeo que aparecerían si no usásemos esta última convención. Por otra parte, cabe notar que con el uso de los ordenadores para los cálculos y análisis de datos es generalmente admisible retener todos los dígitos disponibles en los cálculos intermedios para hacer el redondeo sobre el resultado final.

1,2345	→	1,23
3,247	→	3,25
1250,0	→	1200
0,0315	→	0,032

²Guide to the Expression of Uncertainty in Measurement, de la ISO.

La regla fundamental en el redondeo de las cifras es importante cuando tomamos diferentes medidas para calcular el valor de otra magnitud que depende de una forma conocida de las magnitudes medidas. Supongamos que deseamos conocer el peso total del equipo científico que se va a embarcar en un cohete. El técnico *A* mide la masa de una parte del equipo cuidadosamente y obtiene 72,39 kg, mientras que el técnico *B* mide con menor precisión el resto del equipo arrojando un resultado de 200 kg. ¿Cuál es el peso total del equipo? Debido a la escasa precisión de la segunda medida estamos obligados a admitir que la cifra de 200 kg puede presentar incertidumbres del orden de 100 kg. Así que deberíamos escribir

$$72,39 \text{ kg} + 200 \text{ kg} = 300 \text{ kg}$$

Vemos por tanto que el operando menos preciso determina, en esta operación, el número de cifras significativas con que debemos expresar el resultado final. Obviamente, en este ejemplo deberíamos sugerir al técnico *B* que repitiera su medida con mayor precisión.

Veremos a continuación la forma de evaluar la incertidumbre de las medidas y cómo propagar ésta al resultado final de una magnitud indirecta.

4.4. Incertidumbres de tipo A y de tipo B

La recomendación INC-1 (1980) del B.I.M.P. después de diversas aproximaciones al problema de la evaluación de las incertidumbres de medida, recomienda su clasificación en dos grupos denominados A y B. Hemos de señalar que esta clasificación responde a la diferente forma de *evaluar* la incertidumbre de la medida. De ninguna manera esta clasificación sustituye a los conceptos *aleatorio* y *sistemático* usados ampliamente en la literatura. En este sentido, puede ocurrir que la contribución de un efecto de carácter aleatorio a la incertidumbre pueda evaluarse como de Tipo A en muchos casos y en otros como de Tipo B. Una situación análoga puede darse también en el caso de contribuciones de carácter sistemático a la incertidumbre de la medida.

La intención que subyace a esta clasificación las incertidumbres en Tipo A y Tipo B es indicar las dos formas diferentes de evaluar las componentes de la incertidumbre de modo independiente de la naturaleza de tales contribuciones. Ambos tipos de componentes de la incertidumbre asumen ciertas distribuciones de probabilidad para el mensurando y ambas se cuantifican a través de varianzas o desviaciones típicas. La diferencia es el origen de dichas distribuciones de probabilidad ya que, como veremos, mientras las incertidumbres de tipo A proceden de frecuencias muestrales, las distribuciones de probabilidad de las incertidumbres de tipo B son supuestas por el observador.

Cuando medimos una magnitud *X* consideraremos que se puede describir el conjunto de los resultados experimentales mediante una *variable aleatoria* con una cierta distribución de probabilidad (tal vez conocida). Con el objetivo de no complicar la notación designaremos de igual forma la magnitud *X* que medimos y la variable aleatoria *X* asociada a la población estadística de todas las medidas experimentales.

4.4.1. Evaluación de incertidumbres de tipo A

En los casos en que la medida de una magnitud pueda ser repetida reiteradamente, es posible derivar una distribución de probabilidad *experimental* u *objetiva* de los valores medidos. Consideraremos que el valor *correcto* de la magnitud que deseamos medir se corresponde al valor esperado o media de la población μ (entiéndase por ésta la media de un número infinito de medidas tomadas en las mismas condiciones). Como en la práctica el número de medidas que podemos realizar estará siempre limitado, usaremos los estimadores de los parámetros de la población introducidos en el capítulo 3 de la presente obra para establecer el valor de la magnitud medida y su incertidumbre. El mejor estimador de la media poblacional μ , como ya hemos visto, es la media muestral \bar{x} para n observaciones x_i independientes en iguales condiciones de medida. Por tanto, el valor óptimo que podemos asignar al mensurando es

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (4.2)$$

Dada la naturaleza aleatoria de la medición física, los valores de las observaciones individuales x_i pueden diferir debido a la existencia de variaciones aleatorias de las magnitudes de influencia. Como sabemos, la distribución de frecuencias observadas nos permite estimar la varianza de la distribución de probabilidad de la población, σ^2 . Para ello consideraremos la varianza muestral s^2 (estimador fiel de la varianza poblacional) dada por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (4.3)$$

La estimación de la varianza s^2 y su raíz cuadrada (la desviación típica muestral) s , caracterizan la variabilidad de los valores x_i observados y su dispersión en torno al valor \bar{x} . A su vez sabemos que la varianza $\sigma^2(\bar{x})$ de la media muestral \bar{x} , está relacionada con la varianza poblacional a través de

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n}. \quad (4.4)$$

Por tanto un buen estimador para la incertidumbre típica asociada a la medida vendrá dada por³

$$u_A(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4.5)$$

Esta incertidumbre $u_A(\bar{x})$ se denomina incertidumbre típica de tipo A.

³Utilizamos en este contexto u para la incertidumbre siguiendo la notación recomendada en la GUM.

Ejemplo 4.1

Si tomamos las siguientes medidas en una balanza:

Masa(kg)	1,232	1,229	1,233	1,235	1,231
	1,239	1,233	1,240	1,230	1,234

Evaluar la contribución de tipo A a la incertidumbre de medida.

En este caso tendremos que la media de la muestra viene dada por

$$\bar{M} = \sum_{i=1}^{10} M_i = 1,2336 \text{ kg}$$

y la desviación típica muestral s será

$$s = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (M_i - \bar{M})^2} = 3,6 \cdot 10^{-3} \text{ kg}$$

por lo que finalmente la incertidumbre típica de tipo A es

$$u_A(M) = \frac{s}{\sqrt{10}} = 1,1 \cdot 10^{-3} \text{ kg}$$

4.4.2. Evaluación de incertidumbres de tipo B

Cuando el valor x de una cierta magnitud X no haya sido obtenido mediante medidas repetidas, deberá establecerse a través de toda la información disponible sobre la variabilidad de esta magnitud X una decisión científica sobre la varianza estimada $u^2(x)$ o bien la incertidumbre típica $u(x)$ que debemos atribuir a esta magnitud. El conjunto de la información que podemos usar para establecer esta incertidumbre puede abarcar:

- Resultados de medidas anteriores.
- Conocimientos previos, pruebas y especificaciones de los materiales e instrumentos utilizados.
- Datos suministrados mediante calibración u otros certificados.

Una componente habitual de tipo B de la incertidumbre de una medida es aquella proveniente de la resolución del instrumento de medida. Supongamos que realizamos una medida con un voltímetro cuya cifra menos significativa es la correspondiente al volt y que posee una resolución de 1 V. Con vistas a la

evaluación de la contribución de esta resolución a la incertidumbre de medida debemos asignar una distribución de probabilidad a los valores de este intervalo de resolución del instrumento ⁴. Es usual utilizar una distribución uniforme para evaluar las incertidumbres de tipo B asociadas a la resolución del instrumento de medida. Es de destacar que la asignación de este tipo de distribución se basa en la suposición de que todos los valores en el intervalo de anchura a son igualmente probables. En el caso de que dispongamos de información sobre el comportamiento de esta magnitud o del instrumento de medida que permita refinar dicha hipótesis, la distribución de probabilidad utilizada para la evaluación de las incertidumbres de tipo B puede ser, según los casos, triangular, normal truncada o de cualquier otra naturaleza. En particular, en un instrumento analógico a veces es posible precisar que el centro del intervalo es más probable y seleccionar una distribución triangular. Sin embargo, en este caso consideraremos que dado un valor arrojado por nuestro instrumento (siguiendo las especificaciones del fabricante) el valor de la magnitud que queremos medir se encontrará en un intervalo de anchura total 1 V con una probabilidad idealmente uniforme (Figura 4.4).

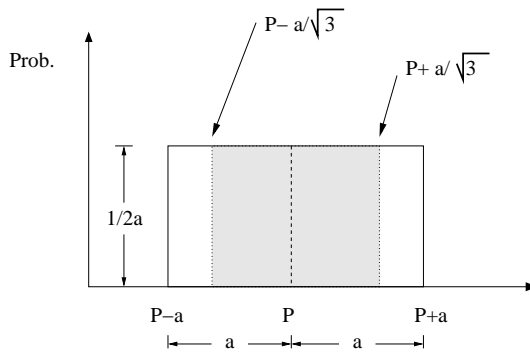


Figura 4.4: Distribución de probabilidad cuadrada asignada a una magnitud con un intervalo de semianchura a .

En este caso debemos asignar una incertidumbre de tipo B a la medida igual a la desviación típica de la distribución de probabilidad teórica que hemos asociado a nuestro aparato de medida. Puesto que si la distribución de probabilidad es cuadrada de semianchura a , la desviación típica es $a/\sqrt{3}$, en nuestro caso la incertidumbre típica es de $1/\sqrt{12}\text{ V}$.

En general, la desviación típica asociada a una distribución de probabilidad uniforme de anchura total $\Delta a = a_+ - a_-$ es igual a $\Delta a/\sqrt{12}$. Siendo a_+ y a_- los valores máximo y mínimo que puede adoptar la magnitud X que hemos medido. Al tener esta magnitud una distribución uniforme, el valor esperado de la magnitud es $x = (a_+ + a_-)/2$ y la varianza asociada será

⁴A diferencia de lo que sucede con las incertidumbres de tipo A, en cuya evaluación intervienen distribuciones experimentales de frecuencias, en el caso de las tipo B las distribuciones de probabilidad utilizadas en la estimación son normalmente supuestas por el observador

$$u_B^2(x) = (a_+ - a_-)^2/12 \quad (4.6)$$

$$u_B(x) = (a_+ - a_-)/\sqrt{12} \quad (4.7)$$

Esta incertidumbre $u_B(x)$ corresponde a la incertidumbre típica de tipo B asociada a la resolución del instrumento de medida.

Hemos de recordar que la distribución de probabilidad usada para la estimación de la incertidumbre de tipo B se adecuará a la magnitud en cuestión pudiendo ser, según los casos, una distribución cuadrada, triangular, normal o de otra naturaleza, según nuestro conocimiento o experiencia sobre el comportamiento de esta magnitud y/o de nuestro instrumento de medida. La importancia de esta elección puede ejemplificarse mediante un cálculo análogo al realizado para la distribución cuadrada que revela que una distribución triangular de semianchura a arroja una incertidumbre de tipo B igual a $u_B = \frac{a}{\sqrt{12}}$, inferior en un factor $\sqrt{2}$ a la que se obtendría de suponer una distribución cuadrada. En caso de suponer una distribución normal pueden adoptarse diferentes convenios, p. ej. hacer coincidir su desviación típica con la semianchura del intervalo en que se considera que puede hallarse la medida real o truncarla de algún modo (multiplicándola por una función cuadrada, restándole una constante y renormalizándola...) para que valga cero fuera de ese intervalo.

Finalmente, suponiendo que ambos tipos de incertidumbres son estadísticamente independientes, el valor total de la incertidumbre de es:

$$u^2 = u_A^2 + u_B^2. \quad (4.8)$$

Esta incertidumbre se denomina incertidumbre combinada y será objeto de análisis detallado en la sección siguiente.

Ejemplo 4.2

Considerando el ejemplo anterior de la medida de una masa, habíamos obtenido como valor medio de 10 medidas 1,2336 kg con una incertidumbre típica de tipo A de $u(\bar{M}) = 1,1 \cdot 10^{-3}$ kg. Supongamos ahora que el fabricante de la báscula nos indica que la exactitud de la medida ha sido establecida en el 1% para masas inferiores a 5 kg. Podremos evaluar cuál es la contribución de incertidumbre de tipo B. Para ello tendremos en cuenta que nuestra medida puede desviarse (por encima o debajo del valor *correcto*) por una cantidad

$$\Delta\bar{M} = 0,01 \times 1,2336 \text{ kg} = 1,2 \cdot 10^{-2} \text{ kg}$$

como consecuencia de esta corrección, asignamos una incertidumbre de tipo B a la medida

$$u_B(\Delta\bar{M}) = 1,2/\sqrt{3} \cdot 10^{-2} \text{ kg} = 6,9 \cdot 10^{-3} \text{ kg}.$$

Puede obtenerse la incertidumbre típica combinada de esta medida como

$$u_c^2 = u_A^2(\bar{M}) + u_B^2(\Delta\bar{M})$$

$$u_c^2 = ((1,1)^2 + (6,9)^2)(10^{-3} \text{ kg})^2,$$

o lo que es lo mismo:

$$u_c = 7,0 \cdot 10^{-3} \text{ kg}.$$

Como vemos en el ejemplo anterior las incertidumbres de tipo A y B se tratan de modo uniforme en la propagación de las incertidumbres para obtener el valor final de la incertidumbre combinada. Esta propagación homogénea o uniforme es preferible a otras elecciones que se pueden encontrar en la literatura y sigue la recomendación INC-1 (1980) del CIPM.

Ejemplo 4.3

Una de las objeciones que aparece cuando se aplica el cálculo de incertidumbres de modo poco riguroso es la medida en que debemos de tener en cuenta la contribución de la resolución instrumental. Un ejemplo típico sería el siguiente: supongamos que hemos medido 10 veces el voltaje de un circuito con un polímetro cuya resolución es 1 V, obteniéndose la siguiente tablas de medidas

Voltaje (V)	11	12	12	12	12
	12	12	12	12	12

Si realizamos el análisis de la incertidumbre de tipo A, obtendremos que ésta tiene un valor $u_A(\bar{V}) = 0,1$ V. Este valor es sustancialmente menor que la incertidumbre real que debemos asignar a nuestra medida. Considerando la contribución de tipo B, ésta será $u_B(\Delta\bar{V}) = 1 \text{ V}/\sqrt{12}$ debido a que asignamos una probabilidad uniforme a la medida en un intervalo cuyo ancho total es 1 V. De aquí obtenemos $u_B(\Delta\bar{V}) = 0,28$ V. De este modo la incertidumbre típica combinada para esta serie de medidas será

$$u_c = \sqrt{[u(\bar{V})]^2 + [u(\Delta\bar{V})]^2} = 0,28 \text{ V}$$

resultado que está mucho más de acuerdo con nuestra percepción de la incertidumbre. Por tanto podremos reflejar nuestra medida como

$$V = (11,90 \pm 0,28) \text{ V}.$$

Es necesario señalar que a la hora de identificar las fuentes de incertidumbre y calcular su contribución no debe tenerse en cuenta la misma contribución dos veces como tipo A y tipo B simultáneamente. Las fuentes de incertidumbre deben de contabilizarse sólo una vez y de modo exclusivo como de tipo A o de tipo B.

4.5. Incertidumbre combinada. Propagación de incertidumbres.

Cuando el resultado de una medición se obtiene a partir de los valores de otras magnitudes, la incertidumbre típica de este resultado se denomina incertidumbre típica combinada, y se denota por u_c . Se determina a partir de las incertidumbres asociadas a cada una de las magnitudes que se usan en el cálculo y constituye la desviación típica estimada de la magnitud calculada indirectamente.

Consideremos una magnitud o mensurando Y que no se ha medido directamente, sino que está determinada a través de otras N magnitudes $\{X_i\}_{i=1}^N$ mediante la relación funcional

$$Y = f(X_1, X_2, \dots, X_N). \quad (4.9)$$

Las magnitudes X_1, X_2, \dots, X_N se suelen denominar *magnitudes de entrada* de la ecuación o relación funcional anterior, mientras que Y suele designarse como la *magnitud de salida*. A su vez, X_1, X_2, \dots, X_N pueden ser consideradas también como mensurandos y depender de otras magnitudes.

Consideremos que la función $Y = f(X_1, X_2, \dots, X_N)$ se puede desarrollar en serie de Taylor y supongamos que las magnitudes X_1, X_2, \dots, X_N sólo tienen componentes de incertidumbre de tipo A, obtenidas mediante la repetición de su medida p veces. Entonces podremos escribir

$$y^i = f(x_1^i, x_2^i, \dots, x_N^i) \quad i = 1, \dots, p, \quad (4.10)$$

donde para cada una de las magnitudes de entrada podremos estimar su valor medio mediante las expresiones

$$\bar{x}_j = \frac{1}{p} \sum_{i=1}^p x_j^i \quad j = 1, \dots, N. \quad (4.11)$$

A su vez consideraremos como el mejor estimador de la magnitud de entrada x_j el valor de \bar{x}_j . La incertidumbre de tipo A de cada una de las magnitudes de entrada puede ser obtenida mediante

$$u^2(\bar{x}_j) = \frac{1}{p(p-1)} \sum_{i=1}^p (x_j^i - \bar{x}_j)^2 \quad j = 1, \dots, N. \quad (4.12)$$

Al desarrollar en serie la función f , podremos hacer la aproximación a primer orden

$$y^i = f(x_1^i, x_2^i, \dots, x_N^i) \simeq f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) + \sum_{j=1}^N \frac{\partial f}{\partial x_j} (x_j^i - \bar{x}_j), \quad (4.13)$$

donde denotamos como $\frac{\partial f}{\partial x_j}$ a la derivada parcial $\frac{\partial f}{\partial X_j}$ evaluada en el punto $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$. El valor esperado para la magnitud Y en esta serie de medidas que denotamos por \bar{y} corresponderá a

$$\bar{y} = \frac{1}{p} \sum_{i=1}^p y^i = \frac{1}{p} \sum_{i=1}^p f(x_1^i, x_2^i, \dots, x_N^i) \quad (4.14)$$

$$\bar{y} \simeq \frac{1}{p} \sum_{i=1}^p \left[f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) + \sum_{j=1}^N \frac{\partial f}{\partial x_j} (x_j^i - \bar{x}_j) \right] = \quad (4.15)$$

$$= \frac{1}{p} \sum_{i=1}^p f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) + \sum_{j=1}^N \frac{\partial f}{\partial x_j} \sum_{i=1}^p (x_j^i - \bar{x}_j). \quad (4.16)$$

Ahora bien, teniendo en cuenta la definición de \bar{x}_j , se cumple que

$$\sum_{i=1}^p (x_j^i - \bar{x}_j) = 0, \quad (4.17)$$

por lo que a primer orden se puede afirmar que

$$\bar{y} \simeq f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) \quad (4.18)$$

Podremos calcular igualmente la incertidumbre de tipo A para la magnitud Y basándonos en los supuestos anteriores. De nuevo consideraremos que el mejor estimador de Y es el valor medio \bar{y} con lo que su incertidumbre es

$$\begin{aligned} u^2(\bar{y}) &= \frac{1}{p(p-1)} \sum_{i=1}^p (y^i - \bar{y})^2 = \frac{1}{p(p-1)} \sum_{i=1}^p \left(\sum_{j=1}^N \frac{\partial f}{\partial x_j} (x_j^i - \bar{x}_j) \right)^2 = \\ &= \frac{1}{p(p-1)} \sum_{i=1}^p \left[\sum_{j=1}^N \left(\frac{\partial f}{\partial x_j} \right)^2 (x_j^i - \bar{x}_j)^2 + \right. \\ &\quad \left. + 2 \sum_{j=1}^{N-1} \sum_{k=j+1}^N \frac{\partial f}{\partial x_j} \frac{\partial f}{\partial x_k} (x_j^i - \bar{x}_j)(x_k^i - \bar{x}_k) \right] \quad (4.19) \end{aligned}$$

Desarrollando la expresión anterior

$$u^2(\bar{y}) = \sum_{j=1}^N \left[\left(\frac{\partial f}{\partial x_j} \right)^2 \frac{1}{p(p-1)} \sum_{i=1}^p (x_j^i - \bar{x}_j)^2 \right] + \quad (4.20)$$

$$+ 2 \sum_{j=1}^{N-1} \sum_{k=j+1}^N \left[\left(\frac{\partial f}{\partial x_j} \right) \left(\frac{\partial f}{\partial x_k} \right) \frac{1}{p(p-1)} \sum_{i=1}^p (x_j^i - \bar{x}_j)(x_k^i - \bar{x}_k) \right] \quad (4.21)$$

y teniendo en cuenta que

$$\begin{aligned} u^2(\bar{x}_j) &= \frac{1}{p(p-1)} \sum_{i=1}^p (x_j^i - \bar{x}_j)^2 \quad y \\ \text{cov}(\bar{x}_j, \bar{x}_k) &= \frac{1}{p(p-1)} \sum_{i=1}^p (x_j^i - \bar{x}_j)(x_k^i - \bar{x}_k) \quad (4.22) \end{aligned}$$

la expresión final para la incertidumbre de la magnitud Y desarrollando a primer orden es

$$u^2(\bar{y}) = \sum_{j=1}^N \left(\frac{\partial f}{\partial x_j} \right)^2 u^2(\bar{x}_j) + 2 \sum_{j=1}^{N-1} \sum_{k=j+1}^N \left(\frac{\partial f}{\partial x_j} \right) \left(\frac{\partial f}{\partial x_k} \right) \text{cov}(\bar{x}_j, \bar{x}_k), \quad (4.23)$$

que ya había sido obtenida como ejercicio en el primer capítulo de esta obra. Como es evidente, la deducción se puede repetir considerando distribuciones de probabilidad supuestas en lugar de distribuciones de frecuencia, por lo que la fórmula es válida también para incertidumbres de tipo B y para la incertidumbre combinada.

Consideraremos a continuación dos posibilidades a la hora de estimar la incertidumbre combinada:

- a) Todas las magnitudes de entrada son *independientes*, es decir no están estadísticamente correlacionadas.
- b) Existe alguna relación entre dos o más magnitudes de entrada. En este caso las magnitudes o son interdependientes o bien, están correlacionadas.

4.5.1. Magnitudes de entrada no correlacionadas.

En este caso las N magnitudes de entrada no tienen ninguna relación entre sí y pueden ser consideradas estadísticamente como variables independientes. En este caso se puede considerar que

$$\text{cov}(x_j, x_k) = 0, \quad (4.24)$$

y por tanto la expresión de la incertidumbre combinada para la magnitud Y es

$$u_c^2(y) = \sum_{j=1}^N \left(\frac{\partial f}{\partial x_j} \right)^2 u^2(x_j). \quad (4.25)$$

Esta expresión se conoce como *ley de propagación de la incertidumbre* para magnitudes no correlacionadas. En ella, la incertidumbre combinada estimada para y es la suma de términos asociados a la incertidumbre estimada de cada variable de entrada x_j . Las derivadas parciales

$$c_j = \left(\frac{\partial f}{\partial x_j} \right). \quad (4.26)$$

se conocen también como coeficientes de sensibilidad, y establecen la influencia que la incertidumbre de x_j tiene sobre la incertidumbre combinada. Con esta notación es habitual escribir

$$u_c^2(y) = \sum_{j=1}^N [c_j u(x_j)]^2, \quad (4.27)$$

y a veces a la componente asociada a la magnitud x_j de la incertidumbre combinada se expresa como

$$u_j(y) = c_j u(x_j), \quad (4.28)$$

por lo que

$$u_c^2(y) = \sum_{j=1}^N u_j^2(y). \quad (4.29)$$

Ejemplo 4.4

En el Ejemplo 4.2 se proporcionó una expresión para la incertidumbre total de una masa con dos contribuciones: una de tipo A y otra de tipo B. En el marco de la *ley de propagación de incertidumbres* podemos pensar que la masa que estamos midiendo se puede escribir como

$$M = \bar{M} + \Delta\bar{M}$$

Por tanto las derivadas parciales correspondientes serán

$$c_{\bar{M}} = \frac{\partial M}{\partial \bar{M}} = 1$$

$$c_{\Delta\bar{M}} = \frac{\partial M}{\partial \Delta\bar{M}} = 1$$

por lo que se puede escribir la incertidumbre combinada típica para la magnitud M como

$$u_c^2(M) = u^2(\bar{M}) + u^2(\Delta\bar{M})$$

que es la expresión que usamos en el Ejemplo 4.2.

En el caso de que la función f que expresa la dependencia funcional sea complicada, o sea necesaria su evaluación mediante métodos numéricos, podemos considerar que la componente de la incertidumbre $u_j(y)$ puede ser calculada como

$$u_j(y) = \frac{1}{2} [f(x_1, x_2, \dots, x_j + u(x_j), \dots, x_N) - f(x_1, x_2, \dots, x_j - u(x_j), \dots, x_N)], \quad (4.30)$$

es decir, corresponde a la mitad de la variación de y debida a tomar variaciones en x_j de $+u(x_j)$ y de $-u(x_j)$. En el caso de que la aproximación a primer orden en las derivadas parciales de la *ley de propagación de incertidumbres* fuese insuficientemente precisa (por ejemplo si las derivadas parciales primeras c_j fuesen nulas), debe considerarse el siguiente orden del desarrollo en serie de Taylor que dará lugar a la expresión

$$u_c^2(y) = \sum_{j=1}^N \left(\frac{\partial f}{\partial x_j} \right)^2 u^2(x_j) + \sum_{i=1}^N \sum_{j=1}^N \left[\frac{1}{2} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 + \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial^3 f}{\partial x_i \partial^2 x_j} \right) \right] u^2(x_i) u^2(x_j) \quad (4.31)$$

Ejemplo 4.5

En un experimento de calorimetría se usa una resistencia para disipar calor en una muestra. Se mide el voltaje V que se aplica a la resistencia, así como su temperatura y se pretende estimar la potencia disipada por ésta. Considerando que la resistencia varía linealmente con la temperatura, esta potencia seguirá una expresión de la forma

$$P = \frac{V^2}{R_0[1 + \alpha(T - T_0)]},$$

donde R_0 es la resistencia para una temperatura de referencia T_0 y α es el coeficiente lineal de temperatura. Para la obtención de la incertidumbre típica combinada debemos calcular las siguientes derivadas parciales

$$\begin{aligned} \frac{\partial P}{\partial V} &= \frac{2V}{R_0[1 + \alpha(T - T_0)]} ; & \frac{\partial P}{\partial R_0} &= -\frac{V^2}{R_0^2[1 + \alpha(T - T_0)]} \\ \frac{\partial P}{\partial \alpha} &= -\frac{V^2(T - T_0)}{R_0[1 + \alpha(T - T_0)]^2} ; & \frac{\partial P}{\partial T} &= -\frac{V^2\alpha}{R_0[1 + \alpha(T - T_0)]^2}, \end{aligned}$$

donde consideramos que la incertidumbre de T_0 tiene una contribución despreciable a la incertidumbre de P . La incertidumbre combinada para la magnitud P será

$$u_c^2(P) = \left(\frac{\partial P}{\partial V} \right)^2 u^2(V) + \left(\frac{\partial P}{\partial R_0} \right)^2 u^2(R_0) + \left(\frac{\partial P}{\partial \alpha} \right)^2 u^2(\alpha) + \left(\frac{\partial P}{\partial T} \right)^2 u^2(T),$$

siendo $u(V)$, $u(R_0)$, $u(\alpha)$ y $u(T)$ las incertidumbres típicas asociadas a las magnitudes V , R_0 , α y T respectivamente, las cuales podrán a su vez tener diversas componentes bien de tipo A o de tipo B.

En el caso de que la magnitud Y tenga una dependencia funcional de la forma $Y = k \cdot X_1^{p_1} \cdot X_2^{p_2} \cdot \dots \cdot X_N^{p_N}$, donde los exponentes p_1, p_2, \dots, p_N son números conocidos (positivos o negativos) de incertidumbre despreciable, entonces la incertidumbre combinada de la magnitud Y se puede expresar como

$$\left(\frac{u_c(y)}{y}\right)^2 = \left(p_1 \frac{u(x_1)}{x_1}\right)^2 + \left(p_2 \frac{u(x_2)}{x_2}\right)^2 \dots + \left(p_N \frac{u(x_N)}{x_N}\right)^2, \quad (4.32)$$

donde al cociente $\left(\frac{u_c(y)}{y}\right)^2$ se conoce como varianza combinada *relativa* de la magnitud Y . También se suele designar a los cocientes $u(y)/|y|$, $u(x_1)/|x_1|$, \dots , $u(x_N)/|x_N|$ como las incertidumbres relativas de las magnitudes Y , X_1 , \dots , X_N respectivamente. Vemos que para esta dependencia funcional la varianza relativa de y es suma de las varianzas relativas de las magnitudes x_i por el cuadrado de sus potencias respectivas. En el caso particular en que $p_i = \pm 1$ entonces la varianza relativa de y es simplemente la suma de las varianzas relativas de las magnitudes de entrada.

4.5.2. Magnitudes de entrada correlacionadas.

En el caso de que las magnitudes de entrada X_i correspondan a variables aleatorias correlacionadas estadísticamente, será necesario tener en cuenta su correlación a través de las covarianzas entre diferentes magnitudes. Por ello debemos usar la forma general de la *ley de propagación para las incertidumbres* recogida en la Ec. (4.23). A veces se usa en lugar de la covarianza el coeficiente de correlación

$$\rho(x_j, x_k) = \frac{\text{cov}(x_j, x_k)}{u(x_j)u(x_k)} \quad (4.33)$$

con lo que la expresión para la incertidumbre combinada será

$$u^2(y) = \sum_{j=1}^N \left(\frac{\partial f}{\partial x_j}\right)^2 u^2(x_j) + 2 \sum_{j=1}^{N-1} \sum_{k=j+1}^N \left(\frac{\partial f}{\partial x_j}\right) \left(\frac{\partial f}{\partial x_k}\right) \rho(x_j, x_k) u(x_j) u(x_k), \quad (4.34)$$

donde debemos recordar que se cumple que los coeficientes de correlación están acotados $-1 \leq \rho(x_j, x_k) \leq 1$. La covarianza entre las magnitudes de entrada no puede ser ignorada cuando se calcula la incertidumbre combinada, y en el caso de que podamos tomar p pares de observaciones independientes y simultáneas, las covarianzas asociadas serán

$$\text{cov}(x_j, x_k) = \frac{1}{p-1} \sum_{i=1}^p (x_j^i - \bar{x}_j)(x_k^i - \bar{x}_k). \quad (4.35)$$

Sin embargo, tal y como hicimos al describir las incertidumbres de tipo A, es habitual usar como estimación de las magnitudes medidas los valores medios de las series de datos, \bar{x}_i , de tal manera que

$$u^2(\bar{x}_i) = \frac{1}{p(p-1)} \sum_{j=1}^p (x_i^j - \bar{x}_i)^2 \quad (4.36)$$

$$\text{cov}(\bar{x}_i, \bar{x}_j) = \frac{1}{p(p-1)} \sum_{k=1}^p (x_i^k - \bar{x}_i)(x_j^k - \bar{x}_j), \quad (4.37)$$

y la expresión de la incertidumbre combinada tomando como estimador de Y el valor \bar{y} es:

$$u_c^2(\bar{y}) = \sum_{j=1}^N \left(\frac{\partial f}{\partial x_j} \right)^2 u^2(\bar{x}_j) + 2 \sum_{j=1}^{N-1} \sum_{k=j+1}^N \left(\frac{\partial f}{\partial x_j} \right) \left(\frac{\partial f}{\partial x_k} \right) \text{cov}(\bar{x}_j, \bar{x}_k). \quad (4.38)$$

Se puede también escribir la fórmula de la propagación de las incertidumbres mediante el uso de la matriz de covarianza definida como

$$\mathbf{Cov} = \begin{pmatrix} u^2(x_1) & \text{cov}(x_1, x_2) & \dots & \text{cov}(x_1, x_N) \\ \text{cov}(x_2, x_1) & u^2(x_2) & \dots & \text{cov}(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(x_N, x_1) & \text{cov}(x_N, x_2) & \dots & u^2(x_N) \end{pmatrix} \quad (4.39)$$

Esta matriz es, por definición, simétrica. Además en el caso de magnitudes no correlacionadas los elementos fuera de la diagonal son nulos y la matriz de covarianzas tiene forma diagonal. Si definimos el vector columna gradiente de la función $f(X_1, X_2, \dots, X_N)$ evaluado en los valores esperados de estas magnitudes,

$$\vec{\nabla} f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{pmatrix}$$

Entonces la incertidumbre combinada puede ser escrita de la siguiente manera en forma matricial

$$u_c^2(y) = [\vec{\nabla} f]^T \cdot \mathbf{Cov} \cdot [\vec{\nabla} f] \quad (4.40)$$

En el caso extremo de que todas las magnitudes estén correlacionadas con coeficientes de correlación $\rho(x_i, x_j) = +1$ entonces la propagación de la incertidumbre se reduce a

$$u_c^2(y) = \left[\sum_{j=1}^N \left(\frac{\partial f}{\partial x_j} \right) u(x_j) \right]^2 \quad (4.41)$$

es decir, la incertidumbre típica combinada será la suma lineal de términos que representan la variación de la magnitud de salida al variar cada x_j por un valor

igual al de su incertidumbre típica $u(x_j)$.⁵

Ejemplo 4.6

Se pretende estimar el valor de una resistencia eléctrica midiendo la diferencia de potencial entre sus bornes y la corriente eléctrica que circula a través de ésta. Se realizan 10 pares de medidas independientes que arrojan los valores

Voltaje (V)	2,323	2,325	2,320	2,331	2,322
	2,330	2,327	2,323	2,329	2,325
Intensidad (mA)	1,451	1,452	1,449	1,457	1,452
	1,463	1,453	1,451	1,459	1,454

De aquí obtenemos los valores medios para las magnitudes medidas V e I ,

$$\bar{V} = 2,3255 \text{ V} \quad \bar{I} = 1,4541 \text{ mA.}$$

El valor estimado de la resistencia $\bar{R} = \frac{\bar{V}}{\bar{I}}$ arroja el valor $\bar{R} = 1\,599,27 \, \Omega$. Las incertidumbres de tipo A asociadas serán

$$u^2(\bar{V}) = \frac{1}{9 \cdot 10} \sum_{i=1}^{10} (V^i - \bar{V})^2 \quad u^2(\bar{I}) = \frac{1}{9 \cdot 10} \sum_{i=1}^{10} (I^i - \bar{I})^2$$

$$u(\bar{V}) = 1,2 \cdot 10^{-3} \text{ V} \quad u(\bar{I}) = 1,4 \cdot 10^{-3} \text{ mA.}$$

La contribución de las incertidumbres de tipo B debido a la resolución instrumental son pequeñas frente a las anteriores. Por otra parte, podremos calcular las covarianzas y la correlación de las magnitudes medidas

$$\text{cov}(\bar{V}, \bar{I}) = \frac{1}{9 \cdot 10} \sum_{i=1}^{10} (V^i - \bar{V})(I^i - \bar{I})$$

$$\text{cov}(\bar{V}, \bar{I}) = 1,37 \cdot 10^{-6} \text{ V} \cdot \text{mA},$$

de lo que obtenemos un coeficiente de correlación

$$r(\bar{V}, \bar{I}) = \frac{\text{cov}(\bar{V}, \bar{I})}{u(\bar{V})u(\bar{I})} = 0,87.$$

Para calcular la resistencia debemos usar la fórmula $R = \frac{V}{I}$ de la cual obtenemos las siguientes derivadas parciales

$$\frac{\partial R}{\partial V} = \frac{1}{I} \quad \frac{\partial R}{\partial I} = -\frac{V}{I^2},$$

de donde obtenemos la incertidumbre combinada mediante la fórmula usual de propagación de incertidumbres,

⁵Esta expresión no debe confundirse con la llamada ley de propagación de *errores*, ya que en (4.41) participan incertidumbres típicas.

$$u_c^2(\bar{R}) = \left(\frac{1}{\bar{I}}\right)^2 u^2(\bar{V}) + \left(\frac{\bar{V}}{\bar{I}^2}\right)^2 u^2(\bar{I}) + 2\left(\frac{1}{\bar{I}}\right)\left(-\frac{\bar{V}}{\bar{I}^2}\right)\text{cov}(\bar{V}, \bar{I}),$$

cuyo resultado es

$$u(\bar{R}) = 0,90 \Omega.$$

Por tanto, el resultado experimental de nuestras medidas puede expresarse de la forma

$$R = (1\,599,27 \pm 0,90) \Omega.$$

4.5.3. Incertidumbre expandida. Factor de cobertura.

Es también común expresar la incertidumbre final U de una magnitud y como un múltiplo entero de la incertidumbre típica combinada $u_c(y)$, esto es $1 \times u_c(y)$, $2 \times u_c(y)$, $3 \times u_c(y)$, etc. Estas nuevas incertidumbres corresponderán a la amplitud de intervalos con un nivel de confianza tanto mayor cuanto lo sea el factor por el que multiplicamos la incertidumbre típica. Este factor multiplicativo se conoce como *factor de cobertura*. Para una magnitud cuya distribución de probabilidad sea normal los factores 1, 2 y 3 se corresponden respectivamente a niveles de confianza del 68,27%, 95,45% y 99,73%. De esta forma expresaríamos nuestro resultado como

$$y - U \leq Y \leq y + U \tag{4.42}$$

incluyendo de modo explícito cuál es el valor del factor de cobertura k usado de modo que $U = k * u_c$. El valor final U usado en la expresión final del resultado experimental se conoce como incertidumbre expandida. En muchos casos se suele considerar que una mejor aproximación a la incertidumbre expandida se obtiene considerando que la magnitud de salida sigue aproximadamente una t de Student con un número de grados de libertad ν , de modo que:

$$U = t_{\alpha/2;\nu} \cdot u_c, \tag{4.43}$$

siendo u_c la incertidumbre típica combinada asociada a la magnitud de salida y $t_{\alpha/2;\nu}$ el percentil de una t de Student con ν grados de libertad que cumple que $p(|t| > t_{\alpha/2;\nu}) = \alpha$.

Cuando la incertidumbre típica combinada de la magnitud de salida sea suma de varias contribuciones a través de los coeficientes de sensibilidad c_i y las incertidumbre típicas $u(x_i)$,

$$u_c^2(y) = c_1^2 u^2(x_1) + c_2^2 u^2(x_2) + \dots + c_n^2 u^2(x_n), \tag{4.44}$$

se define el número de grados efectivos de libertad de Welch-Satterthwaite mediante la relación

$$\frac{u_c^2(y)}{\nu_{eff}} = \frac{c_1^2 u^2(x_1)}{\nu_1} + \frac{c_2^2 u^2(x_2)}{\nu_2} + \dots + \frac{c_n^2 u^2(x_n)}{\nu_n}, \quad (4.45)$$

con lo que

$$\nu_{eff} = \frac{u_c^2(y)}{\frac{c_1^2 u^2(x_1)}{\nu_1} + \frac{c_2^2 u^2(x_2)}{\nu_2} + \dots + \frac{c_n^2 u^2(x_n)}{\nu_n}}. \quad (4.46)$$

Obsérvese que, como consecuencia de esta definición, $\nu_{eff} \leq \sum_{i=1}^n \nu_i$. Además la definición de ν_{eff} implica un resultado analítico no entero que (para su aplicación) supone redondear ν_{eff} al número entero positivo más próximo. Recordemos, por otro lado, que los grados de libertad ν_i corresponden a las componentes de incertidumbre de las magnitudes de entrada. Si la estimación de una componente de incertidumbre es de tipo A con n repeticiones asumiremos normalmente $\nu_i = n - 1$. En el caso de incertidumbres de tipo B es más difícil evaluar el número de grados de libertad efectivos asociados a esta componente de incertidumbre y suele considerarse

$$\nu_i \simeq \frac{1}{2} \frac{u^2(x_i)}{\sigma^2[u(x_i)]}, \quad (4.47)$$

donde $\sigma^2[u(x_i)]$ es la varianza de esta incertidumbre o el grado de variabilidad que es razonable asumirle a nuestra estimación de incertidumbre. Esta estimación está sujeta a nuestro juicio científico basado en la información disponible respecto a la componente de incertidumbre en cuestión.

4.6. Cuestiones y problemas

4.1 Se han realizado cinco medidas del diámetro de una esfera con un calibre que aprecia la centésima de milímetro y los resultados se muestran en la tabla adjunta:

d (mm) 0.10 0.11 0.09 0.10 0.11

Calcular diámetro de la esfera y su incertidumbre. Calcula a su vez el diámetro de la esfera y su incertidumbre.

4.2 Dada una magnitud $y = y(x_1, x_2)$, en la ley de propagación de las incertidumbres $\sigma^2(y) = \left(\frac{\partial y}{\partial x_1}\right)^2 \sigma^2(x_1) + \left(\frac{\partial y}{\partial x_2}\right)^2 \sigma^2(x_2)$. ¿Es esta expresión válida cualesquiera que sean las distribuciones de probabilidad de x_1 y x_2 ? ¿Cuál es el error relativo de la incertidumbre de y si las magnitudes de entrada x_1 y x_2 no son independientes?

4.3 Consideremos las siguientes medidas del espesor de una lámina metálica:

e (mm) 0,13 0,10 0,09 0,14 0,13

¿Cuál es el espesor y la incertidumbre que debemos asociar a la lámina?

4.4 Se mide con un polímetro una resistencia diez veces arrojando los resultados $\{101; 102; 103; 110; 108; 102; 109; 100; 101; 102\}$ en $k\Omega$. Si el fabricante del polímetro nos indica que la lectura en este rango de resistencia puede desviarse del valor correcto hasta $\pm 5\%$, calcular la incertidumbre de la medida.

4.5 En un experimento de calorimetría se usa una resistencia para disipar calor. Si P es la potencia disipada por la resistencia y V el voltaje aplicado

$$P = \frac{V^2}{R_0[1 + \alpha(T - T_0)]}$$

donde R_0 es la resistencia para la temperatura de referencia $T_0=298K$ y α es el coeficiente lineal de temperatura. Se realizan cinco medidas de V y T obteniéndose

$\{(10, 1V; 343K)(10, 3V; 350K)(10, 0V; 349K)(10, 1V; 350K)(10, 2V; 349K)\}$.

Si $R_0= 100 k\Omega$ con una incertidumbre relativa del 2% y $\alpha = 0,0012 K^{-1}$ con una incertidumbre relativa del 10%, calcular P y su incertidumbre típica.

4.6 Dos magnitudes estadísticamente independientes x_1 y x_2 se miden para determinar el valor de una tercera magnitud $y = \ln x_1 + x_2$. ¿Cuál es la relación entre las incertidumbres típicas de las magnitudes x_1 , x_2 e y ?

4.7 Se mide el número de desintegraciones en un minuto de una fuente radiactiva, obteniéndose un contejo de $n = 100$. Calcular la incertidumbre de esta medida. ¿Se trata de una incertidumbre de tipo A o de tipo B?

4.8 Se realizan cien medidas de una magnitud que sigue una distribución de probabilidad normal con desviación típica $\sigma = 2$. Si la media de las cien medidas es 12,1, ¿qué incertidumbre típica asignarías a la medida?

4.9 Se mide la longitud de una varilla metálica a diferentes temperaturas obteniéndose los siguientes resultados

L (mm)	17	25	27	39	41	48
$T - T_0$ (K)	0.0	20.0	42.1	79.0	88.3	100.2

Considerando que $L = L_0 + \alpha(T - T_0)$, determinar L_0 y α mediante un ajuste por mínimos cuadrados si la incertidumbre relativa de L se estima entorno al 10% y la incertidumbre de $T - T_0$ es despreciable.

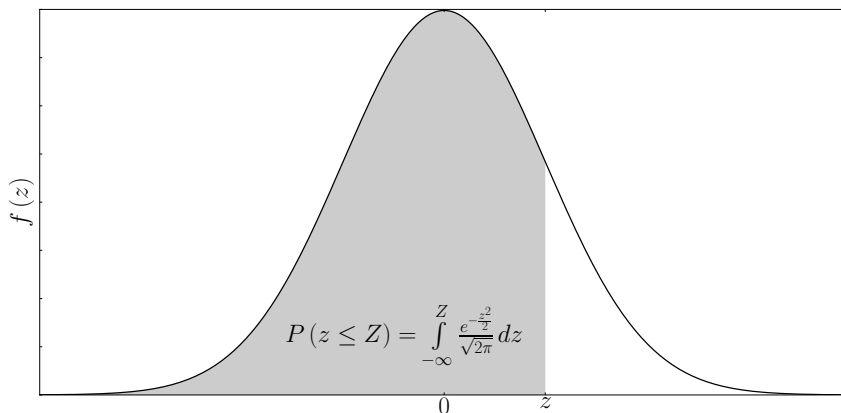
- Determinar la incertidumbre de L_0 y α .
 - Determinar la covarianza $cov(L_0, \alpha)$.
 - Representar los datos con su barra de incertidumbre junto al ajuste.
 - Obtener el valor esperado del modelo lineal para un incremento de temperatura de 200 K y su incertidumbre.
- 4.10 La densidad del aire puede obtenerse (considerando un modelo de gas ideal) del valor de densidad ρ_0 a una temperatura y presión de referencia (T_0, P_0) , para una temperatura y presión arbitraria (T, P) , mediante $\rho = \rho_0 \times \frac{(273.2+T_0) P}{(273.2+T) P_0}$.
- Un laboratorio de referencia nos indica que la densidad del aire es 1.293 kg/m³ para $P_0=101.3$ kPa y $T_0= 20^0\text{C}$, con una incertidumbre de 0.018 kg/m³ con $k=2$. Considerando que nuestra medida de temperatura está afectada de una incertidumbre relativa del 2% y la medida de presión tiene una incertidumbre relativa del 1%, calcula la incertidumbre de la densidad como función de (T, P) .
- 4.11 La tasa de dosis de una fuente puntual se puede considerar como inversa del cuadrado de la distancia d a la misma. Considerando una incertidumbre típica de 1 mm en la determinación de la distancia d , calcula la distancia a la cual la incertidumbre relativa en la tasa de dosis es del 1%.
- 4.12 La actividad de una fuente radiactiva de ^{60}Co disminuye de acuerdo a la ley exponencial de desintegración radiactiva $A(t) = A_0 e^{-\lambda \Delta t}$. El período de semidesintegración de la fuente es $T_{1/2}=1925.5$ d (d=día) con una incertidumbre ($k=1$) $U(T_{1/2})=0.5$ d. La constante λ se obtiene como $\lambda = \frac{\ln(2)}{T_{1/2}}$. La medida del tiempo Δt se realiza con una incertidumbre de 1 minuto. La actividad inicial A_0 era de 2 GBq con una incertidumbre relativa de 1%. Calcula la actividad al cabo de 365 d y su incertidumbre.

Apéndice A

Tablas estadísticas

Las tablas de este apéndice contienen valores de las funciones de distribución (probabilidades de que una variable aleatoria tome un valor menor o igual que el argumento de la función) o sus inversas (que dan los percentiles de la distribución) para algunas distribuciones interesantes para las cuales no existen expresiones en términos de funciones elementales.

A.1. Distribución normal estándar



Z	P(z ≤ Z)	Z	P(z ≤ Z)	Z	P(z ≤ Z)	Z	P(z ≤ Z)
-3,99	0,0000330	-3,97	0,0000359	-3,95	0,0000391	-3,93	0,0000425
-3,91	0,0000461	-3,89	0,0000501	-3,87	0,0000544	-3,85	0,0000591
-3,83	0,0000641	-3,81	0,0000695	-3,79	0,0000753	-3,77	0,0000816
-3,75	0,0000884	-3,73	0,0000957	-3,71	0,0001036	-3,69	0,0001121
-3,67	0,0001213	-3,65	0,0001311	-3,63	0,0001417	-3,61	0,0001531
-3,59	0,0001653	-3,57	0,0001785	-3,55	0,0001926	-3,53	0,0002078
-3,51	0,0002241	-3,49	0,0002415	-3,47	0,0002602	-3,45	0,0002803
-3,43	0,0003018	-3,41	0,0003248	-3,39	0,0003495	-3,37	0,0003758
-3,35	0,0004041	-3,33	0,0004342	-3,31	0,0004665	-3,29	0,0005009
-3,27	0,0005377	-3,25	0,0005770	-3,23	0,0006190	-3,21	0,0006637
-3,19	0,0007114	-3,17	0,0007622	-3,15	0,0008164	-3,13	0,0008740
-3,11	0,0009354	-3,09	0,0010008	-3,07	0,0010703	-3,05	0,0011442
-3,03	0,0012228	-3,01	0,0013062	-2,99	0,0013949	-2,97	0,0014890
-2,95	0,0015889	-2,93	0,0016948	-2,91	0,0018071	-2,89	0,0019262
-2,87	0,0020524	-2,85	0,0021860	-2,83	0,0023274	-2,81	0,0024771
-2,79	0,0026354	-2,77	0,0028028	-2,75	0,0029798	-2,73	0,0031667
-2,71	0,0033642	-2,69	0,0035726	-2,67	0,0037926	-2,65	0,0040246
-2,63	0,0042692	-2,61	0,0045271	-2,59	0,0047988	-2,57	0,0050849
-2,55	0,0053861	-2,53	0,0057031	-2,51	0,0060366	-2,49	0,0063872
-2,47	0,0067557	-2,45	0,0071428	-2,43	0,0075494	-2,41	0,0079763
-2,39	0,0084242	-2,37	0,0088940	-2,35	0,0093867	-2,33	0,0099031
-2,31	0,0104441	-2,29	0,0110107	-2,27	0,0116038	-2,25	0,0122245
-2,23	0,0128737	-2,21	0,0135526	-2,19	0,0142621	-2,17	0,0150034
-2,15	0,0157776	-2,13	0,0165858	-2,11	0,0174292	-2,09	0,0183089
-2,07	0,0192262	-2,05	0,0201822	-2,03	0,0211783	-2,01	0,0222156
-1,99	0,0232955	-1,97	0,0244192	-1,95	0,0255881	-1,93	0,0268034
-1,91	0,0280666	-1,89	0,0293790	-1,87	0,0307419	-1,85	0,0321568
-1,83	0,0336250	-1,81	0,0351479	-1,79	0,0367270	-1,77	0,0383636

Z	$P(z \leq Z)$	Z	$P(z \leq Z)$	Z	$P(z \leq Z)$	Z	$P(z \leq Z)$
-1,75	0,0400592	-1,73	0,0418151	-1,71	0,0436329	-1,69	0,0455140
-1,67	0,0474597	-1,65	0,0494715	-1,63	0,0515507	-1,61	0,0536989
-1,59	0,0559174	-1,57	0,0582076	-1,55	0,0605708	-1,53	0,0630084
-1,51	0,0655217	-1,49	0,0681121	-1,47	0,0707809	-1,45	0,0735293
-1,43	0,0763585	-1,41	0,0792698	-1,39	0,0822644	-1,37	0,0853435
-1,35	0,0885080	-1,33	0,0917591	-1,31	0,0950979	-1,29	0,0985253
-1,27	0,1020423	-1,25	0,1056498	-1,23	0,1093486	-1,21	0,1131394
-1,19	0,1170232	-1,17	0,1210005	-1,15	0,1250719	-1,13	0,1292381
-1,11	0,1334995	-1,09	0,1378566	-1,07	0,1423097	-1,05	0,1468591
-1,03	0,1515050	-1,01	0,1562476	-0,99	0,1610871	-0,97	0,1660232
-0,95	0,1710561	-0,93	0,1761855	-0,91	0,1814113	-0,89	0,1867329
-0,87	0,1921502	-0,85	0,1976625	-0,83	0,2032694	-0,81	0,2089701
-0,79	0,2147639	-0,77	0,2206499	-0,75	0,2266274	-0,73	0,2326951
-0,71	0,2388521	-0,69	0,2450971	-0,67	0,2514289	-0,65	0,2578461
-0,63	0,2643473	-0,61	0,2709309	-0,59	0,2775953	-0,57	0,2843388
-0,55	0,2911597	-0,53	0,2980560	-0,51	0,3050257	-0,49	0,3120669
-0,47	0,3191775	-0,45	0,3263552	-0,43	0,3335978	-0,41	0,3409030
-0,39	0,3482683	-0,37	0,3556912	-0,35	0,3631693	-0,33	0,3707000
-0,31	0,3782805	-0,29	0,3859081	-0,27	0,3935801	-0,25	0,4012937
-0,23	0,4090459	-0,21	0,4168338	-0,19	0,4246546	-0,17	0,4325051
-0,15	0,4403823	-0,13	0,4482832	-0,11	0,4562047	-0,09	0,4641436
-0,07	0,4720968	-0,05	0,4800612	-0,03	0,4880335	-0,01	0,4960106
0,02	0,5079783	0,04	0,5159534	0,06	0,5239222	0,08	0,5318814
0,10	0,5398278	0,12	0,5477584	0,14	0,5556700	0,16	0,5635595
0,18	0,5714237	0,20	0,5792597	0,22	0,5870644	0,24	0,5948349
0,26	0,6025681	0,28	0,6102612	0,30	0,6179114	0,32	0,6255158
0,34	0,6330717	0,36	0,6405764	0,38	0,6480273	0,40	0,6554217
0,42	0,6627573	0,44	0,6700314	0,46	0,6772419	0,48	0,6843863
0,50	0,6914625	0,52	0,6984682	0,54	0,7054015	0,56	0,7122603
0,58	0,7190427	0,60	0,7257469	0,62	0,7323711	0,64	0,7389137
0,66	0,7453731	0,68	0,7517478	0,70	0,7580363	0,72	0,7642375
0,74	0,7703500	0,76	0,7763727	0,78	0,7823046	0,80	0,7881446
0,82	0,7938919	0,84	0,7995458	0,86	0,8051055	0,88	0,8105703
0,90	0,8159399	0,92	0,8212136	0,94	0,8263912	0,96	0,8314724
0,98	0,8364569	1,00	0,8413447	1,02	0,8461358	1,04	0,8508300
1,06	0,8554277	1,08	0,8599289	1,10	0,8643339	1,12	0,8686431
1,14	0,8728568	1,16	0,8769756	1,18	0,8809999	1,20	0,8849303
1,22	0,8887676	1,24	0,8925123	1,26	0,8961653	1,28	0,8997274
1,30	0,9031995	1,32	0,9065825	1,34	0,9098773	1,36	0,9130850
1,38	0,9162067	1,40	0,9192433	1,42	0,9221962	1,44	0,9250663
1,46	0,9278550	1,48	0,9305634	1,50	0,9331928	1,52	0,9357445
1,54	0,9382198	1,56	0,9406201	1,58	0,9429466	1,60	0,9452007
1,62	0,9473839	1,64	0,9494974	1,66	0,9515428	1,68	0,9535213
1,70	0,9554345	1,72	0,9572838	1,74	0,9590705	1,76	0,9607961
1,78	0,9624620	1,80	0,9640697	1,82	0,9656205	1,84	0,9671159
1,86	0,9685572	1,88	0,9699460	1,90	0,9712834	1,92	0,9725711

Z	$P(z \leq Z)$	Z	$P(z \leq Z)$	Z	$P(z \leq Z)$	Z	$P(z \leq Z)$
1,94	0,9738102	1,96	0,9750021	1,98	0,9761482	2,00	0,9772499
2,02	0,9783083	2,04	0,9793248	2,06	0,9803007	2,08	0,9812372
2,10	0,9821356	2,12	0,9829970	2,14	0,9838226	2,16	0,9846137
2,18	0,9853713	2,20	0,9860966	2,22	0,9867906	2,24	0,9874545
2,26	0,9880894	2,28	0,9886962	2,30	0,9892759	2,32	0,9898296
2,34	0,9903581	2,36	0,9908625	2,38	0,9913437	2,40	0,9918025
2,42	0,9922397	2,44	0,9926564	2,46	0,9930531	2,48	0,9934309
2,50	0,9937903	2,52	0,9941323	2,54	0,9944574	2,56	0,9947664
2,58	0,9950600	2,60	0,9953388	2,62	0,9956035	2,64	0,9958547
2,66	0,9960930	2,68	0,9963189	2,70	0,9965330	2,72	0,9967359
2,74	0,9969280	2,76	0,9971099	2,78	0,9972821	2,80	0,9974449
2,82	0,9975988	2,84	0,9977443	2,86	0,9978818	2,88	0,9980116
2,90	0,9981342	2,92	0,9982498	2,94	0,9983589	2,96	0,9984618
2,98	0,9985588	3,00	0,9986501	3,02	0,9987361	3,04	0,9988171
3,06	0,9988933	3,08	0,9989650	3,10	0,9990324	3,12	0,9990957
3,14	0,9991553	3,16	0,9992112	3,18	0,9992636	3,20	0,9993129
3,22	0,9993590	3,24	0,9994024	3,26	0,9994429	3,28	0,9994810
3,30	0,9995166	3,32	0,9995499	3,34	0,9995811	3,36	0,9996103
3,38	0,9996376	3,40	0,9996631	3,42	0,9996869	3,44	0,9997091
3,46	0,9997299	3,48	0,9997493	3,50	0,9997674	3,52	0,9997842
3,54	0,9997999	3,56	0,9998146	3,58	0,9998282	3,60	0,9998409
3,62	0,9998527	3,64	0,9998637	3,66	0,9998739	3,68	0,9998834
3,70	0,9998922	3,72	0,9999004	3,74	0,9999080	3,76	0,9999150
3,78	0,9999216	3,80	0,9999277	3,82	0,9999333	3,84	0,9999385
3,86	0,9999433	3,88	0,9999478	3,90	0,9999519	3,92	0,9999557
3,94	0,9999593	3,96	0,9999625	3,98	0,9999655	4,00	0,9999683

Algunas relaciones útiles a la hora de usar esta tabla son:

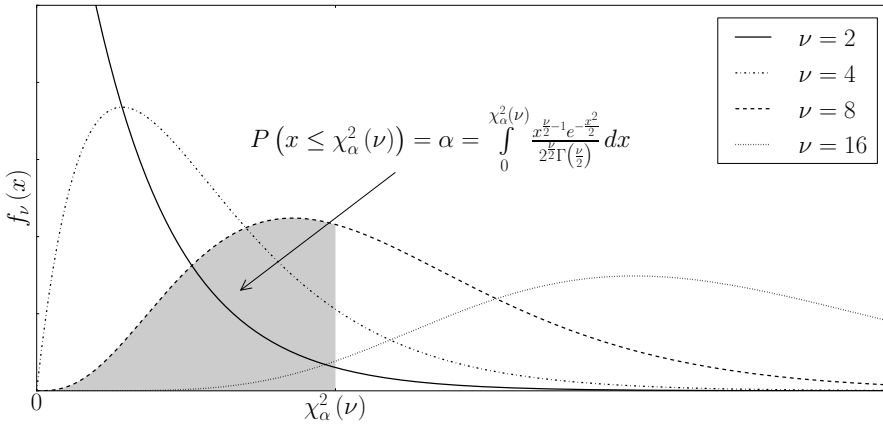
$$P(z > Z) = 1 - P(z \leq Z)$$

$$P(z \leq -Z) = P(z > -Z).$$

Por otro lado, para valores de Z superiores a 4 se puede usar la siguiente aproximación:

$$P(z > Z) \simeq \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}Z} \left(1 - \frac{1}{Z^2}\right).$$

A.2. Percentiles más útiles de la χ^2 de Pearson

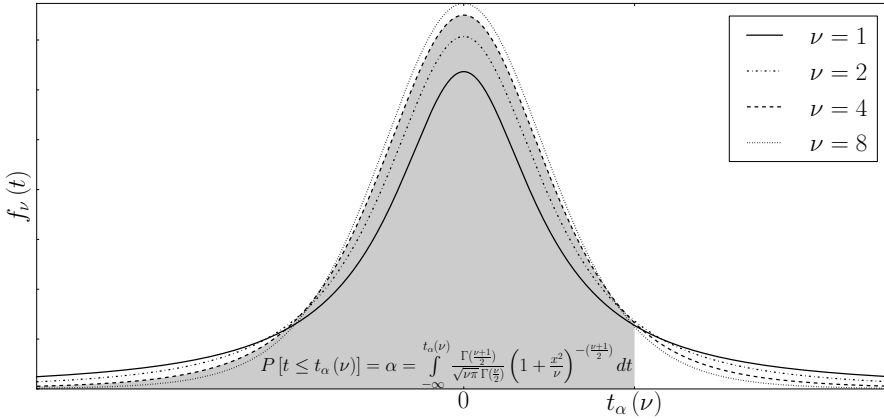


ν	$\chi_{0,75}^2$	$\chi_{0,8}^2$	$\chi_{0,85}^2$	$\chi_{0,9}^2$	$\chi_{0,925}^2$	$\chi_{0,95}^2$	$\chi_{0,975}^2$	$\chi_{0,99}^2$
1	1,323	1,642	2,072	2,706	3,170	3,841	5,024	6,635
2	2,773	3,219	3,794	4,605	5,181	5,991	7,378	9,210
3	4,108	4,642	5,317	6,251	6,905	7,815	9,348	11,345
4	5,385	5,989	6,745	7,779	8,496	9,488	11,143	13,277
5	6,626	7,289	8,115	9,236	10,008	11,070	12,833	15,086
6	7,841	8,558	9,446	10,645	11,466	12,592	14,449	16,812
7	9,037	9,803	10,748	12,017	12,883	14,067	16,013	18,475
8	10,219	11,030	12,027	13,362	14,270	15,507	17,535	20,090
9	11,389	12,242	13,288	14,684	15,631	16,919	19,023	21,666
10	12,549	13,442	14,534	15,987	16,971	18,307	20,483	23,209
11	13,701	14,631	15,767	17,275	18,294	19,675	21,920	24,725
12	14,845	15,812	16,989	18,549	19,602	21,026	23,337	26,217
13	15,984	16,985	18,202	19,812	20,897	22,362	24,736	27,688
14	17,117	18,151	19,406	21,064	22,180	23,685	26,119	29,141
15	18,245	19,311	20,603	22,307	23,452	24,996	27,488	30,578
16	19,369	20,465	21,793	23,542	24,716	26,296	28,845	32,000
17	20,489	21,615	22,977	24,769	25,970	27,587	30,191	33,409
18	21,605	22,760	24,155	25,989	27,218	28,869	31,526	34,805
19	22,718	23,900	25,329	27,204	28,458	30,144	32,852	36,191
20	23,828	25,038	26,498	28,412	29,692	31,410	34,170	37,566
21	24,935	26,171	27,662	29,615	30,920	32,671	35,479	38,932
22	26,039	27,301	28,822	30,813	32,142	33,924	36,781	40,289
23	27,141	28,429	29,979	32,007	33,360	35,172	38,076	41,638
24	28,241	29,553	31,132	33,196	34,572	36,415	39,364	42,980
25	29,339	30,675	32,282	34,382	35,780	37,652	40,646	44,314
26	30,435	31,795	33,429	35,563	36,984	38,885	41,923	45,642
27	31,528	32,912	34,574	36,741	38,184	40,113	43,195	46,963
28	32,620	34,027	35,715	37,916	39,380	41,337	44,461	48,278

ν	$\chi_{0,75}^2$	$\chi_{0,8}^2$	$\chi_{0,85}^2$	$\chi_{0,9}^2$	$\chi_{0,925}^2$	$\chi_{0,95}^2$	$\chi_{0,975}^2$	$\chi_{0,99}^2$
29	33,711	35,139	36,854	39,087	40,573	42,557	45,722	49,588
30	34,800	36,250	37,990	40,256	41,762	43,773	46,979	50,892
31	35,887	37,359	39,124	41,422	42,948	44,985	48,232	52,191
32	36,973	38,466	40,256	42,585	44,131	46,194	49,480	53,486
33	38,058	39,572	41,386	43,745	45,311	47,400	50,725	54,776
34	39,141	40,676	42,514	44,903	46,488	48,602	51,966	56,061
35	40,223	41,778	43,640	46,059	47,663	49,802	53,203	57,342
36	41,304	42,879	44,764	47,212	48,835	50,998	54,437	58,619
37	42,383	43,978	45,886	48,363	50,005	52,192	55,668	59,893
38	43,462	45,076	47,007	49,513	51,173	53,384	56,896	61,162
39	44,539	46,173	48,126	50,660	52,338	54,572	58,120	62,428
40	45,616	47,269	49,244	51,805	53,501	55,758	59,342	63,691
41	46,692	48,363	50,360	52,949	54,662	56,942	60,561	64,950
42	47,766	49,456	51,475	54,090	55,821	58,124	61,777	66,206
43	48,840	50,548	52,588	55,230	56,978	59,304	62,990	67,459
44	49,913	51,639	53,700	56,369	58,134	60,481	64,201	68,710
45	50,985	52,729	54,810	57,505	59,287	61,656	65,410	69,957
46	52,056	53,818	55,920	58,641	60,439	62,830	66,617	71,201
47	53,127	54,906	57,028	59,774	61,589	64,001	67,821	72,443
48	54,196	55,993	58,135	60,907	62,738	65,171	69,023	73,683
49	55,265	57,079	59,241	62,038	63,885	66,339	70,222	74,919
50	56,334	58,164	60,346	63,167	65,030	67,505	71,420	76,154
51	57,401	59,248	61,450	64,295	66,174	68,669	72,616	77,386
52	58,468	60,332	62,553	65,422	67,317	69,832	73,810	78,616
53	59,534	61,414	63,654	66,548	68,458	70,993	75,002	79,843
54	60,600	62,496	64,755	67,673	69,598	72,153	76,192	81,069
55	61,665	63,577	65,855	68,796	70,737	73,311	77,380	82,292
56	62,729	64,658	66,954	69,919	71,874	74,468	78,567	83,513
57	63,793	65,737	68,052	71,040	73,010	75,624	79,752	84,733
58	64,857	66,816	69,149	72,160	74,145	76,778	80,936	85,950
59	65,919	67,894	70,246	73,279	75,279	77,931	82,117	87,166
60	66,981	68,972	71,341	74,397	76,411	79,082	83,298	88,379

Para $\nu > 60$, la distribución puede aproximarse por una gaussiana de media ν y varianza 2ν .

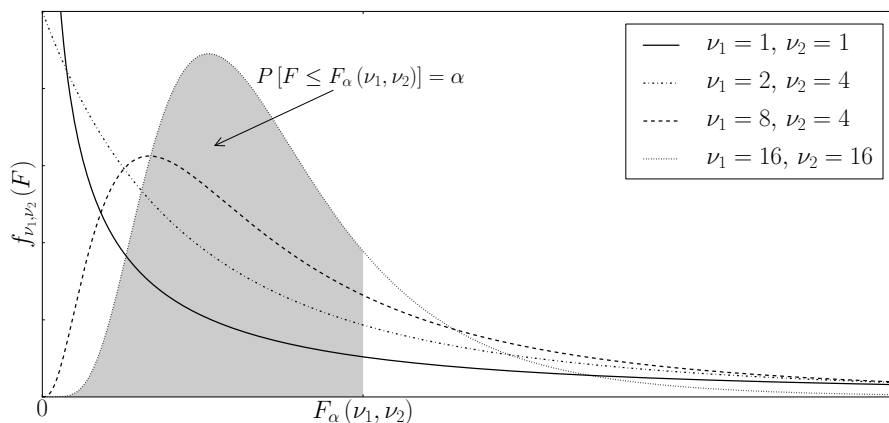
A.3. Percentiles más útiles de la t de Student



ν	$t_{0,01}$	$t_{0,025}$	$t_{0,05}$	$t_{0,1}$	$t_{0,9}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$
1	-31,821	-12,706	-6,314	-3,078	3,078	6,314	12,706	31,821
2	-6,965	-4,303	-2,920	-1,886	1,886	2,920	4,303	6,965
3	-4,541	-3,182	-2,353	-1,638	1,638	2,353	3,182	4,541
4	-3,747	-2,776	-2,132	-1,533	1,533	2,132	2,776	3,747
5	-3,365	-2,571	-2,015	-1,476	1,476	2,015	2,571	3,365
6	-3,143	-2,447	-1,943	-1,440	1,440	1,943	2,447	3,143
7	-2,998	-2,365	-1,895	-1,415	1,415	1,895	2,365	2,998
8	-2,896	-2,306	-1,860	-1,397	1,397	1,860	2,306	2,896
9	-2,821	-2,262	-1,833	-1,383	1,383	1,833	2,262	2,821
10	-2,764	-2,228	-1,812	-1,372	1,372	1,812	2,228	2,764
11	-2,718	-2,201	-1,796	-1,363	1,363	1,796	2,201	2,718
12	-2,681	-2,179	-1,782	-1,356	1,356	1,782	2,179	2,681
13	-2,650	-2,160	-1,771	-1,350	1,350	1,771	2,160	2,650
14	-2,624	-2,145	-1,761	-1,345	1,345	1,761	2,145	2,624
15	-2,602	-2,131	-1,753	-1,341	1,341	1,753	2,131	2,602
16	-2,583	-2,120	-1,746	-1,337	1,337	1,746	2,120	2,583
17	-2,567	-2,110	-1,740	-1,333	1,333	1,740	2,110	2,567
18	-2,552	-2,101	-1,734	-1,330	1,330	1,734	2,101	2,552
19	-2,539	-2,093	-1,729	-1,328	1,328	1,729	2,093	2,539
20	-2,528	-2,086	-1,725	-1,325	1,325	1,725	2,086	2,528
21	-2,518	-2,080	-1,721	-1,323	1,323	1,721	2,080	2,518
22	-2,508	-2,074	-1,717	-1,321	1,321	1,717	2,074	2,508
23	-2,500	-2,069	-1,714	-1,319	1,319	1,714	2,069	2,500
24	-2,492	-2,064	-1,711	-1,318	1,318	1,711	2,064	2,492

Para un número de grados de libertad superior, la gaussiana estándar es una buena aproximación a la t de Student.

A.4. Percentiles más útiles de la F de Fisher



		Percentil $F_{0,9}$							
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	
1	8,5263	5,5383	4,5448	4,0604	3,7759	3,5894	3,4579	3,3603	
2	9,0000	5,4624	4,3246	3,7797	3,4633	3,2574	3,1131	3,0065	
3	9,1618	5,3908	4,1909	3,6195	3,2888	3,0741	2,9238	2,8129	
4	9,2434	5,3426	4,1072	3,5202	3,1808	2,9605	2,8064	2,6927	
5	9,2926	5,3092	4,0506	3,4530	3,1075	2,8833	2,7264	2,6106	
6	9,3255	5,2847	4,0097	3,4045	3,0546	2,8274	2,6683	2,5509	
7	9,3491	5,2662	3,9790	3,3679	3,0145	2,7849	2,6241	2,5053	
8	9,3668	5,2517	3,9549	3,3393	2,9830	2,7516	2,5893	2,4694	
9	9,3805	5,2400	3,9357	3,3163	2,9577	2,7247	2,5612	2,4403	
10	9,3916	5,2304	3,9199	3,2974	2,9369	2,7025	2,5380	2,4163	
11	9,4006	5,2224	3,9067	3,2816	2,9195	2,6839	2,5186	2,3961	
12	9,4081	5,2156	3,8955	3,2682	2,9047	2,6681	2,5020	2,3789	
13	9,4145	5,2098	3,8859	3,2567	2,8920	2,6545	2,4876	2,3640	
14	9,4200	5,2047	3,8776	3,2468	2,8809	2,6426	2,4752	2,3510	
15	9,4247	5,2003	3,8704	3,2380	2,8712	2,6322	2,4642	2,3396	
16	9,4289	5,1964	3,8639	3,2303	2,8626	2,6230	2,4545	2,3295	
17	9,4325	5,1929	3,8582	3,2234	2,8550	2,6148	2,4458	2,3205	
18	9,4358	5,1898	3,8531	3,2172	2,8481	2,6074	2,4380	2,3123	
19	9,4387	5,1870	3,8485	3,2117	2,8419	2,6008	2,4310	2,3050	
20	9,4413	5,1845	3,8443	3,2067	2,8363	2,5947	2,4246	2,2983	
21	9,4437	5,1822	3,8405	3,2021	2,8312	2,5892	2,4188	2,2922	
22	9,4458	5,1801	3,8371	3,1979	2,8266	2,5842	2,4135	2,2867	
23	9,4478	5,1781	3,8339	3,1941	2,8223	2,5796	2,4086	2,2816	
24	9,4496	5,1764	3,8310	3,1905	2,8183	2,5753	2,4041	2,2768	

Percentil $F_{0,9}$								
$\nu_1 \backslash \nu_2$	9	10	11	12	13	14	15	16
1	3,2850	3,2252	3,1765	3,1362	3,1022	3,0732	3,0481	3,0262
2	2,9245	2,8595	2,8068	2,7632	2,7265	2,6952	2,6682	2,6446
3	2,7277	2,6602	2,6055	2,5603	2,5222	2,4898	2,4618	2,4374
4	2,6053	2,5362	2,4801	2,4337	2,3947	2,3614	2,3327	2,3077
5	2,5216	2,4512	2,3940	2,3467	2,3069	2,2730	2,2438	2,2183
6	2,4606	2,3891	2,3310	2,2830	2,2426	2,2081	2,1783	2,1524
7	2,4140	2,3416	2,2828	2,2341	2,1931	2,1582	2,1280	2,1017
8	2,3772	2,3040	2,2446	2,1953	2,1539	2,1185	2,0880	2,0613
9	2,3473	2,2735	2,2135	2,1638	2,1220	2,0862	2,0553	2,0284
10	2,3226	2,2482	2,1878	2,1376	2,0954	2,0593	2,0281	2,0009
11	2,3018	2,2269	2,1660	2,1155	2,0729	2,0366	2,0051	1,9777
12	2,2841	2,2087	2,1474	2,0966	2,0537	2,0171	1,9854	1,9577
13	2,2687	2,1930	2,1313	2,0802	2,0370	2,0001	1,9682	1,9404
14	2,2553	2,1792	2,1173	2,0658	2,0224	1,9853	1,9532	1,9252
15	2,2435	2,1671	2,1049	2,0532	2,0095	1,9722	1,9399	1,9117
16	2,2330	2,1563	2,0938	2,0419	1,9981	1,9605	1,9281	1,8997
17	2,2237	2,1467	2,0839	2,0318	1,9878	1,9501	1,9175	1,8889
18	2,2153	2,1380	2,0750	2,0227	1,9785	1,9407	1,9079	1,8792
19	2,2077	2,1302	2,0670	2,0145	1,9701	1,9321	1,8992	1,8704
20	2,2007	2,1230	2,0597	2,0070	1,9625	1,9243	1,8913	1,8624
21	2,1944	2,1165	2,0530	2,0001	1,9555	1,9172	1,8840	1,8550
22	2,1887	2,1106	2,0469	1,9939	1,9490	1,9106	1,8774	1,8482
23	2,1833	2,1051	2,0412	1,9881	1,9431	1,9046	1,8712	1,8420
24	2,1784	2,1000	2,0360	1,9827	1,9377	1,8990	1,8656	1,8362

Percentil $F_{0,9}$								
$\nu_1 \backslash \nu_2$	17	18	19	20	21	22	23	24
1	3,0070	2,9899	2,9747	2,9610	2,9486	2,9374	2,9271	2,9177
2	2,6239	2,6056	2,5893	2,5746	2,5613	2,5493	2,5383	2,5283
3	2,4160	2,3970	2,3801	2,3649	2,3512	2,3387	2,3274	2,3170
4	2,2858	2,2663	2,2489	2,2333	2,2193	2,2065	2,1949	2,1842
5	2,1958	2,1760	2,1582	2,1423	2,1279	2,1149	2,1030	2,0922
6	2,1296	2,1094	2,0913	2,0751	2,0605	2,0472	2,0351	2,0241
7	2,0785	2,0580	2,0397	2,0233	2,0084	1,9949	1,9826	1,9714
8	2,0379	2,0171	1,9985	1,9819	1,9668	1,9531	1,9407	1,9292
9	2,0047	1,9836	1,9649	1,9480	1,9327	1,9189	1,9063	1,8947
10	1,9770	1,9557	1,9367	1,9197	1,9043	1,8903	1,8775	1,8658
11	1,9535	1,9321	1,9129	1,8956	1,8801	1,8659	1,8530	1,8412
12	1,9333	1,9117	1,8924	1,8750	1,8593	1,8450	1,8319	1,8200
13	1,9158	1,8940	1,8745	1,8570	1,8411	1,8267	1,8136	1,8015
14	1,9004	1,8785	1,8588	1,8412	1,8252	1,8107	1,7974	1,7853
15	1,8868	1,8647	1,8449	1,8271	1,8111	1,7964	1,7831	1,7708
16	1,8747	1,8524	1,8325	1,8146	1,7984	1,7837	1,7703	1,7579

$\nu_1 \backslash \nu_2$	17	18	19	20	21	22	23	24
17	1,8638	1,8414	1,8214	1,8034	1,7871	1,7723	1,7587	1,7463
18	1,8539	1,8314	1,8113	1,7932	1,7768	1,7619	1,7483	1,7358
19	1,8450	1,8224	1,8022	1,7840	1,7675	1,7525	1,7388	1,7263
20	1,8368	1,8142	1,7938	1,7756	1,7590	1,7439	1,7302	1,7175
21	1,8294	1,8066	1,7862	1,7678	1,7512	1,7360	1,7222	1,7095
22	1,8225	1,7997	1,7792	1,7607	1,7440	1,7288	1,7149	1,7021
23	1,8162	1,7932	1,7727	1,7541	1,7374	1,7221	1,7081	1,6953
24	1,8103	1,7873	1,7667	1,7481	1,7312	1,7159	1,7019	1,6890

Percentil $F_{0,95}$								
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8
1	18,5128	10,1280	7,7086	6,6079	5,9874	5,5914	5,3177	5,1174
2	19,0000	9,5521	6,9443	5,7861	5,1433	4,7374	4,4590	4,2565
3	19,1643	9,2766	6,5914	5,4095	4,7571	4,3468	4,0662	3,8625
4	19,2468	9,1172	6,3882	5,1922	4,5337	4,1203	3,8379	3,6331
5	19,2964	9,0135	6,2561	5,0503	4,3874	3,9715	3,6875	3,4817
6	19,3295	8,9406	6,1631	4,9503	4,2839	3,8660	3,5806	3,3738
7	19,3532	8,8867	6,0942	4,8759	4,2067	3,7870	3,5005	3,2927
8	19,3710	8,8452	6,0410	4,8183	4,1468	3,7257	3,4381	3,2296
9	19,3848	8,8123	5,9988	4,7725	4,0990	3,6767	3,3881	3,1789
10	19,3959	8,7855	5,9644	4,7351	4,0600	3,6365	3,3472	3,1373
11	19,4050	8,7633	5,9358	4,7040	4,0274	3,6030	3,3130	3,1025
12	19,4125	8,7446	5,9117	4,6777	3,9999	3,5747	3,2839	3,0729
13	19,4189	8,7287	5,8911	4,6552	3,9764	3,5503	3,2590	3,0475
14	19,4244	8,7149	5,8733	4,6358	3,9559	3,5292	3,2374	3,0255
15	19,4291	8,7029	5,8578	4,6188	3,9381	3,5107	3,2184	3,0061
16	19,4333	8,6923	5,8441	4,6038	3,9223	3,4944	3,2016	2,9890
17	19,4370	8,6829	5,8320	4,5904	3,9083	3,4799	3,1867	2,9737
18	19,4402	8,6745	5,8211	4,5785	3,8957	3,4669	3,1733	2,9600
19	19,4431	8,6670	5,8114	4,5678	3,8844	3,4551	3,1613	2,9477
20	19,4458	8,6602	5,8025	4,5581	3,8742	3,4445	3,1503	2,9365
21	19,4481	8,6540	5,7945	4,5493	3,8649	3,4349	3,1404	2,9263
22	19,4503	8,6484	5,7872	4,5413	3,8564	3,4260	3,1313	2,9169
23	19,4523	8,6432	5,7805	4,5339	3,8486	3,4179	3,1229	2,9084
24	19,4541	8,6385	5,7744	4,5272	3,8415	3,4105	3,1152	2,9005

Percentil $F_{0,95}$								
$\nu_1 \backslash \nu_2$	9	10	11	12	13	14	15	16
1	4,9646	4,8443	4,7472	4,6672	4,6001	4,5431	4,4940	4,4513
2	4,1028	3,9823	3,8853	3,8056	3,7389	3,6823	3,6337	3,5915
3	3,7083	3,5874	3,4903	3,4105	3,3439	3,2874	3,2389	3,1968
4	3,4780	3,3567	3,2592	3,1791	3,1122	3,0556	3,0069	2,9647

$\nu_1 \backslash \nu_2$	9	10	11	12	13	14	15	16
5	3,3258	3,2039	3,1059	3,0254	2,9582	2,9013	2,8524	2,8100
6	3,2172	3,0946	2,9961	2,9153	2,8477	2,7905	2,7413	2,6987
7	3,1355	3,0123	2,9134	2,8321	2,7642	2,7066	2,6572	2,6143
8	3,0717	2,9480	2,8486	2,7669	2,6987	2,6408	2,5911	2,5480
9	3,0204	2,8962	2,7964	2,7144	2,6458	2,5876	2,5377	2,4943
10	2,9782	2,8536	2,7534	2,6710	2,6022	2,5437	2,4935	2,4499
11	2,9430	2,8179	2,7173	2,6347	2,5655	2,5068	2,4564	2,4126
12	2,9130	2,7876	2,6866	2,6037	2,5342	2,4753	2,4247	2,3807
13	2,8872	2,7614	2,6602	2,5769	2,5073	2,4481	2,3973	2,3531
14	2,8647	2,7386	2,6371	2,5536	2,4837	2,4244	2,3733	2,3290
15	2,8450	2,7186	2,6169	2,5331	2,4630	2,4034	2,3522	2,3077
16	2,8276	2,7009	2,5989	2,5149	2,4446	2,3849	2,3335	2,2888
17	2,8120	2,6851	2,5828	2,4987	2,4282	2,3683	2,3167	2,2719
18	2,7980	2,6709	2,5684	2,4841	2,4134	2,3533	2,3016	2,2567
19	2,7854	2,6581	2,5554	2,4709	2,4000	2,3398	2,2880	2,2429
20	2,7740	2,6464	2,5436	2,4589	2,3879	2,3275	2,2756	2,2304
21	2,7636	2,6358	2,5328	2,4479	2,3768	2,3163	2,2642	2,2189
22	2,7541	2,6261	2,5229	2,4379	2,3667	2,3060	2,2538	2,2084
23	2,7453	2,6172	2,5139	2,4287	2,3573	2,2966	2,2443	2,1987
24	2,7372	2,6090	2,5055	2,4202	2,3487	2,2878	2,2354	2,1898

Percentil $F_{0,95}$								
$\nu_1 \backslash \nu_2$	17	18	19	20	21	22	23	24
1	4,4139	4,3807	4,3512	4,3248	4,3009	4,2793	4,2597	4,2417
2	3,5546	3,5219	3,4928	3,4668	3,4434	3,4221	3,4028	3,3852
3	3,1599	3,1274	3,0984	3,0725	3,0491	3,0280	3,0088	2,9912
4	2,9277	2,8951	2,8661	2,8401	2,8167	2,7955	2,7763	2,7587
5	2,7729	2,7401	2,7109	2,6848	2,6613	2,6400	2,6207	2,6030
6	2,6613	2,6283	2,5990	2,5727	2,5491	2,5277	2,5082	2,4904
7	2,5767	2,5435	2,5140	2,4876	2,4638	2,4422	2,4226	2,4047
8	2,5102	2,4768	2,4471	2,4205	2,3965	2,3748	2,3551	2,3371
9	2,4563	2,4227	2,3928	2,3660	2,3419	2,3201	2,3002	2,2821
10	2,4117	2,3779	2,3479	2,3210	2,2967	2,2747	2,2547	2,2365
11	2,3742	2,3402	2,3100	2,2829	2,2585	2,2364	2,2163	2,1979
12	2,3421	2,3080	2,2776	2,2504	2,2258	2,2036	2,1834	2,1649
13	2,3143	2,2800	2,2495	2,2222	2,1975	2,1752	2,1548	2,1362
14	2,2900	2,2556	2,2250	2,1975	2,1727	2,1502	2,1298	2,1111
15	2,2686	2,2341	2,2033	2,1757	2,1508	2,1282	2,1077	2,0889
16	2,2496	2,2149	2,1840	2,1563	2,1313	2,1086	2,0880	2,0691
17	2,2325	2,1977	2,1667	2,1389	2,1138	2,0910	2,0703	2,0513
18	2,2172	2,1823	2,1511	2,1232	2,0980	2,0751	2,0543	2,0353
19	2,2033	2,1683	2,1370	2,1090	2,0837	2,0608	2,0399	2,0207
20	2,1906	2,1555	2,1242	2,0960	2,0707	2,0476	2,0267	2,0075
21	2,1791	2,1438	2,1124	2,0842	2,0587	2,0356	2,0146	1,9953

$\nu_1 \backslash \nu_2$	17	18	19	20	21	22	23	24
22	2,1685	2,1331	2,1016	2,0733	2,0478	2,0246	2,0035	1,9842
23	2,1587	2,1233	2,0917	2,0633	2,0377	2,0144	1,9932	1,9738
24	2,1497	2,1141	2,0825	2,0540	2,0283	2,0050	1,9838	1,9643

Percentil $F_{0,975}$								
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8
1	38,5063	17,4434	12,2179	10,0070	8,8131	8,0727	7,5709	7,2093
2	39,0000	16,0441	10,6491	8,4336	7,2599	6,5415	6,0595	5,7147
3	39,1655	15,4392	9,9792	7,7636	6,5988	5,8898	5,4160	5,0781
4	39,2484	15,1010	9,6045	7,3879	6,2272	5,5226	5,0526	4,7181
5	39,2982	14,8848	9,3645	7,1464	5,9876	5,2852	4,8173	4,4844
6	39,3315	14,7347	9,1973	6,9777	5,8198	5,1186	4,6517	4,3197
7	39,3552	14,6244	9,0741	6,8531	5,6955	4,9949	4,5286	4,1970
8	39,3730	14,5399	8,9796	6,7572	5,5996	4,8993	4,4333	4,1020
9	39,3869	14,4731	8,9047	6,6811	5,5234	4,8232	4,3572	4,0260
10	39,3980	14,4189	8,8439	6,6192	5,4613	4,7611	4,2951	3,9639
11	39,4071	14,3742	8,7935	6,5678	5,4098	4,7095	4,2434	3,9121
12	39,4146	14,3366	8,7512	6,5245	5,3662	4,6658	4,1997	3,8682
13	39,4210	14,3045	8,7150	6,4876	5,3290	4,6285	4,1622	3,8306
14	39,4265	14,2768	8,6838	6,4556	5,2968	4,5961	4,1297	3,7980
15	39,4313	14,2527	8,6565	6,4277	5,2687	4,5678	4,1012	3,7694
16	39,4354	14,2315	8,6326	6,4032	5,2439	4,5428	4,0761	3,7441
17	39,4391	14,2127	8,6113	6,3814	5,2218	4,5206	4,0538	3,7216
18	39,4424	14,1960	8,5924	6,3619	5,2021	4,5008	4,0338	3,7015
19	39,4453	14,1810	8,5753	6,3444	5,1844	4,4829	4,0158	3,6833
20	39,4479	14,1674	8,5599	6,3286	5,1684	4,4667	3,9995	3,6669
21	39,4503	14,1551	8,5460	6,3142	5,1538	4,4520	3,9846	3,6520
22	39,4525	14,1438	8,5332	6,3011	5,1406	4,4386	3,9711	3,6383
23	39,4544	14,1336	8,5216	6,2891	5,1284	4,4263	3,9587	3,6257
24	39,4562	14,1241	8,5109	6,2780	5,1172	4,4150	3,9472	3,6142

Percentil $F_{0,975}$								
$\nu_1 \backslash \nu_2$	9	10	11	12	13	14	15	16
1	6,9367	6,7241	6,5538	6,4143	6,2979	6,1995	6,1151	6,0420
2	5,4564	5,2559	5,0959	4,9653	4,8567	4,7650	4,6867	4,6189
3	4,8256	4,6300	4,4742	4,3472	4,2417	4,1528	4,0768	4,0112
4	4,4683	4,2751	4,1212	3,9959	3,8919	3,8043	3,7294	3,6648
5	4,2361	4,0440	3,8911	3,7667	3,6634	3,5764	3,5021	3,4379
6	4,0721	3,8807	3,7283	3,6043	3,5014	3,4147	3,3406	3,2767
7	3,9498	3,7586	3,6065	3,4827	3,3799	3,2934	3,2194	3,1556
8	3,8549	3,6638	3,5118	3,3880	3,2853	3,1987	3,1248	3,0610
9	3,7790	3,5879	3,4358	3,3120	3,2093	3,1227	3,0488	2,9849

$\nu_1 \backslash \nu_2$	9	10	11	12	13	14	15	16
10	3,7168	3,5257	3,3736	3,2497	3,1469	3,0602	2,9862	2,9222
11	3,6649	3,4737	3,3215	3,1975	3,0946	3,0078	2,9337	2,8696
12	3,6209	3,4296	3,2773	3,1532	3,0502	2,9633	2,8890	2,8249
13	3,5832	3,3917	3,2393	3,1150	3,0119	2,9249	2,8506	2,7863
14	3,5504	3,3588	3,2062	3,0819	2,9786	2,8915	2,8170	2,7526
15	3,5217	3,3299	3,1772	3,0527	2,9493	2,8621	2,7875	2,7230
16	3,4963	3,3044	3,1515	3,0269	2,9234	2,8360	2,7614	2,6968
17	3,4737	3,2816	3,1286	3,0039	2,9003	2,8128	2,7380	2,6733
18	3,4534	3,2612	3,1081	2,9832	2,8795	2,7919	2,7170	2,6522
19	3,4351	3,2428	3,0896	2,9646	2,8607	2,7730	2,6980	2,6331
20	3,4185	3,2261	3,0728	2,9477	2,8437	2,7559	2,6808	2,6158
21	3,4035	3,2109	3,0575	2,9322	2,8282	2,7403	2,6651	2,6000
22	3,3897	3,1970	3,0434	2,9181	2,8139	2,7260	2,6507	2,5855
23	3,3770	3,1843	3,0306	2,9052	2,8009	2,7128	2,6374	2,5721
24	3,3654	3,1725	3,0187	2,8932	2,7888	2,7006	2,6252	2,5598

Percentil $F_{0,975}$								
$\nu_1 \backslash \nu_2$	17	18	19	20	21	22	23	24
1	5,9781	5,9216	5,8715	5,8266	5,7863	5,7498	5,7166	5,6864
2	4,5597	4,5075	4,4613	4,4199	4,3828	4,3492	4,3187	4,2909
3	3,9539	3,9034	3,8587	3,8188	3,7829	3,7505	3,7211	3,6943
4	3,6083	3,5587	3,5147	3,4754	3,4401	3,4083	3,3794	3,3530
5	3,3820	3,3327	3,2891	3,2501	3,2151	3,1835	3,1548	3,1287
6	3,2209	3,1718	3,1283	3,0895	3,0546	3,0232	2,9946	2,9685
7	3,0999	3,0509	3,0074	2,9686	2,9338	2,9023	2,8738	2,8478
8	3,0053	2,9563	2,9128	2,8740	2,8392	2,8077	2,7791	2,7531
9	2,9291	2,8801	2,8365	2,7977	2,7628	2,7313	2,7027	2,6766
10	2,8664	2,8172	2,7737	2,7348	2,6998	2,6682	2,6396	2,6135
11	2,8137	2,7645	2,7209	2,6819	2,6469	2,6152	2,5865	2,5603
12	2,7689	2,7196	2,6758	2,6368	2,6017	2,5699	2,5411	2,5149
13	2,7302	2,6808	2,6369	2,5978	2,5626	2,5308	2,5019	2,4756
14	2,6964	2,6469	2,6030	2,5638	2,5285	2,4966	2,4677	2,4413
15	2,6667	2,6171	2,5731	2,5338	2,4984	2,4665	2,4374	2,4110
16	2,6404	2,5907	2,5465	2,5071	2,4717	2,4396	2,4105	2,3840
17	2,6168	2,5670	2,5228	2,4833	2,4478	2,4157	2,3865	2,3599
18	2,5956	2,5457	2,5014	2,4618	2,4262	2,3940	2,3648	2,3381
19	2,5764	2,5265	2,4821	2,4424	2,4067	2,3745	2,3452	2,3184
20	2,5590	2,5089	2,4645	2,4247	2,3890	2,3567	2,3273	2,3005
21	2,5431	2,4930	2,4484	2,4086	2,3728	2,3404	2,3109	2,2840
22	2,5285	2,4783	2,4337	2,3938	2,3579	2,3254	2,2959	2,2690
23	2,5151	2,4648	2,4201	2,3801	2,3442	2,3116	2,2821	2,2551
24	2,5027	2,4523	2,4076	2,3675	2,3315	2,2989	2,2693	2,2422

Percentil $F_{0,99}$								
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8
1	98,5025	34,1162	21,1977	16,2582	13,7450	12,2464	11,2586	10,5614
2	99,0000	30,8165	18,0000	13,2739	10,9248	9,5466	8,6491	8,0215
3	99,1662	29,4567	16,6944	12,0600	9,7795	8,4513	7,5910	6,9919
4	99,2494	28,7099	15,9770	11,3919	9,1483	7,8466	7,0061	6,4221
5	99,2993	28,2371	15,5219	10,9670	8,7459	7,4604	6,6318	6,0569
6	99,3326	27,9107	15,2069	10,6723	8,4661	7,1914	6,3707	5,8018
7	99,3564	27,6717	14,9758	10,4555	8,2600	6,9928	6,1776	5,6129
8	99,3742	27,4892	14,7989	10,2893	8,1017	6,8400	6,0289	5,4671
9	99,3881	27,3452	14,6591	10,1578	7,9761	6,7188	5,9106	5,3511
10	99,3992	27,2287	14,5459	10,0510	7,8741	6,6201	5,8143	5,2565
11	99,4083	27,1326	14,4523	9,9626	7,7896	6,5382	5,7343	5,1779
12	99,4159	27,0518	14,3736	9,8883	7,7183	6,4691	5,6667	5,1114
13	99,4223	26,9831	14,3065	9,8248	7,6575	6,4100	5,6089	5,0545
14	99,4278	26,9238	14,2486	9,7700	7,6049	6,3590	5,5589	5,0052
15	99,4325	26,8722	14,1982	9,7222	7,5590	6,3143	5,5151	4,9621
16	99,4367	26,8269	14,1539	9,6802	7,5186	6,2750	5,4766	4,9240
17	99,4404	26,7867	14,1146	9,6429	7,4827	6,2401	5,4423	4,8902
18	99,4436	26,7509	14,0795	9,6096	7,4507	6,2089	5,4116	4,8599
19	99,4465	26,7188	14,0480	9,5797	7,4219	6,1808	5,3840	4,8327
20	99,4492	26,6898	14,0196	9,5526	7,3958	6,1554	5,3591	4,8080
21	99,4516	26,6635	13,9938	9,5281	7,3722	6,1324	5,3364	4,7856
22	99,4537	26,6396	13,9703	9,5058	7,3506	6,1113	5,3157	4,7651
23	99,4557	26,6176	13,9488	9,4853	7,3309	6,0921	5,2967	4,7463
24	99,4575	26,5975	13,9291	9,4665	7,3127	6,0743	5,2793	4,7290

Percentil $F_{0,99}$								
$\nu_1 \backslash \nu_2$	9	10	11	12	13	14	15	16
1	10,0443	9,6460	9,3302	9,0738	8,8616	8,6831	8,5310	8,3997
2	7,5594	7,2057	6,9266	6,7010	6,5149	6,3589	6,2262	6,1121
3	6,5523	6,2167	5,9525	5,7394	5,5639	5,4170	5,2922	5,1850
4	5,9943	5,6683	5,4120	5,2053	5,0354	4,8932	4,7726	4,6690
5	5,6363	5,3160	5,0643	4,8616	4,6950	4,5556	4,4374	4,3359
6	5,3858	5,0692	4,8206	4,6204	4,4558	4,3183	4,2016	4,1015
7	5,2001	4,8861	4,6395	4,4410	4,2779	4,1415	4,0259	3,9267
8	5,0567	4,7445	4,4994	4,3021	4,1399	4,0045	3,8896	3,7910
9	4,9424	4,6315	4,3875	4,1911	4,0297	3,8948	3,7804	3,6822
10	4,8491	4,5393	4,2961	4,1003	3,9394	3,8049	3,6909	3,5931
11	4,7715	4,4624	4,2198	4,0245	3,8640	3,7299	3,6162	3,5185
12	4,7059	4,3974	4,1553	3,9603	3,8001	3,6662	3,5527	3,4552
13	4,6496	4,3416	4,0999	3,9052	3,7452	3,6115	3,4981	3,4007
14	4,6008	4,2932	4,0518	3,8573	3,6975	3,5639	3,4506	3,3533
15	4,5581	4,2509	4,0096	3,8154	3,6557	3,5222	3,4089	3,3117
16	4,5204	4,2134	3,9724	3,7783	3,6187	3,4852	3,3720	3,2748

$\nu_1 \backslash \nu_2$	9	10	11	12	13	14	15	16
17	4,4869	4,1801	3,9392	3,7452	3,5857	3,4523	3,3391	3,2419
18	4,4569	4,1503	3,9095	3,7156	3,5561	3,4228	3,3096	3,2124
19	4,4299	4,1234	3,8827	3,6888	3,5294	3,3961	3,2829	3,1857
20	4,4054	4,0990	3,8584	3,6646	3,5052	3,3719	3,2587	3,1615
21	4,3831	4,0769	3,8363	3,6425	3,4832	3,3498	3,2367	3,1394
22	4,3628	4,0566	3,8161	3,6224	3,4630	3,3297	3,2165	3,1192
23	4,3441	4,0380	3,7976	3,6038	3,4445	3,3111	3,1979	3,1006
24	4,3269	4,0209	3,7805	3,5868	3,4274	3,2940	3,1808	3,0835

Percentil $F_{0,99}$								
$\nu_1 \backslash \nu_2$	17	18	19	20	21	22	23	24
1	8,2854	8,1849	8,0960	8,0166	7,9454	7,8811	7,8229	7,7698
2	6,0129	5,9259	5,8489	5,7804	5,7190	5,6637	5,6136	5,5680
3	5,0919	5,0103	4,9382	4,8740	4,8166	4,7649	4,7181	4,6755
4	4,5790	4,5003	4,4307	4,3688	4,3134	4,2636	4,2184	4,1774
5	4,2479	4,1708	4,1027	4,0421	3,9880	3,9392	3,8951	3,8550
6	4,0146	3,9386	3,8714	3,8117	3,7583	3,7102	3,6667	3,6272
7	3,8406	3,7653	3,6987	3,6396	3,5867	3,5390	3,4959	3,4568
8	3,7054	3,6305	3,5644	3,5056	3,4530	3,4057	3,3629	3,3239
9	3,5971	3,5225	3,4567	3,3981	3,3458	3,2986	3,2560	3,2172
10	3,5082	3,4338	3,3682	3,3098	3,2576	3,2106	3,1681	3,1294
11	3,4338	3,3596	3,2941	3,2359	3,1837	3,1368	3,0944	3,0558
12	3,3706	3,2965	3,2311	3,1730	3,1209	3,0740	3,0316	2,9931
13	3,3162	3,2422	3,1769	3,1187	3,0667	3,0199	2,9775	2,9389
14	3,2689	3,1949	3,1296	3,0715	3,0195	2,9727	2,9303	2,8917
15	3,2273	3,1533	3,0880	3,0300	2,9779	2,9311	2,8887	2,8502
16	3,1904	3,1165	3,0512	2,9931	2,9411	2,8943	2,8519	2,8133
17	3,1575	3,0836	3,0183	2,9602	2,9082	2,8613	2,8189	2,7803
18	3,1280	3,0541	2,9887	2,9306	2,8786	2,8317	2,7892	2,7506
19	3,1013	3,0274	2,9620	2,9039	2,8518	2,8049	2,7624	2,7238
20	3,0771	3,0031	2,9377	2,8796	2,8274	2,7805	2,7380	2,6993
21	3,0550	2,9810	2,9156	2,8574	2,8052	2,7583	2,7157	2,6770
22	3,0348	2,9607	2,8953	2,8370	2,7849	2,7378	2,6953	2,6565
23	3,0161	2,9421	2,8766	2,8183	2,7661	2,7191	2,6765	2,6377
24	2,9990	2,9249	2,8594	2,8010	2,7488	2,7017	2,6591	2,6203

Bibliografía

Bevington, P. R., Robinson, D. K., *Data reduction and error analysis for the physical sciences*, 3ª Ed., McGrawHill, 2002.

Brandt, S., *Data analysis: Statistical and computational methods for scientists and engineers*. Springer, 1999.

Canavos, G. C., *Probabilidad y Estadística. Aplicaciones y Métodos*. McGraw Hill, 1998.

Centro Español de Metrología, *Guía para la Expresión de la Incertidumbre de Medida*, Ministerio de Fomento, 2000.

De la Horra Navarro, J. *Estadística Aplicada* Ed. Díaz de Santos.

Jaynes, E. T., *Information Theory and Statistical Mechanics*, *Physical Review*, vol. 106, p. 620, 1957.

Jaynes, E. T., *Information Theory and Statistical Mechanics II*, *Physical Review*, vol. 108, p. 171, 1957.

Kolmogorov, A. N., *Foundations of the theory of probability*, Chelsea, 1956.

Le Bellac, M., *Quantum and statistical field theory*, Oxford Science Publications, 1992.

National Institute of Standards and Technology (USA), *Guide for the Use of the International System of Units (SI)*, NIST special publication 811, 1995 Edition. <http://physics.nist.gov/Pubs/SP811/>

Peña Sánchez de Rivera. D. *Estadística. Modelos y métodos*, vol. I. Alianza Editorial, 1989.

Peña, D., Romo, J. *Introducción a la estadística para las ciencias sociales*. McGrawHill, 2003.

Popper, K. R. , *Teoría cuántica y el cisma en física*, Tecnos, 1985.

Quesada, V., Isidoro, A., López, L. A., *Curso y ejercicios de estadística*. Alhambra Universidad, 1998.

Sánchez del Río, C. *Análisis de errores*. Eudema, Madrid, 1989.

Shannon, C. E., *A mathematical theory of communication*, *The Bell System technical journal*, vol. 27, p. 379, 623, 1948.

Wonnacott, T. H., Wonnacott, R. J., *Introducción a la estadística*, Limusa, 1997.