

# **Teoría de probabilidades y estadística matemática**

**Gert Maibaum**

# **Teoría de probabilidades y estadística matemática**

# **Teoría de probabilidades y estadística matemática**

**Gert Maibaum**

**Tomada de la edición en alemán de la editorial Deutscher Verlag der Wissenschaften, Berlín,  
1976.**

**Traducción: Lic. Marta Álvarez Pérez**

**Edición: Prof. Martha Entralgo Flórez**

**Ilustración: Martha Tresancos Espín**

**Primera reimpresión, 1988**

**La presente edición se realiza en virtud de la licencia No. 15 del 12 de diciembre de 1987,  
otorgada por el Centro Nacional de Derecho de Autor, de conformidad con lo dispuesto en el  
Artículo 37 de la Ley No. 14 de Derecho de Autor de 28 de diciembre de 1977**

**SNLC:RA 01.13560.0**

## Nota a la edición en español

La presente obra es una traducción del libro *Wahrscheinlichkeitstheorie und mathematische Statistik* de Gert Maibaum, que forma parte de la serie *Mathematik für Lehrer* (abreviadamente MfL), cuyo objetivo principal consiste en brindar una bibliografía adecuada a los estudiantes que se forman como profesores de Matemática en la República Democrática Alemana.

Este libro, publicado en 1976, expone de forma rigurosamente exacta y desde posiciones acordes con nuestra concepción científica del mundo, los conceptos y métodos fundamentales de la teoría de probabilidades y la estadística matemática. Por esta razón, y porque responde a las exigencias en cuanto a la formación en la disciplina Probabilidades y Estadística que deben tener los estudiantes de la Licenciatura en Educación, especialidad Matemática, se ha decidido la publicación de esta obra en nuestro país para que sirva de texto básico, lo cual no excluye su utilización por otro círculo de lectores.

Esperamos que esta obra sea acogida favorablemente y que constituya un útil instrumento en manos de nuestros estudiantes.

DIRECCIÓN DE FORMACIÓN Y PERFECCIONAMIENTO DE PERSONAL PEDAGÓGICO

## Prefacio

El presente tomo 11 de la Colección de textos de estudio *Mathematik für Lehrer* ofrece una introducción a la teoría de probabilidades y la estadística matemática, disciplinas que poseen una gran significación para las más diversas esferas de aplicación e investigación científica, razón por la cual han entrado a formar parte de la formación matemática en la escuela media superior ampliada.

Este libro, en correspondencia con el objetivo general de la serie, está destinado, principalmente, a servir de texto básico en la formación de profesores de Matemática, pero además, debe ser apropiado para los estudiantes de otras especialidades que durante su estudio establezcan contacto con el Cálculo de probabilidades y la Estadística, o con ramas que empleen sus métodos y procedimientos. Por último, este texto debe brindarle a los profesores en ejercicio un acceso seguro y racional a la Teoría de probabilidades y a la Estadística matemática, así como un medio de consulta útil para la preparación y realización de cursos y círculos de interés sobre esta temática.

En esta obra se utilizan siete, de un total de 13 capítulos, para exponer la Teoría de probabilidades; los primeros tres capítulos abarcan el Cálculo de probabilidades, mientras que los capítulos 4 hasta el 7 se dedican al tratamiento de variables aleatorias y alcanzan su punto culminante con la formulación de proposiciones acerca de la Ley de los Grandes Números y del Teorema integral de De Moivre-Laplace. A continuación del capítulo 8 sobre Estadística descriptiva, se da respuesta a las principales interrogantes de la Estadística matemática en los capítulos 9 hasta el 11, donde las estimaciones puntuales y por intervalo de confianza, así como las pruebas de significación constituyen los puntos clave. El capítulo 12 contiene algunas tablas; por una parte se debe dar con esto una visión numérica de algunas distribuciones de probabilidad y, por otra, se agrupan aquí para la realización práctica de estimaciones por intervalo de confianza y pruebas de significación, percentiles frecuentemente utilizados en las distribuciones de probabilidad de los estadígrafos correspondientes. Con el capítulo 13 se da un pequeño bosquejo de la historia del Cálculo de Probabilidades. Por último, hay que señalar la bibliografía al final del libro, pues aquí se encuentran también algunos consejos que deben servir para la elección de literatura adecuada (por ejemplo, para la aplicación de métodos estadísticos en la investigación pedagógica o para la realización de cursos y círculos de interés sobre el Cálculo de probabilidades).

# Índice

0.	<b>Introducción .....</b>	11
1.	<b>Sucesos aleatorios .....</b>	13
1.1	<b>Experimentos aleatorios .....</b>	13
1.2	<b>Sucesos aleatorios .....</b>	14
1.3	<b>Operaciones entre sucesos aleatorios .....</b>	17
1.3.1	<b>Suma de sucesos .....</b>	18
1.3.2	<b>Producto de sucesos .....</b>	19
1.3.3	<b>Suceso contrario o complementario .....</b>	20
1.3.4	<b>Diferencia de sucesos .....</b>	21
1.3.5	<b>Diferencia simétrica de sucesos .....</b>	22
1.4	<b>Álgebras de sucesos .....</b>	22
1.5	<b>Álgebras de sucesos y álgebras de conjuntos .....</b>	24
2.	<b>Probabilidad .....</b>	26
2.1	<b>Frecuencia relativa .....</b>	27
2.2	<b>Definición clásica de probabilidad .....</b>	29
2.3	<b>Definición geométrica de probabilidad .....</b>	32
2.4	<b>Definición axiomática de probabilidad .....</b>	35
2.5	<b>Leyes de cálculo para probabilidades .....</b>	37
3.	<b>Probabilidad condicionada .....</b>	40
3.1	<b>Definición de probabilidad condicionada .....</b>	41
3.2	<b>Teorema de la multiplicación para probabilidades .....</b>	43
3.3	<b>Independencia de sucesos aleatorios .....</b>	45
3.4	<b>Fórmula de la probabilidad total .....</b>	47
3.5	<b>Fórmula de Bayes .....</b>	49
4.	<b>Variables aleatorias discretas .....</b>	51
4.1	<b>Definición general de variable aleatoria .....</b>	51
4.2	<b>Definición de variable aleatoria discreta .....</b>	55
4.3	<b>Características numéricas de las variables aleatorias discretas .....</b>	58
4.4	<b>Distribución discreta uniforme .....</b>	63
4.5	<b>Distribución binomial .....</b>	64
4.6	<b>Distribución hipergeométrica .....</b>	69
4.7	<b>Distribución de Poisson .....</b>	71
5.	<b>Variables aleatorias continuas .....</b>	74
5.1	<b>Definición de variable aleatoria continua .....</b>	74

Me he esforzado mucho por presentar los conceptos y proposiciones fundamentales de la Teoría de probabilidades de forma matemáticamente exacta, pero a la vez intuitiva. El objetivo esencial de los capítulos sobre Estadística matemática está en la explicación y fundamentación de las principales formas de deducción de esta disciplina. En su totalidad, la exposición está hecha, de modo tal, que la aplicación práctica no debe ofrecer dificultad alguna. Además, se introdujeron por esto numerosos ejemplos de las más diversas ramas. A causa de la extensión se tuvo que renunciar a una parte especialmente dedicada a ejercicios, que mostrara la amplia aplicación de la Teoría de probabilidades y de la Estadística matemática. El lector interesado puede encontrar también en la bibliografía referencias al respecto.

Quisiera aprovechar la ocasión para agradecer efusivamente a mi estimado maestro, Herr Profesor Dr. rer. nat. habil. P.H. Müller, quien ha revisado todo el manuscrito de forma sumamente crítica y me ha dado numerosas y valiosas indicaciones, tanto para la concepción y estructuración del libro, como también para su redacción definitiva. Además, es para mí un agradable deber agradecer a los editores de la serie *Mathematik für Lehrer* —en particular al editor coordinador, Herr Profesor Dr. sc. nat. W. Engel— y a la empresa nacionalizada Deutscher Verlag der Wissenschaften— especialmente a Frl. Dipl.-Math. E. Arndt y a la redactora de este libro, Frau Dipl. —Math. K. Bratz— por la grata cooperación, ayuda y competente asesoramiento. A continuación quisiera agradecer cordialmente a los cajistas de la empresa nacionalizada Druckhaus "Máximo Gorki" en Altenburg por el cuidadoso trabajo realizado por ellos. Por último, tengo que agradecer a Frl. I. Tittel y a mi esposa; ambas me han ayudado mucho en la confección del manuscrito.

Espero que el libro responda a las necesidades. Aceptaré con gusto cualquier indicación proveniente del círculo de lectores.

Dresden, febrero de 1976

GERT MAIBAUM

5.2	Características numéricas de las variables aleatorias continuas .....	77
5.3	Distribución continua uniforme .....	80
5.4	Distribución normal .....	81
5.5	Distribución exponencial .....	87
5.6	Distribución $\chi_2$ , $t$ y $F$ .....	89
5.6.1	Distribución $\chi_1$ .....	90
5.6.2	Distribución $t$ .....	92
5.6.3	Distribución $F$ .....	93
6.	Vectores aleatorios .....	94
6.1	Definición general de vector aleatorio .....	95
6.2	Vectores aleatorios discretos .....	97
6.3	Vectores aleatorios continuos .....	102
6.4	Independencia de variables aleatorias .....	106
6.5	Distribución de funciones de variables aleatorias .....	110
7.	Teoremas límites .....	117
7.1	Desigualdad de Chebyshov .....	118
7.2	Tipos de convergencia en la Teoría de probabilidades .....	120
7.3	Teoremas de Bernoulli y de Poisson (Ley de los grandes números) .....	124
7.4	Generalización de la Ley de los grandes números .....	126
7.5	Teorema local de De Moivre-Laplace .....	129
7.6	Teorema central del límite .....	132
8.	Estadística descriptiva .....	136
8.1	Métodos para el estudio de una característica medible .....	136
8.2	Medidas estadísticas para el estudio de una característica medible .....	140
8.2.1	Medidas de tendencia central .....	140
8.2.2	Medidas de dispersión .....	141
8.3	Métodos para el estudio de dos características medibles .....	142
8.4	Medidas estadísticas para el estudio de dos características medibles .....	146
9.	Conceptos fundamentales de la Estadística matemática .....	146
9.1	Tareas que se plantea la Estadística matemática .....	146
9.2	Población y muestra .....	148
9.3	Teorema fundamental de la Estadística matemática .....	150
9.4	Estadigráficos .....	153
10.	Introducción a la Teoría de la estimación .....	156
10.1	Tareas que se plantea la Teoría de la estimación .....	156
10.2	Estimadores puntuales (propiedades) .....	158
10.3	Sobre la construcción de estimadores puntuales .....	165
10.4	Ejemplos importantes de estimadores puntuales .....	170
10.4.1	Estimador puntual para un valor esperado desconocido .....	170
10.4.2	Estimadores puntuales para una varianza desconocida .....	171
10.4.3	Estimador puntual para una probabilidad desconocida .....	171
10.4.4	Estimador puntual para una función de distribución desconocida .....	172
10.4.5	Estimador puntual para un coeficiente de correlación desconocido .....	172
10.5	Estimaciones por intervalo de confianza .....	173
10.6	Ejemplos importantes de estimaciones por intervalo de confianza .....	177
10.6.1	Intervalos de confianza para los parámetros de una distribución normal .....	178
10.6.2	Intervalo de confianza para una probabilidad desconocida .....	180
10.6.3	Intervalo de confianza para una función de distribución desconocida .....	181
11.	Introducción a la teoría de la docimia de hipótesis .....	183
11.1	Tareas que se plantea la teoría de la docimia de hipótesis .....	183
11.2	Conceptos fundamentales de la teoría de la docimia de hipótesis .....	185
11.3	Procedimiento general para realizar una docimia de significación .....	189
11.4	Ejemplos importantes de docimas paramétricas .....	193
11.4.1	Docimia $t$ simple .....	195

11.4.2	Dócima $t$ doble .....	195
11.4.3	Dócima $\chi_2$ .....	196
11.4.4	Dócima $F$ .....	197
11.4.5	Dócima para una probabilidad desconocida .....	197
11.5	Ejemplos importantes de dócimas no paramétricas .....	198
11.5.1	Dócima de ajuste $\chi_2$ .....	199
11.5.2	Dócima de Kolmogorov .....	200
11.5.3	Dócima de homogeneidad $\chi_2$ .....	201
11.5.4	Dócima para dos distribuciones .....	201
11.5.5	Dócima de independencia $\chi_2$ .....	202
11.6	Ejemplo de aplicación .....	203
12.	<b>Tablas de algunas distribuciones importantes</b> .....	205
12.1	Tabla de la distribución binomial .....	205
12.2	Tabla de la distribución de Poisson .....	207
12.3	Tabla de la distribución normal .....	211
12.4	Tabla de la distribución $\chi_2$ .....	214
12.5	Tabla de la distribución $t$ .....	216
12.6	Tabla de la distribución $F$ .....	217
13.	<b>Breve bosquejo de la historia del cálculo de probabilidades</b> .....	222
	<b>Bibliografía</b> .....	226

## 0. Introducción

La Teoría de probabilidades y la Estadística matemática, son disciplinas matemáticas relativamente jóvenes por sí mismas, donde la Teoría de probabilidades, como teoría independiente —que incluye a su vez numerosas disciplinas especiales y campos de aplicación— y como fundamento de la Estadística matemática, posee una significación particular.

La *Teoría de Probabilidades* proporciona modelos matemáticos para la descripción de fenómenos sujetos a influjos casuales, y tiene como objetivo esencial la comprensión matemática de las regularidades de los fenómenos aleatorios.

La Teoría de probabilidades se construye de forma axiomática, de acuerdo con un procedimiento probado y muy utilizado hoy en día, y se sirve en gran medida de los métodos y resultados del Análisis.

La *Estadística matemática* proporciona, sobre la base de la Teoría de probabilidades, métodos mediante los cuales se puede obtener información sobre las distintas poblaciones a investigar, utilizando datos muestrales aleatorios; con esto se da origen también a métodos de ajuste de un modelo matemático, que considere efectos aleatorios, al proceso real correspondiente, sobre la base de datos concretos. El desarrollo de dispositivos electrónicos de alta potencia para el procesamiento de datos, exige la aplicación de métodos de la Estadística matemática, en particular de los métodos de análisis estadístico (por ejemplo, los análisis de correlación, regresión, varianza y análisis factorial), en los más diversos dominios de la práctica.

En los últimos decenios se desarrollaron numerosas disciplinas que se ocupan con interrogantes especiales de la Teoría de probabilidades y de la aplicación de métodos teórico-probabilísticos y estadísticos en distintas ciencias naturales y sociales (entre otras, en la pedagogía y la sicología), en la medicina, la técnica y la economía. Podemos citar como ejemplos, las teorías de la confiabilidad, la reposición, los juegos, la decisión, la información, la teoría ergódica, el diseño de experimentos, la biometría, la teoría del control estadístico de la calidad y la de la simulación por el método de Monte Carlo. Además, los métodos teórico-probabilísticos se utilizan de forma creciente y exitosamente en la ciencia militar, en el marco de la investigación de operaciones, de la toma de decisiones en los procesos económicos y en la cibernetica.

**La Teoría de probabilidades y la Estadística matemática, incluyendo sus disciplinas especiales y sus dominios de aplicación** (todas las ramas del saber que se ocupan en lo esencial del tratamiento matemático de fenómenos aleatorios) son conocidas en los últimos tiempos con el nombre de *estocásticas* (*στόχος*: el objetivo, la suposición; griego).

Junto a los fines de aplicación de la Teoría de probabilidades (por ejemplo, en la investigación de la confiabilidad de sistemas sobre la base de la de sus componentes individuales, en la determinación de las dimensiones de equipos de servicio o en la realización de controles de calidad en el marco de producciones masivas), se debe destacar también la significación de esta disciplina para el dominio de las ciencias naturales. Con las formaciones de conceptos y métodos de la Teoría de probabilidades es posible describir matemáticamente numerosos fenómenos (por ejemplo, los problemas que se relacionan con el movimiento de las partículas elementales, las leyes de Mendel en la biología, las leyes de los gases en la química y la física) de una forma aún más ajustada a la realidad objetiva, interpretar los resultados existentes de un modo nuevo y mucho más concluyente y, además, obtener proposiciones nuevas de gran valor cognoscitivo.

La aplicación práctica de la Teoría de probabilidades y de la Estadística matemática se basa en el convencimiento de que el grado de indeterminación de la ocurrencia de sucesos aleatorios se puede determinar, en cada caso, de forma objetiva, mediante un número: la probabilidad. Para ello se parte, en correspondencia con la realidad objetiva, de que a los fenómenos dependientes de la casualidad, así como a los procesos que transcurren de forma determinista, les son inherentes ciertas regularidades y de que la casualidad no significa ausencia total de reglas o caos. En este contexto se debe destacar que el concepto matemático *probabilidad*, que define en forma objetiva y cuantitativa la probabilidad de un suceso aleatorio, se diferencia del concepto de lo *probable*, utilizado en el lenguaje común, que tiene generalmente fuertes caracteres subjetivos y con el cual muchas veces solo se consideran proposiciones cualitativas. No obstante, se demuestra que las ideas subjetivas sobre la probabilidad de un suceso aleatorio se aproximan más y más a las relaciones objetivas que constituyen la esencia del concepto matemático probabilidad, en la medida en que aumenta el arsenal de nuestras experiencias.

Ahora nos dedicaremos a la construcción sistemática de la Teoría de probabilidades. Su representación se realiza en el marco de siete capítulos; los primeros tres capítulos abarcan la materia que se designa usualmente como *Cálculo de probabilidades*.

# 1. Sucesos aleatorios

En este capítulo nos ocuparemos de los *sucesos aleatorios*, que son aquellos que pueden presentarse bajo determinadas condiciones, pero no de forma obligatoria; nosotros los concebiremos como resultados de *experimentos aleatorios*, que son los que tienen un desenlace incierto en el marco de distintas posibilidades. Junto a la explicación detallada de estos y otros conceptos, trataremos en este capítulo las *operaciones entre sucesos aleatorios*. Por último, llegaremos a conocer el concepto *álgebra de sucesos*, de gran importancia para la construcción axiomática de la Teoría de probabilidades. Analizaremos también la relación entre álgebras de sucesos, álgebras de Boole y álgebras de conjuntos.

## 1.1 Experimentos aleatorios

Entendemos por *experimento aleatorio* aquel cuyo resultado es incierto en el marco de distintas posibilidades y se puede repetir un número de veces arbitrario (al menos mentalmente), manteniendo las mismas condiciones exteriores que caracterizan a dicho experimento.

### Ejemplos

1. El lanzamiento de una moneda es un experimento aleatorio. Los posibles resultados de este experimento están caracterizados por "estrella arriba" y "escudo arriba"
2. La tirada única de un dado después de agitarlo en un cubilete es un experimento aleatorio. Los posibles resultados de este experimento están caracterizados por el número que aparece en la cara superior del dado.
3. Las tiradas de un dado después de agitarlo en un cubilete pueden considerarse como un experimento aleatorio. Si solo nos interesamos porque aparezca el número seis, este experimento tiene  $n+1$  resultados. (Las veces que aparezca el número seis es una llamada variable aleatoria discreta que puede aceptar los  $n+1$  valores 0, 1, 2, ...,  $n$ )
4. La extracción al azar de una muestra de  $n$  objetos de una población (por ejemplo, la producción diaria de una fábrica) de  $N$  objetos, que contiene un número  $M$  de defec-

tuosos, puede entenderse como un experimento aleatorio. Aquí se realiza una extracción (sin reposición) de la muestra y cada uno de los  $N$  objetos en total tiene la misma oportunidad de ser sacado. Si solo nos interesamos por el número de objetos defectuosos en la muestra, este experimento tiene  $n+1$  desenlaces, en el caso que se cumpla  $M \geq n$ . (El número de objetos defectuosos es también una variable aleatoria discreta, cuya distribución de probabilidad desempeña una importante función en el control estadístico de la calidad.)

5. Toda medición (por ejemplo, de una longitud, un ángulo, un tiempo, una magnitud física), puede concebirse como un experimento aleatorio. De una parte, las mediciones realizadas en un mismo objeto son, por lo general, diferentes a causa de las insuficiencias del observador para llevarlas a cabo con precisión una y otra vez. Por otra parte, las mediciones realizadas en varios objetos iguales conducen también a resultados distintos, como consecuencia de las diferencias existentes entre estos.

Por tanto, en un experimento aleatorio existen influencias que no son consideradas en su descripción, es decir, en la enumeración de las condiciones que lo caracterizan y que conducen a que el resultado de este sea incierto en el marco de distintas posibilidades.

En la explicación anterior hemos también destacado, que los experimentos aleatorios pueden repetirse –al menos mentalmente– un número de veces arbitrario. Esta condición permite el estudio de aquellas regularidades, que solo pueden reconocerse mediante un número elevado de repeticiones del experimento aleatorio correspondiente. (Expresamos también esta particularidad diciendo que los fenómenos en que se investigan tales regularidades son masivos.) El estudio de las regularidades que se presentan en los fenómenos aleatorios es el objetivo principal de la Teoría de probabilidades.

## 1.2 Sucesos aleatorios

Designaremos por *suceso aleatorio* un resultado de un experimento aleatorio. Por consiguiente, este puede presentarse bajo las condiciones que caracterizan al experimento aleatorio y puede no presentarse.

Describimos frecuentemente un suceso aleatorio mediante la ilustración de la situación en que se presenta. Por lo general designamos los sucesos aleatorios con letras mayúsculas latinas, que en algunos casos pueden estar provistas de índices.

Ejemplos. Nos remitiremos a los ejemplos de 1.1:

1.  $A$  ... El escudo aparece arriba.
2.  $A_k$  ... El número obtenido al tirar el dado es igual a  $k$  ( $k=1, \dots, 6$ ).  
 $B$  ... El número obtenido al tirar el dado es par.
3.  $A_k$  ... Las veces que aparece el número seis al realizar  $n$  tiradas del dado es igual a  $k$  ( $k=0, 1, 2, \dots, n$ ).
4.  $A_k$  ... El número de los objetos defectuosos en la muestra aleatoria es igual a  $k$  ( $k=0, 1, 2, \dots, n$ ).
5.  $A$  ... La magnitud que se mide está entre los límites de tolerancia.

En las consideraciones sobre sucesos aleatorios queremos referirnos a aquellos que pueden concebirse como casos especiales de sucesos aleatorios: sucesos seguros y sucesos imposibles.

Los sucesos seguros son los que se presentan obligatoriamente bajo las condiciones que caracterizan al experimento aleatorio considerado; los sucesos imposibles son los que no se pueden presentar nunca.

Designaremos, de forma única, los sucesos seguros con  $\Omega$  (se lee: omega mayúscula) y los sucesos imposibles, con  $\emptyset$  (con el símbolo del conjunto vacío).

Ejemplo. El experimento aleatorio consiste en la tirada única de dos dados después de agitarlos en un cubilete. Un suceso seguro es, por ejemplo, que la suma de los números obtenidos sea menor o igual que 12: un suceso imposible es, digamos, que la suma de los números obtenidos sea menor que 2.

A menudo se pueden ilustrar los sucesos aleatorios por medio de subconjuntos sobre la recta numérica o en el plano.

#### Ejemplos

1. El experimento aleatorio consiste en rotar un disco al cual se ha fijado un indicador. Los infinitos resultados imaginables de este experimento son las posiciones que puede tener el indicador cuando el disco permanece quieto. Cada una de estas posiciones puede caracterizarse mediante la amplitud del ángulo  $\varphi$  formado entre el eje positivo de las  $x$  y el indicador (fig. 1).

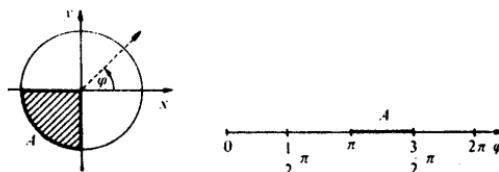


Figura 1

De esta forma, todo suceso  $A$  relacionado con este experimento aleatorio puede describirse por medio del conjunto  $\tilde{A}$  de aquellas amplitudes de ángulos  $\varphi$  que son "convenientes" para el suceso considerado, y decimos esto en el sentido de que el suceso  $A$  se presenta si y solo si la posición del indicador cuando el disco no se mueve se describe por una de las amplitudes de ángulos del conjunto  $\tilde{A}$ . Si, por ejemplo, el suceso  $A$  consiste en que el indicador permanezca quieto en el tercer cuadrante, le asociamos a este suceso el intervalo de  $\pi$  a  $\frac{3\pi}{2}$  sobre el eje  $\varphi$ , o sea, el conjunto

$$\tilde{A} = \left\{ \varphi : \pi \leq \varphi \leq \frac{3\pi}{2} \right\} \text{ (ver fig. 1).}$$

2. El experimento aleatorio consiste en tirar sobre un disco con diez circunferencias concéntricas de radios  $r_1 > r_2 > \dots > r_{10} > 0$  (fig. 2).

Todo suceso  $A$ , relacionado con este experimento, puede describirse mediante el conjunto  $\tilde{A}$  de todos los puntos "convenientes" en el plano  $x, y$  para el suceso considerado, y decimos convenientes en el sentido de que  $A$  se presenta si y solo si el tiro acierta sobre un punto de  $\tilde{A}$ . Si, por ejemplo, el suceso  $A$  es que el tiro disparado sea certero, se describe este suceso por medio del conjunto

$$\tilde{A} = \{(x, y) : x^2 + y^2 \leq r_i^2\}.$$

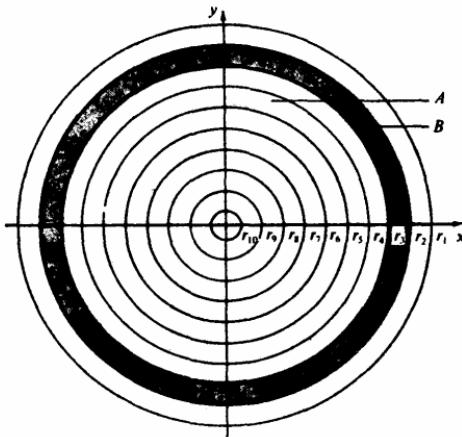


Figura 2

### El conjunto

$$\tilde{B} = \{(x, y) : r_3^2 < x^2 + y^2 \leq r_2^2\}$$

representa al suceso **B** que se presenta si y solo si el tiro acierta en el anillo circular limitado por las circunferencias de radios  $r_2$  y  $r_3$ .

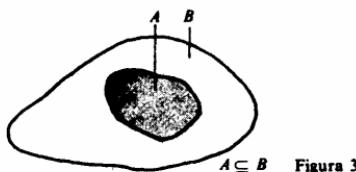
Para consideraciones generales se ilustran también los sucesos aleatorios mediante conjuntos de puntos en el plano. Posteriormente analizaremos más exactamente la estrecha relación entre los sucesos aleatorios y los conjuntos (ver 1.5).

A continuación queremos definir una relación entre sucesos aleatorios con la cual se pueda después concebir también la igualdad de sucesos aleatorios en forma matemática. Además, nos imaginaremos siempre que los sucesos aleatorios observados pertenecen a un determinado experimento aleatorio.

**Definición 1.** Si a la ocurrencia del suceso aleatorio **A** está siempre unida la ocurrencia del suceso aleatorio **B**, escribimos

$$A \subseteq B,$$

y se lee: **A** entraña **B**, **A** implica **B** o **A** es una parte de **B** (fig. 3).



$A \subseteq B$  Figura 3

Luego utilizamos aquí un símbolo de la teoría de conjuntos (ver MfL Tomo 1, 1.5); la figura 3 debe recordarnos el comportamiento correspondiente en conjuntos. (Se puede hacer corresponder a un sistema de sucesos, perteneciente a un experimento aleatorio, un sistema de subconjuntos de un conjunto universo, de forma tal que la relación  $A \subseteq B$  exista para sucesos aleatorios  $A$ , y  $B$  si y solo si el conjunto asociado al suceso  $A$  es un subconjunto del asociado al suceso  $B$ . En particular, se hace corresponder al suceso seguro el conjunto universo y al suceso imposible, el conjunto vacío (ver 1.5).

Ejemplo. Tirada de un dado.

- $A \dots$  El número obtenido al tirar el dado es igual a 6 ( $A = \{6\}$ ). }  
 $B \dots$  El número obtenido al tirar el dado es par ( $B = \{2, 4, 6\}$ ). }  $\Rightarrow A \subseteq B$

Con la definición 1 se confirma enseguida que para todo suceso aleatorio  $A$  se cumplen las proposiciones siguientes:

$$\phi \subseteq A, A \subseteq A, A \subseteq \Omega. \quad (1)$$

Si con el suceso  $A$  se presenta siempre el suceso  $B$  y el  $B$  implica al suceso  $C$ , entonces el suceso  $A$  entraña evidentemente al suceso  $C$ . Expresado en fórmulas:

$$A \subseteq B, B \subseteq C \Rightarrow A \subseteq C. \quad (2)$$

Llegamos ahora a la definición de la igualdad de sucesos aleatorios.

Definición 2. Dos sucesos aleatorios  $A$  y  $B$  se llaman *iguales* ( $A = B$ ) si tanto el suceso  $A$  implica al suceso  $B$  ( $A \subseteq B$ ) como también a la inversa, el suceso  $B$  implica al suceso  $A$  ( $B \subseteq A$ ).

Esta definición contempla que dos sucesos aleatorios se consideran iguales si y solo si en cada repetición se presentan siempre ambos sucesos o no se presentan.

Si dos sucesos aleatorios  $A$  y  $B$  no son iguales, expresamos esto a través de  $A \neq B$ .

Por último, destacamos que la relación  $\subseteq$  es reflexiva y transitiva a causa de (1) y (2), y antisimétrica en virtud de la definición 2, es decir, que la relación  $\subseteq$  es una relación de orden parcial (ver MfL Tomo 1, 1.5.). En lugar de  $A \subseteq B$  escribimos también  $B \supseteq A$ .

### 1.3 Operaciones entre sucesos aleatorios

En este epígrafe tratamos las operaciones entre sucesos aleatorios, cuya aplicación es muy conveniente y con frecuencia conduce a una formulación muy clara de distintos hechos. Aquí se presentan símbolos de operaciones conocidos del tratamiento de la teoría de conjuntos (ver MfL Tomo 1, 1.4). Aclaramos que si se sustituyen los sucesos que aparecen por conjuntos, surgen siempre de las proposiciones siguientes (sobre sucesos) proposiciones verdaderas de la teoría de conjuntos y viceversa, se obtiene de las proposiciones correspondientes de la teoría de conjuntos proposiciones verdaderas sobre sucesos aleatorios, si se sustituyen los conjuntos que aparecen por esos sucesos. (La fundamentación de esto lo proporciona un teorema sobre el isomorfismo entre las álgebras de sucesos y las álgebras de conjuntos, que trataremos en el epígrafe 1.5.) Las figuras dadas a continuación de las siguientes definiciones de las operaciones entre sucesos aleatorios deben servir

para recordar las definiciones de las operaciones correspondientes con conjuntos. Todos los ejemplos de este epígrafe se refieren, para mayor sencillez, al experimento aleatorio consistente en la tirada única de un dado.

### 1.3.1 Suma de sucesos

**Definición 1.** Si  $A$  y  $B$  son sucesos aleatorios, entonces designamos al suceso que ocurre si y solo si al menos uno de los sucesos  $A$  y  $B$  ocurre, por

$$A \cup B$$

y se lee:  $A$  o  $B$ , suma de  $A$  y  $B$  o  $A$  unido con  $B$  (fig. 4).

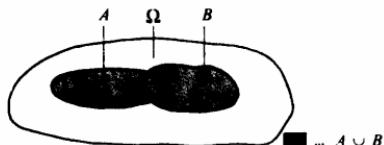


Figura 4

**Ejemplo.** Tirada de un dado.

$A\dots$  El número obtenido es par ( $A=\{2,4,6\}$ ).

$B\dots$  El número obtenido es mayor o igual que 3 ( $B=\{3,4,5,6\}$ ).

$A \cup B\dots$  El número obtenido es distinto de 1 ( $A \cup B=\{2,3,4,5,6\}$ ).

Las siguientes proposiciones son fáciles de comprobar:

$$A \cup \emptyset = A, A \cup A = A, A \cup \Omega = \Omega, \quad (1)$$

$$A \subseteq A \cup B, B \subseteq A \cup B. \quad (2)$$

$$A \cup B = B \cup A \text{ (comutatividad),} \quad (3)$$

$$A \cup (B \cup C) = (A \cup B) \cup C \text{ (asociatividad).} \quad (4)$$

Sobre la base de la validez de la ley asociativa se puede definir la suma de  $n (n \geq 2)$  sucesos aleatorios de la forma siguiente.

**Definición 2.** Si  $A_1, A_2, \dots, A_n$  son sucesos aleatorios, entonces designamos al suceso que ocurre si y solo si al menos uno de los sucesos  $A_i$  ( $i=1,2,\dots, n$ ) ocurre, por

$$A_1 \cup A_2 \cup \dots \cup A_n$$

o también con

$$\bigcup_{i=1}^n A_i.$$

Generalizando, podemos designar al suceso que ocurre si y solo si al menos un suceso de la sucesión (infinita)  $A_1, A_2, \dots$  de sucesos  $A_i$  ( $i=1,2,\dots$ ) ocurre, por

$$A_1 \cup A_2 \cup \dots$$

o también con

$$\bigcup_{i=1}^{\infty} A_i.$$

### 1.3.2 Producto de sucesos

Definición 3. Si  $A$  y  $B$  son sucesos aleatorios, entonces designamos al suceso que ocurre si y solo si tanto  $A$  como  $B$  ocurre, por

$$A \cap B$$

y se lee:  $A$  y  $B$ , producto de  $A$  y  $B$  o intersección de  $A$  y  $B$  (fig. 5).

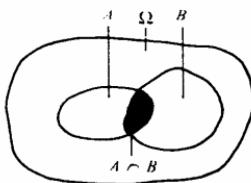


Figura 5

Ejemplo. Tirada de un dado.

$A$  ... El número obtenido es par ( $A = \{2, 4, 6\}$ ).

$B$  ... El número obtenido es menor que 3 ( $B = \{1, 2\}$ ).

$A \cap B$  ... El número obtenido es igual a 2 ( $A \cap B = \{2\}$ ).

Las proposiciones siguientes son también fáciles de verificar:

$$A \cap \emptyset = \emptyset, A \cap A = A, A \cap \Omega = A. \quad (5)$$

$$A \cap B \subseteq A, A \cap B \subseteq B. \quad (6)$$

$$A \cap B = B \cap A \text{ (comutatividad).} \quad (7)$$

$$A \cap (B \cap C) = (A \cap B) \cap C \text{ (asociatividad).} \quad (8)$$

Sobre la base de la validez de la ley asociativa podemos definir el producto de  $n$  ( $n \geq 2$ ) sucesos aleatorios de la forma siguiente.

Definición 4. Si  $A_1, A_2, \dots, A_n$  son sucesos aleatorios, entonces designamos al suceso que ocurre si y solo si cada uno de los sucesos  $A_i$  ( $i=1, 2, \dots, n$ ) ocurre, por

$$A_1 \cap A_2 \cap \dots \cap A_n$$

o también por

$$\bigcap_{i=1}^n A_i.$$

Generalizando, podemos designar al suceso que ocurre si y solo si cada uno de los sucesos de la sucesión (infinita)  $A_1, A_2, \dots$  de sucesos  $A_i$  ( $i=1, 2, \dots$ ) ocurre, mediante

$$A_1 \cap A_2 \cap \dots$$

o también

$$\bigcap_{i=1}^{\infty} A_i.$$

Aquí queremos introducir aún dos conceptos sobre los cuales volveremos posteriormente.



**Definición 5.** Dos sucesos aleatorios  $A$  y  $B$  se llaman *mutuamente excluyentes*, si se cumple

$$A \cap B = \emptyset.$$

$A \cap B = \emptyset$  significa en cuanto al contenido, que la ocurrencia común de los sucesos  $A$  y  $B$  es imposible. Se dice también que  $A$  y  $B$  son incompatibles o que  $A$  y  $B$  son disjuntos (fig. 6).

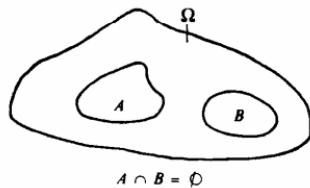


Figura 6

**Definición 6.** Un conjunto  $\{A_1, A_2, \dots, A_n, \dots\}$  de sucesos aleatorios  $A_i \neq \emptyset$  se llama un *sistema completo de sucesos*, si se cumple

$$\begin{aligned} A_i \cap A_k &= \emptyset \quad (i \neq k), \\ A_1 \cup A_2 \cup \dots \cup A_n \cup \dots &= \Omega. \end{aligned}$$

**Ejemplo.** Tirada de un dado.

$A_1, \dots$  El número obtenido al tirar el dado es igual a  $i$  ( $i=1,2,3,4,5,6$ ).  
 $\{A_1, A_2, A_3, A_4, A_5, A_6\}$  es un sistema completo de sucesos.

De modo general, si consideramos un experimento aleatorio que tiene siempre como resultado la ocurrencia de exactamente uno de los sucesos aleatorios  $A_1, A_2, \dots, A_n, \dots$ , entonces el conjunto de estos resultados forma un sistema completo de sucesos.

### 1.3.3 Suceso contrario o complementario

**Definición 7.** Si  $A$  es un suceso aleatorio, entonces designamos al suceso que ocurre si y solo si  $A$  no ocurre, por  $\bar{A}$  y llamamos a este el *suceso contrario o complementario de  $A$*  (fig. 7).

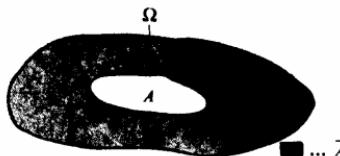


Figura 7

**Ejemplo.** Tirada de un dado.

$A \dots$  El número obtenido es menor e igual que 3 ( $A=\{1,2,3\}$ ).

$\bar{A} \dots$  El número obtenido es mayor que 3 ( $\bar{A}=\{4,5,6\}$ ).

Evidentemente para un suceso  $A$  cualquiera se cumplen las relaciones

$$A \cup \bar{A} = \Omega \quad \text{y} \quad A \cap \bar{A} = \emptyset. \tag{9}$$

Por tanto, si  $A$  es un suceso aleatorio que no es imposible ni seguro, es decir,  $A \neq \emptyset$ ,  $A \neq \Omega$ , entonces el conjunto  $\{A, \bar{A}\}$  es un sistema completo de sucesos.

Además, se verifica directamente la validez de las proposiciones

$$\bar{\phi} = \Omega, \quad \bar{\Omega} = \emptyset, \quad \bar{(\bar{A})} = A. \quad (10)$$

Seguidamente escribiremos algunas otras proposiciones, que no son difíciles de comprobar:

$$A \subseteq B \Rightarrow \bar{B} \subseteq \bar{A}, \quad (11)$$

$$\overline{A \cap \bar{B}} = \bar{A} \cup \bar{B}, \text{ más general: } \overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \bar{A}_i, \quad (12)$$

$$\overline{A \cup \bar{B}} = \bar{A} \cap \bar{B}, \text{ más general: } \overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \bar{A}_i. \quad (13)$$

A continuación damos fórmulas para la descomposición de la suma de dos sucesos aleatorios en sucesos mutuamente excluyentes dos a dos (fig. 8).

$$A \cup B = A \cup (B \cap \bar{A}), \quad (14)$$

$$A \cup B = B \cup (A \cap \bar{B}), \quad (15)$$

$$A \cup B = (A \cap \bar{B}) \cup (\bar{A} \cap B) \cup (A \cap B). \quad (16)$$

Dejamos al lector la fácil comprobación de lo anterior.

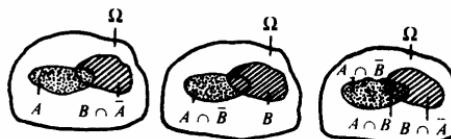


Figura 8

### 1.3.4 Diferencia de sucesos

Definición 8. Si  $A$  y  $B$  son sucesos aleatorios, entonces designamos al suceso que ocurre si y solo si el suceso  $A$ , pero no el suceso  $B$ , ocurre, por

$$A \setminus B$$

y se lee:  $A$  y no  $B$ , diferencia de  $A$  y  $B$ ,  $A$  menos  $B$  (fig. 9).

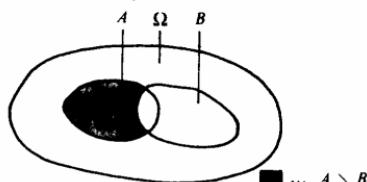


Figura 9

Ejemplo. Tirada de un dado.

$A \dots$  El número obtenido es par ( $A = \{2, 4, 6\}$ ).

$B \dots$  El número obtenido es menor e igual que 3 ( $B = \{1, 2, 3\}$ ).

$A \setminus B \dots$  El número obtenido es igual a 4 ó a 6 ( $A \setminus B = \{4, 6\}$ ).

$B \setminus A \dots$  El número obtenido es igual a 1 ó a 3 ( $B \setminus A = \{1, 3\}$ ).

Ya que la operación  $\setminus$  se puede expresar sobre la base de la relación

$$A \setminus B = A \cap \bar{B} \quad (17)$$

mediante las operaciones  $\cap$  y  $\neg$ , podemos renunciar a otras explicaciones. Llamamos la atención de que para la operación  $\setminus$  no se cumple trivialmente la ley conmutativa (ver ejemplo anterior).

### 1.3.5 Diferencia simétrica de sucesos

Definición 9. Si  $A$  y  $B$  son sucesos aleatorios, entonces designamos al suceso que ocurre si y solo si  $A$  o  $B$ , pero no ambos sucesos ocurren, por

$$A \Delta B$$

y se lee: exactamente uno de los sucesos  $A$  y  $B$ , diferencia simétrica de  $A$  y  $B$  (fig. 10).

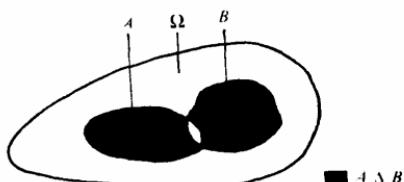


Figura 10

Ya que la operación  $\Delta$  se puede expresar sobre la base de la relación

$$A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cap \bar{B}) \cup (B \cap \bar{A}) \quad (18)$$

mediante las operaciones  $\cap$ ,  $\cup$  y  $\neg$ , renunciamos también a otras discusiones al respecto. Solo queremos señalar que se cumple la conmutatividad para la operación  $\Delta$ .

## 1.4 Álgebras de sucesos

Un álgebra de sucesos es un conjunto de sucesos aleatorios que, hablando sin mucho rigor, contiene, además de los sucesos interesados directamente en relación con un experimento aleatorio, a todos aquellos que resultan de estos mediante la aplicación de las operaciones tratadas. La fijación exacta de este concepto es el contenido de la definición siguiente.

Definición 1. Un conjunto  $A$  de sucesos aleatorios se llama un álgebra de sucesos, si posee las propiedades siguientes:

1. El suceso seguro pertenece a  $A$ :  $\Omega \in A$ .

2. Si dos sucesos aleatorios pertenecen a  $\mathbf{A}$ , este contiene también su suma:

$$A \in \mathbf{A}, B \in \mathbf{A} \Rightarrow A \cup B \in \mathbf{A}.$$

3. Para todo suceso aleatorio perteneciente a  $\mathbf{A}$ , este contiene también al suceso complementario:

$$A \in \mathbf{A} \Rightarrow \bar{A} \in \mathbf{A}.$$

Si  $\mathbf{A}$  contiene infinitos elementos, posee también la propiedad siguiente:

4. Para toda sucesión de sucesos aleatorios perteneciente a  $\mathbf{A}$ , este contiene también su suma:

$$A_i \in \mathbf{A} \quad (i=1, 2, \dots) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathbf{A}.$$

De las propiedades mencionadas en la definición 1 resultan fácilmente otras propiedades.

**Corolario.** Sea  $\mathbf{A}$  un álgebra de sucesos. Entonces  $\mathbf{A}$  posee además las propiedades siguientes:

1. El suceso imposible pertenece a  $\mathbf{A}$ :  $\emptyset \in \mathbf{A}$ .

2. Si dos sucesos aleatorios pertenecen a  $\mathbf{A}$ , este contiene también su producto, su diferencia y su diferencia simétrica:

$$A \in \mathbf{A}, B \in \mathbf{A} \Rightarrow A \cap B \in \mathbf{A}, A \setminus B \in \mathbf{A}, A \Delta B \in \mathbf{A}.$$

3. Para toda sucesión de sucesos aleatorios pertenecientes a  $\mathbf{A}$ , este contiene también su producto:

$$A_i \in \mathbf{A} \quad (i=1, 2, \dots) \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathbf{A}.$$

#### Demostración

1. Se cumple  $\bar{\Omega} = \emptyset$  (ver 1.3 (10)). De las propiedades 1 y 3 del álgebra de sucesos resulta que  $\emptyset \in \mathbf{A}$ .

2. Se cumplen las siguientes identidades:

$$A \cap B = \overline{\bar{A} \cup \bar{B}} \quad (\text{ver 1.3 (13)}),$$

$$A \setminus B = A \cap \bar{B} \quad (\text{ver 1.3 (17)}),$$

$$A \Delta B = (A \cap \bar{B}) \cup (B \cap \bar{A}) \quad (\text{ver 1.3 (18)}).$$

Si  $A$  y  $B$  son elementos del álgebra de sucesos  $\mathbf{A}$ , entonces resulta, sobre la base de las propiedades 2 y 3 del álgebra de sucesos, que  $A \cap B \in \mathbf{A}$  y de aquí (aplicando de nuevo las propiedades 2 y 3), que

$$A \setminus B \in \mathbf{A} \text{ y } A \Delta B \in \mathbf{A},$$

$$A \cap B \in \mathbf{A} \text{ y } A \Delta B \in \mathbf{A}.$$

3. Se cumple  $\bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} \bar{A}_i$  (ver 1.3 (12).) Si  $A_i$  ( $i=1, 2, \dots$ ) son elementos del álgebra de sucesos  $\mathbf{A}$ , entonces resulta a consecuencia de la propiedad 3 del álgebra de sucesos  $\bar{A}_i \in \mathbf{A}$  ( $i=1, 2, \dots$ ).

Considerando la propiedad 4 se obtiene  $\bigcup_{i=1}^{\infty} \bar{A}_i \in \mathbf{A}$ , y por último, en virtud de la propiedad 3  $\bigcup_{i=1}^{\infty} \bar{A}_i \in \mathbf{A}$ , es decir, por la relación dada al principio se cumple  $\bigcap_{i=1}^{\infty} A_i \in \mathbf{A}$ .

Un álgebra de sucesos es, por consiguiente, un conjunto de sucesos aleatorios, con la propiedad de que la aplicación de las operaciones introducidas en 1.3 a los elementos de este conjunto, proporcionan siempre elementos de este conjunto.

Concluimos este epígrafe con la definición del llamado suceso elemental y con una observación sobre la estructura matemática del álgebra de sucesos.

**Definición 2.** Sea  $A$  un álgebra de sucesos. Un suceso  $A \in A$  se llama *suceso elemental* (con respecto a  $A$ ) si no existe un suceso  $B \in A$ ,  $B \neq \emptyset$  y  $B \neq A$ , tal que se cumpla  $B \subseteq A$ . En caso contrario  $A$  se llama *suceso compuesto*.



**Corolario.** Las siguientes proposiciones son equivalentes:

1.  $A \in A$  es un suceso elemental.
2.  $A \in A$  no se puede representar de la forma  $A = B \cup C$  con  $B \in A$ ,  $C \in A$ ,  $B \neq A$  y  $C \neq A$ .
3.  $A \in A$  está constituido de modo que para todo  $B \in A$  se cumple  $A \cap B = \emptyset$  o  $A \subseteq B$ .

Desde el punto de vista de la estructura matemática, un álgebra de sucesos es un álgebra de Boole. Antes de fundamentar esto recordemos la definición de un álgebra de Boole.

**Definición 3.** Sea  $M$  un conjunto sobre el cual están definidas dos operaciones  $+$  y  $\cdot$  (es decir, funciones que asocian a cada dos elementos  $x \in M$  y  $y \in M$  los elementos  $x + y$  y  $x \cdot y$  pertenecientes a  $M$ ).  $M$  se llama un *álgebra de Boole*, si se satisfacen las proposiciones siguientes para cualesquiera elementos  $x, y, z$  de  $M$ :

1.  $x + y = y + x$ ,  $x \cdot y = y \cdot x$  (comutatividad).
2.  $x + (y + z) = (x + y) + z$ ,  $x \cdot (y \cdot z) = (x \cdot y) \cdot z$  (asociatividad).
3.  $x + (x \cdot y) = x$ ,  $x \cdot (x + y) = x$  (absorción).
4.  $x + (y \cdot z) = (x + y) \cdot (x + z)$  (distributividad).
5. Existen elementos  $0$  y  $e$  en  $M$  con  $x \cdot 0 = 0$  y  $x + e = e$ .
6. Para todo  $x \in M$  existe un  $x' \in M$  (el llamado complemento de  $x$ ) con  $x \cdot x' = 0$  y  $x + x' = e$ .

**Corolario 3.** Toda álgebra de sucesos es un álgebra de Boole.

**Demuestra.** Como operación  $+$  empleamos  $\cup$  y como operación  $\cdot$ ,  $\cap$  sobre un álgebra de sucesos  $A$ . Entonces se cumplen las proposiciones 1 hasta 4 de la definición 3. Como elemento neutro respecto a la adición ( $+$ ) utilizamos el suceso imposible  $\emptyset$ ; como elemento neutro de la multiplicación ( $\cdot$ ), el suceso seguro y, por último, empleamos como complemento de  $A \in A$  el suceso complementario  $\bar{A}$  correspondiente a  $A$ . Estos elementos poseen las propiedades exigidas en la definición 3 y pertenecen todos a  $A$ . Con esto  $A$  es, por tanto, un álgebra de Boole.

## 1.5 Álgebras de sucesos y álgebras de conjuntos

Ahora estudiaremos la estrecha relación que existe entre los sucesos aleatorios y los conjuntos, más exactamente entre las álgebras de sucesos y las álgebras de conjuntos. Para ello recordemos la definición de un álgebra de conjuntos.

**Definición 1.** Un sistema  $A$  de subconjuntos de un conjunto universo  $\Omega$  se llama un álgebra de conjuntos (sobre  $\Omega$ ), si posee las propiedades siguientes:

1. El conjunto universo  $\Omega$  pertenece a  $A$ :  $\Omega \in A$ .
2. Si dos subconjuntos de  $\Omega$  pertenecen a  $A$ , este contiene también su unión:  
$$A \in A, B \in A \Rightarrow A \cup B \in A.$$
3. Para todo subconjunto de  $\Omega$  perteneciente a  $A$ , este contiene también su complemento respecto al conjunto universo:  
$$A \in A \Rightarrow \bar{A} \in A.$$

Si, además, la siguiente condición 4 se satisface, entonces  $\mathbf{A}$  se llama una  $\sigma$ -álgebra de subconjuntos de  $\Omega$  y el par  $[\Omega, \mathbf{A}]$  se llama un espacio medible.

4. Para toda sucesión de subconjuntos pertenecientes a  $\mathbf{A}$ , este contiene también su unión:

$$A_i \in \mathbf{A} \quad (i = 1, 2, \dots) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathbf{A}.$$

**Corolario 1.** Toda álgebra de conjuntos es un álgebra de Boole.

**Demostración.** Se desarrolla análoga a la demostración del corolario 3 (1.4).

El siguiente teorema de M.H. Stone proporciona la relación anunciada entre álgebras de sucesos y álgebras de conjuntos.

**Teorema 1.** Para toda álgebra de sucesos se puede indicar un álgebra de conjuntos isomorfa.

Tenemos que renunciar a la demostración de este profundo teorema, pero todavía queremos explicar un poco su contenido.

Si  $\mathbf{A}$  es un álgebra de sucesos, entonces existe un conjunto universo  $\tilde{\Omega}$  y un álgebra  $\tilde{\mathbf{A}}$  de subconjuntos de este conjunto  $\tilde{\Omega}$  con las propiedades siguientes:

1. Existe una aplicación biunívoca de  $\mathbf{A}$  sobre  $\tilde{\mathbf{A}}$ .
2. Al suceso seguro  $\Omega$  le corresponde el conjunto universo  $\tilde{\Omega}$  y al suceso imposible el conjunto vacío.
3. Si designamos con  $\tilde{C}$  el conjunto ( $\in \tilde{\mathbf{A}}$ ) asociado al suceso  $C \in \mathbf{A}$ , entonces a la suma de los sucesos  $A$  y  $B$  (es decir, al suceso  $A \cup B$ ) le corresponde la unión de los conjuntos  $\tilde{A}$  y  $\tilde{B}$  (es decir, el subconjunto  $\tilde{A} \cup \tilde{B}$  de  $\tilde{\Omega}$ ), al producto de los sucesos  $A$  y  $B$  (es decir, al suceso  $A \cap B$ ), la intersección de los conjuntos  $\tilde{A}$  y  $\tilde{B}$  (es decir, el subconjunto  $\tilde{A} \cap \tilde{B}$  de  $\tilde{\Omega}$ ), y al suceso  $\tilde{A}$  el conjunto complementario de  $\tilde{A}$  respecto a  $\tilde{\Omega}$  (es decir, el subconjunto  $\tilde{A}$  de  $\tilde{\Omega}$ ).
4. Si a la ocurrencia del suceso  $A$  ( $\in \mathbf{A}$ ) está siempre unida también la ocurrencia del suceso  $B$  ( $\in \mathbf{A}$ ) (es decir, se cumple  $A \subseteq B$ ), entonces  $\tilde{A}$  es un subconjunto de  $\tilde{B}$  (es decir, se cumple  $\tilde{A} \subseteq \tilde{B}$ ).

Por tanto, podemos considerar siempre en lugar de un álgebra de sucesos  $\mathbf{A}$ , el álgebra de conjuntos isomorfa existente según el teorema anterior, y saber cómo las operaciones entre los sucesos aleatorios se expresan como operaciones entre los conjuntos asociados. (Por lo demás, hemos ya anticipado esto mediante el uso de los mismos símbolos para las operaciones. Con esto queda claro que las reglas de cálculo para operar con sucesos aleatorios siempre llevan implícitas las reglas de cálculo para operar con conjuntos, y viceversa.) En las exposiciones posteriores no partiremos en muchas ocasiones de un álgebra de sucesos, sino del álgebra de conjuntos isomorfa a ella, sobre la base del teorema de M.H. Stone. Aquí supondremos siempre que se trata de una  $\sigma$ -álgebra. Además, queremos simplificar la escritura, de modo que designaremos al álgebra de sucesos y a la  $\sigma$ -álgebra correspondiente con el mismo símbolo  $\mathbf{A}$ . De acuerdo con esto, nombraremos a los sucesos y a los conjuntos asociados con el mismo símbolo; en particular, designaremos también con  $\Omega$  al conjunto universo asociado al suceso seguro  $\Omega$  (cuyos elementos se nombran muchas veces sucesos elementales).

Por tanto, el punto de partida de nuestras consideraciones posteriores será un álgebra de sucesos  $\mathbf{A}$  o un espacio medible  $[\Omega, \mathbf{A}]$ .

## 2. Probabilidad

En este capítulo nos dedicaremos al *concepto probabilidad*, que constituye el concepto central y fundamental de la Teoría de probabilidades y también de la Estadística matemática. Aquí caracterizamos al concepto probabilidad mediante axiomas, de acuerdo con un procedimiento usual hoy en día en la matemática moderna (epígrafe 2.4). Para la formación del sistema de axiomas partiremos de las propiedades comunes de la *frecuencia relativa* (epígrafe 2.1) y del así llamado *concepto clásico de probabilidad* (epígrafes 2.2 y 2.3). El concepto clásico de probabilidad se basa en la –en realidad no universalmente aplicable– definición clásica de probabilidad, que en realidad no es universalmente aplicable, y según la cual la probabilidad de un suceso aleatorio es igual al cociente del número de resultados del experimento “convenientes” para el suceso observado, entre el número total de posibles resultados; en una relación semejante se dice que un resultado del experimento es conveniente para un suceso, cuando este implica la ocurrencia del suceso considerado. Las consideraciones sobre la frecuencia relativa deben convencernos, en particular, de que el grado de indeterminación de la ocurrencia de un suceso aleatorio se puede concebir siempre de forma objetiva mediante un número. En este contexto llamamos la atención de que el concepto probabilidad utilizado en el lenguaje común muestra con frecuencia caracteres subjetivos y que con este sólo se intenta dar en muchas ocasiones una proposición cualitativa con respecto al propio convencimiento de la ocurrencia de una situación determinada.

Se calcularon probabilidades antes de que existiera una construcción axiomática del Cálculo de probabilidades (por ejemplo, en el marco de la estadística poblacional, en problemas de aseguramiento y también en juegos de azar). No obstante, el desarrollo impenetrable de la técnica y de las ciencias naturales desde el comienzo de nuestro siglo situó al cálculo de probabilidades exigencias elevadas. De aquí se desprendió la necesidad de construir el Cálculo de probabilidades, y con esto la Estadística matemática, como una disciplina matemática rigurosamente fundamentada. La solución de este problema, uno de los 23 grandes problemas de la matemática nombrados por el famoso matemático alemán D. Hilbert (1862-1943) en el Segundo Congreso Internacional de Matemáticos en París (1900), fue lograda por el importante matemático soviético A.N. Kolmogorov (nacido en 1903), quien publicó en 1933 una construcción axiomática de Cálculo de probabilidades, que se ha convertido en la base de todos los libros de texto modernos existentes, sobre la Teoría de probabilidades.

Es interesante que D. Hilbert en su conferencia en el año 1900 en París considerara al Cálculo de probabilidades como un capítulo de la física, en el cual los métodos matemáticos desempeñan un papel sobresaliente. Solo por medio de la fundamentación axiomática del Cálculo de probabilidades y la explicación de los conceptos fundamentales ligados a este por A.N. Kolmogorov se integra el cálculo de probabilidades al edificio de la matemática de forma armónica y como una valiosa disciplina especial.

## 2.1 Frecuencia relativa

Designemos por  $A$  un suceso aleatorio que está en relación con un experimento aleatorio cualquiera (por ejemplo,  $A$  puede ser obtener un 6 cuando se tira un dado una sola vez). Repitamos este experimento  $n$ -veces, independientemente una vez de otra, y contemos cuántas veces ocurre el suceso  $A$  en estos experimentos. Si  $A$  ocurre en total  $m$  veces, entonces  $m$  se llama *frecuencia absoluta* de  $A$  y  $\frac{m}{n}$ , *frecuencia relativa* de  $A$  en estos  $n$  experimentos.

En general, queremos designar la frecuencia absoluta de  $A$  en  $n$  experimentos con  $F_n(A)$  y la frecuencia relativa de  $A$  en  $n$  experimentos, con  $f_n(A)$ . Los valores para la frecuencia absoluta  $F_n(A)$  de un suceso  $A$  en  $n$  experimentos, pueden ser los  $n+1$  números  $0, 1, 2, \dots, n-1, n$  y para la frecuencia relativa  $f_n(A)$ , los números  $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$ . La frecuencia absoluta o relativa en una serie de experimentos concreta no se puede predecir con seguridad; las frecuencias absoluta y relativa son medidas dependientes de la casualidad, llamadas variables aleatorias (nosotros las clasificaremos más tarde como variables aleatorias discretas y determinaremos la distribución de probabilidad que les pertenece).

Seguidamente escribiremos algunas propiedades de la frecuencia relativa, cuya demostración dejamos al lector.

### Corolario 1

1.  $0 \leq f_n(A) \leq 1$ .
2.  $f_n(\Omega) = 1$ .
3.  $f_n(A \cup B) = f_n(A) + f_n(B)$  para  $A \cap B = \emptyset$ .
4.  $f_n(\emptyset) = 0$ .
5.  $f_n(\bar{A}) = 1 - f_n(A)$ .
6.  $f_n(A \cup B) = f_n(A) + f_n(B) - f_n(A \cap B)$ .
7. De  $A \subseteq B$  resulta  $f_n(A) \leq f_n(B)$ .

Observemos en relación con las propiedades 2 y 4, que de  $f_n(A) = 1$  o  $f_n(A) = 0$  no se puede deducir que  $A$  sea un suceso seguro o imposible.

Podemos concebir la correspondencia  $A \rightarrow f_n(A)$  ( $n$  es un número natural fijo) como una función que a cada suceso aleatorio  $A$ , que está en relación con el experimento aleatorio observado, le hace corresponder un número situado entre cero y uno, mostrándose las propiedades principales de esta función en el corolario 1. El dominio de definición de esta

función es, por tanto, un conjunto de sucesos aleatorios; queremos suponer siempre que se trata de un álgebra de sucesos.

En relación con el corolario 1 se debe hacer hincapié en una cuestión importante para la forma de proceder en la caracterización axiomática del concepto probabilidad: toda función real  $f$  definida sobre un álgebra de sucesos que posea las propiedades 1, 2 y 3, posee también las propiedades 4, 5, 6 y 7. Aquí queremos demostrar esto solo en un ejemplo; mostraremos que de las propiedades 2 y 3 resulta la propiedad 5: se cumple  $A \cap \bar{A} = \emptyset$  y por la propiedad 3,  $f(A \cup \bar{A}) = f(\bar{A}) + f(A)$ . A cada causa de que  $A \cup \bar{A} = \Omega$  se cumple, por la propiedad 2, la relación  $f(A \cup \bar{A}) = 1$ . Luego, se cumple  $1 = f(\bar{A}) + f(A)$ , es decir, se cumple  $f(\bar{A}) = 1 - f(A)$ .

Analizaremos ahora hasta dónde la frecuencia relativa de un suceso (en una serie de  $n$  repeticiones de un mismo experimento, realizadas independientemente una de otra), es una medida apropiada para el grado de indeterminación de la ocurrencia de este suceso.

Para determinar un valor concreto de la frecuencia relativa se tiene que realizar primero una serie de experimentos semejante; por lo demás se obtendrá generalmente un valor distinto al repetir la serie de experimentos considerada. Pero si se llevan a cabo largas series de repeticiones independientes de un mismo experimento y se indaga cada vez la frecuencia relativa del suceso aleatorio considerado, se comprueba que estos números se diferencian poco unos de otros, es decir, que la frecuencia relativa muestra una cierta *estabilidad*. Luego, las frecuencias relativas del suceso  $A$  varían ligeramente, por lo general alrededor de un cierto valor que frecuentemente desconocemos. Queremos llamar a este valor la *probabilidad* del suceso  $A$ . Está claro que no podemos calcular la probabilidad de un suceso por esta vía, sino solo obtener un valor estimado para esa probabilidad. Sin embargo, con esto hemos logrado el convencimiento de que el grado de indeterminación de la ocurrencia de un suceso aleatorio se puede caracterizar de forma objetiva mediante un número.

**Ejemplo.** Tomamos este ejemplo de la literatura. Científicos significativos como, por ejemplo, el Conde de Buffon (1707-1788), creador de un método teórico-probabilístico para la determinación aproximada del número  $\pi$ , y K. Pearson (1857-1936), fundador de una famosa escuela en la rama de la Estadística matemática en Inglaterra, estudiaron el efecto de la estabilización de la frecuencia relativa, en el ejemplo de la tirada de la moneda, entre otros. Sea  $A$  el suceso "número arriba".

	Número de tiradas de la moneda: $n$	Frecuencia absoluta de $A: F_n(A)$	Frecuencia relativa de $A: f_n(A) = \frac{F_n(A)}{n}$
DE BUFFON	4 040	2 048 (2 020)	0,5080
K. PEARSON	12 000	6 019 (6 000)	0,5016
K. PEARSON	24 000	12 012 (12 000)	0,5005

Esperamos que aproximadamente en la mitad de todas las tiradas de la moneda ocurra el suceso  $A$ . En la tercera columna de la tabla anterior hemos indicado los valores esperados entre paréntesis. La tabla muestra claramente que lo que esperábamos se satisface tanto mejor cuanto mayor es el número de tiradas realizadas.

Por último, queremos analizar la interrogante de si para *toda* serie de experimentos concreta, la sucesión ( $f_n(A)$ ) de las frecuencias relativas  $f_n(A)$  de un suceso  $A$  converge hacia un límite común  $f(A)$  cuando  $n \rightarrow \infty$ . (Si este fuera el caso se podría definir sencillamente

la probabilidad de un suceso aleatorio como el límite de la sucesión de las frecuencias relativas.) Pero esto no es así. Por un lado, solo es posible crear una sucesión *finita* de frecuencias relativas, de modo que no se puede decidir nunca si existe la convergencia de la sucesión investigada, convergencia entendida en el sentido de la de las sucesiones numéricas. Por otro lado, aún si no se presta atención a esta circunstancia, se puede pensar también que no tiene que existir una convergencia de la sucesión  $\{f_n(A)\}$ . Si se cumpliera que  $\lim_{n \rightarrow \infty} f_n(A) = f(A)$ , entonces existiría para todo  $\epsilon > 0$  un número natural  $n_0$ , tal que

$|f_n(A) - f(A)| < \epsilon$  para todo  $n \geq n_0$ . Pero recurriendo al ejemplo anterior es fácil imaginar que el suceso "número arriba" no ocurre ni una sola vez en series de experimentos muy largas, de modo que la inecuación  $|f_n(A) - f(A)| < \epsilon$  para un número suficientemente pequeño  $\epsilon > 0$  no se cumple para todo  $n$  a partir de un cierto índice  $n_0$ . (A decir verdad un caso semejante nos parece muy "improbable".)

Una formulación matemática precisa del efecto de estabilización de la frecuencia relativa se realiza más tarde por otro camino con el tratamiento de la Ley de los Grandes Números.

## 2.2 Definición clásica de probabilidad

Mucho antes de la fundamentación axiomática del Cálculo de probabilidades, se calcularon probabilidades de sucesos aleatorios. La definición de probabilidad en la cual se basaban dichos cálculos se conoce hoy como definición clásica de probabilidad que estudiaremos en este epígrafe.

Sea el punto de partida un experimento aleatorio con un número *finito* de resultados igualmente posibles, es decir, que no se diferencian con respecto al grado de indeterminación de la ocurrencia. Todo suceso aleatorio  $A$  en relación con el experimento aleatorio considerado, se puede caracterizar por la enumeración de aquellos resultados que son favorables para este suceso, es decir, que provocan su ocurrencia. Si designamos con  $g(A)$  su número y con  $k (< \infty)$  el de todos los resultados, entonces la razón de  $g(A)$  y  $k$  proporciona una idea sobre el grado de seguridad de la aparición del suceso aleatorio  $A$ . En el marco de la llamada *definición clásica de probabilidad*, a este cociente se le llama probabilidad del suceso aleatorio  $A$  y se designa con  $P(A)$ :

$$P(A) = \frac{g(A)}{k} = \frac{\text{número de los resultados favorables para } A}{\text{número total de los resultados}}. \quad (1)$$

**Observación.** Con frecuencia, en la literatura se encuentran formulaciones que solo se diferencian de esta en que en lugar de la palabra *resultados* se utilizan las palabras *posibilidades* o *casos*. La fórmula (1) se debe al matemático francés P.S. Laplace (1749-1827); el principio sobre el cual se basa la fórmula (1) se nombra con frecuencia *Principio de los casos igualmente posibles de Laplace*.

**Ejemplo.** En un recipiente se encuentran 150 piezas troqueladas, de las cuales 21 no tienen una medida adecuada. El experimento aleatorio consiste en la extracción de una pieza, teniendo cada una de ellas la misma oportunidad de ser tomada. Calculemos la probabilidad de que la pieza extraída aleatoriamente de esta forma, tenga las medidas correctas (suceso  $A$ ).

Número de resultados posibles: 150

Número de los resultados favorables para  $A$ :  $150 - 21 = 129$

Con esto se obtiene

$$P(A) = \frac{g(A)}{k} = \frac{129}{150} = \frac{43}{50} = 0,86 = 86\%.$$

La aplicación de la definición clásica de probabilidad está permitida solo en el marco de determinados experimentos aleatorios. Queremos reflexionar sobre cómo se reflejan las condiciones de los experimentos aleatorios en propiedades (adicionales) de las álgebras de sucesos. Designemos con  $\mathbf{A}$  al álgebra de sucesos correspondiente a un experimento aleatorio con un número finito de resultados  $A_1, A_2, \dots, A_k$  igualmente posibles, que deben concebirse como sucesos elementales de dicha álgebra de sucesos. Todo suceso aleatorio arbitrario  $A \in \mathbf{A}$ ,  $A \neq \emptyset$ , se puede expresar como la suma de aquellos sucesos elementales  $A_i$  que implican a  $A$ , es decir, para los cuales se cumple que  $A_i \subseteq A$ . Para hallar la probabilidad del suceso  $A$  es necesario conocer solo, junto al número total  $k$  de los sucesos elementales, el número de los sucesos elementales  $A_i$  que implican a  $A$ . Con esto está claro que a cada suceso aleatorio  $A \in \mathbf{A}$  está asociado de forma unívoca mediante (1) un número real, o sea, que por medio de (1) está definida una función real sobre  $\mathbf{A}$ . En particular se cumple a causa de

$$g(A_1) = g(A_2) = \dots = g(A_k) = 1$$

la relación

$$P(A_1) = P(A_2) = \dots = P(A_k) = \frac{1}{k}, \quad (2)$$

es decir, la condición de que los resultados sean igualmente posibles se refleja en que los sucesos elementales  $A_i (i=1, 2, \dots, k)$  tienen la misma probabilidad.

A continuación enunciaremos algunas propiedades y reglas de cálculo para el concepto clásico de probabilidad, y con esto para la función  $A \rightarrow P(A)$  sobre  $\mathbf{A}$  dada por (1), cuya demostración dejaremos al lector (ver 2.1, corolario 1).

Corolario 1

1.  $0 \leq P(A) \leq 1$ .
2.  $P(\Omega) = 1$ .
3.  $P(A \cup B) = P(A) + P(B)$  para  $A \cap B = \emptyset$ .
4.  $P(\emptyset) = 0$ .
5.  $P(\bar{A}) = 1 - P(A)$ .
6.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .
7. De  $A \subseteq B$  resulta  $P(A) \leq P(B)$ .

Como suplemento de las propiedades 2 y 4 aclaramos que de  $P(A) = 1$  o  $P(A) = 0$  se deduce que  $A = \Omega$  o  $A = \emptyset$ . Un suceso aleatorio  $A$  tiene, por consiguiente, la probabilidad uno o cero si y solo si es un suceso seguro o imposible.

Además, se debe llamar la atención de que es suficiente demostrar las proposiciones 1 hasta 3, ya que como fue explicado en el epígrafe 2.1, toda función real definida sobre un álgebra de sucesos que posea las propiedades 1 hasta 3, posee también las propiedades 4 hasta 7.

A la definición clásica de probabilidad, corresponde una significación especial, porque sobre esta base se pueden calcular probabilidades. El cálculo de las probabilidades que nos

interesan, o sea, el cálculo del número de los casos posibles y del de los convenientes en cada ocasión, se efectúa, por lo general, con los métodos de la combinatoria (ver MFL, Tomo 1,3.6). Esto no es siempre muy sencillo.

### Ejemplos

1. Calculemos la probabilidad para ganar la lotería \* en "5 de 35" (suceso  $G$ ), es decir, para acertar tres números (suceso  $A$ ), cuatro (suceso  $B$ ) o cinco (suceso  $C$ ). Se cumple

$$k = \binom{35}{5} = \frac{35 \cdot 34 \cdot 33 \cdot 32 \cdot 31}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 324\,632,$$

$$g(A) = \binom{5}{3} \binom{30}{2} = \frac{5 \cdot 4}{1 \cdot 2} \cdot \frac{30 \cdot 29}{1 \cdot 2} = 4\,350,$$

$$g(B) = \binom{5}{4} \binom{30}{1} = \frac{5}{1} \cdot \frac{30}{1} = 150,$$

$$g(C) = \binom{5}{5} \binom{30}{0} = 1 \cdot 1 = 1.$$

Con esto obtenemos

$$P(A) = \frac{g(A)}{k} = \frac{4\,350}{324\,632} \approx 0,0134 \text{ (probabilidad de obtener tres),}$$

$$P(B) = \frac{g(B)}{k} = \frac{150}{324\,632} \approx 0,0005 \text{ (probabilidad de obtener cuatro),}$$

$$P(C) = \frac{g(C)}{k} = \frac{1}{324\,632} \approx 0,000\,003 \text{ (probabilidad de obtener cinco).}$$

Ahora, se cumple que  $G = A \cup B \cup C$  siendo los sucesos  $A, B$  y  $C$  mutuamente excluyentes dos a dos. Por tanto, se cumple que  $P(G) = P(A) + P(B) + P(C)$  (ver corolario 1, proposición 3) y obtenemos finalmente  $P(G) \approx 0,014$  (probabilidad de una ganancia).

2. Se eligen de forma aleatoria  $n$  personas (aleatoria en el sentido de que cada persona tiene la misma oportunidad de ser elegida) de un conjunto grande de estas (por ejemplo, del conjunto de los habitantes actuales de la ciudad de Dresden) y se anotan las fechas de sus cumpleaños. Nos interesaremos por la probabilidad de que por lo menos dos de estas personas cumplan años el mismo día (suceso  $A$ ). En la solución de este problema suponemos adicionalmente que las personas que han nacido el 29 de febrero de un año bisiesto no han sido elegidas de modo que tenemos que calcular en total solo con 365 días. Además, suponemos que la probabilidad de que una persona elegida de forma aleatoria cumpla años un día determinado, es igual para los 365 días, luego es igual a  $\frac{1}{365}$ .

Indagamos primero el número  $k$  de los posibles resultados del experimento, consistiendo un posible resultado en elegir  $n$  días (no necesariamente distintos) de los 365. El número de estas posibilidades es igual (considerando la sucesión) a  $k = \frac{365 \cdot 365 \cdots 365}{n \text{ factores}} = 365^n$  (por lo demás se cumple que para  $n > 4$ ,  $k = 365^n$  es mayor que un billón).

\* Juego de lotería televisivo en la República Democrática Alemana.

Para el cálculo de la probabilidad buscada tenemos que averiguar ahora el número  $g(A)$  de los resultados favorables para  $A$ . Es mucho más conveniente calcular primero el número  $g(\bar{A})$  de los desenlaces favorables para  $\bar{A}$ . El suceso  $\bar{A}$  consiste en que entre las  $n$  personas elegidas no haya dos o más que cumplan años el mismo día, es decir, en que cada una de las  $n$  personas cumpla años un día distinto al de todos los demás. El número de los resultados favorables para  $\bar{A}$  es igual (considerando de nuevo la sucesión) a

$$g(\bar{A}) = \frac{365 \cdot 364 \cdots (365-(n-1))}{n \text{ factores}} = \binom{365}{n} n!$$

De aquí obtenemos que

$$P(\bar{A}) = \frac{g(\bar{A})}{k} = \frac{\binom{365}{n} n!}{365^n},$$

de donde resulta, según una fórmula anterior (ver corolario 1, proposición 5), la probabilidad buscada

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{\binom{365}{n} n!}{365^n}.$$

En la tabla siguiente damos, para distintas  $n$ , la probabilidad de que entre  $n$  personas, por lo menos dos cumplan años el mismo día.

$n$	10	20	22	23	24	30	40	50
$P(A)$	0,12	0,41	0,48	0,51	0,54	0,71	0,89	0,97

(Para  $n > 365$  se obtiene naturalmente que  $P(A) = 1$ .)

## 2.3 Definición geométrica de probabilidad

La fórmula (1) indicada en el epígrafe 2.2 para el cálculo de probabilidades de sucesos aleatorios es solo aplicable cuando el experimento aleatorio considerado posee un número finito de resultados igualmente posibles. Ahora, existe una serie de experimentos aleatorios que no satisfacen estas condiciones, pero para los cuales se puede indicar, de forma semejante, una fórmula para el cálculo de las probabilidades que nos interesan. Siempre y cuando pueda interpretarse el experimento aleatorio como el modelo de la tirada aleatoria de un punto sobre un dominio básico  $E$  cualquiera del espacio euclíadiano  $n$ -dimensional, donde la palabra aleatoria debe entenderse de modo que:

1. El punto lanzado pueda caer sobre todo punto arbitrario de  $E$  y

2. los sucesos  $A$  y  $B$ , a los cuales corresponden dominios parciales de igual medida (por ejemplo, intervalos de igual longitud, conjuntos de puntos en el plano de igual área, cuerpos en el espacio tridimensional de igual volumen), posean también la misma probabilidad, se calcula la probabilidad de un suceso  $A$ , que esté en relación con un experimento semejante, según la fórmula

$$P(A) = \frac{m(A)}{m(E)} = \frac{\text{Medida del dominio parcial de } E \text{ correspondiente al suceso } A}{\text{Medida del dominio básico } E} \quad (1)$$

(definición geométrica de probabilidad (fig. 11).

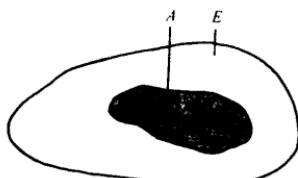


Figura 11

Por tanto, la probabilidad de un suceso es independiente de la configuración especial y de la situación del dominio parcial que representa al suceso  $A$ ; ella es proporcional a la medida (o sea, proporcional a la longitud, al área, al volumen) de este dominio parcial. Formulado de otra manera, la probabilidad de un suceso es, por consiguiente, igual a la razón de las medidas del dominio parcial conveniente para el suceso y del dominio básico. En esta formulación de la definición geométrica de probabilidad se muestra claramente la analogía con la definición clásica de probabilidad. El principio de los casos igualmente posibles de Laplace, sobre el cual se basa la definición clásica de probabilidad, se manifiesta en esta definición geométrica al establecer que los sucesos a los cuales corresponden dominios parciales de igual medida poseen la misma probabilidad.

**Ejemplo.** Dos personas acuerdan encontrarse en un lugar determinado entre las 12 pm y la 1 am. Cada una de las personas elige el momento de llegada, independientemente una de otra. Sin embargo, ambas se comprometen a estar con seguridad entre las 12 pm y la 1 am en el lugar acordado; no se hacen indicaciones más precisas con respecto al momento del arribo. Ahora, ellas concertan que en caso necesario, cada una espere a la otra 15 min, pero que después se vaya. Calculemos la probabilidad de que ambas personas se encuentren. Para el cálculo de la probabilidad buscada tomemos por base la definición geométrica de probabilidad.

Designemos los tiempos de llegada de las dos personas con  $x$  y  $y$ , respectivamente (por ejemplo, ambos medidos en minutos y fracciones de minutos después de las 12 pm) y representémoslos como puntos en el plano (fig. 12).

El suceso  $A$ , consistente en que ambas personas se encuentren, es descrito por medio del conjunto  $\{(x, y) : 0 \leq x \leq 60, 0 \leq y \leq 60, |x-y| \leq 15\}$ . De la figura 12 inferimos directamente que

$$m(A) = 60 \cdot 60 - 2 \cdot \frac{45 \cdot 45}{2}, \quad m(E) = 60 \cdot 60$$

y obtenemos con esto para la probabilidad buscada

$$P(A) = \frac{m(A)}{m(E)} = 1 - \left(\frac{3}{4}\right)^2 = \frac{7}{16}.$$

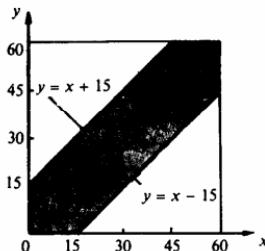


Figura 12

La probabilidad del encuentro con 15 min de espera es, por tanto, algo menor que 0,5. Dejamos al lector que verifique que, por ejemplo, la probabilidad del encuentro con 30 min de espera es igual a 0,75. Además, el lector puede deducir fácilmente una relación general entre la probabilidad del encuentro y el tiempo de espera.

Obsérvese que a los sucesos aleatorios a los cuales corresponde un dominio parcial, que posee una dimensión más pequeña que el dominio básico  $E$  (por ejemplo, un punto sobre una recta numérica, una recta en el plano, un plano en el espacio), les corresponde la probabilidad cero.

La definición geométrica de probabilidad dio motivo en épocas anteriores a todo tipo de falsos entendimientos, equivocos y críticas; esta condujo incluso en cierta medida a un rechazo del cálculo de probabilidades como disciplina científica. Para fundamentar esto se hizo referencia a problemas cuya solución es dependiente del método utilizado, es decir, que conducen a distintos resultados con métodos de solución diferentes. La causa de esto no radica en cualesquiera contradicciones del concepto geométrico de probabilidad, sino en la insuficiente precisión en el planteamiento del problema. Traemos un ejemplo que es conocido en la literatura como la paradoja de Bertrand; este proviene, como otros muchos ejemplos semejantes, del matemático francés J. Bertrand (1822-1900).

**Problema.** En una circunferencia se traza de forma aleatoria (arbitraria) una cuerda. ¿Cuál es la probabilidad de que su longitud supere la del lado de un triángulo equilátero inscrito en la circunferencia (suceso  $A$ )?

**Solución 1.** Fijemos una dirección de la cuerda y observemos un diámetro perpendicular a dicha dirección (fig. 13). El suceso  $A$  ocurre si y solo si la cuerda corta al diámetro entre  $\frac{r}{2}$  y  $\frac{3r}{2}$ .

Luego se cumple

$$P(A) = \frac{m(A)}{m(E)} = \frac{\frac{3r}{2} - \frac{r}{2}}{2r} = \frac{1}{2}.$$

**Solución 2.** Fijemos un punto final de la cuerda sobre la circunferencia, tracemos la tangente a la circunferencia en este punto y dibujemos un triángulo equilátero inscrito en ella con un vértice en dicho punto (fig. 14). El suceso  $A$  ocurre si y solo si la cuerda cae en el sector angular del ángulo del medio. Luego se cumple

$$P(A) = \frac{m(A)}{m(E)} = \frac{\frac{\pi}{3}}{\pi} = \frac{1}{3}.$$

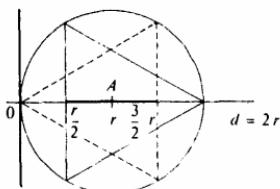


Figura 13

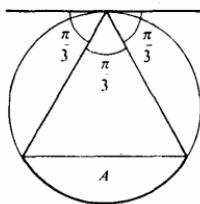


Figura 14

**Solución 3.** La longitud de la cuerda se obtiene de forma unívoca de la situación del punto medio de esta. Si  $\rho$  es la distancia del centro de la circunferencia al punto medio de la cuerda y  $l$  designa la longitud de la cuerda, entonces se cumple que  $l=2\sqrt{r^2-\rho^2}$  (fig. 15). El suceso  $A$  ocurre si y sólo si  $l \geq \sqrt{3}r$  ( $\sqrt{3}r$ =longitud del lado de un triángulo equilátero inscrito en la circunferencia), o sea, si se cumple  $\rho \leq \frac{r}{2}$ . Luego se cumple

$$P(A) = \frac{m(A)}{m(E)} = \frac{\left(\frac{r}{2}\right)^2\pi}{r^2\pi} = \frac{1}{4}.$$

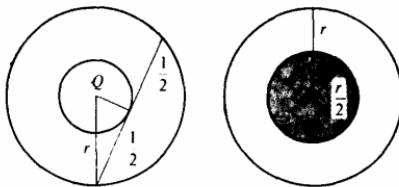


Figura 15

En el planteamiento del problema no está fijado qué se entiende por el trazado *aleatorio* de una cuerda. En las soluciones dadas esto fue concebido cada vez de manera diferente. En la solución 1 se partió del modelo de la tirada aleatoria de un punto sobre un intervalo de la longitud  $2r$ ; en la 2, del lanzamiento aleatorio de un punto sobre un intervalo de la longitud  $\pi$ , y en la 3, de la tirada aleatoria de un punto sobre la superficie de un círculo con radio  $r$ , entendiendo cada vez la palabra *aleatoria* tal como se indica en la definición geométrica de probabilidad. Las tres soluciones dadas no son, por tanto, soluciones del problema anterior, sino de otros 3 problemas distintos entre sí; el problema mismo no es, sin precisión de lo que se entiende por *trazado aleatorio* de una cuerda, soluble en la forma dada.

## 2.4 Definición axiomática de probabilidad

De las reflexiones sobre el efecto de estabilización de la frecuencia relativa extrajimos en el epígrafe 2.1 la conclusión de que el grado de indeterminación de la ocurrencia de un suceso  $A$ , se puede caracterizar de forma objetiva mediante un número, llamado la probabilidad del suceso  $A$  y designado con  $P(A)$ . En los epígrafes 2.2 y 2.3 hemos dado –para el caso en que el experimento aleatorio satisface ciertas propiedades adicionales

(que restringen bastante su aplicación) – fórmulas para el cálculo de probabilidades. Una fórmula aplicable en todos los casos para el cálculo de probabilidades no existe y no puede tampoco existir. Por eso, para la construcción sucesiva del cálculo de probabilidades, queremos tomar por base algunas suposiciones (axiomas) que se traducen en propiedades y reglas de cálculo, relativas al concepto de probabilidad y que reconoceremos como válidas sin demostración. Aquí partiremos naturalmente de las experiencias acumuladas hasta ahora por nosotros, o sea, construiremos el sistema de axiomas del cálculo de probabilidades de las propiedades comunes de la frecuencia relativa y de los conceptos clásico y geométrico de probabilidad.

Para la formulación del sistema de axiomas partiremos de un álgebra de sucesos A. Decimos que sobre A está definida una *probabilidad P* (o una *medida de probabilidad*), si P es una función con las propiedades señaladas en los siguientes axiomas.

**Axioma 1.** A todo suceso aleatorio  $A \in A$  le corresponde de forma única un número  $P(A)$ , la llamada probabilidad de A, y se cumple que

$$0 \leq P(A) \leq 1.$$

Con el axioma 1 se establece, por tanto, el dominio de definición y la imagen de la función P; P es una función real definida sobre un álgebra de sucesos con valores entre cero y uno. El axioma 1 lleva implícito también que todo suceso aleatorio posee una probabilidad bien determinada.

**Axioma 2.** La probabilidad del suceso seguro es igual a uno:

$$P(\Omega) = 1 \text{ (axioma de normación).}$$

El suceso seguro es siempre, según definición, un elemento del álgebra de sucesos A, es decir, un elemento del dominio de definición de la función. El axioma 2 dice que el valor de la función P para el argumento  $\Omega$  es igual a uno.

**Axioma 3.** Dados dos sucesos aleatorios mutuamente excluyentes del álgebra de sucesos considerada, la probabilidad de que ocurra uno de ellos es igual a la suma de las probabilidades de estos sucesos:

$$A \in A, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B) \text{ (axioma de adición).}$$

Observemos al respecto que un álgebra de sucesos al cual pertenezcan los sucesos aleatorios A y B contiene también, según definición, a  $A \cup B$ , o sea, que junto con A y B también  $A \cup B$  pertenece al dominio de definición de la función P.

Utilizando solamente el axioma 3 se puede demostrar con el principio de inducción completa la proposición siguiente:

**Corolario 1.** Dados  $n$  ( $n \geq 2$ ) sucesos aleatorios mutuamente excluyentes dos a dos del álgebra de sucesos considerada, la probabilidad de que ocurra uno de ellos es igual a la suma de las probabilidades de estos sucesos:

$$\left. \begin{array}{l} A_j \in A \quad (j=1, 2, \dots, n), \\ A_i \cap A_k = \emptyset \quad (i \neq k; i, k = 1, 2, \dots, n) \end{array} \right\} \Rightarrow P \left( \bigcup_{j=1}^n A_j \right) = \sum_{j=1}^n P(A_j).$$

Una regla de cálculo correspondiente, para la probabilidad de la suma de un conjunto infinito numerable de sucesos aleatorios incompatibles dos a dos, no se puede demostrar con el axioma 3; no obstante, subordinamos también al concepto general de probabilidad la validez de una regla de cálculo semejante de forma conveniente.

**Axioma 4.** Dado un conjunto infinito numerable de sucesos aleatorios mutuamente excluyentes dos a dos del álgebra de sucesos considerada, la probabilidad de que ocurra uno de ellos es igual a la suma de las probabilidades de estos sucesos:

$$A_j \in \mathbb{A} \quad (j=1,2,\dots), \quad A_i \cap A_k = \emptyset \quad (i \neq k : i, k = 1, 2, \dots), \quad \left\} \Rightarrow P \left( \bigcup_{j=1}^{\infty} A_j \right) = \sum_{j=1}^{\infty} P(A_j). \right.$$

Advertimos que un álgebra de sucesos a la cual pertenezcan los sucesos  $A_j (j=1,2,\dots)$  contiene también, según definición, a  $\bigcup_{j=1}^{\infty} A_j$ , o sea, al igual que  $A_j (j=1,2,\dots)$ , también  $\bigcup_{j=1}^{\infty} A_j$  pertenece al dominio de definición de la función  $P$ . El concepto álgebra de sucesos está fijado de tal modo, que todos los sucesos que aparecen en los axiomas y en las proposiciones del epígrafe 2.5, que se deducen de estos, pertenecen al álgebra de sucesos, es decir, al dominio de definición de la función  $P$ .

La propiedad expresada en el axioma 4 se designa como  $\sigma$ -aditividad de la medida de probabilidad  $P$ . Esta conduce a una propiedad de continuidad en el sentido siguiente.

**Teorema 1.** Sea  $(A_j)$  una sucesión de sucesos aleatorios  $A_j \in \mathbb{A} (j=1,2,\dots)$ .

a) Si se cumple  $A_1 \subseteq A_2 \subseteq \dots$ , entonces  $P \left( \bigcup_{j=1}^{\infty} A_j \right) = \lim_{j \rightarrow \infty} P(A_j)$ .

b) Si se cumple que  $A_1 \supseteq A_2 \supseteq \dots$ , entonces  $P \left( \bigcap_{j=1}^{\infty} A_j \right) = \lim_{j \rightarrow \infty} P(A_j)$ .

No demostraremos este teorema, pero lo comentaremos un poco. Si  $(A_j)$  es una sucesión de subconjuntos (de un conjunto universo  $\Omega$ ), entonces las sucesiones con  $A_1 \subseteq A_2 \subseteq \dots$  y  $A_1 \supseteq A_2 \supseteq \dots$  son convergentes en el sentido del límite algebraico conjuntista, y se cumple que

$$\lim_{k \rightarrow \infty} A_k = \bigcup_{j=1}^{\infty} A_j \text{ y } \lim_{k \rightarrow \infty} A_k = \bigcap_{j=1}^{\infty} A_j$$

respectivamente. Luego, las proposiciones contenidas en el teorema significan la validez de  $P(\lim_{j \rightarrow \infty} A_j) = \lim_{j \rightarrow \infty} P(A_j)$ . Esto es equivalente a la continuidad de  $P$ .

Los axiomas 1 hasta 3 proporcionan que se pueden demostrar en el caso en que se aplique la definición clásica de probabilidad (ver 2.2, coloquio 1, proposiciones 1 hasta 3). Asimismo son válidas proposiciones semejantes para la función  $f_n$ , que hace corresponder a cada suceso aleatorio  $A \in \mathbb{A}$  la frecuencia relativa de la ocurrencia de  $A$  en  $n$  repeticiones realizadas independientes unas de otras del experimento aleatorio observado (ver 2.1, coloquio 1, proposiciones 1 hasta 3). No formularemos como axiomas para el concepto general de probabilidad las otras propiedades comunes establecidas para la frecuencia relativa y el concepto clásico de probabilidad, porque ellas se pueden deducir de los axiomas 1 hasta 3 (ver 2.5). Tampoco exigiremos que  $A$  sea un suceso seguro cuando se cumpla que  $P(A) = 1$ , ya que esta proposición no es verdadera en el marco de la definición geométrica de probabilidad (ver 2.3). En este contexto introduciremos dos conceptos.

**Definición 1.** Si se cumple que  $P(A) = 1$  ( $P(A) = 0$ ), entonces se llama al suceso aleatorio  $A (\in \mathbb{A})$  un *suceso casi seguro* (*suceso casi imposible*.)

A continuación damos las definiciones de dos conceptos frecuentemente utilizados en la teoría de probabilidades.

**Definición 2.** Si  $A$  es un álgebra de sucesos y  $P$  una probabilidad sobre  $A$ , entonces se llama al par  $[A, P]$  una familia de probabilidades.

A causa de la estrecha relación entre las álgebras de sucesos y los espacios medibles, verificada en el epígrafe 1.5, se puede partir también en la introducción axiomática del concepto *probabilidad* de un espacio medida  $[\Omega, A]$ . Entonces se denomina a una función  $P$  definida sobre la  $\sigma$ -álgebra  $A$  de subconjuntos del conjunto universo  $\Omega$ , una medida de probabilidad, si esta posee las propiedades expresadas en los axiomas 1 hasta 4.

**Definición 3.** Si  $[\Omega, A]$  es un espacio medible y  $P$  una medida de probabilidad sobre  $A$ , entonces a la terna  $[\Omega, A, P]$  se le llama *espacio de probabilidad*.

En investigaciones teórico-probabilísticas actuales se parte generalmente de un espacio de probabilidad.

## 2.5 Leyes de cálculo para probabilidades

Formularemos y demostraremos en este epígrafe proposiciones para el cálculo con probabilidades, que resultan directamente de los axiomas del Cálculo de probabilidad y que corresponden a las propiedades 4 hasta la 9 del colorario 1 de los epígrafes 2.1 y 2.2. Aquí hacemos la abstracción de que existe una familia de probabilidades  $[A, P]$ , es decir, que existe un álgebra de sucesos  $A$  sobre la cual está definida una función  $P$  que satisface los axiomas 1 hasta 4. (Naturalmente podemos partir también de un espacio de probabilidad  $[\Omega, A, P]$ , o sea, de un conjunto universo  $\Omega$ , una  $\sigma$ -álgebra de subconjuntos de  $\Omega$  y de una función  $P$  definida sobre  $A$ , que posee las propiedades expresadas en los axiomas 1 hasta 4.)

**Teorema 1.** La probabilidad del suceso imposible es igual a cero.

$$P(\emptyset) = 0. \quad (1)$$

**Demostración.** Se cumple que  $\emptyset \in \Omega$  (ver 1.4, corolario 1, proposición 1), o sea, que el suceso imposible pertenece al dominio de definición de  $P$ . A causa de que  $\emptyset \cap \emptyset = \emptyset$ , se cumple, según el axioma 3, que

$$P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset) = 2P(\emptyset).$$

Como  $\emptyset \cup \emptyset = \emptyset$ , se cumple que  $P(\emptyset \cup \emptyset) = P(\emptyset)$  y con esto que  $P(\emptyset) = 2P(\emptyset)$ , de donde se obtiene (1).

**Teorema 2.** Para todo suceso aleatorio  $A \in A$  se cumple que

$$P(\bar{A}) = 1 - P(A). \quad (2)$$

**Demostración.** Si  $A \in A$ , entonces se cumple también que  $\bar{A} \in A$  (ver 1.4, definición 1), es decir, al igual que  $A$ , pertenece también  $\bar{A}$  al dominio de definición de  $P$ . Ahora, se cumplen las proposiciones  $A \cap \bar{A} = \emptyset$  y  $A \cup \bar{A} = \Omega$  (ver 1.3 (9)). De los axiomas 3 y 2 resulta que  $P(A \cup \bar{A}) = P(A) + P(\bar{A})$  y que  $P(A \cup \bar{A}) = 1$ , de donde se obtiene que  $1 = P(A) + P(\bar{A})$  y con esto (2).

**Teorema 3.** Para sucesos aleatorios cualesquiera  $A \in \mathcal{A}$  y  $B \in \mathcal{A}$  se cumple que

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (3)$$

**Demostración.** Se cumplen las siguientes ecuaciones:

$$A \cup B = A \cup (B \cap \bar{A}) \quad (\text{ver 1.3 (14)}),$$

$$A \cup B = B \cup (A \cap \bar{B}) \quad (\text{ver 1.3 (15)}),$$

$$A \cup B = (A \cap \bar{B}) \cup (B \cap \bar{A}) \cup (A \cap B) \quad (\text{ver 1.3 (16)});$$

donde los sumandos situados a la derecha son en todos los casos mutuamente excluyentes dos a dos (fig. 8). De la aplicación del axioma 3 y del corolario dado a continuación de este se obtiene que

$$P(A \cup B) = P(A) + P(B \cap \bar{A}),$$

$$P(A \cup B) = P(B) + P(A \cap \bar{B}),$$

$$P(A \cup B) = P(A \cap \bar{B}) + P(B \cap \bar{A}) + P(A \cap B).$$

Si formamos la diferencia entre la suma de las dos primeras ecuaciones y la tercera ecuación, se obtiene (3).

**Teorema 4.** Si la ocurrencia del suceso aleatorio  $A \in \mathcal{A}$  implica la ocurrencia del suceso aleatorio  $B \in \mathcal{A}$  (o sea, si se cumple que  $A \subseteq B$ ), entonces se cumple que  $P(A) \leq P(B)$ .

**Demostración.** Se cumple (fig. 16) que

$$B = A \cup (B \cap \bar{A}) \text{ con } A \cap (B \cap \bar{A}) = \emptyset.$$

Del axioma 3 se obtiene que  $P(B) = P(A) + P(B \cap \bar{A})$ . Según el axioma 1 se cumple que  $P(B \cap \bar{A}) \geq 0$ , con lo cual resulta que  $P(B) \geq P(A)$ .

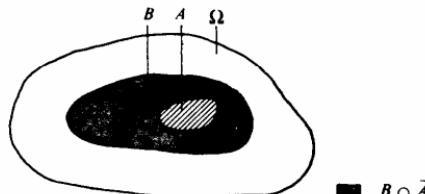


Figura 16

**Teorema 5.** Si el conjunto  $\{A_1, A_2, \dots, A_n, \dots\}$  es un sistema completo de sucesos aleatorios, entonces se cumple que

$$\sum_n P(A_n) = 1.$$

**Demostración.** Según la premisa se cumple (ver 1.3, definición 6) que

$$\bigcup_n A_n = \Omega, \quad A_j \cap A_k = \emptyset \quad (j \neq k).$$

La aplicación del corolario dado a continuación del axioma 3 o la aplicación del axioma 4, proporciona, bajo la consideración del axioma 2, la proposición de este teorema.

### 3. Probabilidad condicionada

Introduciremos en este capítulo el concepto *probabilidad condicionada* (epígrafe 3.1) y obtendremos de esto una fórmula para el cálculo de la probabilidad del producto de sucesos aleatorios (*teorema de la multiplicación*, epígrafe 3.2). Sobre esta base trataremos en el epígrafe 3.3 el concepto *independencia de sucesos aleatorios*, extraordinariamente importante para todo el Cálculo de probabilidades. Por último, estudiaremos dos fórmulas útiles para numerosas interrogantes prácticas, la *fórmula de la probabilidad total* (epígrafe 3.4) y la *fórmula de Bayes* (epígrafe 3.5). En cada ocasión consideraremos un ejemplo en el cual esté presente una situación típica para la aplicación de estas fórmulas.

#### 3.1 Definición de probabilidad condicionada

Partiremos de un experimento aleatorio que nos imaginamos descrito matemáticamente por una familia de probabilidades  $[A, P]$ , es decir, por un álgebra de sucesos  $A$  y una probabilidad  $P$  definida sobre ella. El número  $P(A)$  indica, por tanto, la probabilidad de la ocurrencia del suceso  $A \in A$  en el marco de las condiciones que caracterizan al experimento aleatorio observado. Añadimos aún mentalmente a estas condiciones la de que el suceso aleatorio  $B \in A$  ocurre y entonces el grado de indeterminación de la ocurrencia del suceso  $A$  se describirá, por lo general, mediante un número distinto de  $P(A)$ .

Designaremos posteriormente este número con  $P(A|B)$  y lo llamaremos probabilidad (condicionada) de  $A$  bajo la condición  $B$ . La definición matemática de probabilidad (condicionada) de  $A$  bajo la condición  $B$  queremos hacerla de modo que se corresponda con las ideas relativas al contenido de este concepto, explicadas anteriormente. Para ello realizaremos algunas reflexiones previas con respecto a la frecuencia relativa y al concepto clásico de probabilidad.

Si en  $n$  repeticiones realizadas independientemente unas de otras del experimento aleatorio observado se presenta  $m$  veces el suceso  $B$  y  $l$  veces el suceso  $A \cap B$ , entonces se cum-

ple para la frecuencia relativa  $f_n(A|B)$  de la ocurrencia de  $A$  en los  $m$  experimentos en los cuales  $B$  ocurre, la relación

$$f_n(A|B) = \frac{\frac{l}{n}}{\frac{m}{n}} = \frac{f_n(A \cap B)}{f_n(B)}. \quad (1)$$

Si el experimento aleatorio observado posee  $k (< \infty)$  resultados y estos son igualmente posibles, entonces se cumple para la probabilidad  $P(A|B)$  del suceso  $A$  bajo la condición de que el suceso  $B$  ocurra, según la definición clásica, la relación

$$P(A|B) = \frac{\frac{g(A \cap B)}{k}}{\frac{g(B)}{k}} = \frac{g(A \cap B)}{g(B)} = \frac{P(A \cap B)}{P(B)} \quad (2)$$

denotando  $g(C)$ , como antes, el número de los resultados que provocan la presencia del suceso  $C$ .

Las relaciones (1) y (2) son la base para la siguiente definición general de probabilidad condicionada.

**Definición 1.** Sea  $A$  un álgebra de sucesos,  $P$  una probabilidad sobre  $A$  y  $B \in A$  un suceso aleatorio de probabilidad positiva ( $P(B) > 0$ ). Entonces se llama a

$$P(A|B) = P \frac{(A \cap B)}{P(B)} \quad (3)$$

la probabilidad (condicionada) del suceso  $A \in A$  bajo la condición (o también bajo la hipótesis)  $B$  o abreviadamente la probabilidad condicionada de  $A$  respecto a  $B$  (fig. 17).

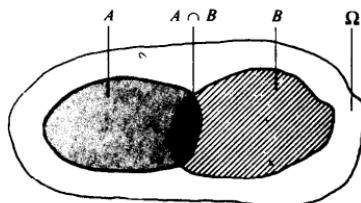


Figura 17

**Ejemplo.** Un sistema se compone de tres máquinas I, II y III dispuestas en serie; el sistema falla si y solo si lo hace una de las máquinas, suponiendo que dos máquinas cualesquier no pueden fallar al mismo tiempo. La probabilidad de que, en caso de desperfecto del sistema, la causa radique en la máquina I sea igual a  $p(0 \leq p \leq 1)$ ; para la máquina II, igual a  $q(q \geq 0, p+q \leq 1)$  y para la máquina III, igual a  $1-(p+q)$  (fig. 18).

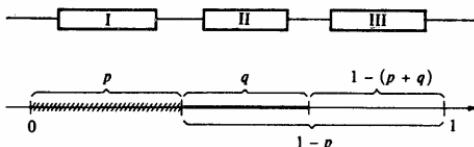


Figura 18

Supongamos ahora que el sistema de máquinas no funciona y que se ha buscado en vano un defecto en la máquina I. Calculemos la probabilidad de que la causa del desperfecto radique en la máquina II. Para ello introduzcamos los sucesos siguientes:

- A... La causa del desperfecto radica en la máquina II.
- B... La causa del desperfecto no radica en la máquina I.

Luego hay que determinar  $P(A|B)$ . Según (3) se tiene que  $P(A|B) = P \frac{(A \cap B)}{P(B)}$ . Ahora, se cumple que  $A \subseteq B$  y, por consiguiente,  $A \cap B = A$ . Con esto

$$P(A|B) = \frac{P(A)}{P(B)}.$$

Con  $P(A) = q$  y  $P(B) = 1 - P(\bar{B}) = 1 - p$  (fig. 18), obtenemos

$$P(A|B) = \frac{q}{1-p}.$$

Indicamos algunas inferencias directas de (3), que fundamentan más ampliamente la conveniencia de la definición 1.

**Corolario 1.** Si a la ocurrencia del suceso aleatorio  $B \in \mathcal{A}$ ,  $P(B) > 0$ , está siempre unida la ocurrencia del suceso aleatorio  $A \in \mathcal{A}$  ( $B \subseteq A$ ), entonces se cumple  $P(A|B) = 1$ .

**Corolario 2.** Si  $A \in \mathcal{A}$  y  $B \in \mathcal{A}$  son sucesos aleatorios mutuamente excluyentes ( $A \cap B = \emptyset$ ) y se cumple que  $P(B) > 0$ , entonces se tiene que  $P(A|B) = 0$ .

La probabilidad condicionada  $P(A|B)$  de  $A$  con respecto a  $B$ , puede ser menor, mayor y también igual a la probabilidad (incondicionada)  $P(A)$ . (Nos ocuparemos más detalladamente en el epígrafe 3.3 con el caso de la igualdad.)

**Ejemplo.** Tirada de un dado.

B... El número obtenido es par  $\left( P(B) = \frac{3}{6} = \frac{1}{2} \right)$ .

a) A... El número obtenido no es mayor que 3:

$$\left( P(A) = \frac{3}{6} = \frac{1}{2} \right) P(A|B) = \frac{1}{3} < P(A).$$

b) A... El número obtenido es igual a 2, 3 o 4:

$$\left( P(A) = \frac{3}{6} = \frac{1}{2} \right) P(A|B) = \frac{2}{3} > P(A).$$

c) A... El número obtenido es igual a 1 o 2:

$$\left( P(A) = \frac{2}{6} = \frac{1}{3} \right) P(A|B) = \frac{1}{3} = P(A).$$

Llamamos también la atención de que la probabilidad condicionada  $P(A|B)$  de  $A$  con respecto a  $B$  se debe diferenciar exactamente de la probabilidad condicionada  $P(B|A)$  de  $B$  con respecto a  $A$  y también de la probabilidad  $P(A \cap B)$  de la ocurrencia simultánea de los sucesos  $A$  y  $B$ .

**Ejemplo.** Tirada de un dado.

$A\dots$  El número obtenido al tirar el dado no es mayor que 4.

$B\dots$  El número obtenido al tirar el dado es igual a 3, 5 o 6.

$$P(A) = \frac{4}{6} = \frac{2}{3}, \quad P(B) = \frac{3}{6} = \frac{1}{2},$$

$$P(A \cap B) = \frac{1}{6}, \quad P(A|B) = \frac{1}{3}, \quad P(B|A) = \frac{1}{4}.$$

La correspondencia

$$A \rightarrow P(A|B), \quad A \in \mathcal{A} \quad (4)$$

es una función definida sobre el álgebra de sucesos  $\mathcal{A}$  para un suceso fijo  $B \in \mathcal{A}$  de probabilidad positiva  $P(B) > 0$ . Designemos esta función con  $P_B$ ; se cumple por tanto que

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

El siguiente teorema, cuya demostración recomendamos mucho al lector, contiene propiedades esenciales de la función  $P_B$ .

**Teorema 1.** Sea  $[\mathcal{A}, P]$  una familia de probabilidades y  $B \in \mathcal{A}$  un suceso aleatorio de probabilidad positiva. La función  $P_B$  definida por (4) posee todas las propiedades que se expresan en los axiomas 1 hasta 4 (épígrafe 2.4), es decir,  $[\mathcal{A}, P_B]$  es también una familia de probabilidades.

La probabilidad condicionada  $P_B$  posee también, a causa de la validez del teorema 1, todas las propiedades que fueron demostradas para la probabilidad (incondicionada)  $P$  (ver 2.5, teoremas 1 hasta 5).

Por último, advertimos que se puede interpretar la probabilidad (incondicionada) como probabilidad condicionada con respecto al suceso seguro; se cumple para todo suceso aleatorio  $A \in \mathcal{A}$  que

$$P(A|\Omega) = \frac{P(A \cap \Omega)}{P(\Omega)} = \frac{P(A)}{1} = P(A). \quad (5)$$

### 3.2 Teorema de la multiplicación para probabilidades

Trataremos en este capítulo el cálculo de la probabilidad del producto de dos sucesos aleatorios  $A$  y  $B$ . Para ello supongamos que  $A$  y  $B$  poseen probabilidades positivas. (En caso contrario se cumple, en virtud de  $A \cap B \subseteq A$  y  $A \cap B \subseteq B$ , la relación  $P(A \cap B) = 0$  (ver 2.5, teorema 4), de modo que entonces toda investigación ulterior es innecesaria). La proba-

bilidad  $P(A \cap B)$  se presentó en el epígrafe 3.1 en la definición de la probabilidad condicionada. Despejando la ecuación (3) de 3.1 obtenemos la proposición siguiente:

**Teorema 1.(Teorema de la multiplicación)**

Sean  $A$  y  $B$  sucesos aleatorios con probabilidades positivas. Entonces se cumple que

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A). \quad (1)$$

La probabilidad del producto de dos sucesos aleatorios con probabilidades positivas es, por tanto, igual a la probabilidad condicionada de un suceso respecto al otro por la probabilidad (incondicionada) del otro.

De (1) se obtiene directamente la siguiente relación, que necesitaremos más tarde:

$$\frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}. \quad (2)$$

La aplicación de la fórmula (1) para el cálculo de la probabilidad de la ocurrencia común de dos sucesos presupone, en particular el conocimiento de una de las probabilidades condicionadas que aparecen en (1). En problemas concretos es posible obtener frecuentemente probabilidades condicionadas mediante reflexiones que se basan en la interpretación del contenido del concepto probabilidad condicionada.

**Ejemplo.** En una cajita se encuentran 10 fusibles, entre los cuales hay 4 defectuosos. Se extraen sucesivamente dos fusibles, no reponiéndose el fusible tomado al inicio antes de haber extraído el segundo y teniendo cada fusible la misma posibilidad de ser tomado; calculemos la probabilidad de que los fusibles extraídos estén en buenas condiciones (suceso  $A$ ). Para ello introduciremos los sucesos siguientes:

$A_1 \dots$  El fusible tomado en la extracción número  $i$  está en buenas condiciones ( $i=1,2$ ).

Entonces se cumple que  $A = A_1 \cap A_2$ , y, por tanto, que  $P(A) = P(A_1 \cap A_2)$ . Utilizaremos para el cálculo de esta probabilidad la fórmula (1) en la forma

$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1).$$

Se cumple, utilizando la definición clásica de probabilidad, que

$$P(A_1) = \frac{6}{10} = \frac{3}{5}, \quad P(A_2|A_1) = \frac{5}{9}.$$

Con esto

$$P(A) = \frac{3}{5} \cdot \frac{5}{9} = \frac{1}{3}.$$

(Se puede obtener también este resultado directamente por medio de la definición clásica de probabilidad:

$$P(A) = \frac{\binom{6}{2} \binom{4}{0}}{\binom{10}{2}} = \frac{6 \cdot 5}{1 \cdot 2} \cdot \frac{1 \cdot 2}{10 \cdot 9} = \frac{1}{3}.$$

A continuación indicamos una fórmula para el cálculo de la probabilidad de un producto de  $n (\geq 2)$  sucesos aleatorios.

**Teorema 2.** Sean  $A_1, A_2, \dots, A_n$  sucesos aleatorios con

$$P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0.$$

Entonces se cumple que

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_n) P(A_2 | A_1) \dots P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (3)$$

Dejamos al lector la demostración de esta proposición; esta se debe realizar sobre la base del teorema 1 con ayuda del principio de inducción completa.

### 3.3 Independencia de sucesos aleatorios

Sean  $A$  y  $B$  sucesos aleatorios con probabilidades positivas. En el tratamiento de la probabilidad condicionada hemos advertido que esta puede ser también igual a la probabilidad (incondicionada) ( $P(A|B) = P(A)$ ). La adición de la condición *el suceso  $B$  ocurre* a las condiciones que caracterizan al experimento aleatorio observado, no tiene en este caso influencia sobre la probabilidad del suceso  $A$ , o sea, el suceso  $A$  es en este sentido independiente del suceso  $B$ . Ahora, se infiere de  $P(A|B) = P(A)$  la relación  $P(B|A) = P(B)$  (ver 3.1 (2)), es decir, si  $A$  es independiente de  $B$  en el sentido anterior, entonces  $B$  es también, en el mismo sentido, independiente de  $A$  y se cumple que  $P(A \cap B) = P(A) \cdot P(B)$ . (ver 3.1, teorema 1). Utilizaremos esta relación para la definición matemática de la independencia de dos sucesos aleatorios.

**Definición 1.** Dos sucesos aleatorios  $A$  y  $B$  se llaman *independientes* (uno de otro) (también: *estocásticamente independientes*), si se cumple que

$$P(A \cap B) = P(A) \cdot P(B), \quad (1)$$

o sea, si la probabilidad del producto de los sucesos es igual al producto de las probabilidades de dichos sucesos.

**Observación.** En esta definición no hemos prestado atención a la limitación, dada desde un inicio, de que  $A$  y  $B$  posean probabilidades positivas. Dos sucesos aleatorios, de los cuales uno por lo menos posee la probabilidad cero, se pueden concebir como independientes uno de otro según la definición 1, ya que siempre se satisface (1).

Los conceptos *mutuamente excluyentes* e *independientes* se deben diferenciar rigurosamente. La exclusión mutua de dos sucesos  $A$  y  $B$  significa que  $A \cap B = \emptyset$ , y por tanto se cumple que  $P(A \cap B) = 0$ . Por el contrario, la independencia significa que  $P(A \cap B) = P(A) \cdot P(B)$ . Por consiguiente, dos sucesos mutuamente excluyentes de probabilidad positiva no son independientes uno de otro.

**Corolario 1.** Si los sucesos  $A$  y  $B$  son independientes uno de otro, entonces también lo son los sucesos  $\bar{A}$  y  $B$ ,  $A$  y  $\bar{B}$ , y también los sucesos  $\bar{A}$  y  $\bar{B}$ .

**Demostración.** Es suficiente demostrar que de la independencia de  $A$  y  $B$  resulta la de  $\bar{A}$  y  $B$ ; lo restante se aclara con esto. Sean por tanto  $A$  y  $B$  independientes, es decir, sea  $P(A \cap B) = P(A) \cdot P(B)$ . De  $B = (A \cap B) \cup (\bar{A} \cap B)$  y de  $(A \cap B) \cap (\bar{A} \cap B) = \emptyset$  resulta, según el axioma 3, la relación  $P(B) = P(A \cap B) + P(\bar{A} \cap B)$ ; con  $P(A \cap B) = P(A) \cdot P(B)$  obtenemos de esto

$$P(\bar{A} \cap B) = P(B) - P(A)P(B) = (1 - P(A))P(B) = P(\bar{A}) \cdot P(B), \text{ o sea, } \bar{A} \text{ y } B \text{ son independientes uno de otro.}$$

El ejemplo siguiente debe ilustrar no solo el concepto independencia de dos sucesos, sino también preparar la ampliación de la definición de independencia al caso de más de dos sucesos.

**Ejemplo.** Tiremos dos dados una vez –imaginemos los dados numerados– y observemos los sucesos siguientes:

- A ... El número obtenido con el dado 1 es impar.
- B ... El número obtenido con el dado 2 es par.
- C ... Los números obtenidos son ambos pares o impares.

Supongamos que los 36 resultados posibles del lanzamiento de dos dados son igualmente probables. Entonces obtenemos (mediante la definición clásica de probabilidad) que

$$P(A) = P(B) = P(C) = \frac{18}{36} = \frac{1}{2},$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{9}{36} = \frac{1}{4}.$$

Los sucesos A, B y C son, por tanto, independientes dos a dos. Sin embargo, se cumple por ejemplo que  $P(C|A \cap B) = 0 \neq P(C)$ , es decir, el suceso C no es independiente del suceso  $A \cap B$ . Por consiguiente, no designaremos a los sucesos A, B y C como completamente independientes unos de otros.

**Definición 2.** Los sucesos aleatorios  $A_1, A_2, \dots, A_n$  se llaman *completamente independientes* (entre sí), si para todo número natural  $k \leq n$  y para números naturales cualesquiera  $i_1, \dots, i_k$ , con  $1 \leq i_1 < \dots < i_k \leq n$  se cumple la relación

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}). \quad (2)$$

Los sucesos aleatorios  $A_1, A_2, \dots, A_n, \dots$  de una sucesión infinita se llaman completamente independientes si para todo número natural  $n$  los sucesos  $A_1, A_2, \dots, A_n$  son completamente independientes.

**Corolario 2.** Si los sucesos  $A_1, A_2, \dots, A_n$  son completamente independientes, entonces son también independientes dos a dos.

Esta proposición se obtiene directamente de la definición 2. Como muestra el ejemplo anterior, el recíproco es falso, es decir, de la independencia mutua (dos a dos) no resulta la independencia completa.

Para finalizar este epígrafe, queremos indicar un teorema que proporciona ideas interesantes sobre las familias de probabilidades y sobre el concepto independencia.

**Teorema 1. (Lema de Borel-Cantelli)**

Sea  $(A, P)$  una familia de probabilidades y  $(A_n)_{n \in \mathbb{N}}$  una sucesión de sucesos aleatorios  $A_n \in A$ . Con  $A_\infty$  denotamos al suceso aleatorio que tiene lugar si y solo si ocurre un número infinito de sucesos de la sucesión  $(A_n)_{n \in \mathbb{N}}$ .

a) Si se cumple que  $\sum_{n=0}^{\infty} P(A_n) < \infty$ , entonces  $P(A_\infty) = 0$ , o sea, a lo sumo un número finito de sucesos de la sucesión  $(A_n)_{n \in \mathbb{N}}$  ocurre con probabilidad 1.

b) Si se cumple que  $\sum_{n=0}^{\infty} P(A_n) = \infty$  y los sucesos  $A_1, A_2, \dots$  son independientes dos a dos, entonces se cumple que  $P(A_\infty) = 1$ .

Este teorema, que no queremos demostrar, desempeña una función importante en la demostración de las leyes fuertes de los grandes números. Sin embargo, queremos fundamentar por lo menos que la proposición de este teorema es razonable, o sea, que se cumple  $A_{\infty} \in A$ . Esto resulta en virtud de las propiedades de un álgebra de sucesos (ver 1.4, definición 1 y corolario 1) sobre la base de la relación

$$A_{\infty} = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k. \quad (\text{Si } A_1, A_2, \dots \text{ son subconjuntos de un conjunto universo } \Omega, \text{ entonces})$$

$$A_{\infty} = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

es el llamado *límite superior de la sucesión*  $(A_n)_{n \in \mathbb{N}}$ : se cumple que  $x \in A_{\infty}$  si y solo si  $x$  es elemento de un número infinito de subconjuntos  $A_n$ .

### 3.4 Fórmula de la probabilidad total

La fórmula de la probabilidad total sirve para el cálculo de la probabilidad  $P(B)$  de un suceso aleatorio  $B$  a partir de las probabilidades  $P(A_i)$  de un sistema completo  $\{A_1, A_2, \dots, A_n\}$  de sucesos  $A_i$  (ver 1.3, definición 6) y, de las probabilidades condicionadas  $P(B|A_i)$  del suceso  $B$  con respecto a  $A_i$  ( $i=1, 2, \dots, n$ ).

**Teorema 1. (Fórmula de la probabilidad total)**

Sea  $[A, P]$  una familia de probabilidades y  $\{A_1, A_2, \dots, A_n\}$  un conjunto de sucesos aleatorios  $A_i \in A$  mutuamente excluyentes dos a dos y con probabilidades positivas ( $i=1, 2, \dots, n$ ), cuya suma es el suceso seguro. Entonces se cumple para todo suceso aleatorio  $B \in A$  que

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i). \quad (1)$$

**Observación.** La fórmula (1) se llama fórmula de la probabilidad total o también completa porque con ella se puede calcular la probabilidad (incondicionada) de un suceso  $B$  a partir de sus probabilidades condicionadas, que en este contexto se designa como probabilidad total o completa (fig. 19).

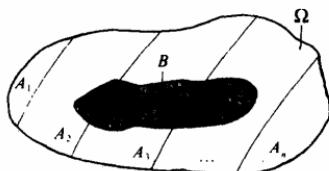


Figura 19

**Demostración.** En virtud de las condiciones impuestas a los sucesos  $A_1, A_2, \dots, A_n$ , el suceso  $B$  ocurre al menos con uno de estos sucesos. Luego, el suceso  $B$  puede representarse como suma de  $n$  sucesos mutuamente excluyentes dos a dos  $B \cap A_i$ ,  $i=1, 2, \dots, n$  (fig. 19).

$$B = \bigcup_{i=1}^n (B \cap A_i).$$

De aquí resulta (ver 2.4, corolario 1)

$$P(B) = \sum_{i=1}^n P(B \cap A_i).$$

La aplicación del teorema de la multiplicación proporciona por último (ver 3.2, teorema 1)

$$P(B) = \sum_{i=1}^n P(B|A_i) P(A_i),$$

o sea, se cumple (1).

**Ejemplo.** Observemos un modelo sencillo de un sistema de trasmisión de noticias, consistente en una fuente de noticias, un canal interferido y un receptor (fig. 20). La fuente envía exactamente una de las señales  $x_1, x_2, \dots, x_n$ ; esta se transmite por el canal y se convierte en una de las señales  $y_1, y_2, \dots, y_m$ , que a su vez, se recibe por el receptor. Describamos la fuente mediante las probabilidades  $p_i > 0$  de la ocurrencia de las señales  $x_i$  ( $i=1, 2, \dots, n$ ), y el canal interferido, por las probabilidades  $p_{ij}$  de la transición de la señal  $x_i$  en la señal  $y_j$  ( $i=1, 2, \dots, n; j=1, 2, \dots, m$ ). Nos interesamos por las probabilidades  $q_j$  de la ocurrencia de las señales  $y_j$  ( $j=1, 2, \dots, m$ ) en el receptor.

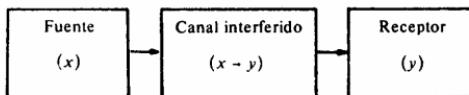


Figura 20

Introducimos los sucesos siguientes:

$A_i \dots$  La fuente envía la señal  $x_i$  ( $i=1, 2, \dots, n$ ).

$B_j \dots$  El receptor recibe la señal  $y_j$  ( $j=1, 2, \dots, m$ ).

Entonces se cumple que  $A_i \cap A_k = \emptyset$  ( $i \neq k$ ),  $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$ . Además, se dan los números  $p_i = P(A_i)$  mayores que 0 ( $i=1, 2, \dots, n$ ), y también los números  $p_{ij} = P(B_j|A_i)$  ( $i=1, 2, \dots, n; j=1, 2, \dots, m$ ). Para  $q_j = P(B_j)$  obtenemos con esto, sobre la base de la fórmula de la probabilidad total,

$$P(B) = \sum_{i=1}^n P(B_j|A_i) P(A_i), \text{ por tanto } q_j = \sum_{i=1}^n p_{ij} p_i (j=1, 2, \dots, m).$$

Reunamos los números  $p_1, p_2, \dots, p_n$  en una matriz  $p$  de una fila y los números  $p_{11}, \dots, p_{nm}$  en una matriz  $P$ . Entonces se cumple para la matriz  $q$  de una sola fila, formada por los números  $q_1, q_2, \dots, q_m$ , la relación  $q = pP$ , entendiéndose la multiplicación que se encuentra en el miembro derecho de esta ecuación como multiplicación de dos matrices.

**Ejemplo numérico.**  $n=m=3$ ,  $p=(0,5; 0,3; 0,2)$

$$P = \begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,3 & 0,5 & 0,2 \\ 0,3 & 0 & 0,7 \end{pmatrix}$$

(Por ejemplo, la señal  $x_3$  se convierte en  $y_1$  con la probabilidad 0,3 y en  $y_3$ , con la probabilidad 0,7). Con esto se obtiene  $q = pP = (0,5; 0,25; 0,25)$ .

### 3.5 Fórmula de Bayes

La fórmula de Bayes sirve para el cálculo de las probabilidades condicionadas  $P(A_k|B)$  de los sucesos  $A_k$  de un sistema completo  $\{A_1, A_2, \dots, A_n\}$  de sucesos con respecto a un suceso  $B$  de probabilidad positiva ( $k=1, 2, \dots, n$ ), a partir de las probabilidades  $P(A_i)$  y de las probabilidades condicionadas  $P(B|A_i)$  ( $i=1, 2, \dots, n$ ).

**Teorema 1.** (Fórmula de Bayes). Sea  $[A, P]$  una familia de probabilidades,  $\{A_1, A_2, \dots, A_n\}$  un conjunto de sucesos aleatorios  $A_i \in A$ , mutuamente excluyentes dos a dos y con probabilidades positivas ( $i=1, 2, \dots, n$ ), cuya suma es el suceso seguro, y  $B \in A$ , un suceso aleatorio con probabilidad positiva. Entonces se cumple que

$$P(A_k|B) = \frac{\sum_{i=1}^n P(B|A_i) P(A_i)}{\sum_{i=1}^n P(B|A_i) P(A_i)} \quad (k=1, 2, \dots, n) \quad (1)$$

**Demostración.** Se cumple (ver 3.2 (2)) que

$$\frac{P(A_k|B)}{P(A_k)} = \frac{P(B|A_k)}{P(B)} \quad (k=1, 2, \dots, n).$$

De aquí resulta

$$P(A_k|B) = \frac{P(B|A_k) P(A_k)}{P(B)} \quad (k=1, 2, \dots, n).$$

Como las condiciones para la aplicación de la fórmula de la probabilidad total se satisfacen (ver 3.4, teorema 1), obtenemos con esto

$$P(A_k|B) = \frac{\sum_{i=1}^n P(B|A_i) P(A_i)}{\sum_{i=1}^n P(B|A_i) P(A_i)} \quad (k=1, 2, \dots, n),$$

o sea, se cumple (1).

**Ejemplo.** Continuamos con el ejemplo del epígrafe 3.4 y nos interesamos ahora por la probabilidad  $r_{jk}$  de que la señal  $x_k$  haya sido la enviada una vez que se ha recibido ya la señal  $y_j$ . Con las notaciones anteriores se tiene que  $r_{jk} = P(A_k|B_j)$ . Por medio de la fórmula de Bayes obtenemos

$$r_{jk} = P(A_k|B_j) = \frac{P(B_j|A_k) P(A_k)}{P(B_j)} = \frac{p_{kj} p_k}{q_j}$$

$$(k=1, 2, \dots, n; j=1, 2, \dots, m),$$

donde los números  $q_j$  están dados por  $q_j = \sum_{i=1}^n p_{ij} p_i$  ( $j=1, 2, \dots, m$ ).

**Ejemplo numérico.** Utilicemos los datos del ejemplo numérico del epígrafe 3.4 y obtenemos

$$(r_{jk})_{\substack{j=1,2,3 \\ k=1,2,3}} = \begin{pmatrix} 0,70 & 0,18 & 0,12 \\ 0,40 & 0,60 & 0 \\ 0,20 & 0,24 & 0,56 \end{pmatrix}$$

(Por ejemplo, se cumple que  $r_{12} = \frac{p_{11}p_2}{q_1} = \frac{0,2 \cdot 0,3}{0,25} = 0,24$ , es decir, la probabilidad de que la señal  $x_1$  haya sido enviada cuando se recibió la señal  $y_1$  es de 0,24.)

Queremos fundamentar un poco la significación de la fórmula de Bayes. Para ello podemos partir de la consideración de un experimento aleatorio en el cual, en cada oportunidad, ocurre exactamente uno de los sucesos aleatorios  $A_1, A_2, \dots, A_n$ . Imaginemos que no es posible una observación directa del experimento con respecto a la ocurrencia de los sucesos  $A_1, A_2, \dots, A_n$ , pero que las probabilidades de estos sucesos son conocidas o que existen valores estimados para ellas. (En esta relación se denominan también las probabilidades  $P(A_i)$  ( $i=1, 2, \dots, n$ ) como *probabilidades a priori*.) Si se puede observar ahora la ocurrencia del suceso  $B$  en la realización del experimento, se procura utilizar esta información en la toma de la decisión sobre cuál de los sucesos  $A_1, A_2, \dots, A_n$  ocurre en el experimento. Para ello se calcularán las probabilidades condicionadas  $P(A_k|B)$  de los sucesos  $A_k$  ( $k=1, 2, \dots, n$ ) con respecto a  $B$  según la fórmula de Bayes. (En este contexto se denominan también las probabilidades  $P(A_k|B)$  ( $k=1, 2, \dots, n$ ) como *probabilidades a posteriori*.)

Una regla de decisión posible y muy clara consiste en que ante la presencia del suceso  $B$  se considere como ocurrido aquel de los sucesos  $A_k$  ( $k=1, 2, \dots, n$ ) que tiene la mayor probabilidad bajo la hipótesis de que *el suceso  $B$  ocurre*; por tanto, se elige entre los sucesos  $A_k$  ( $k=1, 2, \dots, n$ ) aquel que, dando por sentado a  $B$ , tiene mayor probabilidad. Naturalmente, esta decisión no está exenta de error, pero se puede indicar la probabilidad de una decisión falsa. Sobre este principio de decisión se basan muchas reflexiones, particularmente de la Estadística matemática; el principio se debe a un clérigo inglés, Thomas Bayes (fallecido en 1763), pero fue solo conocido y aplicable después de una nueva formulación hecha por P.S. Laplace.

**Ejemplo.** Si aplicamos el principio de decisión descrito al modelo considerado de un sistema de trasmisión de noticias, esto significa que ante la recepción de la señal  $y_1$ , consideramos como enviada aquella señal  $x_1$ , para la cual la probabilidad  $r_{11}$  es el máximo del conjunto de los números  $r_{ik}$  ( $k=1, 2, \dots, n$ ), es decir, que tiene la mayor probabilidad de haber sido enviada. Para el ejemplo numérico esto significa, que ante la recepción de las señales  $y_1, y_2$  y  $y_3$ , se decidió por  $x_1, x_2$  y  $x_3$ , respectivamente. (Estas tres decisiones están provistas de errores; la probabilidad de una decisión falsa asciende a 0,3 para la deducción de  $y_1$  a  $x_1$ , 0,4 para la de  $y_2$  a  $x_2$  y a 0,44 para la de  $y_3$  a  $x_3$ .)

## 4. Variables aleatorias discretas

El concepto *variable aleatoria* tiene una significación central en la Teoría de probabilidades y sus aplicaciones. Por medio de variables aleatorias se describen numéricamente algunas características de los fenómenos aleatorios. Así se describe, por ejemplo, el número de artículos defectuosos en una muestra aleatoria de la producción diaria de una fábrica, el número de partículas emitidas por una sustancia radiactiva en un tiempo determinado, la duración de un bombillo o el resultado de un proceso de medición cualquiera en la técnica. Frecuentemente la realización de un experimento aleatorio sirve para emitir un valor numérico de una variable aleatoria. En la naturaleza del fenómeno radica el que se puedan observar distintos valores de las variables aleatorias en repeticiones del experimento aleatorio. Para la caracterización teórico-probabilística de una variable aleatoria, no es suficiente la indicación del conjunto de los valores imaginables; son mucho más necesarias las probabilidades de aquellos sucesos aleatorios que están en relación con la variable aleatoria considerada, por ejemplo, las probabilidades con las cuales la variable aleatoria acepta determinados valores o valores de determinados intervalos.

En este capítulo queremos trabajar con las llamadas *variables aleatorias discretas*, cuya característica común consiste en que pueden aceptar un número finito o infinito numerable de valores; en el capítulo 5 nos ocuparemos de las llamadas variables aleatorias continuas, cuyos valores imaginables cubren un intervalo.

A estas consideraciones queremos anteponer la definición general de *variable aleatoria*, que requiere del concepto espacio de probabilidad, y la definición de *función de distribución de una variable aleatoria*.

### 4.1 Definición general de variable aleatoria

Los epígrafes siguientes contienen muchos ejemplos y motivaciones para los conceptos que se introducen aquí de forma general, de modo que se obtendrá pronto una cierta familiarización con estos conceptos.

**Definición 1.** Sea  $[\Omega, A, P]$  un espacio de probabilidad. Una función real  $X$  definida sobre  $\Omega$  ( $\omega \in \Omega \rightarrow X(\omega) \in \mathbb{R}$ ) se llama una variable aleatoria (sobre  $[\Omega, A, P]$ ), si para todo número real  $x$  se cumple que

$$\{\omega \in \Omega : X(\omega) < x\} \in A.$$

Para evitar falsos entendimientos que pudieran resultar de la denominación *variable aleatoria* llamamos la atención expresamente de que una variable aleatoria  $X$  (sobre un espacio de probabilidad  $[\Omega, A, P]$ ) es una función, es decir, que indicando la variable independiente  $\omega (\in \Omega)$  está fijado únicamente el valor  $X(\omega) (\in \mathbb{R})$  de la variable aleatoria  $X$ . La aleatoriedad radica solo en la elección de la variable independiente  $\omega \in \Omega$  y esta elección se realiza según la medida de probabilidad  $P$ .

Queremos ahora seguir explicando la definición 1. Para ello escribiremos abreviadamente en lugar de  $\{\omega \in \Omega : X(\omega) < x\}$  solo  $(X < x)$ , de forma correspondiente, en lugar de  $\{\omega \in \Omega : a \leq X(\omega) < b\}$  y  $\{\omega \in \Omega : X(\omega) = c\}$  escribiremos  $(a \leq X < b)$  y  $(X = c)$ , respectivamente. La definición 1 dice entonces que, para una variable aleatoria  $X$ , cada uno de los conjuntos  $(X < x)$ ,  $x \in \mathbb{R}$ , pertenece a la  $\sigma$ -álgebra  $A$  de los subconjuntos del conjunto  $\Omega$ , es decir, que cada uno de estos conjuntos pertenece al dominio de definición de  $P$ . (De aquí se obtiene fácilmente que también cada uno de los conjuntos  $(a \leq X < b)$  y  $(X = c)$  pertenece también al dominio de definición de  $P$ .) Por esto es razonable hablar de la probabilidad de que una variable aleatoria  $X$  acepte un valor menor que  $x (x \in \mathbb{R})$ . Para esta probabilidad, o sea, para  $P(\{\omega \in \Omega : X(\omega) < x\})$  escribimos abreviadamente  $P(X < x)$ .

**Definición 2.** Sea  $[\Omega, A, P]$  un espacio de probabilidad y  $X$  una variable aleatoria sobre  $[\Omega, A, P]$ . La función  $F_x$  definida por

$$F_x(x) = P(X < x), \quad x \in \mathbb{R} \tag{1}$$

se llama *función de distribución* de la variable aleatoria  $X$ .

El valor de la función de distribución  $F_x$  de una variable aleatoria  $X$  en el lugar  $x$  es, por tanto, según definición, igual a la probabilidad de que la variable aleatoria  $X$  acepte un valor que sea menor que  $x$ .

Por medio de la función de distribución de una variable aleatoria se pueden expresar las probabilidades de casi todos los sucesos aleatorios que están en relación con esta variable aleatoria. Así se cumple, por ejemplo, que

$$P(a \leq X < b) = F_x(b) - F_x(a); \tag{2}$$

dejamos al lector la demostración de esta propiedad.

Sobre la base de los axiomas de la Teoría de probabilidades se pueden demostrar las propiedades de una función de distribución  $F$ , enumeradas en el teorema siguiente.

**Teorema 1.** Sea  $F$  la función de distribución de una variable aleatoria. Entonces se cumple:

1. Para todo  $x \in \mathbb{R}$ ,  $0 \leq F(x) \leq 1$ .
2.  $F$  es monótona creciente ( $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$ ).
3.  $F$  es continua por la izquierda ( $\lim_{x \rightarrow x_0^-} F(x) = F(x_0)$ ).
4.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow +\infty} F(x) = 1$ .

**Demostración.** Consideremos que  $X$  designa una variable aleatoria con la función de distribución  $F$ .

1. Como  $F(x)$  indica la probabilidad de un suceso aleatorio, se cumple que  $0 \leq F(x) \leq 1$  (ver 2.4, axioma 1).

2. De  $x_1 < x_2$ , resulta  $(X < x_1) \subseteq (X < x_2)$  y de aquí (ver 2.5, teorema 4)  $P(X < x_1) \leq P(X < x_2)$  es decir,  $F(x_1) \leq F(x_2)$ .

3. Si  $(x_n)$  es una sucesión monótona creciente de números reales  $x_n < a$  con  $\lim_{n \rightarrow \infty} x_n = a$ , entonces se cumple que  $(X < x_n) \subseteq (X < x_{n+1})$  y  $\bigcup_{n=1}^{\infty} (X < x_n) = (X < a)$ . De aquí resulta (ver 2.4, teorema 1) que  $P(X < a) = \lim_{n \rightarrow \infty} P(X < x_n)$ , o sea,  $F(a) = \lim_{n \rightarrow \infty} F(x_n)$ , con lo cual está demostrada la continuidad por la izquierda de  $F$ .

4. La existencia de los límites señalados resulta de la monotonía y del acotamiento de  $F$  (proposiciones 1 y 2); además, se cumple evidentemente que  $0 \leq \lim_{x \rightarrow -\infty} F(x) \leq \lim_{x \rightarrow +\infty} F(x) \leq 1$ . Por tanto, es suficiente demostrar que se cumple  $\lim_{n \rightarrow \infty} F(-n) = 0$  y  $\lim_{n \rightarrow \infty} F(n) = 1$ , recorriendo  $n$  el conjunto de los números naturales. Para ello consideremos los sucesos mutuamente excluyentes dos a dos  $A_j = (-j-1 \leq X < j)$ , ( $j=0, \pm 1, \pm 2, \dots$ ). Entonces se cumple (ver 2.4, axiomas 2 y 4) que

$$1 = P(\Omega) = P\left(\bigcup_{j=-\infty}^{\infty} A_j\right) = \sum_{j=-\infty}^{\infty} P(A_j) = \lim_{n \rightarrow \infty} \sum_{j=-n+1}^n P(A_j).$$

En virtud de (2) se cumple que

$$P(A_j) = P(j-1 \leq X < j) = F(j) - F(j-1)$$

y, por consiguiente,

$$\lim_{n \rightarrow \infty} \sum_{j=-n+1}^n P(A_j) = \lim_{n \rightarrow \infty} \sum_{j=-n+1}^n (F(j) - F(j-1)) = \lim_{n \rightarrow \infty} (F(n) - F(-n)).$$

Luego, se cumple en total que  $\lim_{n \rightarrow \infty} F(n) - \lim_{n \rightarrow \infty} F(-n) = 1$ .

Como la diferencia de dos números situados entre cero y uno puede tener el valor uno, solo si el minuendo es igual a uno y el sustraendo igual a cero, resulta de aquí que

$$\lim_{n \rightarrow \infty} F(n) = 1 \text{ y } \lim_{n \rightarrow \infty} F(-n) = 0,$$

con lo cual todo está demostrado. Además podemos afirmar que la propiedad 1 resulta directamente de las propiedades 2 y 4.

**Observación.** Las propiedades indicadas en el teorema 1 son características en el sentido de que, para cada función  $F$  que tenga estas propiedades existe una variable aleatoria  $X$ , cuya función de distribución  $F_X$  coincide con la función  $F$ .

Por último, queremos señalar la validez de la ecuación

$$P(X=c) = F_X(c+0) - F_X(c); \quad (3)$$

aquí designa  $F_X(c+0)$  el límite por la derecha de la función de distribución  $F_X$  de la variable aleatoria  $X$  en el punto  $c$ . Por tanto, si  $c$  es un punto de continuidad de la función de distribución de  $X$ , entonces  $X$  acepta el valor  $c$  con la probabilidad cero, o sea, el suceso  $(X=c)$  es un suceso casi imposible.

Con (3) se comprueba la validez de las ecuaciones siguientes:

$$P(a < X < b) = F_X(b) - F_X(a+0), \quad (4)$$

$$P(a < X \leq b) = F_X(b+0) - F_X(a+0), \quad (5)$$

$$P(a \leq X \leq b) = F_X(b+0) - F_X(a), \quad (6)$$

que en unión con (1) muestran cómo se calcula, mediante la función de distribución  $F_Y$ , la probabilidad de que la variable aleatoria  $X$  acepte un valor de un intervalo arbitrario dado.

Ahora queremos tratar brevemente las funciones de variables aleatorias. Primero nos ocuparemos de la igualdad de variables aleatorias. Las variables aleatorias son funciones y, por tanto, ya está definida en principio la igualdad de dos de ellas. En la Teoría de probabilidades es conveniente y usual definir un concepto igualdad un poco más general, el cual considere la particularidad del dominio de definición común (conjunto universo de un espacio de probabilidad) de una forma adecuada.

**Definición 3:** Dos variables aleatorias  $X$  y  $Y$  definidas sobre un espacio de probabilidad común  $[\Omega, \mathcal{A}, P]$  se denominan *iguales* (simbólicamente:  $X = Y$ ), si se cumple que

$$P(\{\omega \in \Omega : X(\omega) = Y(\omega)\}) = 1, \quad (7)$$

o sea, si el suceso ( $X = Y$ ) es casi seguro.

**Teorema 2.** Sea  $[\Omega, \mathcal{A}, P]$  un espacio de probabilidad,  $X$  una variable aleatoria (sobre  $[\Omega, \mathcal{A}, P]$ ) y  $g$  una función real continua definida sobre el eje real. Entonces la función  $g(X)$  definida por

$$[g(X)](\omega) = g(X(\omega)), \quad \omega \in \Omega \quad (8)$$

es también una variable aleatoria (sobre  $[\Omega, \mathcal{A}, P]$ ).

Renunciaremos a la demostración de este teorema; pero queremos expresar aún, para algunas funciones especiales  $g$ , la función de distribución de  $Y = g(X)$  mediante la función de distribución de  $X$ .

**Teorema 3.** Sea  $X$  una variable aleatoria con la función de distribución  $F_X$ .

1. Para  $Y = aX + b$  ( $a \neq 0$  real,  $b$  real) se cumple que

$$F_Y(x) = F_X\left(\frac{x-b}{a}\right) \quad \text{para } a > 0, \quad (9)$$

$$F_Y(x) = 1 - F_X\left(\frac{x-b}{a} + 0\right) \quad \text{para } a < 0. \quad (10)$$

2. Para  $Y = X^2$  se cumple que

$$F_Y(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ F_X(\sqrt{x}) - F_X(-\sqrt{x} + 0) & \text{para } x > 0. \end{cases} \quad (11)$$

3. Para  $Y = |X|$  se cumple que

$$F_Y(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ F_X(x) - F_X(-x + 0) & \text{para } x > 0. \end{cases} \quad (12)$$

**Demostración.** Se emplean las ecuaciones (1) hasta (6).

1. Sea  $a > 0$ . Entonces se cumple que

$$\begin{aligned} F_Y(x) &= P(Y < x) = P(aX + b < x) \\ &= P\left(X < \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right). \end{aligned}$$

o sea, (9). En el caso de que  $a < 0$  se obtiene que

$$F_Y(x) = P(aX + b < x) = P\left(X > \frac{x-b}{a}\right) = 1 - P\left(X \leq \frac{x-b}{a}\right) = 1 - F_X\left(\frac{x-b}{a} + 0\right).$$

o sea, (10).

2. Para  $x \leq 0$  se cumple que  $F_Y(x) = P(X^2 < x) = 0$ . Para  $x > 0$  se obtiene que

$$\begin{aligned} F_Y(x) &= P(X^2 < x) = P(|X| < \sqrt{x}) \\ &= P(-\sqrt{x} < X < \sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x} + 0), \end{aligned}$$

o sea, (11).

3. Para  $x \leq 0$  se cumple que  $F_Y(x) = P(|X| < x) = 0$ . Para  $x > 0$  se obtiene que

$$F_Y(x) = P(|X| < x) = P(-x < X < x) = F_X(x) - F_X(-x + 0),$$

o sea, (12).

Queremos concluir nuestras consideraciones sobre variables aleatorias, con un señalamiento referente a que el espacio de probabilidad tomado por base para una variable aleatoria no se presenta frecuentemente de forma explícita. Para investigaciones teórico-probabilísticas de variables aleatorias, en casos de aplicación, son esenciales las distribuciones de probabilidad de las variables aleatorias consideradas, que están caracterizadas por las funciones de distribución.

Por último, advertimos que en algunos libros de texto la función de distribución  $F_X$  de una variable aleatoria  $X$  no se introduce como aquí mediante la definición 2, por  $F_X(x) = P(X < x)$ , sino por  $F_X(x) = P(X \leq x)$ .

## 4.2 Definición de variable aleatoria discreta

**Definición 1.** Una variable aleatoria se llama *discreta*, si puede aceptar un número finito o infinito numerable de valores, es decir, si el dominio de valores es un conjunto a lo sumo numerable.

Desde el punto de vista del Cálculo de probabilidades podemos considerar una variable aleatoria discreta como dada, si están dados los distintos valores  $x_k$  de la variable aleatoria  $X$  y las llamadas probabilidades individuales  $p_k = P(X = x_k)$ , con las cuales la variable aleatoria  $X$  acepta estos valores. En casos concretos se mencionan por conveniencia sólo aquellos valores  $x_k$ , para los cuales la probabilidad individual correspondiente  $p_k$  es positiva; sin embargo, no queremos acordar esto rigurosamente, para que no resulten dificultades innecesarias en las consideraciones teóricas.

Se caracteriza una variable aleatoria discreta  $X$ , que acepta los valores  $x_k$  con las probabilidades  $p_k$ , por la llamada *tabla de distribución*.

$Y :$	<table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"> <tr> <td><math>x_1</math></td><td><math>x_2</math></td><td><math>x_3</math></td><td><math>\dots</math></td></tr> <tr> <td><math>p_1</math></td><td><math>p_2</math></td><td><math>p_3</math></td><td><math>\dots</math></td></tr> </table>	$x_1$	$x_2$	$x_3$	$\dots$	$p_1$	$p_2$	$p_3$	$\dots$	$, \quad p_k = P(X = x_k)$
$x_1$	$x_2$	$x_3$	$\dots$							
$p_1$	$p_2$	$p_3$	$\dots$							

(1)

que, si es posible, se representa también gráficamente (fig. 21).

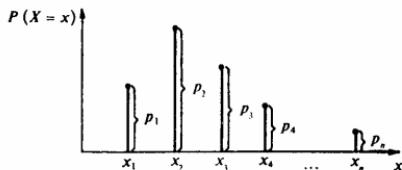


Figura 21

El teorema siguiente muestra, entre otras cosas, que mediante la tabla de distribución se fija realmente la función de distribución de la variable aleatoria considerada.

**Teorema 1.** Sea  $X$  una variable aleatoria discreta con la tabla de distribución (1). Entonces se cumplen las proposiciones siguientes:

1.  $p_k \geq 0, \quad \sum_k p_k = 1.$
2.  $F_x(x) = \sum_{k: x_k < x} p_k$ , extendiéndose la sumatoria sobre todas aquellas  $k$  para las cuales se cumple que  $x_k < x$ .
3. La función de distribución  $F_x$  es una función escalonada que posee en los lugares  $x_k$  saltos de la altura  $p_k$ .

Dejamos la demostración sencilla de este teorema al lector; esta se obtiene de los axiomas del Cálculo de probabilidades y mediante referencia a la definición de función de distribución. No hemos excluido en la definición 1 el caso de que la variable aleatoria  $X$  pueda aceptar solo un único valor  $x_1$ ; ella aceptaría entonces este valor con la probabilidad 1. La tabla de distribución perteneciente a esta variable aleatoria  $X$  y la función de distribución tienen la forma sencilla siguiente:

$$X: \begin{cases} x_1 \\ 1 \end{cases} \quad P(X=x_1)=1; \quad F_x(x) = \begin{cases} 0 & \text{para } x \leq x_1 \\ 1 & \text{para } x > x_1 \end{cases} \quad (\text{fig. 22}).$$

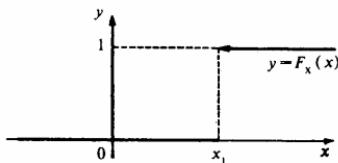


Figura 22

Se dice también que  $X$  posee una *distribución puntual* (en el punto  $x_1$ ). Por consiguiente, una variable aleatoria distribuida en un punto posee siempre, independientemente del resultado del experimento, un mismo valor. Este caso puede concebirse como caso extremo de lo casual.

Concluiremos este epígrafe con un ejemplo.

**Ejemplo.** La probabilidad de que un cazador acierte un objetivo es de 0,4 en cada tiro. Se acuerda que solo en caso de no acertar con el primer tiro se tire una segunda vez.

Si entonces el objetivo tampoco es acertado, se dispara una tercera y hasta una cuarta vez, en caso de no dar en el blanco con el tercer tiro. Independientemente de si el cuarto tiro fue certero o no, no se dispara después ninguna otra vez. Designemos con  $X$  el número de los tiros disparados por los cazadores;  $X$  es una variable aleatoria discreta. Los valores posibles de esta variable aleatoria son los números 1, 2, 3 y 4. Calculemos ahora las probabilidades individuales  $p_k = P(X=k)$  para  $k=1, 2, 3$  y 4. Para ello introduzcamos los sucesos siguientes:

$A_i \dots$  El tiro número  $i$  es certero ( $i=1, 2, 3, 4$ ).

Se cumple que  $P(A_1) = 0,4$  y  $P(\bar{A}_1) = 0,6$ . Además, los sucesos  $A_1, A_2, A_3$  y  $A_4$  son completamente independientes (ver 3.3, definición 2). Así, por ejemplo, la probabilidad del suceso *dar en el blanco con el tercer tiro* es igual a la probabilidad de este suceso bajo la condición de que los tiros anteriores fueran certeros; por tanto, en esta reflexión no posee ninguna significación el que, por ejemplo, no se disparen otros tiros en caso de dar en el blanco con el primero.

Expresemos los sucesos  $(X=1), (X=2), (X=3)$  y  $(X=4)$  mediante los sucesos  $A_1, A_2, A_3$  y  $A_4$ .

$$\begin{aligned} (X=1) &= A_1, \\ (X=2) &= \bar{A}_1 \cap A_2, \\ (X=3) &= \bar{A}_1 \cap \bar{A}_2 \cap A_3, \\ (X=4) &= \bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3. \end{aligned}$$

Luego, se muestra que no necesitamos para esto al suceso  $A_4$ .

Considerando la independencia de los sucesos  $A_1, A_2, A_3$  y  $A_4$  obtenemos

$$p_1 = P(X=1) = P(A_1) = 0,4,$$

$$p_2 = P(X=2) = P(\bar{A}_1 \cap A_2) = P(\bar{A}_1)P(A_2) = 0,6 \cdot 0,4 = 0,24,$$

$$p_3 = P(X=3) = P(\bar{A}_1 \cap \bar{A}_2 \cap A_3) = P(\bar{A}_1)P(\bar{A}_2)P(A_3) = 0,6 \cdot 0,6 \cdot 0,4 = 0,144,$$

$$p_4 = P(X=4) = P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = P(\bar{A}_1)P(\bar{A}_2)P(\bar{A}_3) = 0,6 \cdot 0,6 \cdot 0,6 = 0,216.$$

(El cálculo de  $p_4$  habríamos podido hacerlo más sencillo, ya que los sucesos  $(X=1), (X=2), (X=3)$  y  $(X=4)$  forman un sistema completo de sucesos y con esto se cumple que  $p_1 + p_2 + p_3 + p_4 = 1$ ).

La tabla de distribución de la variable aleatoria  $X$  tiene, por consiguiente, la forma siguiente (comparar con fig. 23):

1	2	3	4
0,4	0,24	0,144	0,216

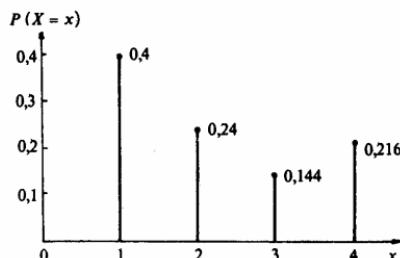


Figura 23

Para la función de distribución  $F_x$  se obtiene (fig. 24)

$$F_x(x) = p(X < x) = \begin{cases} 0 & \text{para } x \leq 1, \\ p_1 = 0,4 & \text{para } 1 < x \leq 2, \\ p_1 + p_2 = 0,64 & \text{para } 2 < x \leq 3, \\ p_1 + p_2 + p_3 = 0,784 & \text{para } 3 < x \leq 4, \\ p_1 + p_2 + p_3 + p_4 = 1 & \text{para } x > 4. \end{cases}$$

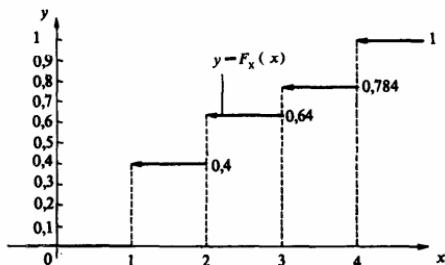


Figura 24

### 4.3 Características numéricas de las variables aleatorias discretas

En muchas ocasiones no estamos muy interesados por el conocimiento completo de todas las probabilidades individuales de una variable aleatoria discreta, sino mucho más por ciertas magnitudes denominadas también características, que siempre proporcionan alguna información sobre la variable aleatoria y su distribución de probabilidad. En este epígrafe trataremos el valor esperado y la varianza de variables aleatorias discretas. El valor esperado y la varianza, pertenecen a los llamados *momentos* de una variable aleatoria.

**Definición 1.** Sea  $X$  una variable aleatoria discreta que toma los valores  $x_k$  con las probabilidades  $p_k$ . Entonces el número  $EX$  definido por

$$EX = \sum_k x_k p_k \quad (1)$$

se llama *valor esperado* de la variable aleatoria  $X$ ; aquí se supone que la serie situada en el miembro derecho de (1) converge absolutamente, o sea, que se cumple que

$\sum_k |x_k| p_k < \infty$ . (Esta condición se satisface trivialmente en el caso que  $X$  posea solo un número finito de valores, de modo que a toda variable aleatoria discreta con un número finito de valores le corresponde, según (1), un valor esperado.)

Por consiguiente, el valor esperado de una variable aleatoria discreta es la media pesada de todos los valores  $x_k$  de  $X$ , empleándose como peso de todo valor  $x_k$  la probabilidad individual correspondiente  $p_k$ . (Aquí no se presenta explícitamente la división por la suma de todos los pesos, usual para la media pesada, ya que esta suma es igual a uno.)

La tabla de distribución de una variable aleatoria discreta que toma los valores  $x_k$  con las probabilidades  $p_k$ , se ilustra bien como un sistema de masas puntuales que posee en los lugares  $x_k$  masas  $p_k$  (y tiene, por tanto, la masa total uno). En esta ilustración corresponde al valor esperado de la variable aleatoria el centro de gravedad del sistema de masas puntuales.

**Ejemplo.** Calculemos para la variable aleatoria  $X$  considerada en el ejemplo del epígrafe 4.2 el valor esperado:

$$EX = \sum_k x_k p_k = 1 \cdot 0,4 + 2 \cdot 0,24 + 3 \cdot 0,144 + 4 \cdot 0,216 = 2,176.$$

Como muestra el ejemplo, el valor esperado no es, comúnmente, un valor de la variable aleatoria considerada. Aún cuando el valor esperado sea un valor de la variable aleatoria, este no será, por lo general, uno de los valores de esta, que en comparación con los otros tiene la mayor probabilidad y que por eso uno esperaría más. Estos valores se denominan *valores modales*. La razón para denominar a  $EX$  valor esperado se debe ver en que la media aritmética de los valores observados de la variable aleatoria es aproximadamente igual al valor esperado, satisfaciéndose esto tanto mejor, cuanto mayor sea el número de los valores observados utilizados para la formación de la media (ver 7.4).

Los teoremas siguientes contienen proposiciones, que son útiles para el cálculo con valores esperados.

**Teorema 1.** Sea  $X$  una variable aleatoria discreta con el valor esperado  $EX$ , y  $a$  y  $b$  sean números reales cualesquiera. Entonces se cumple que

$$E(aX+b) = aEX+b. \quad (2)$$

**Demuestra**ción. Si la variable aleatoria  $X$  toma los valores  $x_k$  con las probabilidades  $p_k$ , entonces la variable aleatoria  $Y=aX+b$  acepta los valores  $y_k=ax_k+b$  con las probabilidades  $p_k$ . Por tanto, se cumple que

$$EY = E(aX+b) = \sum_k y_k p_k = \sum_k (ax_k+b)p_k = a \sum_k x_k p_k + b \sum_k p_k$$

Con  $EX = \sum_k x_k p_k$  y  $\sum_k p_k = 1$ , resulta de aquí la afirmación.

Luego, se cumple en particular ( $a=1$ ,  $b=-EX$ ) que

$$E(X-EX) = 0; \quad (3)$$

el paso de la variable aleatoria  $X$  a la  $X-EX$  se llama *centrar*.

**Teorema 2.** Sea  $X$  una variable aleatoria discreta que toma los valores  $x_k$  con las probabilidades  $p_k$  y  $g$ , una función real continua definida sobre el eje real. Si la serie  $\sum_k g(x_k)p_k$  converge absolutamente (es decir, si  $\sum_k |g(x_k)|p_k < \infty$ ), entonces se cumple que

$$Eg(X) = \sum_k g(x_k)p_k. \quad (4)$$

Dejamos la demostración al lector. Para  $g(x)=x$  se cumple el teorema 2 sobre la base de la definición 1. Para  $g(x)=(x-c)^j$  y  $g'(x)=|x-c|^j$  ( $j$  un número natural arbitrario,  $c$  un número real cualquiera) se obtiene respectivamente con (4) que

$$E(X-c)^j = \sum_k (x_k - c)^j p_k \quad (5)$$

y

$$E|X-c| = \sum_k |x_k - c| p_k \quad (6)$$

siempre y cuando la serie situada a la derecha de (6) sea convergente.

Variables aleatorias con el mismo valor esperado pueden diferenciarse considerablemente en las tablas de distribución, ya que el valor esperado no ofrece ninguna información de cómo se desvian los valores individuales de la variable aleatoria del valor esperado. La llamada varianza es la medida más utilizada de la desviación de los valores respecto al valor promedio de la variable aleatoria, que se describe por el valor esperado.

**Definición 2.** Sea  $X$  una variable aleatoria discreta con el valor esperado  $EX$ , que toma los valores  $x_k$  con las probabilidades  $p_k = P(X=x_k)$ . Entonces el número  $D^2X$  definido por

$$D^2X = E(X-EX)^2 = \sum_k (x_k - EX)^2 p_k \quad (7)$$

se llama *varianza* (también *dispersión*) de la variable aleatoria  $X$ , donde se supone la convergencia de la serie situada en el miembro derecho de (7) (o sea,  $\sum_k (x_k - EX)^2 p_k < \infty$ ). (Esta condición se satisface trivialmente en el caso de que  $X$  posea sólo un número finito de valores, de modo que, a toda variable aleatoria discreta con un número finito de valores le corresponde según (7) una varianza.) El número

$$\sigma_x = \sqrt{D^2X} \quad (8)$$

se llama *desviación estándar* (o *desviación típica*) de la variable aleatoria  $X$ .

La varianza de una variable aleatoria  $X$  es, por tanto, la media pesada de los cuadrados de las desviaciones de los valores  $x_k$  de  $X$ , del valor esperado  $EX$  de esta variable aleatoria discreta, siendo utilizadas de nuevo como pesos las probabilidades individuales con las cuales se presentan estos valores.

Si se ilustra una variable aleatoria discreta  $X$  (valor esperado  $EX$ , varianza  $D^2X$ ) como un sistema de masas puntuales (con el centro de gravedad  $EX$ ), entonces corresponde a la varianza  $D^2X$  el momento de inercia de este sistema con respecto a un eje que pasa por el centro de gravedad.

**Ejemplo.** Calculemos para la variable aleatoria  $X$ , considerada en el ejemplo del epígrafe 4.2, la varianza y la desviación estándar; para ello emplearemos  $EX=2,176$ :

$$\begin{aligned} D^2X &= \sum_k (x_k - EX)^2 p_k \\ &= (1 - 2,176)^2 \cdot 0,4 + (2 - 2,176)^2 \cdot 0,24 + (3 - 2,176)^2 \cdot 0,144 + \\ &\quad (4 - 2,176)^2 \cdot 0,216 \\ &\approx 2,257 \end{aligned}$$

$$\sigma_x = \sqrt{D^2X} \approx \sqrt{2,257} \approx 1,503.$$

La fórmula contenida en el teorema siguiente se recomienda con frecuencia para el cálculo de la varianza.

**Teorema 3.** Sea  $X$  una variable aleatoria discreta con valor esperado  $EX$  y varianza  $D^2X$ , que toma los valores  $x_k$  con las probabilidades  $p_k$ . Entonces existe  $EX^2$ , y se cumple que

$$D^2X = \sum_k x_k^2 p_k - \left( \sum_k x_k p_k \right)^2 = EX^2 - (EX)^2. \quad (8)$$

**Demostración.** Utilizando (7), (1) y  $\sum_k p_k = 1$  se obtiene

$$\begin{aligned} D^2X &= \sum_k (x_k - EX)^2 p_k = \sum_k (x_k^2 - 2x_k EX + (EX)^2) p_k \\ &= \sum_k x_k^2 p_k - 2(EX) \sum_k x_k p_k + (EX)^2 \sum_k p_k \\ &= \sum_k x_k^2 p_k - \left( \sum_k x_k p_k \right)^2; \end{aligned}$$

el resto se obtiene con (4), si se hace  $g(x) = x^2$ .

Si se ilustra una variable aleatoria discreta como un sistema de masas puntuales con la masa total uno, entonces el teorema 3 reproduce el hecho bien conocido en la Mecánica y denominado como teorema de Steiner, según el cual, el momento de inercia de un sistema semejante de masas puntuales respecto a un eje que pasa por el origen, es igual a la suma del momento de inercia con respecto a un eje que pasa por el centro de gravedad y el cuadrado de la distancia del centro de gravedad al origen. Por esta razón, se denomina también en la Teoría de probabilidades la proposición del teorema 3 como teorema de Steiner.

Veamos ahora una proposición que se corresponde bien con nuestras ideas acerca del contenido del concepto varianza.

**Teorema 4.** La varianza de una variable aleatoria discreta es igual a cero, si y solo si la variable aleatoria posee una distribución puntual.

Dejamos la demostración al lector; ella se obtiene directamente de (7).

**Teorema 5.** Sea  $X$  una variable aleatoria discreta con la varianza  $D^2X$ , y sean  $a$  y  $b$  números reales cualesquiera. Entonces se cumple que

$$D^2(aX+b) = a^2 D^2X. \quad (10)$$

**Demostración.** Con (7) y (2) se obtiene

$$\begin{aligned} D^2(aX+b) &= E(aX+b - E(aX+b))^2 \\ &= E(aX+b - aEX - b)^2 \\ &= E(a^2(X-EX)^2) \\ &= a^2 E(X-EX)^2 = a^2 D^2X. \end{aligned}$$

Luego, se cumplen en particular las ecuaciones

$$D^2(-X) = D^2X, \quad (11)$$

y

$$D^2\left(\frac{X}{\sqrt{D^2X}}\right) = 1. \quad (12)$$

El paso de la variable aleatoria  $X$  a la  $\frac{X}{\sqrt{D^2X}}$  se llama *normar*.

Para la variable aleatoria  $Z = \frac{X-EX}{\sqrt{D^2X}}$  se cumple, por tanto, que  $EZ=0$  y  $D^2Z=1$ ; el paso de  $X$  a  $\frac{X-EX}{\sqrt{D^2X}}$  se llama *estandarizar*.

Las características tratadas hasta ahora: valor esperado y varianza, pertenecen a los denominados momentos. A continuación traemos la definición de los momentos.

**Definición 3.** Sea  $X$  una variable aleatoria discreta que toma los valores  $x_k$  con las probabilidades  $p_k$ ; además, sea  $j$  un número natural y  $c$ , un número real arbitrario. Entonces los números

$$\mu_j(c) = E(X-c)^j = \sum_k (x_k - c)^j p_k \quad (13)$$

y

$$a_j(c) = E|X-c|^j = \sum_k |x_k - c|^j p_k \quad (14)$$

se llaman, respectivamente, *momento ordinario* y *momento absoluto de orden j con respecto a c*, suponiéndose la convergencia absoluta de la serie situada a la derecha en (13) (o sea, la convergencia de la serie situada a la derecha en (14)). Para  $c=0$  se habla de *momentos iniciales* y para  $c=EX$ , de *momentos centrales* (suponiéndose la existencia de  $EX$ ).

A simple vista se observa que se cumplen las ecuaciones  $\mu_1(0)=EX$ ,  $\mu_2(0)=EX^2$ ,  $a_2(0)=EX^2$  y  $\mu_2(EX)=D^2X=a_2(EX)$ . La ecuación (9) plantea que  $\mu_j(EX)=\mu_j(0)-[\mu_1(0)]^j$ .

Aún queremos dar y demostrar una inecuación sobre momentos.

**Teorema 6.** Sea  $X$  una variable aleatoria discreta con la varianza  $D^2X$  y  $c$  un número real arbitrario. Entonces se cumple que

$$D^2X \leq \mu_2(c); \quad (15)$$

aquí se establece el símbolo de igualdad si y solo si se hace  $c=EX$ .

**Demostración.** Utilicemos (13), (1),  $\sum_k p_k=1$ , (9) y obtenemos que

$$\begin{aligned} \mu_2(c) &= E(X-c)^2 = \sum_k (x_k - c)^2 p_k = \sum_k (x_k^2 - 2cx_k + c^2) p_k \\ &= \sum_k x_k^2 p_k - 2c \sum_k x_k p_k + c^2 \sum_k p_k \\ &= EX^2 - 2cEX + c^2 \\ &= EX^2 - (EX)^2 + (EX)^2 - 2cEX + c^2 \\ &= D^2X + (EX - c)^2 \geq D^2X. \end{aligned}$$

de donde se obtiene la proposición del teorema 6.

El teorema 6 muestra que la varianza es el más pequeño de los momentos de segundo orden. El lector debiera comparar esta proposición con la correspondiente sobre momentos de inercia.

El teorema siguiente, sin demostración, contiene algunas otras proposiciones sobre momentos, utilizando para los momentos iniciales ordinarios de orden  $j$  la notación  $m_j$  ( $m_j=\mu_j(0)$ ); para los momentos centrales ordinarios de orden  $j$ , la notación  $\mu_j$  ( $\mu_j=\mu_j(EX)$ ) y para los momentos iniciales absolutos de orden  $j$ , la notación  $\beta_j$  ( $\beta_j=a_j(0)$ ).

**Teorema 7.** Se cumplen las proposiciones siguientes:

1.  $m_{2j} = \beta_{2j}$ , más general,  $\mu_{2j}(c) = a_{2j}(c)$ .

2. Si existe  $\beta_n$ , entonces existe también  $\beta_l$  para  $0 < l < j$ , y se cumple la inecuación  $\sqrt{\beta_l} \leq \sqrt{\beta_j}$ .

3.  $\mu_j = \sum_{l=2}^j (-1)^{j-l} \binom{j}{l} m_l m_1^{l-1} + (-1)^{j-1} (j-1) m'_1$  ( $j=2, 3, \dots$ ). (Para  $j=2$  proporciona esto  $\mu_2 = m_1 - m'_1$ , es decir, la ecuación (9)).

Las características derivadas de los momentos, dadas en la siguiente definición, son de importancia para la apreciación de una distribución de probabilidad.

**Definición 4.** Sea  $X$  una variable aleatoria discreta con varianza positiva. Entonces se llama

$$\sigma_x = \frac{\sigma_x}{EX} = \frac{\sqrt{\mu_2}}{m_1} \quad (\text{coeficiente de variación}), \quad (16)$$

$$\gamma = \frac{E(X-EX)^3}{\sigma_x^3} = \frac{\mu_3}{\sqrt{(\mu_2)^3}} \quad (\text{coeficiente de asimetría}), \quad (17)$$

$$\eta = \frac{E(X-EX)^4}{\sigma_x^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3 \quad (\text{curtosis}). \quad (18)$$

aquí se supone la existencia de los momentos que aparecen y que  $EX \neq 0$  en (16).

El coeficiente de variación es una medida de dispersión referida al valor esperado. El coeficiente de asimetría se muestra como una medida para la asimetría de una distribución de probabilidad, denominándose una variable aleatoria  $X$  con la función de distribución  $F$  simétrica (con respecto a  $a$ ), si existe un número  $a$  tal que  $P(X < a-x) = P(X > a+x)$ , o sea, si se cumple que  $F(a-x) = 1 - F(a+x+0)$  para todo número real  $x$ . Por último, la curtosis se utiliza como una medida para la desviación de una distribución de probabilidad de la distribución normal (tratada en 5.4). (Para la distribución normal se cumple  $\eta=0$ .)

## 4.4 Distribución discreta uniforme

En este y en los siguientes epígrafes trataremos algunas distribuciones de probabilidad especiales de variables aleatorias discretas.

**Definición 1.** Una variable aleatoria discreta  $X$  con los valores  $x_1, x_2, \dots, x_n$  se denomina *uniformemente distribuida*, si se cumple que

$$p_k = P(X=x_k) = \frac{1}{n} \quad (k=1, 2, \dots, n). \quad (1)$$

Se dice también, entonces, que  $X$  posee una distribución discreta uniforme (en los valores  $x_1, x_2, \dots, x_n$ ).

Una variable aleatoria discreta distribuida uniformemente está caracterizada, por tanto, porque solo puede tomar un número finito de valores, que tienen todos la misma probabilidad. Evidentemente no puede existir una distribución uniforme en un número infinito numerable de valores.

En casos de aplicación se considera distribuida uniformemente una variable aleatoria con un número finito de valores, si ésta –expresado de forma intuitiva– no prefiere ninguno de sus valores. Así se acepta, por ejemplo, que el número que resulta al tirar un dado es una variable aleatoria distribuida uniformemente (en los números 1 hasta 6), así como que los números emitidos en Tele-Lotto también poseen una distribución uniforme.

Para el valor esperado  $EX$  de una variable aleatoria distribuida uniformemente en los valores  $x_1, x_2, \dots, x_n$  se obtiene (ver 4.3 (1)) que

$$EX = \frac{1}{n} \sum_{k=1}^n x_k. \quad (2)$$

luego se obtiene la media aritmética de los valores; para la varianza se cumple (ver 4.3 (9)) que

$$D^2X = \frac{1}{n} \sum_{k=1}^n x_k^2 - \left( \frac{1}{n} \sum_{k=1}^n x_k \right)^2. \quad (3)$$

## 4.5 Distribución binomial

La distribución binomial es una distribución discreta que posee gran significación práctica. Además, representa un medio auxiliar apropiado para la investigación de regularidades de fenómenos aleatorios, que son de importancia fundamental para la teoría de probabilidades y para su aplicación práctica.

Definición 1. Sea  $n$  un número natural arbitrario y  $p$ , un número situado entre cero y uno. Una variable aleatoria  $X$  que tome los valores  $0, 1, 2, \dots, n$  se denomina *distribuida binomialmente con los parámetros  $n$  y  $p$* , si se cumple que

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (1)$$

para  $k=0, 1, 2, \dots, n$ . Se dice también que  $X$  posee una *distribución binomial con los parámetros  $n$  y  $p$* .

Antes de que investiguemos de forma más exacta la distribución binomial, queremos ocuparnos de su existencia. El punto de partida lo constituye un suceso aleatorio  $A$ , que se presenta en el resultado de un determinado experimento aleatorio con la probabilidad  $P(A)=p$ . El número (aleatorio)  $F_n(A)$ , de la ocurrencia de  $A$  en  $n$  repeticiones realizadas independientemente unas de otras del experimento aleatorio considerado, es una variable aleatoria discreta con los  $n+1$  valores  $0, 1, 2, \dots, n$ . Ahora queremos calcular las probabilidades

$$p_k = P(F_n(A)=k) \text{ para } k=0, 1, 2, \dots, n.$$

El suceso  $(F_n(A)=k)$  ocurre si y solo si en la serie de experimentos descrita, el suceso  $A$  ocurre  $k$  veces y el  $\bar{A}$ ,  $(n-k)$  veces. Toda sucesión de sucesos semejante posee, a causa de la independencia de cada uno de los experimentos, la probabilidad  $p^k (1-p)^{n-k}$ . Como

existen  $\binom{n}{k}$  sucesiones de resultados, para los cuales aparece  $k$  veces  $A$  y  $(n-k)$  veces  $\bar{A}$ , se obtiene

$$P(F_n(A)=k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (2)$$

La frecuencia absoluta, concebida como variable aleatoria, de la ocurrencia del suceso  $A$  ( $P(A)=p$ ) en  $n$  repeticiones independientes del experimento tomado por base posee, por consiguiente, una distribución binomial con los parámetros  $n$  y  $p$  (ver 2.1).

Para destacar la dependencia de cada una de las probabilidades  $P(X=k)$  de una variable aleatoria distribuida binomialmente con los parámetros  $n$  y  $p$ , de estos parámetros, se utiliza ocasionalmente la notación  $b(k; n, p)$ ,

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (3)$$

El nombre de distribución binomial se basa en que cada una de las probabilidades  $b(k; n, p)$  para  $k=0, 1, 2, \dots, n$  son los sumandos del desarrollo del binomio  $[(1-p)+p]^n$ , con lo cual se aclara también la relación  $\sum_{k=0}^n b(k; n, p) = 1$ .

La distribución binomial se debe a Jacobo Bernoulli (1654-1705), que fue uno de los primeros en tratar la teoría de probabilidades. Jacobo Bernoulli y su igualmente famoso hermano Juan Bernoulli (1667-1748) pertenecen a los más significativos discípulos de G.W. Leibniz (1646-1716). Jacobo Bernoulli fue profesor desde 1687 hasta su fallecimiento en la Universidad de Basilea. Él escribió *Ars conjectandi* (publicado póstumamente en 1713), uno de los primeros libros sobre el Cálculo de probabilidades; este contiene proposiciones fundamentales, en particular, sobre la distribución binomial. Por eso se encuentra con frecuencia la distribución binomial bajo el nombre de distribución de Bernoulli, y más aún la denominación del esquema de experimentos descrito anteriormente (repeticiones independientes de un mismo experimento) como *esquema de Bernoulli*.

**Ejemplo.** En una fábrica se producen piezas troqueladas. El productor ha asegurado que las piezas con dimensiones adecuadas representan el 90 %. Se extraen ahora 20 piezas de la producción continua y entre estas solo se encuentran 15 con dimensiones adecuadas. Queremos ocuparnos con la interrogante de si está justificado poner en duda los informes del productor con respecto al porcentaje de piezas con dimensiones adecuadas, sobre la base de la muestra. Para ello consideramos la variable aleatoria  $X$ , que indica el número (aleatorio) de piezas con dimensiones no adecuadas en una muestra de tamaño  $n=20$ . Supongamos, de acuerdo con el informe del productor, que la probabilidad de producir una pieza con dimensiones no adecuadas sea igual a 0,10 (=10 %); entonces la variable aleatoria  $X$  posee una distribución binomial con los parámetros  $n=20$  y  $p=0,10$ . Cada una de las probabilidades  $P(X=k)$  de esta variable aleatoria  $X$  se deben calcular, por tanto, según la fórmula

$$P(X=k) = b(k; 20, 0,10) = \binom{20}{k} 0,10^k (1-0,10)^{20-k} \quad (k=0, 1, 2, \dots, 20)$$

Obtenemos la tabla de distribución

0	1	2	3	4	5	6	7
0,122	0,270	0,285	0,190	0,090	0,032	0,009	0,002

y  $P(X=k) < 0,0005$  para  $k=8, 9, \dots, 20$  (ver tabla 1 (12.1) y fig. 25). Con esto se demuestra que el resultado descrito anteriormente de la muestra (5 piezas con dimensiones no adecuadas en la muestra aleatoria de 20 piezas), suponiendo que  $p=0,10$ , posee una probabilidad que es aproximadamente igual a 0,03 (= 3 %). Por tanto, sobre la base de esta muestra se pondrán seriamente en duda los informes del productor. Si se quiere estimar la probabilidad  $p$  de producir una pieza con dimensiones no adecuadas, sobre la base de la muestra independientemente de los informes del productor, entonces se utilizará como valor estimado  $\hat{p}$  la frecuencia relativa de la presencia de piezas con dimensiones no adecuadas en la muestra, es decir, se utilizará el número  $\hat{p} = \frac{5}{20} = \frac{1}{4} = 0,25$  (25 %). (Se reflexiona fácilmente que  $\hat{p}$  es aquel número para el cual la función  $p \rightarrow b(5; 20, p)$  acepta el máximo, o sea, que  $\hat{p}$  es aquel valor para el cual es mayor la probabilidad de obtener una muestra como la extraída.)

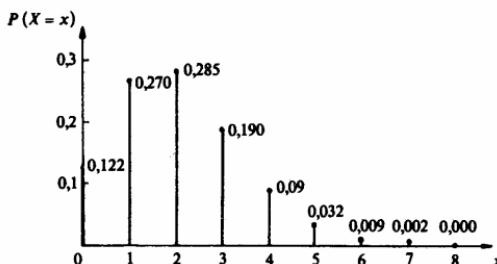


Figura 25

La gran significación práctica de la distribución binomial se muestra ya en este ejemplo. En general, podemos afirmar que el número aleatorio de las piezas defectuosas (o de las distinguidas por alguna otra propiedad) en una muestra de tamaño  $n$ , tomada de una producción continua cuyo porcentaje de piezas desecharables es de  $100p\%$ , posee una distribución binomial con los parámetros  $n$  y  $p$ . También el número aleatorio de las piezas defectuosas en una muestra de tamaño  $n$ , tomada de una población finita (por ejemplo, de la producción diaria de una fábrica), con un porcentaje de desecho de 100%, posee una distribución binomial con los parámetros  $n$  y  $p$ , si la extracción de cada una de las piezas se realiza consecutivamente y antes de cada extracción se repone de nuevo la pieza tomada anteriormente. (Una muestra tomada de esta forma se llama una muestra con reposición. Se debe prestar atención a que en una muestra sin reposición, el número aleatorio de las piezas defectuosas no posee una distribución binomial, sino una llamada distribución hipergeométrica; de esta distribución nos ocuparemos en el próximo epígrafe.)

Para el cálculo práctico de probabilidades de variables aleatorias distribuidas binomialmente, son importantes las proposiciones señaladas en el teorema siguiente.

#### Teorema 1. Se cumplen las ecuaciones

$$b(k; n, p) = b(n-k; n, 1-p), \quad (4)$$

$$b(k+1; n, p) = \frac{n-k}{k+1} \cdot \frac{p}{1-p} b(k; n, p), \quad (5)$$

$$b(k-1; n, p) = \frac{k}{n-k+1} \cdot \frac{1-p}{p} b(k; n, p). \quad (6)$$

Las demostraciones de las fórmulas indicadas son fáciles de realizar mediante el empleo de la definición de los coeficientes del binomio y utilizando (3). La fórmula (4) muestra que para hacer tablas nos podemos limitar al caso  $0 < p \leq 0,5$ ; las fórmulas (5) y (6) son fórmulas para el cálculo recursivo de  $b(k+1; n, p)$  y  $b(k-1; n, p)$  a partir de  $b(k; n, p)$ . Por lo demás, se debe tener en cuenta que el cálculo de  $b(k; n, p)$  tropieza con dificultades, particularmente para  $n$  grandes y  $p$  pequeñas; con posterioridad conoceremos fórmulas de aproximación, convenientes precisamente para estos casos.

Nos dedicaremos ahora a la determinación del valor esperado y de la varianza de variables aleatorias distribuidas binomialmente.

**Teorema 2.** Sea  $X$  una variable aleatoria distribuida binomialmente con los parámetros  $n$  y  $p$ . Entonces se cumple que

$$EX = np, \quad (7)$$

$$D^2X = np(1-p), \quad (8)$$

$$\sigma_x = \sqrt{np(1-p)}. \quad (9)$$

**Demostración.** Demostraremos solo (7); la fórmula (8) se obtiene a través de cálculos análogos y (9) se obtiene directamente de (8). Para el valor esperado tenemos que

$$\begin{aligned} EX &= \sum_{k=0}^n k P(X=k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\ &= np [p + (1-p)]^{n-1} = np. \end{aligned}$$

Así vemos que, en concordancia con nuestras ideas sobre este contenido, el valor esperado de la frecuencia absoluta  $F_n(A)$  de la ocurrencia de  $A$  en  $n$  repeticiones independientes de un experimento, es igual al producto del número  $n$  de experimentos por la probabilidad  $P(A)$  de este suceso, y que la varianza para  $p=0$  y  $p=1$  es igual a cero y para  $p=\frac{1}{2}$ , es máxima.

El teorema siguiente da información sobre el coeficiente de variación  $\delta$ , el coeficiente de asimetría  $\gamma$  y la curtosis  $\eta$  de una distribución binomial.

**Teorema 3.** Sea  $X$  una variable aleatoria distribuida binomialmente con los parámetros  $n$  y  $p$ . Entonces se cumple que

$$\delta = \sqrt{\frac{1-p}{np}}, \quad (10)$$

$$\gamma = \frac{1-2p}{\sqrt{np(1-p)}}, \quad (11)$$

$$\eta = \frac{1-6p(1-p)}{np(1-p)}. \quad (12)$$

Renunciaremos a la demostración de (11) y (12); (10) se aclara sobre la base de (7) y (9). Observamos que en el caso  $p = \frac{1}{2}$ ,  $\gamma$  es igual a cero. En este caso, se cumple que  $P(X=k) = P(X=n-k)$ , lo cual es equivalente a la simetría de la distribución binomial con los parámetros  $n$  y  $p = \frac{1}{2}$ .

Para finalizar las consideraciones sobre la distribución binomial, queremos destacar una relación fundamental entre la frecuencia relativa de un suceso en  $n$  experimentos (ver 2.1) y la probabilidad de este.

**Teorema 4.** Sea  $A$  un suceso aleatorio que se presenta en el desarrollo de un determinado experimento con la probabilidad  $P(A)$ . Además, designe  $f_n(A)$  la frecuencia relativa (concebida como variable aleatoria) de la ocurrencia de  $A$  en  $n$  repeticiones realizadas independientemente unas de otras de este experimento. Entonces se cumple que

$$Ef_n(A) = P(A), \quad (13)$$

$$Df_n(A) \rightarrow 0 \text{ para } n \rightarrow \infty. \quad (14)$$

**Demostración.** Designemos con  $F_n(A)$  la frecuencia absoluta (concebida como variable aleatoria) de la ocurrencia de  $A$  en un esquema de Bernoulli. Según reflexiones anteriores  $F_n(A)$  está distribuida binomialmente con los parámetros  $n$  y  $p = P(A)$ . Sobre la base de (7) y (8) se cumple, por tanto,  $EF_n(A) = np$  y  $D^2F_n(A) = np(1-p)$ . Entre la frecuencia absoluta  $F_n(A)$  y la frecuencia relativa  $f_n(A)$  existe la relación  $f_n(A) = \frac{F_n(A)}{n}$ . De aquí se obtiene (ver 4.3 (2) y (10) con  $a = \frac{1}{n}$  y  $b = 0$ ),

$$Ef_n(A) = E\left(\frac{F_n(A)}{n}\right) = \frac{1}{n} EF_n(A) = \frac{1}{n} np = p = P(A),$$

$$Df_n(A) = D^2\left(\frac{F_n(A)}{n}\right) = \frac{1}{n^2} D^2F_n(A) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n} \rightarrow 0 \text{ (} n \rightarrow \infty \text{)}.$$

Las relaciones (13) y (14) muestran que entre la probabilidad de un suceso aleatorio, introducida axiomáticamente, y las frecuencias relativas de este suceso, halladas de forma práctica, existen nexos muy estrechos. La validez de las relaciones señaladas constituye un motivo suficiente para estimar la probabilidad de un suceso aleatorio mediante frecuencias relativas; este valor estimado representará tanto mejor un valor aproximado de la probabilidad cuanto mayor sea el número de los experimentos realizados. La posibilidad de estimar probabilidades de modo razonable hace de la teoría de probabilidades una disciplina matemática de aplicación práctica.

## 4.6 Distribución hipergeométrica

La distribución hipergeométrica es una distribución discreta, que posee gran significación práctica, sobre todo en el control estadístico de la calidad.

**Definición 1.** Sean  $N$ ,  $M$  y  $n$  números naturales con  $M \leq N$  y  $n \leq N$ . Una variable aleatoria  $X$  que posee como valores los números naturales  $k$  con  $k \leq n$ ,  $k \leq M$ ,  $n-k \leq N-M$  (luego, estos son los números  $k = \max(0, n-(N-M), \dots, \min(M, n))$ ), se denomina *distribuida hipergeométricamente* si se cumple que

$$P(X=k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}. \quad (1)$$

Se dice entonces también que  $X$  posee una *distribución hipergeométrica*.

Hemos advertido ya en el epígrafe anterior que la distribución hipergeométrica se presenta en relación con muestras aleatorias, sin reposición; queremos explicar esto de forma más exacta.

Un lote de mercancías contiene  $N$  objetos, entre los que se encuentran  $M$  defectuosos (o distinguidos por alguna otra propiedad). Tomemos sucesivamente del lote, de forma aleatoria y sin reposición o de una vez, que es lo mismo,  $n$  objetos; en este contexto la frase *de forma aleatoria* significa que todas las muestras posibles tienen la misma probabilidad. Si designamos con  $X$  el número, concebido como variable aleatoria, de los objetos defectuosos en una muestra extraída de este modo, entonces un número natural  $k$  es evidentemente un valor de  $X$  si y solo si  $k \leq n$ ,  $k \leq M$  y  $n-k \leq N-M$ . Para el cálculo de las probabilidades  $P(X=k)$  fijemos que el suceso  $(X=k)$  ocurre si y solo si de los  $M$  objetos defectuosos existentes están contenidos  $k$  de ellos en la muestra aleatoria (para esto existen  $\binom{M}{k}$  posibilidades), y si de los  $N-M$  sin defectuosos están contenidos  $n-k$  en la muestra (para esto existen  $\binom{N-M}{n-k}$  posibilidades). Como existen en total  $\binom{N}{n}$  posibilidades de escoger  $n$  objetos de  $N$  de ellos, se obtiene precisamente para  $P(X=k)$ , aplicando la definición clásica de probabilidad, la ecuación (1), o sea,  $X$  está distribuida hipergeométricamente. Llamamos la atención de que el número (aleatorio) de los objetos defectuosos en una muestra aleatoria con reposición está distribuido binomialmente con los parámetros  $n$  y  $p = \frac{M}{N}$ .

**Ejemplo.** Sea  $N=100$ ,  $M=5$  y  $n=10$ . Designe  $X$  el número (aleatorio) de los objetos defectuosos en una muestra aleatoria.

- a) con reposición,
- b) sin reposición.

Calculemos para cada caso la probabilidad  $P(X=1)$ .

a)  $P(X=1) = b(1; 10, 0,05) = \binom{10}{1} 0,05 (1-0,05)^9 \approx 0,32$ .

$$\text{b) } P(X=1) = \frac{\binom{5}{1} \binom{100-5}{10-1}}{\binom{100}{10}} = \frac{\binom{5}{1} \binom{95}{9}}{\binom{100}{10}} \approx 0,34.$$

Nos asalta entonces la idea, de que cada una de las probabilidades de la distribución hipergeométrica y binomial no se diferencian esencialmente, si el tamaño de la muestra  $n$  es pequeña en relación con el tamaño  $N$  del lote de mercancías ( $n \ll N$ ). En este caso, por ejemplo, la no reposición de un objeto defectuoso tiene una influencia muy pequeña sobre la distribución de probabilidad para la próxima extracción. (En esta relación es interesante la proposición siguiente: también en una muestra sin reposición la probabilidad de extraer un objeto defectuoso es igual para las distintas extracciones; esta es igual a  $p = \frac{M}{N}$ .)

El teorema siguiente afirma la suposición anteriormente señalada.

**Teorema 1.** Se cumple para  $k=0, 1, 2, \dots, n$

$$\lim_{\substack{N \rightarrow \infty \\ M \rightarrow \infty \\ \frac{M}{N} = p = \text{const.}}} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}. \quad (2)$$

Renunciaremos a la demostración, que no es difícil. Del teorema 1 inferimos que se puede sustituir en el caso  $n \ll N$  las probabilidades  $P(X=k)$  de una variable aleatoria distribuida hipergeométricamente por las probabilidades  $b(k; n, p)$  de una variable aleatoria distribuida binomialmente, haciéndose  $p = \frac{M}{N}$ .

Por último, indicaremos el valor esperado y la varianza de una variable aleatoria distribuida hipergeométricamente.

**Teorema 2.** Sea  $X$  una variable aleatoria distribuida hipergeométricamente. Entonces se cumple, con  $p = \frac{M}{N}$ , que

$$EX = np, \quad (3)$$

$$D^2X = np(1-p) \frac{N-n}{N-1}. \quad (4)$$

Dejamos la demostración de esto al lector. Comparemos aún el valor esperado y la varianza del número (aleatorio) de los objetos defectuosos en una muestra sin reposición (distribución hipergeométrica), con los parámetros correspondientes en una muestra con reposición (distribución binomial, ver 4.5 (7) y (8)). Como se aprecia, los valores esperados son iguales con ambos métodos de extracción de la muestra. Por el contrario, la varianza en una muestra sin reposición es menor que en una con reposición ( $np(1-p) \frac{N-n}{N-1} < np(1-p)$  para  $1 < n \leq N$ ), pero para  $N$  grande la diferencia es pequeña ( $\lim_{N \rightarrow \infty} np(1-p) \frac{N-n}{N-1} = np(1-p)$ ), como era de esperar también sobre la base del teorema 1.

## 4.7 Distribución de Poisson

La distribución de Poisson es una distribución discreta en un número infinito numerable de valores; esta desempeña una importante función como distribución límite de la distribución binomial, en particular, para el cálculo numérico de las probabilidades  $b(k; n, p)$  cuando  $n$  es grande y  $p$  pequeña.

**Definición 1.** Sea  $\lambda$  un número positivo arbitrario. Una variable aleatoria  $X$ , que puede tomar los valores 0, 1, 2, ..., se denomina *distribuida según Poisson con el parámetro  $\lambda$* , si se cumple que

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (1)$$

para  $k=0, 1, 2, \dots$ . Se dice entonces que  $X$  posee una *distribución de Poisson con el parámetro  $\lambda$* .

La evidencia de que mediante (1) está definida una probabilidad, se obtiene directamente aplicando el desarrollo en serie de la función exponencial  $e^x = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}; -\infty < \lambda < \infty$ .

Con el objetivo de destacar la dependencia del parámetro  $\lambda$  de las probabilidades  $P(X=k)$  de una variable aleatoria  $X$ , que posee una distribución de Poisson con parámetro  $\lambda$ , se utiliza ocasionalmente la notación  $p(k; \lambda)$  para estas probabilidades

$$p(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}. \quad (2)$$

La distribución de Poisson se debe a S.D. Poisson (1781-1840), matemático francés extraordinariamente productivo, cuyo nombre está unido a numerosos conceptos de la matemática (por ejemplo, la integral de Poisson y la ecuación de Poisson en la teoría de los potenciales).

Indicaremos ahora el valor esperado y la varianza de una variable aleatoria distribuida según Poisson con el parámetro  $\lambda$ ; aquí también se aclarará la función del parámetro  $\lambda$ .

**Teorema 1.** Sea  $X$  una variable aleatoria distribuida según Poisson con el parámetro  $\lambda > 0$ . Entonces se cumple que

$$EX = \lambda, \quad (3)$$

$$D^2X = \lambda. \quad (4)$$

**Demostración.** Solo demostraremos (3); el lector debe demostrar (4) como ejercicio. Se cumple que

$$\begin{aligned} EX &= \sum_k x_k p_k = \sum_k k p(k; \lambda) = \sum_{k=0}^{\infty} k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} = \lambda. e^{\lambda} e^{-\lambda} = \lambda. \end{aligned}$$

El siguiente teorema ofrece más información sobre la influencia del parámetro  $\lambda$  en la distribución de Poisson.

**Teorema 2.** Sea  $X$  una variable aleatoria distribuida según Poisson con el parámetro  $\lambda > 0$ . Entonces se cumple que

$$\delta = \frac{1}{\sqrt{\lambda}} \text{ (coeficiente de variación).} \quad (5)$$

$$\gamma = \frac{1}{\sqrt{\lambda}} \text{ (coeficiente de asimetría).} \quad (6)$$

$$\eta = \frac{1}{\lambda} \text{ (curtosis).} \quad (7)$$

El siguiente teorema muestra una relación entre la distribución binomial y la de Poisson.

**Teorema 3.** (Teorema límite de Poisson). Se cumple para  $k=0, 1, 2, \dots$  que

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np = \lambda = \text{const.}}} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (8)$$

Demostración. Con  $p = \frac{\lambda}{n}$  se cumple que

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n \cdot n \cdots n} \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

De aquí se obtiene directamente (8), para  $n \rightarrow \infty$ ,  $p \rightarrow 0$  y  $np = \lambda = \text{constante}$  con  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$ .

El teorema (3) muestra que se pueden sustituir las probabilidades  $b(k; n, p)$  de una variable aleatoria distribuida binomialmente con los parámetros  $n$  y  $p$ , por las  $p(k; \lambda)$  de una variable aleatoria distribuida según Poisson con el parámetro  $\lambda = np$ , en el caso de un número  $n$  grande y uno  $p$  pequeño; para  $n \gg 1$  y  $p \ll 1$  se cumple, por tanto, que

$$b(k; n, p) \approx p(k; \lambda) \text{ con } \lambda = np. \quad (9)$$

Como los números  $b(k; n, p)$  son difíciles de calcular, especialmente para el caso  $n \gg 1$  y  $p \ll 1$ , la relación (9) es muy útil para la determinación numérica de probabilidades de la distribución binomial. Para el cálculo de las probabilidades de la distribución de Poisson, que se necesitan también en la aplicación de (9), son convenientes las fórmulas recursivas dadas en el siguiente teorema.

**Teorema 4.** Se cumplen las relaciones

$$p(k+1; \lambda) = \frac{\lambda}{k+1} p(k; \lambda), \quad k \geq 0 \quad (10)$$

$$p(k-1; \lambda) = \frac{k}{\lambda} p(k; \lambda), \quad k \geq 1. \quad (11)$$

Las demostraciones se obtienen directamente de (2).

Las probabilidades de la distribución de Poisson se encuentran en tablas para valores de  $\lambda$  moderadamente grandes (ver tabla 2 (12.2), allí  $\lambda \leq 20$ ); para mayores valores de  $\lambda$  conoceremos posteriormente fórmulas de aproximación.

Nos ocuparemos ahora con la cuestión de cuáles de las variables aleatorias, que se presentan en casos de aplicación, poseen una distribución de Poisson.

Si se puede interpretar una variable aleatoria  $X$  (con un modelo) como el número de ocurrencias de un suceso aleatorio  $A$  en una larga serie de experimentos independientes, en los cuales el suceso  $A$  tiene siempre una probabilidad pequeña, entonces  $X$  puede concebirse de forma aproximada como distribuida según Poisson. La fundamentación matemática de esto radica en que el número (aleatorio) de la ocurrencia de un suceso  $A$  en  $n$  repeticiones realizadas independientemente unas de otras de un mismo experimento, posee una distribución binomial con los parámetros  $n$  y  $p$ , y que en el caso  $n \gg 1$  y  $p \ll 1$  se cumple la proposición (9). (A causa de que  $p \ll 1$  se denomina también con frecuencia la distribución de Poisson como distribución de los sucesos raros, una denominación evidentemente poco acertada.) Aquí se establece, de forma conveniente, el parámetro  $\lambda$  igual a la media aritmética de los valores observados de la variable aleatoria (ver para esto (3) y 4.3, observación antes del teorema 1). Por último, nombraremos algunos ejemplos concretos de variables aleatorias, que pueden aceptarse distribuidas según Poisson de acuerdo con el modelo anteriormente ilustrado: el número (aleatorio) de llamadas que llegan a una central telefónica durante un determinado lapso, el número de roturas de los hilos que ocurren en una hilandería, para una determinada clase de tejido, dentro de un período de tiempo dado; el número de átomos de una sustancia radiactiva que se descomponen en un intervalo de tiempo fijado, etcétera.

Concluimos este epígrafe con un ejemplo.

**Ejemplo.** Una carga de simientes se vende en paquetes. Cada paquete contiene (alrededor de) 1 000 semillas. De pruebas anteriores es conocido que (aproximadamente) el 0,5 % de las semillas no pertenecen a la clase de las simientes. Calculemos la probabilidad de que en un paquete (aleatoriamente elegido) hayan más de cinco semillas que no pertenezcan a la clase de las simientes (suceso  $B$ ).

Para ello designe  $X$  el número (aleatorio) de semillas que no pertenecen a la clase de las simientes en un paquete. Se supone, de acuerdo con los datos, que  $X$  está binomialmente distribuida con los parámetros  $n=1\,000$  y  $p=0,005$ . Se cumple entonces que

$$\begin{aligned} P(B) &= P(X > 5) = 1 - P(X \leq 5) = 1 - \sum_{k=0}^5 P(X=k) \\ &= 1 - \sum_{k=0}^5 b(k; 1\,000, 0,005). \end{aligned}$$

Utilizamos (9) con  $\lambda=np=1\,000 \cdot 0,005=5$  y obtenemos

$$P(B) \approx 1 - \sum_{k=0}^5 p(k; 5) \approx 1 - 0,616 = 0,384$$

(ver tabla 2(12.2)).

## 5. Variables aleatorias continuas

En este capítulo queremos tratar las variables aleatorias continuas, cuya característica común consiste en que el dominio de valores es un intervalo (estando también permitido el conjunto  $\mathbb{R}$ ). En relación con variables aleatorias continuas nos interesa particularmente que la variable aleatoria considerada tome valores de un intervalo arbitrario dado. La probabilidad de que una variable aleatoria continua tome un valor determinado cualquiera, es siempre igual a cero, de modo que no se puede caracterizar la distribución de probabilidad de una variable aleatoria continua indicando probabilidades particulares. Luego, las variables aleatorias continuas se caracterizan por el hecho de que la probabilidad de tomar valores de un intervalo cualquiera se obtiene como el área entre el eje  $x$  y la llamada densidad de probabilidad sobre el intervalo considerado. Esto conduce, por tanto, a la aplicación del concepto de integral y en especial, a la utilización de integrales impropias.

Observe el lector la analogía de las definiciones, fórmulas y proposiciones de este capítulo con las correspondientes del capítulo 4; estas solo se diferencian con frecuencia en que en lugar del símbolo de sumatoria y de la probabilidad particular están el símbolo de integral y la diferencial de la función de distribución, respectivamente.

Utilizando una teoría general de la integración y la medida, se puede tratar al mismo tiempo variables aleatorias discretas y continuas. De esta forma se pueden representar de forma única, mediante integrales adecuadas, las probabilidades, el valor esperado, la varianza y los momentos de orden superior que nos interesan, obteniéndose, naturalmente, tanto en el caso discreto como continuo, las definiciones, fórmulas y proposiciones dadas en este libro.

### 5.1 Definición de variable aleatoria continua

**Definición 1.** Una variable aleatoria  $X$  se llama *continua*, si existe una función  $f_X$  no negativa definida sobre el conjunto  $\mathbb{R}$  de los números reales, al menos continua a trozos, de modo que

$$P(a \leq X \leq b) = \int_a^b f_X(x) \, dx \quad (1)$$

para todos los números reales  $a$  y  $b$  con  $a \leq b$  (fig. 26).

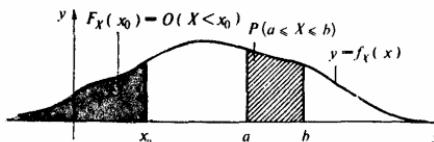


Figura 26

Desde el punto de vista del Cálculo de probabilidades, podemos entender que una variable aleatoria continua  $X$  está dada cuando conocemos la función  $f_x$ . La función  $f_x$  se llama *densidad de probabilidad* (también: *densidad de distribución*, *densidad o función de densidad*) de la variable aleatoria  $X$ . El teorema siguiente muestra que mediante la función de densidad está fijada realmente la función de distribución de la variable aleatoria considerada (ver 4.2, teorema 1).

**Teorema 1.** Sea  $X$  una variable aleatoria continua con la función de densidad  $f_x$ . Entonces se cumplen las proposiciones siguientes:

1.  $f_x(x) \geq 0$  para todo  $x \in \mathbb{R}$ ,  $\int_{-\infty}^{\infty} f_x(x) dx = 1$ .

2.  $F_x(x) = \int_{-\infty}^x f_x(t) dt$  (fig. 27).

3. La función de distribución  $F_x$  es una función continua, que es diferenciable en todos los puntos de continuidad de  $f_x$ , cumpliéndose  $F'_x(x) = f_x(x)$ .

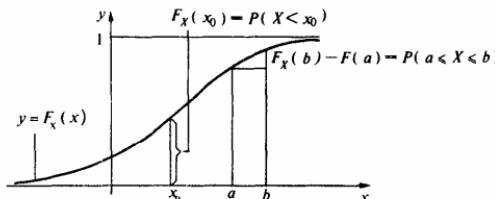


Figura 27

También aquí dejamos la demostración al lector; se debe observar que para una variable aleatoria continua  $X$  y para un número real cualquiera  $c$ , se cumple que (ver 4.1 (3)).

$$P(X=c) = \int_c^c f_x(x) dx = 0.$$

Veamos ahora un ejemplo.

**Ejemplo.** Consideremos la función (fig. 28), dada por

$$f(x) = \begin{cases} \frac{2}{b-a} \left(1 - \frac{2}{b-a} \left|x - \frac{a+b}{2}\right|\right) & \text{para } a \leq x \leq b, \\ 0 & \text{para los demás.} \end{cases}$$

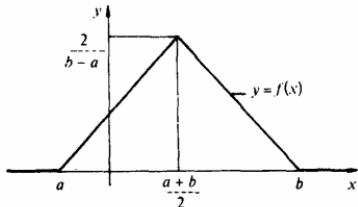


Figura 28

Esta función es no negativa y se cumple que  $\int_{-\infty}^{\infty} f(x) dx = 1$  (fig. 28). Si una variable aleatoria continua  $X$  posee esta función  $f$  como función de densidad ( $f_x = f$ ), entonces se cumple que, por ejemplo,

$$P(X \leq a) = 0, \quad P\left(a \leq X \leq \frac{a+b}{2}\right) = P\left(\frac{a+b}{2} \leq X \leq b\right) = \frac{1}{2},$$

$$P(X \geq b) = 1.$$

Para la función de distribución  $F$  correspondiente a esta variable aleatoria (fig. 29) se obtiene que

$$F(x) = P(X < x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{para } x \leq a, \\ 2\left(\frac{x-a}{b-a}\right)^2 & \text{para } a \leq x \leq \frac{a+b}{2}, \\ 1 - 2\left(\frac{b-x}{b-a}\right)^2 & \text{para } \frac{a+b}{2} \leq x \leq b, \\ 1 & \text{para } x \geq b. \end{cases}$$

La distribución de probabilidad caracterizada por la densidad de probabilidad  $f$  o la función de distribución  $F$ , se denomina *distribución triangular*.

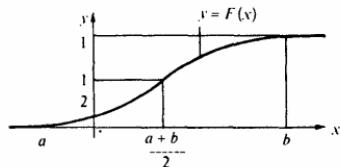


Figura 29

A continuación damos para algunas funciones especiales  $g$ , la relación entre la densidad de probabilidad  $f_x$  de una variable aleatoria continua  $X$  y la  $f_y$  de la variable aleatoria  $Y=g(X)$ .

**Teorema 2.** Sea  $X$  una variable aleatoria continua con la función de densidad  $f_x$ .

1. La variable aleatoria  $Y=ax+b$  ( $a \neq 0$ ,  $b$  reales) posee la función de densidad

$$f_y(x) = \frac{1}{|a|} f_x\left(\frac{x-b}{a}\right), \quad -\infty < x < \infty. \quad (2)$$

2. La variable aleatoria  $Y=X^2$  posee la función de densidad  $f_Y$ .

$$f_Y(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ \frac{f_X(\sqrt{x}) + f_X(-\sqrt{x})}{2\sqrt{x}} & \text{para } x > 0. \end{cases} \quad (3)$$

3. La variable aleatoria  $Y=|X|$  posee la función de densidad  $f_Y$ .

$$f_Y(x) = \begin{cases} 0 & \text{para } x \leq 0 \\ f_X(x) + f_X(-x) & \text{para } x > 0. \end{cases} \quad (4)$$

La demostración de este teorema se obtiene fácilmente con el teorema 3 del epígrafe 4.1, aplicando la proposición 3 del teorema 1.

## 5.2 Características numéricas de las variables aleatorias continuas

Trataremos en este epígrafe el valor esperado y la varianza como características numéricas importantes de las variables aleatorias continuas. Observe el lector las analogías con las definiciones y proposiciones correspondientes del epígrafe 4.3 sobre las características numéricas de las variables aleatorias discretas.

**Definición 1.** Sea  $X$  una variable aleatoria continua con la densidad de probabilidad  $f_X$ . Entonces el número  $EX$  definido por

$$EX = \int_{-\infty}^{\infty} xf_X(x) dx \quad (1)$$

se llama *valor esperado* de la variable aleatoria  $X$ ; aquí se supone que la integral situada en el miembro derecho de (1) converge absolutamente; o sea, se cumple que

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty.$$

**Ejemplo.** Calculemos para la variable aleatoria  $X$ , considerada en el ejemplo del epígrafe 5.1, el valor esperado:

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} xf_X(x) dx = \int_a^b x \frac{2}{b-a} \left( 1 - \frac{2}{b-a} \left| x - \frac{a+b}{2} \right| \right) dx \\ &= \int_a^{\frac{a+b}{2}} x \frac{2}{b-a} \left( 1 - \frac{2}{b-a} \left( -x + \frac{a+b}{2} \right) \right) dx \\ &\quad + \int_{\frac{a+b}{2}}^b x \frac{2}{b-a} \left( 1 - \frac{2}{b-a} \left( x - \frac{a+b}{2} \right) \right) dx = \frac{a+b}{2}. \end{aligned}$$

Los teoremas siguientes son útiles para el cálculo con valores esperados.

**Teorema 1.** Sea  $X$  una variable aleatoria continua con el valor esperado  $EX$  y sean  $a \neq 0$  y  $b$ , números reales cualesquiera. Entonces se cumple que

$$E(aX+b) = aEX+b. \quad (2)$$

**Demostración.** Si la variable aleatoria  $X$  posee la densidad de probabilidad  $f_x$ , entonces la variable aleatoria  $Y=aX+b$  posee la densidad de probabilidad  $f_y$ ,

$$f_y(x) = \frac{1}{|a|} f_x\left(\frac{x-b}{a}\right)$$

(ver 5.1, teorema 2, proposición 1). Con esto obtenemos aplicando (1) y  $\int_{-\infty}^{\infty} f_X(t) dt = 1$

$$\begin{aligned} EY &= E(aX+b) = \int_{-\infty}^{\infty} xf_y(x) dx = \int_{-\infty}^{\infty} x \frac{1}{|a|} f_x\left(\frac{x-b}{a}\right) dx \\ &= \int_{-\infty}^{\infty} (at+b)f_X(t) dt = a \int_{-\infty}^{\infty} tf_X(t) dt + b \int_{-\infty}^{\infty} f_X(t) dt \\ &= aEX + b. \end{aligned}$$

(En el cálculo se debe realizar una diferenciación de casos con respecto al signo de  $a$ .)

Por tanto, se cumple en particular para una variable aleatoria continua  $X$ , la relación

$$E(X-EX) = 0. \quad (3)$$

**Teorema 2.** Sea  $X$  una variable aleatoria continua con la densidad de probabilidad  $f_x$  y  $g$  una función real continua definida sobre el eje real. Si la integral  $\int_{-\infty}^{\infty} |g(x)| f_x(x) dx$  converge absolutamente (es decir, si se cumple que  $\int_{-\infty}^{\infty} |g(x)| f_x(x) dx < \infty$ ), entonces se cumple que

$$Eg(X) = \int_{-\infty}^{\infty} g(x) f_x(x) dx. \quad (4)$$

Renunciaremos a la exposición (por lo demás no muy sencilla) de la demostración. Sin embargo, observamos que para  $g(x) = x$  se cumple el teorema 2 sobre la base de la definición 1.

El cálculo del valor esperado  $Eg(X)$  sin recurrir al teorema 2, tendría que realizarse con la fórmula  $Eg(X) = \int_{-\infty}^{\infty} y f_{g(X)}(y) dy$ , lo cual exige, por consiguiente, el conocimiento de la densidad de probabilidad  $f_{g(X)}$  de la variable aleatoria  $g(X)$  (ver demostración del teorema 1). Esto no es necesario utilizando (4), mediante la cual se simplifica considerablemente en muchas ocasiones el cálculo de  $Eg(X)$ ; de aquí se desprende la importancia del teorema 2.

Para  $g(x) = (x-c)^j$  y  $g_j(x) = |x-c|^j$  ( $j$  un número natural cualquiera y  $c$  un número real arbitrario), se obtiene según (4)

$$E(X-c)^j = \int_{-\infty}^{\infty} (x-c)^j f_x(x) dx \quad (5)$$

y

$$E|X-c| = \int_{-\infty}^{\infty} |x-c| f_X(x) dx \quad (6)$$

respectivamente, siempre y cuando la integral situada en el miembro derecho de (6) sea convergente.

Definición 2. Sea  $X$  una variable aleatoria continua con el valor esperado  $EX$  y la densidad de probabilidad  $f_X$ . Entonces el número  $D^2X$  definido por

$$D^2X = E(X-EX)^2 = \int_{-\infty}^{\infty} (x-EX)^2 f_X(x) dx \quad (7)$$

se llama *varianza (dispersión)* de la variable aleatoria  $X$ , suponiéndose la convergencia de la integral situada en el miembro derecho de (7). El número

$$\sigma_X = \sqrt{D^2X} \quad (8)$$

se llama la *desviación estándar* de la variable aleatoria  $X$ .

Ejemplo. Calculemos la varianza para la variable aleatoria considerada en el ejemplo del epígrafe 5.1; aquí emplearemos  $EX = \frac{a+b}{2}$ :

$$\begin{aligned} D^2X &= \int_{-\infty}^{\infty} (x-EX)^2 f_X(x) dx = \int_a^b \left( x - \frac{a+b}{2} \right)^2 \frac{2}{b-a} \left( 1 - \frac{2}{b-a} \left| x - \frac{a+b}{2} \right| \right) dx \\ &= 2 \int_0^{\frac{b-a}{2}} t^2 \frac{2}{b-a} \left( 1 - \frac{2}{b-a} \right) dt = \frac{1}{24} (b-a)^2. \end{aligned}$$

Los teoremas siguientes son útiles para el cálculo de la varianza.

Teorema 3. Sea  $X$  una variable aleatoria continua con el valor esperado  $EX$ , la varianza  $D^2X$  y la densidad de probabilidad  $f_X$ . Entonces existe  $EX^2$  y se cumple que

$$D^2X = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left( \int_{-\infty}^{\infty} x f_X(x) dx \right)^2 = EX^2 - (EX)^2. \quad (9)$$

La demostración de este teorema se realiza de forma análoga a la del teorema 3(4.3). (Formalmente se tiene que sustituir  $\sum_k$  por  $\int_{-\infty}^{\infty}$ ,  $x_k$  por  $x$  y  $p_k$  por  $f_X(x) dx$ .)

Teorema 4. Sea  $X$  una variable aleatoria continua con la varianza  $D^2X$  y sean  $a \neq 0$  y  $b$  números reales cualesquiera. Entonces se cumple que

$$D^2(aX+b) = a^2 D^2X. \quad (10)$$

La demostración del teorema 5(4.3) es válida para aquí también.

Por consiguiente, para una variable aleatoria continua  $X$  se cumplen también las relaciones

$$D^2(-X) = D^2X \quad (11)$$

y

$$D^2 \left( \frac{X}{\sqrt{D^2 X}} \right) = 1. \quad (12)$$

Como en el caso de las variables aleatorias discretas, se utiliza también para las continuas el concepto *centrar* para el paso de  $X$  a  $X - EX$ , el de *normar* para el de  $X$  a  $\frac{X}{\sqrt{D^2 X}}$  y el de *estandarizar* para el de  $X$  a  $\frac{X - EX}{\sqrt{D^2 X}}$ .

Por último queremos advertir que el valor esperado y la varianza, como para el caso de las variables aleatorias discretas, son momentos especiales que caracterizaremos en la definición siguiente.

**Definición 3.** Sea  $X$  una variable aleatoria continua con la densidad de probabilidad  $f_X$ ,  $j$  un número natural y  $c$  un número real. Entonces se llaman

$$\mu_j(c) = E(X - c)^j = \int_{-\infty}^{\infty} (x - c)^j f_X(x) dx \quad (13)$$

y

$$\alpha_j(c) = E|X - c|^j = \int_{-\infty}^{\infty} |x - c|^j f_X(x) dx \quad (14)$$

los momentos ordinario y absoluto de orden  $j$  con respecto a  $c$  respectivamente, suponiéndose la convergencia de la integral situada a la derecha en (14). Para  $c=0$  se habla de *momentos iniciales* y para  $c=EX$  de *momentos centrales* (se supone la existencia de  $EX$ ).

Las proposiciones sobre momentos dadas a continuación de la definición 3 (4.3), se cumplen también para variables aleatorias continuas. De igual modo que para las variables aleatorias discretas, se definen para las continuas las características numéricas derivadas de los momentos: *coeficiente de variación*, *coeficiente de asimetría* y *curtosis* (ver 4.3, definición 4).

### 5.3 Distribución continua uniforme

En este y en los siguientes epígrafes trataremos algunas distribuciones de probabilidad especiales de variables aleatorias continuas.

**Definición 1.** Una variable aleatoria continua  $X$  se denomina *distribuida uniformemente* (sobre el intervalo  $[a, b]$ ,  $a < b$ ), si la densidad de probabilidad  $f_X$  tiene la forma

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b, \\ 0 & \text{para los demás.} \end{cases} \quad (1)$$

Se dice también que  $X$  posee una *distribución uniforme* (sobre el intervalo  $[a, b]$ ) o una *distribución rectangular* (fig. 30).

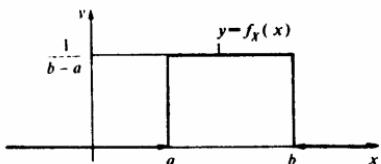


Figura 30

Para la función de distribución  $F_X$  (fig. 31) se obtiene

$$F_X(x) = P(X < x) = \int_{-\infty}^x f_X(t) dt = \begin{cases} 0 & \text{para } x \leq a, \\ \frac{x-a}{b-a} & \text{para } a \leq x \leq b, \\ 1 & \text{para } x \geq b. \end{cases} \quad (2)$$

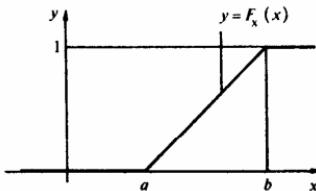


Figura 31

Para el valor esperado  $EX$  se obtiene

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \quad (3)$$

y para la varianza se tiene

$$D^2X = \int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx = \int_a^b \left( x - \frac{a+b}{2} \right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}. \quad (4)$$

Para una variable aleatoria continua existe una distribución uniforme, si y solo si esta toma valores de subintervalos de igual longitud pertenecientes a su dominio de valores y que es a su vez un intervalo, con igual probabilidad. En casos de aplicación se acepta que una variable aleatoria está distribuida uniformemente, si ésta -hablando sin mucha precisión- no prefiere ninguno de los subintervalos de igual longitud (de su dominio de valores).

## 5.4 Distribución normal

La distribución normal es una distribución de variables aleatorias continuas, que se utiliza mucho en las aplicaciones del Cálculo de probabilidades. Pero antes de referirnos a esto, queremos caracterizar la distribución normal mediante la densidad de probabilidad correspondiente e investigarla detalladamente.

**Definición 1.** Sea  $\mu$  un número real y  $\sigma$  un número positivo. Una variable aleatoria continua se denomina *distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$* , si la densidad de probabilidad  $f_x$  tiene la forma

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty. \quad (1)$$

Se dice también que  $X$  posee una *distribución normal con los parámetros  $\mu$  y  $\sigma^2$*  o una *distribución  $N(\mu, \sigma^2)$*  (fig. 32).

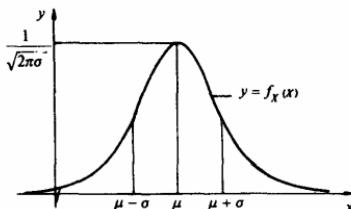


Figura 32

La demostración de que mediante (1) está definida realmente una densidad de probabilidad, se basa fundamentalmente sobre la ecuación

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}.$$

Para la densidad de probabilidad de una variable aleatoria distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$ , se utiliza generalmente la notación  $\phi$ , donde la dependencia de  $\mu$  y  $\sigma^2$  queda expresada en la forma

$$\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty \quad (2)$$

La influencia de los parámetros  $\mu$  y  $\sigma^2$  sobre la situación y la forma de la curva dada por (2), se reconoce de la figura 32; la curva es simétrica con respecto a la recta  $x=\mu$ , posee puntos de inflexión en  $\mu-\sigma$  y  $\mu+\sigma$  y tiene en  $x=\mu$  un máximo con el valor de la

función  $\frac{1}{\sqrt{2\pi}\sigma}$ .

Para la función de distribución  $F_x$  de una variable aleatoria  $X$ , distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$ , se cumple que

$$F_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \quad (3)$$

La integración de la función que está en (3) bajo el símbolo de integral no es realizable sobre un intervalo cerrado, pero se puede indicar con la exactitud requerida un valor aproximado de la integral anterior para todo  $x$ , con métodos apropiados de la matemática práctica.

Para la función de distribución de una variable aleatoria distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$ , se utiliza generalmente la notación  $\Phi$ , donde de forma análoga a (2), la dependencia de  $\mu$  y  $\sigma^2$  queda expresada en la forma

$$\Phi(x; \mu, \sigma^2) = \int_{-\infty}^x \phi(t; \mu, \sigma^2) dt = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \quad (4)$$

El teorema siguiente pone de manifiesto la significación teórico-probabilística de los parámetros  $\mu$  y  $\sigma^2$ .

**Teorema 1.** Sea  $X$  una variable aleatoria distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$ . Entonces se cumple que

$$EX = \mu, \quad (5)$$

$$D^2X = \sigma^2. \quad (6)$$

**Demostración.** Con  $t = \frac{x-\mu}{\sigma}$  y  $\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$  se obtiene que

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} x \phi(x; \mu, \sigma^2) dx = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \mu. \end{aligned}$$

De esta expresión y con

$$\int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$$

se obtiene que

$$\begin{aligned} D^2X &= \int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx = \int_{-\infty}^{\infty} (x - \mu)^2 \phi(x; \mu, \sigma^2) dx \\ &= \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt = \sigma^2. \end{aligned}$$

El teorema siguiente se refiere a momentos de orden superior de la distribución normal y a características numéricas derivadas de los momentos.

**Teorema 2.** Sea  $X$  una variable aleatoria distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$ . Entonces se cumple que

$$\mu_{2k+1}(EX) = E(X - EX)^{2k+1} = 0, \quad k=1, 2, \dots, \quad (7)$$

$$\mu_{2k}(EX) = E(X - EX)^{2k} = 1 \cdot 3 \cdots (2k-1) \sigma^{2k}, \quad k=1, 2, \dots, \quad (8)$$

$$\theta = \frac{\sigma}{\mu} \text{ (coeficiente de variación)}, \quad (9)$$

$$\gamma_0 = 0 \text{ (coeficiente de asimetría)}, \quad (10)$$

$$\eta_0 = 0 \text{ (curtosis)}, \quad (11)$$

donde se supone en (9) que  $\mu \neq 0$ .

El lector puede realizar independientemente la demostración sencilla de estas fórmulas. Añadimos, que una variable aleatoria distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$  es simétrica con respecto a  $x=\mu$  y aseguramos que todos los momentos de orden impar referidos a  $\mu$ , así como el coeficiente de asimetría, son iguales a cero. La curtosis está definida, precisamente, de modo que esta característica numérica sea igual a cero para el caso especial de la distribución normal.

Trataremos ahora la distribución  $N(0,1)$ . Queremos denotar con  $\varphi$  la densidad de probabilidad de una variable aleatoria distribuida normalmente con los parámetros 0 y 1, y con  $\Phi$ , la función de distribución correspondiente. Se cumple (figs. 33 y 34), por tanto, que

$$\varphi(x) = \varphi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty, \quad (12)$$

$$\Phi(x) = \Phi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad -\infty < x < \infty. \quad (13)$$

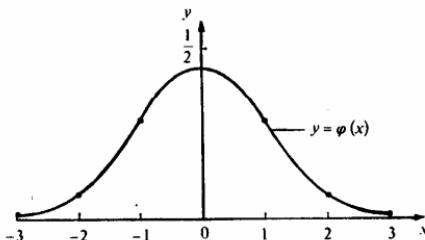


Figura 33

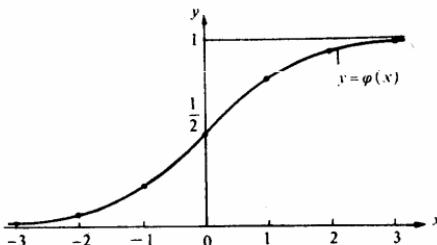


Figura 34

La función  $\Phi$  (y además  $\varphi$ ) está tabulada (ver tabla 3 (12.3)); a causa de

$$\varphi(-x) = \varphi(x), \quad -\infty < x < \infty, \quad (14)$$

$$\Phi(-x) = 1 - \Phi(x), \quad -\infty < x < \infty, \quad (15)$$

nos podemos limitar en este caso a argumentos  $x$  no negativos.

Calculemos ahora la probabilidad de que una variable aleatoria  $X$  distribuida normalmente con los parámetros 0 y 1, tome valores entre  $-k$  y  $+k$  ( $k$ : número natural). Se cumple que:

$$P(|X| < k) = P(-k < X < k) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1. \quad (16)$$

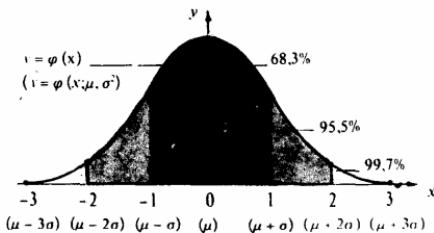


Figura 35

Aquí hemos utilizado (15) y  $P(X=c)=0$  ( $X$ , variable aleatoria continua y  $c$ , número real). Para  $k=1,2,3$  obtenemos, por consiguiente, (ver tabla 3(12.3) y fig. 35).

$$P(|X|<1) \approx 0.683 = 68.3\%, \quad (17)$$

$$P(|X|<2) \approx 0.955 = 95.5\%, \quad (18)$$

$$P(|X|<3) \approx 0.997 = 99.7\%. \quad (19)$$

La relación (19) expresa que es *prácticamente seguro*, que una variable aleatoria distribuida normalmente con los parámetros  $\mu=0$  y  $\sigma^2=1$  tome solo valores entre  $-3$  y  $+3$ . Observe el lector que la probabilidad de que una variable aleatoria distribuida normalmente con los parámetros 0 y 1 tome valores de un intervalo arbitrario dado, es positiva, pero que es prácticamente imposible que una tal variable aleatoria tome valores de un intervalo disjunto con  $\{x : x \in \mathbb{R} \wedge -3 < x < 3\}$ .

Mostraremos ahora como se pueden calcular los valores  $\Phi(x; \mu, \sigma^2)$  de la función de distribución de una variable aleatoria distribuida normalmente con parámetros cualesquiera  $\mu$  y  $\sigma^2$ , sobre la base de los valores  $\Phi(x)$  de la función de distribución  $\Phi$  de una variable aleatoria distribuida normalmente con los parámetros  $\mu=0$  y  $\sigma^2=1$ .

**Teorema 3.** Para todo número real  $x$  se cumple que

$$\phi(x; \mu, \sigma^2) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right), \quad (20)$$

$$\Phi(x; \mu, \sigma^2) = \Phi\left(\frac{x-\mu}{\sigma}\right). \quad (21)$$

#### Demostración

$$\begin{aligned} \phi(x; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \\ &= \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right), \end{aligned}$$

$$\begin{aligned} \Phi(x; \mu, \sigma^2) &= \int_{-\infty}^x \phi(t; \mu, \sigma^2) dt = \frac{1}{\sigma} \int_{-\infty}^x \varphi\left(\frac{t-\mu}{\sigma}\right) dt \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(u) du = \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

De aquí se obtiene fácilmente la proposición siguiente:

**Teorema 4.** Si  $X$  posee una distribución  $N(\mu, \sigma^2)$ , entonces  $\frac{X-\mu}{\sigma}$  posee una distribución  $N(0,1)$ .

Demostración

$$\begin{aligned} F_{\frac{x-\mu}{\sigma}}(x) &= P\left(\frac{X-\mu}{\sigma} < x\right) = P(X < x\sigma + \mu) \\ &= \Phi(x\sigma + \mu; \mu, \sigma^2) = \Phi\left(\frac{x\sigma + \mu - \mu}{\sigma}\right) = \Phi(x). \end{aligned}$$

(Observemos que en virtud de  $EX=\mu$  y  $D^2X=\sigma^2$ , la variable aleatoria  $\frac{X-\mu}{\sigma}$  posee siempre el valor esperado cero y la varianza uno; la proposición fundamental del teorema 4 consiste en que si  $X$  está distribuida normalmente, entonces  $\frac{X-\mu}{\sigma}$  también lo está.)

Estas proposiciones permiten calcular de forma sencilla, utilizando una tabla para  $\Phi$ , la probabilidad de que una variable aleatoria  $X$  distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$  tome un valor de un intervalo arbitrario. Se cumple que

$$P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \quad (22)$$

En particular, obtenemos para un número natural  $k$  cualquiera que

$$P(|X-\mu| < k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1, \quad (23)$$

(ver (16)), de donde se obtiene para  $k=1,2,3$ , utilizando (17), (18) y (19)

$$P(|X-\mu| < \sigma) \approx 0,683 = 68,3\%, \quad (24)$$

$$P(|X-\mu| < 2\sigma) \approx 0,955 = 95,5\%, \quad (25)$$

$$P(|X-\mu| < 3\sigma) \approx 0,997 = 99,7\%. \quad (26)$$

Luego, es prácticamente seguro que una variable aleatoria distribuida normalmente con los parámetros  $\mu$  y  $\sigma^2$  tome solo valores entre  $\mu-3\sigma$  y  $\mu+3\sigma$ , o sea, que estén a una distancia del valor esperado  $\mu$  menor que el triple de la desviación estándar  $\sigma$ . Esta regla se llama *regla 3 σ* (ver fig. 35).

Queremos tratar ahora la *existencia de la distribución normal*. Para muchas variables aleatorias que aparecen en planteamientos de problemas prácticos, se muestra (por ejemplo, sobre la base de los valores observados de la variable aleatoria considerada especialmente) que la distribución de probabilidad se puede describir muy bien a través de una distribución normal. Una característica común de estas variables aleatorias consiste frecuentemente, en que estas se obtienen mediante superposición aditiva de un número elevado de efectos aleatorios, independientes unos de otros, teniendo cada uno una influencia insignificante sobre la variable aleatoria considerada, en comparación con la suma de los otros efectos. Posteriormente daremos la fundamentación matemática de que tales variables aleatorias puedan concebirse, en buena aproximación, distribuidas normalmente (ver 7.6). Aquí solo queremos informar que los errores de observación en un proceso de medición (por ejemplo, en mediciones de longitud) y las propiedades de un producto, en una fabricación en serie, que se pueden describir numéricamente (por ejemplo, la resis-

tencia a la compresión de cubos de hormigón o del contenido de botellas llenadas automáticamente), se pueden concebir como variables aleatorias distribuidas normalmente.

Ejemplo. En una cepilladora de metales se producen discos y se investiga su grosor  $X$ . Sobre la base de las experiencias existentes, se supone que  $X$  está distribuida normalmente y que para una determinada graduación de la máquina posee el valor esperado  $EX=\mu=10$  mm y la varianza  $D^2X=\sigma^2=(0,02 \text{ mm})^2$ . Un disco tiene las medidas adecuadas y, por tanto, está en condiciones de ser utilizado, si su grosor está entre 9,97 y 10,05 mm. Calculemos la probabilidad de que un disco posea las medidas adecuadas; para ello utilizaremos (22), (15) y la tabla 3(12.3):

$$P(9,97 < X < 10,05) = \Phi\left(\frac{10,05 - 10}{0,02}\right) - \Phi\left(\frac{9,97 - 10}{0,02}\right)$$

$$= \Phi(2,5) - \Phi(-1,5) = \Phi(2,5) + \Phi(1,5) - 1 \approx 0,927.$$

Considerando los límites de tolerancia dados y la simetría de la distribución normal, es evidentemente más conveniente elegir una graduación de la máquina con  $\mu=10,1$  mm. Para una varianza fija  $\sigma^2=(0,02 \text{ mm})^2$  se obtiene el valor 0,955 para la probabilidad buscada, lo que puede confirmar directamente el lector con (25).

Queremos concluir nuestras consideraciones sobre la distribución normal con algunas observaciones acerca de la historia de esta distribución tan nombrada y utilizada hoy en día. Se puede tomar como fecha de nacimiento de la distribución normal el 12 de noviembre de 1733; ese día se publicó un pequeño escrito de A. De Moivre (1667-1754, matemático relevante que fue desterrado de Francia y que en Londres se ocupó en dar indicaciones a los jugadores de azar), en el cual la distribución normal, incluyendo su ecuación de definición, se deducía como distribución límite de la distribución binomial. Las aplicaciones prácticas se obtuvieron solo mediante las investigaciones astronómicas intensivas de P.S. Laplace (1749-1827, en 1812 apareció su gran obra sobre el Cálculo de probabilidades) y C.F. Gauss (1777-1855) dentro de la teoría de los errores de observación, con lo cual la distribución normal fue redescubierta. Por esto, en los países de habla germana se designa la gráfica de la densidad de probabilidad de la distribución normal como curva de la campana de Gauss. La llamada integral del error de Gauss

$$G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (27)$$

se relaciona con la función de distribución  $\Phi$  de la distribución  $N(0,1)$  mediante las ecuaciones

$$G(x) = 2\Phi(x\sqrt{2}) - 1,$$

$$\Phi(x) = \frac{1}{2} + \frac{1}{2} G\left(\frac{x}{\sqrt{2}}\right). \quad (28)$$

A la divulgación de la distribución normal contribuyó decisivamente el científico belga A. Quetelet (1796-1874), quien fue activo en numerosos campos, y se considera como descubridor de la distribución normal para la Biometría y de quien provino también el nombre de *distribución normal*. Esta denominación dio motivo a todo tipo de interpretaciones erróneas. Uno de los méritos de K. Pearson (1857-1936, quien se ocupó además intensivamente de la historia de la distribución normal), es haber comprobado que en la naturaleza existen variables aleatorias que no están distribuidas normalmente y que esto no es algo anormal.

## 5.5 Distribución exponencial

La distribución exponencial es una distribución de variables aleatorias continuas, que se presenta en casos de aplicación, en particular, en la descripción de tiempos y de diferen-

cias de tiempo dependientes de la causalidad. Desde el punto de vista matemático, la distribución exponencial se caracteriza por ser muy fácil de manejar.

**Definición 1.** Sea  $\alpha$  un número positivo. Una variable aleatoria continua  $X$  se denomina *distribuida exponencialmente con el parámetro  $\alpha$* , si la densidad de probabilidad  $f_X$  tiene la forma

$$f_X(x) = \begin{cases} 0 & \text{para } x \leq 0 \\ \alpha e^{-\alpha x} & \text{para } x > 0. \end{cases} \quad (1)$$

Se dice también que  $X$  posee una *distribución exponencial con el parámetro  $\alpha$*  (fig. 36).

(El lector debe reflexionar si mediante (1) está definida realmente una distribución de probabilidad, es decir, si se cumple en particular que  $\int_{-\infty}^{\infty} f_X(x) dx = 1$ .)

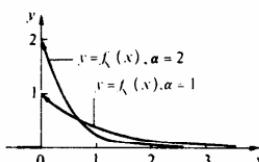


Figura 36

Para la función de distribución  $F_X$  de una variable aleatoria  $X$  distribuida exponencialmente con el parámetro  $\alpha$  (fig. 37), se cumple que

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \begin{cases} 0 & \text{para } x \leq 0, \\ 1 - e^{-\alpha x} & \text{para } x \geq 0. \end{cases} \quad (2)$$

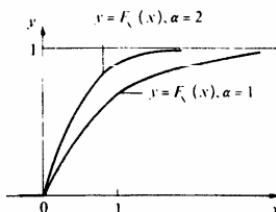


Figura 37

Ahora damos el valor esperado y la varianza de una variable aleatoria distribuida exponencialmente con el parámetro  $\alpha > 0$  donde se muestra también la significación teórica probabilística del parámetro  $\alpha$ .

**Teorema 1.** Sea  $X$  una variable aleatoria distribuida exponencialmente con el parámetro  $\alpha > 0$ . Entonces se cumple que

$$EX = \frac{1}{\alpha}, \quad (3)$$

$$D^2X = \left( \frac{1}{\alpha} \right)^2. \quad (4)$$

**Demostración.** Sólo demostraremos (3); la demostración de (4) se desarrolla de forma similar. Se cumple que

$$\begin{aligned} \int_0^b xae^{-ax} dx &= -xe^{-ax} \Big|_0^b + \int_0^b e^{-ax} dx \\ &= -b e^{-ab} - \frac{1}{a} e^{-ab} + \frac{1}{a}. \end{aligned}$$

Con  $\lim_{b \rightarrow \infty} (-b e^{-ab}) = \lim_{b \rightarrow \infty} \left( -\frac{1}{a} e^{-ab} \right) = 0$  obtenemos que

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} xae^{-ax} dx = \lim_{b \rightarrow \infty} \int_0^b xae^{-ax} dx = \frac{1}{a}.$$

Si  $X_1$  y  $X_2$  están distribuidas exponencialmente con los parámetros  $a_1$  y  $a_2$ , respectivamente, entonces se cumplen en caso de que  $a_1 < a_2$  las inecuaciones  $EX_1 > EX_2$  y  $D^2X_1 > D^2X_2$ . Estas proposiciones coinciden bien con la idea de la distribución exponencial, que se logra con la figura 36.

**Ejemplo.** Calculemos la probabilidad de que una variable aleatoria  $X$ , distribuida exponencialmente con el parámetro  $a > 0$ , tome un valor que sea menor que el valor esperado. Con (3) y (2) se obtiene que

$$P(X < EX) = P\left(X < \frac{1}{a}\right) = F_X\left(\frac{1}{a}\right) = 1 - e^{-\frac{1}{a}} = 1 - e^{-1} = 0,63.$$

Esta probabilidad es, por consiguiente, independiente de  $a$  y es mayor que 0,5.

Para concluir, queremos nombrar algunas variables aleatorias que se presentan en casos de aplicación, cuya distribución de probabilidad se describe frecuentemente mediante una distribución exponencial: duración de llamadas telefónicas, diferencia de tiempo entre la ocurrencia de interrupciones en un parque de máquinas o, más general, entre el encuentro de clientes en una instalación de servicios, tiempo de vida de elementos de contacto, así como de seres vivientes, etc. Aquí se hará, de modo conveniente, el parámetro  $a$  igual al inverso de la media aritmética de los valores observados de la variable aleatoria considerada en cada ocasión (ver (3) y 4.3, observación antes del teorema 1).

## 5.6 Distribución $\chi^2$ , $t$ y $F$

En este epígrafe presentaremos otras distribuciones de probabilidad de variables aleatorias continuas, que desempeñan una función en la estadística matemática y que en esta relación se denominan *distribuciones de prueba*; se trata de las distribuciones  $\chi^2$ ,  $t$  y  $F$ . Aquí caracterizaremos en cada ocasión la distribución por medio de la densidad de probabilidad e indicaremos el valor esperado y la varianza. Renunciaremos a las demostraciones; el lector interesado las encontrará en otra bibliografía.

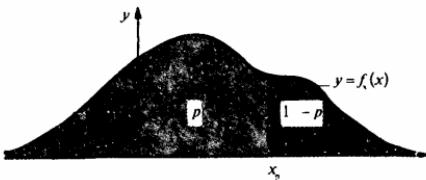
Para la realización práctica de procedimientos estadísticos frecuentemente se necesita para un valor  $p$  dado ( $0 < p < 1$ ) un valor  $x_p$  de la variable aleatoria  $X$  correspondiente,

para el cual la probabilidad de que  $X$  tome valores mayores que  $x_p$  sea igual a  $1-p$  ( $P(X > x_p) = 1-p$ ). Tales valores se denominan percentiles de orden  $p$ , cuya caracterización exacta, utilizando la función de distribución  $F_x$ , es el objeto de la definición siguiente.

Definición 1. Sea  $X$  una variable aleatoria continua (densidad de probabilidad  $f_x$ , función de distribución  $F_x$ ) y  $p$  un número situado entre cero y uno. Entonces un número  $x_p$  se llama *percentil de orden p*, si se cumple que (fig. 38)

$$F_X(x_p) = p.$$

Un percentil de orden  $p = \frac{1}{2}$  se llama *mediana*.



**Figura 38**

Para las distribuciones de prueba que se tratan a continuación, en el capítulo 12 se dan algunos percentiles.

### 5.6.1 Distribución $\chi^2$

Definición 2. Sea  $m$  un número natural. Una variable aleatoria continua  $X$  se denomina *distribuida  $\chi^2$  con  $m$  grados de libertad*, si la densidad de probabilidad  $f_x$  tiene la forma

$$f_x(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ \frac{1}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} & \text{para } x > 0. \end{cases} \quad (2)$$

Se dice también que  $X$  posee una *distribución  $\chi^2$*  con  $m$  *grados de libertad* (fig. 39). Denotamos el percentil de orden  $p$  de la distribución  $\chi^2$  con  $m$  grados de libertad con  $\chi^2_{m,p}$ .

En (2)  $\Gamma$  es la llamada función gamma completa definida por

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \quad z > 0. \quad (3)$$

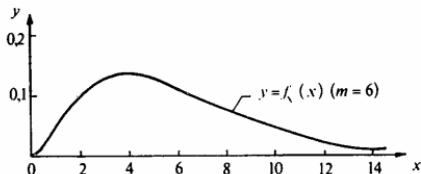


Figura 39

La función gamma se debe a L. Euler (1707-1783), el matemático más productivo, al menos del siglo XVIII. Aunque Euler perdió la vista de un ojo en 1735 y en 1766 quedó completamente ciego, escribió en total 886 manuscritos, entre los cuales se encuentra un número asombroso de libros de texto.

Para nuestros intereses es suficiente conocer las proposiciones siguientes sobre la función gamma. Se cumple que

$$\Gamma(z) = (z-1)\Gamma(z-1), \text{ para } z > 1, \quad (4)$$

$$\Gamma(1) = 1, \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad (5)$$

de donde se obtiene en particular que

$$\Gamma(m) = (m-1)!, \text{ para } m \geq 1, m \in \mathbb{N}. \quad (6)$$

El teorema siguiente trata sobre el valor esperado y la varianza de la distribución  $\chi^2$  con  $m$  grados de libertad; aquí se aclara también la influencia de  $m$ .

**Teorema 1.** Si  $X$  posee una distribución  $\chi^2$  con  $m$  grados de libertad, entonces se cumple que

$$EX = m, \quad (7)$$

$$D^2X = 2m. \quad (8)$$

Advertimos aún que la distribución  $\chi^2$  con  $m=2$  grados de libertad es una distribución exponencial con el parámetro  $a = \frac{1}{2}$  (ver 5.5).

La distribución  $\chi^2$  está en estrecha relación con la distribución normal. Para mostrarlo demostraremos la siguiente proposición especial.

**Teorema 2.** Sea  $X$  una variable aleatoria con una distribución  $N(0,1)$ . Entonces la variable aleatoria  $Y = X^2$  posee una distribución  $\chi^2$  con un grado de libertad.

**Demostración.** Se cumple (ver 5.1, teorema 2, proposición 2) que

$$f_Y(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ \frac{f_X(\sqrt{x}) + f_X(-\sqrt{x})}{2\sqrt{x}} & \text{para } x > 0. \end{cases}$$

Con  $f_X(t) = \psi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$  y  $\psi(-t) = \psi(t)$  se obtiene de aquí

$$f_Y(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ \frac{2\psi(\sqrt{x})}{2\sqrt{x}} = \frac{1}{\sqrt{2\pi}} \cdot \frac{e^{-\frac{x}{2}}}{\sqrt{x}} = \frac{1}{2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)} x^{\frac{1}{2}-1} e^{-\frac{x}{2}} & \text{para } x > 0. \end{cases}$$

con lo cual está demostrada la proposición del teorema.

La distribución  $\chi^2$  fue descubierta en 1876 por R. Helmert (como distribución de la suma de cuadrados de variables aleatorias independientes con distribución  $N(0,1)$ ) y vuelta a hallar en 1900 por K. Pearson, fundador en Inglaterra de una escuela de Estadística matemática de altos rendimientos: por eso esta distribución se denomina de *Helmert* o de *Helmert-Pearson*.

### 5.6.2 Distribución $t$

**Definición 3.** Sea  $m$  un número natural. Una variable aleatoria continua  $X$  se denomina *distribuida t con m grados de libertad*, si la densidad de probabilidad  $f_X$  tiene la forma

$$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi m} \Gamma\left(\frac{m}{2}\right)} \cdot \frac{1}{\left(1 + \frac{x^2}{m}\right)^{\frac{m+1}{2}}}, \quad -\infty < x < \infty. \quad (9)$$

Se dice también que  $X$  posee una *distribución t con m grados de libertad* (fig. 40). Denotamos el percentil de orden  $p$  de la distribución  $t$  con  $m$  grados de libertad con  $t_{m,p}$ .

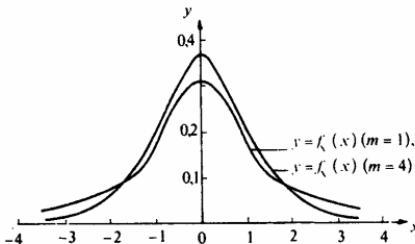


Figura 40

En (9),  $\Gamma$  es de nuevo el símbolo para la función gamma completa. Observemos que la densidad de la distribución  $t$  con  $m$  grados de libertad es una función par ( $f_X(-x) = f_X(x)$ , para todo  $x \in \mathbb{R}$ ), cuya representación gráfica no se diferencia sustancialmente de la curva de la campana de Gauss para  $m$  grande (ver fig. 33).

Para  $m=1$  obtenemos especialmente (fig. 40) la función de densidad  $f_X$

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \quad -\infty < x < \infty; \quad (10)$$

la distribución de probabilidad determinada por ella se denomina también, en honor de A.L. Cauchy (1789-1857), *distribución de Cauchy*.

El teorema siguiente se refiere al valor esperado y la varianza de la distribución  $t$  con  $m$  grados de libertad.

**Teorema 3.** Si  $X$  posee una distribución  $t$  con  $m$  grados de libertad, entonces se cumple que

$$EX=0, \quad m \geq 2, \quad (11)$$

$$D^2X = \frac{m}{m-2}, \quad m \geq 3. \quad (12)$$

Añadimos que una variable aleatoria que tenga una distribución  $t$  con  $m$  grados de libertad posee solo momentos de orden  $k \leq m-1$ . Por tanto, la distribución de Cauchy no posee, en particular, ningún valor esperado.

La distribución  $t$  fue descubierta e investigada (1908) por W.S. Gosset (1876-1937), quien publicaba bajo el seudónimo Student; por esta razón se encuentra también la distribución  $t$  con el nombre de *distribución de Student*.

### 5.6.3 Distribución $F$

**Definición 4.** Sean  $m_1$  y  $m_2$  números naturales. Una variable aleatoria continua  $X$  se denomina *distribuida F con  $(m_1, m_2)$  grados de libertad*, si la densidad de probabilidad  $f_1$  tiene la forma

$$f_1(x) = \begin{cases} \frac{\Gamma\left(\frac{m_1+m_2}{2}\right)m_1^{\frac{m_1}{2}}m_2^{\frac{m_2}{2}}}{\Gamma\left(\frac{m_1}{2}\right)\Gamma\left(\frac{m_2}{2}\right)} \cdot \frac{x^{\frac{m_1}{2}-1}}{(m_2+m_1x)^{\frac{m_1+m_2}{2}}} & \text{para } x > 0, \\ 0 & \text{para } x \leq 0. \end{cases} \quad (13)$$

Se dice también que  $X$  posee una *distribución F con  $(m_1, m_2)$  grados de libertad* (fig. 41). Denotamos el percentil de orden  $p$  de la distribución  $F$  con  $(m_1, m_2)$  grados de libertad  $F_{m_1, m_2, p}$ .

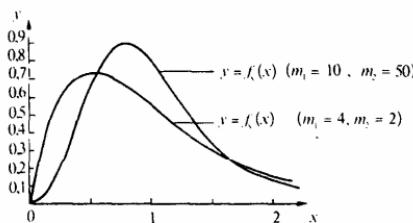


Figura 41

**Teorema 4.** Si  $X$  posee una distribución  $F$  con  $(m_1, m_2)$  grados de libertad, entonces se cumple que

$$EX = \frac{m_2}{m_2 - 2}, \quad (m_2 \geq 3), \quad (14)$$

$$D^2X = \frac{2m_2^2(m_1+m_2-2)}{m_1(m_2-2)^2(m_2-4)}. \quad (m_2 \geq 5). \quad (15)$$

Observemos que el valor esperado no depende de  $m_1$  y que  $EX \approx 1$  para  $m_2 >> 1$ . Además, añadimos que para  $m_2 \leq 2$  no existe valor esperado y para  $m_2 \leq 4$  no existe varianza.

La distribución  $F$  se debe a R.A. Fisher (1890-1962), uno de los representantes más conocidos de la Estadística matemática en Inglaterra, quien además trabajó en el campo de la teoría de la información matemática.

## 6. Vectores aleatorios

Los vectores aleatorios son aquellos cuyas componentes son variables aleatorias. Estos se utilizan para representar, desde un punto de vista matemático, algunas características que se pueden describir numéricamente en un fenómeno aleatorio. Así, por ejemplo, la longitud, ancho y altura de una pieza de trabajo en forma de cubo, producida automáticamente, y la talla y peso de un hombre, se pueden describir por medio de un vector aleatorio.

Después de la definición general y la caracterización teórico-probabilística de un vector aleatorio (epígrafe 6.1), trataremos en el epígrafe 6.2 los llamados *vectores aleatorios discretos* lo cual realizaremos apoyándonos en el tratamiento de las variables aleatorias discretas (ver 4.2 y 4.3), y en el epígrafe 6.3 nos ocuparemos de los denominados *vectores aleatorios continuos*, para lo cual partiremos de los estudios sobre variables aleatorias continuas (ver 5.1 y 5.2).

Las características numéricas para la comprensión de la dependencia mutua, de la relación entre las componentes de un vector aleatorio, son de especial interés; estudiaremos, en particular, los llamados *coeficientes de correlación* para la dependencia lineal entre dos variables aleatorias. En el epígrafe 6.4 trataremos el concepto *independencia de variables aleatorias*, que constituye un concepto central de toda la teoría de probabilidades. Aquí también deduciremos consecuencias de la independencia, que resultan muy útiles para el trabajo práctico con variables aleatorias independientes. Por último, se realiza en el epígrafe 6.5 la caracterización de la distribución de probabilidad para la suma, diferencia, producto y cociente de dos variables aleatorias continuas independientes; los teoremas señalados aquí se necesitarán especialmente en la parte correspondiente a la Estadística matemática.

### 6.1 Definición general de vector aleatorio

Realizaremos la exposición de este epígrafe de forma análoga a como lo hicimos en el epígrafe 4.1; en caso necesario el lector puede orientarse otra vez por allí.

Definición 1. Sea  $[\Omega, \mathcal{A}, P]$  un espacio de probabilidad y sean  $X_1, X_2, \dots, X_n$  ( $n \geq 2$ ) variables aleatorias sobre  $[\Omega, \mathcal{A}, P]$ . Entonces, el  $n$ -uplo  $(X_1, X_2, \dots, X_n)$  se llama *vector aleatorio* ( $n$ -dimensional sobre  $[\Omega, \mathcal{A}, P]$ ).

Nos dedicaremos a continuación a la caracterización de la distribución de probabilidad de un vector aleatorio. Para ello, sean  $x_1, x_2, \dots, x_n$  números reales cualesquiera. Como las  $X_k$  son variables aleatorias, se cumple que  $(X_k < x_k) \in \mathcal{A}$  ( $k = 1, 2, \dots, n$ ).  $\mathcal{A}$  es una  $\sigma$ -álgebra, de modo que se cumple en particular la relación  $\bigcap_{k=1}^n (X_k < x_k) \in \mathcal{A}$ . En virtud de

$$\begin{aligned}\{\omega \in \Omega : X_k(\omega) < x_k\} &= \bigcap_{k=1}^n \{\omega \in \Omega : X_k(\omega) < x_k\} \\ &= \bigcap_{k=1}^n (X_k < x_k)\end{aligned}$$

resulta que  $\{\omega \in \Omega : X_1(\omega) < x_1, \dots, X_n(\omega) < x_n\} \in \mathcal{A}$ .

Si denotamos abreviadamente el subconjunto  $\{\omega \in \Omega : X_1(\omega) < x_1, \dots, X_n(\omega) < x_n\}$  de  $\Omega$  por  $(X_1 < x_1, \dots, X_n < x_n)$ , entonces es razonable hablar de la probabilidad del suceso aleatorio  $(X_1 < x_1, \dots, X_n < x_n)$ ; para esta probabilidad escribiremos de forma abreviada  $P(X_1 < x_1, \dots, X_n < x_n)$ .

Definición 2. Sea  $[\Omega, \mathcal{A}, P]$  un espacio de probabilidad y  $(X_1, X_2, \dots, X_n)$  un vector aleatorio. La función  $F_{(X_1, X_2, \dots, X_n)}$  definida por

$$F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n) \quad (1)$$

$$(x_k \in \mathbb{R}, k = 1, 2, \dots, n).$$

se denomina *función de distribución del vector aleatorio*  $(X_1, X_2, \dots, X_n)$  o *función de distribución conjunta de las variables aleatorias*  $X_1, X_2, \dots, X_n$ .

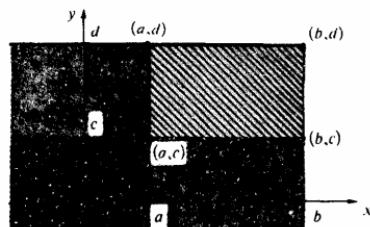


Figura 42

La función de distribución de un vector aleatorio  $n$ -dimensional es, por tanto, una función real de  $n$  variables reales. Por medio de la función de distribución de un vector aleatorio se pueden expresar las probabilidades de casi todos los sucesos aleatorios que están en relación con este. Así, por ejemplo, se cumple en el caso  $n=2$  (fig. 42)

$$P(a \leq X < b, c \leq Y < d) = F_{(X, Y)}(b, d) - F_{(X, Y)}(b, c) - F_{(X, Y)}(a, d) + F_{(X, Y)}(a, c). \quad (2)$$

En el teorema siguiente resumiremos las propiedades de la función de distribución de un vector aleatorio.

**Teorema 1.** Sea  $F$  la función de distribución de un vector aleatorio  $n$ -dimensional. Entonces se cumple:

1. Para todo  $x_k \in \mathbb{R}$  ( $k=1, 2, \dots, n$ ) es  $0 \leq F(x_1, x_2, \dots, x_n) \leq 1$ .
2.  $F$  es monótona creciente en toda variable  $x_k$ .
3.  $F$  es continua por la izquierda en toda variable  $x_k$ .
4.  $\lim_{x_k \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = 0$  ( $k=1, 2, \dots, n$ ),  $\lim_{x_k \rightarrow +\infty} F(x_1, x_2, \dots, x_n) = 1$ .

$$\begin{array}{c} \square \\ x_k \rightarrow -\infty \\ \square \\ x_k \rightarrow +\infty \end{array}$$

La demostración se desarrolla de acuerdo con la del teorema 1(4.1); la dejamos al lector.

Como muestra el ejemplo siguiente, las proposiciones señaladas en el teorema 1 no son suficientes para que una función  $F$ , con estas propiedades, sea la función de distribución de un vector aleatorio.

Ejemplo. Consideremos la función dada por

$$F(x, y) = \begin{cases} 0 & \text{para } x+y \leq 0, \\ 1 & \text{para } x+y > 0. \end{cases}$$

Evidentemente  $F$  posee todas las propiedades señaladas en el teorema 1. Pero se cumple que

$$F(1, 1) - F(1, 0) - F(0, 1) + F(0, 0) = 1 - 1 - 1 + 0 = -1;$$

luego en virtud de (2),  $F$  no puede ser la función de distribución de un vector aleatorio de dimensión  $n=2$ .

El lector interesado puede informarse sobre las condiciones suplementarias que aseguran que una función de varias variables sea función de distribución de un vector aleatorio.

En los capítulos correspondientes a la Estadística matemática trataremos en muchas ocasiones funciones de un vector aleatorio  $(X_1, X_2, \dots, X_n)$ , por ejemplo, las funciones  $g(X_1, X_2, \dots, X_n) = X_1 + X_2 + \dots + X_n$  y  $g(X_1, X_2, \dots, X_n) = X_1^2 + X_2^2 + \dots + X_n^2$ . Ya que nos interesarán, en particular, por la distribución de probabilidad de estas funciones, es importante conocer una clase de funciones  $g$  lo suficientemente grande para la cual la función  $g(X_1, X_2, \dots, X_n)$ , definida sobre  $\Omega$  por  $[g(X_1, X_2, \dots, X_n)](\omega) = g(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$ , sea una variable aleatoria, o sea, posea una distribución de probabilidad. Para ello damos el siguiente teorema sin demostración:

**Teorema 2.** Sea  $[\Omega, \mathcal{A}, P]$  un espacio de probabilidad,  $(X_1, X_2, \dots, X_n)$  un vector aleatorio  $n$ -dimensional (sobre  $[\Omega, \mathcal{A}, P]$ ) y  $g$ , una función real continua definida sobre el conjunto de todos los  $n$ -uplos de números reales. Entonces la función  $g(X_1, X_2, \dots, X_n)$  definida sobre  $\Omega$  por

$$[g(X_1, X_2, \dots, X_n)](\omega) = g(X_1(\omega), X_2(\omega), \dots, X_n(\omega)) \quad (3)$$

es una variable aleatoria (sobre  $[\Omega, \mathcal{A}, P]$ ).

En especial, para las funciones  $g$  dadas por

$$\begin{aligned} g(x_1, x_2, \dots, x_n) &= x_1 + x_2 + \dots + x_n, \\ g(x_1, x_2, \dots, x_n) &= x_1^2 + x_2^2 + \dots + x_n^2. \end{aligned}$$

o

$$g(x_1, x_2, \dots, x_n) = x_1 \cdot x_2 \cdots x_n,$$

las funciones  $g(X_1, X_2, \dots, X_n)$  definidas sobre  $\Omega$  son variables aleatorias.

A continuación nos limitaremos al caso  $n=2$ ; por lo tanto, trataremos los vectores aleatorios bidimensionales  $(X, Y)$ . Muchas veces es de interés, por ejemplo, la distribución de probabilidad de la variable aleatoria  $X$  en el marco del vector aleatorio  $(X, Y)$ . Se cumple (ver 2.4, teorema 1) que

$$\begin{aligned} F_X(x) &= P(X < x) = P(X < x, Y < \infty) \\ &= \lim_{y \rightarrow \infty} P(X < x, Y < y) = \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y). \end{aligned}$$

**Definición 2.** La función de distribución  $F_X$  dada por

$$F_X(x) = \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y) \quad (4)$$

se llama *función de distribución marginal* de  $X$ , de la distribución conjunta de  $X$  y  $Y$ ; la distribución de probabilidad caracterizada se llama *distribución marginal de  $X$*  de la distribución conjunta de  $X$  y  $Y$ . (Una definición correspondiente existe para la función de distribución marginal  $F_Y$  de  $Y$ , de la distribución conjunta de  $X$  y  $Y$ .)

Concluiremos este epígrafe con la observación, de que para un vector aleatorio  $n$ -dimensional se pueden considerar evidentemente  $\binom{n}{k}$  distribuciones marginales de vectores aleatorios  $k$ -dimensionales ( $k=1, 2, \dots, n-1$ ).

## 6.2 Vectores aleatorios discretos

**Definición 1.** Un vector aleatorio se llama *discreto*, si puede tomar un número finito o infinito numerable de valores.

En las explicaciones posteriores nos limitaremos al caso de un vector aleatorio bidimensional.

Desde el punto de vista del Cálculo de probabilidades, podemos considerar un vector aleatorio bidimensional  $(X, Y)$  como dado, si están dados a su vez todos los valores  $(x_i, y_k)$  del vector aleatorio y las probabilidades particulares correspondientes

$$p_{ik} = P(X=x_i, Y=y_k), \quad (1)$$

con las cuales el vector aleatorio  $(X, Y)$  toma estos valores. Por ello, se puede caracterizar también un vector aleatorio bidimensional  $(X, Y)$  por la llamada tabla de distribución.

$X \backslash Y$	$y_1$	$y_2$	$y_3$	$\dots$	
$x_1$	$p_{11}$	$p_{12}$	$p_{13}$	$\dots$	$p_{1\cdot}$
$x_2$	$p_{21}$	$p_{22}$	$p_{23}$	$\dots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	$\dots$	$p_{\cdot \cdot} = 1$

(Aclararemos más tarde el significado de  $p_{ik}$  y  $p_{\cdot \cdot}$ .)

Para las probabilidades  $p_{ik}$  se cumple que

$$p_{ik} \geq 0, \quad \sum_{i,k} p_{ik} = 1. \quad (3)$$

Los valores de la función de distribución ( $F_{(X,Y)}$ ) se obtienen de las probabilidades  $p_{ik}$  según

$$F_{(X,Y)}(x,y) = P(X < x, Y < y) = \sum_{\substack{i, k \\ i x_i < x \\ k y_k < y}} P(X=x_i, Y=y_k) = \sum_{\substack{i, k \\ i x_i < x \\ k y_k < y}} p_{ik}, \quad (4)$$

extendiéndose la sumatoria sobre todos los  $i$  y  $k$  para los cuales se cumple que  $x_i < x$  y  $y_k < y$ .

Ahora queremos caracterizar las distribuciones marginales de un vector aleatorio discreto ( $X, Y$ ). La distribución marginal de  $X$  es una distribución discreta;  $X$  toma los valores  $x_i$  con las probabilidades

$$p_i = \sum_k p_{ik} = \sum_k P(X=x_i, Y=y_k). \quad (5)$$

De igual forma la distribución marginal de  $Y$  es una distribución discreta;  $Y$  toma los valores  $y_k$  con las probabilidades

$$p_k = \sum_i p_{ik} = \sum_i P(X=x_i, Y=y_k). \quad (6)$$

En la tabla de distribución (2) hemos registrado en la última columna los números  $p_i$  y en la última fila los que caracterizan las distribuciones marginales de  $X$  y  $Y$ .

Seguidamente nos referiremos a algunas características numéricas para vectores aleatorios discretos bidimensionales ( $X, Y$ ). Junto al valor esperado y la varianza de las variables aleatorias  $X$  y  $Y$ , en caso de que existan, nos interesa, en especial, una medida para expresar la dependencia mutua de las variables aleatorias  $X$  y  $Y$ . Trataremos la llamada covarianza y, sobre esta base, el denominado coeficiente de correlación. Pero primariamente anotaremos una fórmula para el cálculo del valor esperado de una función de un vector aleatorio discreto, de donde se obtienen fórmulas para el valor esperado y la varianza de una suma de variables aleatorias.

**Teorema 1.** Sea  $(X, Y)$  un vector aleatorio discreto, que toma los valores  $(x_i, y_k)$  con las probabilidades  $p_{ik}$ , y  $g$  una función real continua definida sobre el conjunto de todos los pares de números reales. Si la serie  $\sum_{i,k} g(x_i, y_k) p_{ik}$  converge absolutamente (o sea, si  $\sum_{i,k} |g(x_i, y_k)| p_{ik} < \infty$ ), entonces se cumple

$$Eg(X, Y) = \sum_{i,k} g(x_i, y_k) p_{ik} \quad (7)$$

(ver 4.3, teorema 2).

Renunciaremos a la exposición de la demostración de este teorema.

Para  $g(x, y) = x$  y  $g(x, y) = y$  obtenemos especialmente

$$EX = \sum_i x_i p_i \text{ y } EY = \sum_k y_k p_k, \quad (8)$$

es decir, los valores esperados de las variables aleatorias  $X$  y  $Y$  respectivamente, en el marco de la distribución conjunta de  $X$  y  $Y$ , siempre y cuando las series indicadas en (8) converjan absolutamente.

Bajo una condición correspondiente se obtiene para  $g(x,y) = (x - EX)^2 + g(x,y) = (y - EY)^2$ , la varianza de las variables aleatorias  $X$  y  $Y$  respectivamente, en el marco de la distribución conjunta de  $X$  y  $Y$ ,

$$D^2X = \sum_i (x_i - EX)^2 p_i \text{ y } D^2Y = \sum_k (y_k - EY)^2 p_k. \quad (9)$$

Trataremos el caso  $g(x,y) = x + y$ .

**Teorema 2.** Sea  $(X, Y)$  un vector aleatorio discreto. Entonces se cumple que

$$E(X + Y) = EX + EY. \quad (10)$$

suponiéndose la existencia de los valores esperados señalados en el miembro derecho de (10).

**Demostración.** La función dada por  $g(x,y) = x + y$  satisface todas las condiciones nombradas en el teorema 1. Por consiguiente, se cumple (7) y con esto

$$\begin{aligned} E(X + Y) &= \sum_{i,k} (x_i + y_k) p_{ik} = \sum_{i,k} x_i p_{ik} + \sum_{i,k} y_k p_{ik} \\ &= \sum_i x_i p_i + \sum_k y_k p_k = EX + EY. \end{aligned}$$

La validez de la proposición siguiente se obtiene directamente de aquí con el principio de inducción completa.

**Corolario 1.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias discretas con los valores esperados  $EX_1, EX_2, \dots, EX_n$ . Entonces se cumple que

$$E(X_1 + X_2 + \dots + X_n) = EX_1 + EX_2 + \dots + EX_n. \quad (11)$$

Observemos que para el cálculo del valor esperado de una suma de variables aleatorias discretas, no se necesita su distribución conjunta; para ello es suficiente el conocimiento de las distribuciones de probabilidad de cada una de las variables aleatorias. Para la varianza esto se comporta de otra forma.

**Teorema 3.** Sea  $(X, Y)$  un vector aleatorio discreto. Entonces se cumple que

$$D^2(X + Y) = D^2X + D^2Y + 2(EXY - (EX)(EY)), \quad (12)$$

suponiéndose la existencia de los sumandos en el miembro derecho de (12).

**Demostración.** Utilizando  $D^2Z = EZ^2 - (EZ)^2$  (ver 4.3, teorema 3) y el corolario 1, obtenemos

$$\begin{aligned} D^2(X + Y) &= E(X + Y)^2 - (E(X + Y))^2 \\ &= E(X^2 + 2XY + Y^2) - (EX + EY)^2 \\ &= EX^2 + 2EXY + EY^2 - (EX)^2 - 2(EX)(EY) - (EY)^2 \\ &= D^2X + D^2Y + 2(EXY - (EX)(EY)). \end{aligned}$$

**Definición 2.** Sea  $(X, Y)$  un vector aleatorio discreto, que toma los valores  $(x_i, y_k)$  con las probabilidades  $p_{ik}$ . Entonces el número definido por

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = \sum_{i,k} (x_i - EX)(y_k - EY)p_{ik} \quad (13)$$

se denomina covarianza de  $X$  y  $Y$ ; aquí se supone, junto a la existencia de  $EX$  y  $EY$ , la convergencia absoluta de la serie situada en el miembro derecho de (13).

Debemos observar en (13) que, a causa de la continuidad de la función dada por  $g(x, y) = (x - EX)(y - EY)$ , la función  $(X - EX)(Y - EY)$  definida sobre  $\Omega$  es una variable aleatoria y que para su valor esperado se cumple, sobre la base de las condiciones en la definición 2 y según (7), la relación

$$E(X - EX)(Y - EY) = \sum_{i,k} (x_i - EX)(y_k - EY)p_{ik}.$$

Se comprueba fácilmente que se cumple

$$\text{cov}(X, Y) = EXY - (EX)(EY), \quad (14)$$

de modo que (12) se puede escribir también en la forma

$$D^2(X+Y) = D^2X + D^2Y + 2\text{cov}(X, Y). \quad (15)$$

Evidentemente se cumple que  $\text{cov}(X, X) = D^2X$ . La matriz (simétrica)

$$\begin{pmatrix} D^2X & \text{cov}(X, Y) \\ \text{cov}(X, Y) & D^2Y \end{pmatrix} \quad (16)$$

se denomina matriz de covarianza del vector aleatorio  $(X, Y)$ . En general, la matriz  $(b_{ij})$ ,  $b_{ij} = \text{cov}(X_i, X_j)$ , asociada a un vector aleatorio  $n$ -dimensional,  $(X_1, X_2, \dots, X_n)$ , se llama matriz de covarianza; en la diagonal principal están las varianzas de las componentes del vector aleatorio ( $b_{ii} = \text{cov}(X_i, X_i) = D^2X_i$ ).

**Definición 3.** Sea  $(X, Y)$  un vector aleatorio discreto que toma los valores  $(x_i > y_k)$  con las probabilidades  $p_{ik}$ . Entonces el número definido por

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{D^2X} \sqrt{D^2Y}} = \frac{\sum_{i,k} (x_i - EX)(y_k - EY)p_{ik}}{\sqrt{\sum_i (x_i - EX)^2 p_i} \sqrt{\sum_k (y_k - EY)^2 p_k}} \quad (17)$$

se denomina coeficiente de correlación de  $X$  y  $Y$ ; aquí se supone la convergencia absoluta de las series que aparecen en (17) y, además, que  $D^2X > 0$  y  $D^2Y > 0$ .

El teorema siguiente trata sobre las propiedades del coeficiente de correlación.

**Teorema 4.** Sea  $(X, Y)$  un vector aleatorio discreto con el coeficiente de correlación  $\rho(X, Y)$ .

1. Se cumple que  $|\rho(X, Y)| \leq 1$ .
2. Se cumple que  $|\rho(X, Y)| = 1$  si y solo si existen números  $a \neq 0$  y  $b$ , tales que  $Y = aX + b$ .

**Demostración.** Consideraremos las variables aleatorias que se derivan de  $X$  y  $Y$  mediante estandarización

$$X_0 = \frac{X - EX}{\sqrt{D^2X}} \quad \text{y} \quad Y_0 = \frac{Y - EY}{\sqrt{D^2Y}}$$

Como  $EX_0=EY_0=0$  se cumple que

$$\begin{aligned}\text{cov}(X_0, Y_0) &= EX_0Y_0=E\left(\frac{X-EX}{\sqrt{D^2X}}\right)\left(\frac{Y-EY}{\sqrt{D^2Y}}\right) \\ &= \frac{E(X-EX)(Y-EY)}{\sqrt{D^2X}\sqrt{D^2Y}} \\ &= \rho(X, Y).\end{aligned}$$

Con  $D^2X_0=D^2Y_0=1$  obtenemos con esto (ver (15))

$$\begin{aligned}D^2(X_0 \pm Y_0) &= D^2X_0 + D^2Y_0 \pm 2 \text{ cov}(X_0, Y_0) \\ &= 2(1 \pm \rho(X, Y)).\end{aligned}\quad (*)$$

1. Como la varianza de una variable aleatoria es un número no negativo, resulta de (\*):  $1 \pm \rho(X, Y) \geq 0$ , luego  $\rho(X, Y) \geq -1$  y  $\rho(X, Y) \leq 1$ , o sea,  $\rho(X, Y) \in [-1, 1]$ .

2.a) Si se cumple que  $\rho(X, Y) = \pm 1$ , entonces se cumple, según (\*),  $D^2(X_0 \mp Y_0) = 0$ . La variable aleatoria  $X_0 \mp Y_0$  posee, por tanto, una distribución puntual única (ver 4.3, teorema 4). En virtud de

$$E(X_0 \mp Y_0) = EX_0 \mp EY_0 = 0 \mp 0 = 0,$$

resulta  $P(X_0 \mp Y_0 = 0) = 1$ , es decir, se cumple que  $Y_0 = \pm X_0$ , o expresado de otra manera,  $Y = aX + b$  con

$$a = \pm \frac{\sqrt{D^2Y}}{\sqrt{D^2X}} \quad \text{y} \quad b = EY \mp EX \frac{\sqrt{D^2Y}}{\sqrt{D^2X}}.$$

b) Si se cumple que  $Y = aX + b$  ( $a, b$  reales), entonces se cumple que  $EY = aEX + b$  (ver 4.3, teorema 1),  $D^2Y = a^2D^2X$  (ver 4.5, teorema 5) y con esto

$$\begin{aligned}|\rho(X, Y)| &= \frac{|\text{cov}(X, Y)|}{\sqrt{D^2X}\sqrt{D^2Y}} \\ &= \frac{|E(X-EX)(aX+b-aEX-b)|}{\sqrt{D^2X}\sqrt{a^2D^2X}} \\ &= \frac{|a| |E(X-EX)^2|}{|a| \sqrt{D^2X} \sqrt{D^2X}} = \frac{D^2X}{D^2X} = 1.\end{aligned}$$

Con esto está demostrado completamente el teorema 4.

El teorema 4 expresa que el coeficiente de correlación es un número situado entre  $-1$  y  $+1$  que mide la dependencia lineal de dos variables aleatorias, existiendo dependencia lineal si y solo si el valor absoluto del coeficiente de correlación es igual a uno. Retrocederemos al caso  $\rho=0$  en el epígrafe 6.4; de todas formas, de  $\rho=0$  no resulta que entre las variables aleatorias  $X$  y  $Y$  no pueda existir una dependencia funcional, es decir, una relación de la forma  $Y=g(X)$ .

**Ejemplo.**  $X$  toma los valores  $-1, 0$  y  $+1$  con la probabilidad  $\frac{1}{3}$ . Entonces se cumple que  $EX=0$  y  $D^2X>0$ . Hagamos ahora  $Y=X^2$ ; se cumple que  $D^2Y>0$ . La variable aleatoria  $X \cdot Y=X^3$  toma entonces cada una de los valores  $-1, 0$  y  $+1$  con la probabilidad  $\frac{1}{3}$ , de modo que se cumple que  $EX^3=0$ . Con esto (ver (14)).

$$\text{cov}(X, Y) = EXY - (EX)(EY) = EX^3 - 0 = 0 - 0 = 0$$

y, por tanto,  $\rho(X, Y)=0$ . Sin embargo, existe una dependencia funcional entre  $X$  y  $Y(Y=X^2)$ .

### 6.3 Vectores aleatorios continuos

Nos limitaremos también a considerar vectores aleatorios bidimensionales; con esto se aclara cómo se debe tratar el caso general.

**Definición 1.** Un vector aleatorio  $(X, Y)$  se llama *continuo*, si existe una función continua no negativa  $f_{(x,y)}$  definida sobre el conjunto de todos los pares de números reales, tal que se cumple que

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{(x,y)}(x, y) dy dx \quad (1)$$

para todos los números reales  $a, b, c$  y  $d$  con  $a \leq b$  y  $c \leq d$ .

La distribución de probabilidad de un vector aleatorio continuo  $(X, Y)$  está prefijada por la función  $f_{(x,y)}$ , que se denomina *densidad de probabilidad* (densidad de distribución, densidad o función de densidad) del vector aleatorio  $(X, Y)$  o *densidad de probabilidad conjunta* de las variables aleatorias  $X$  y  $Y$ . Los valores de la función de distribución  $F_{(x,y)}$  se obtienen sobre la base de la densidad de probabilidad  $f_{(x,y)}$ , según

$$F_{(x,y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{(x,y)}(u, v) dv du. \quad (2)$$

La relación (2) entre la función de distribución  $F_{(x,y)}$  y la densidad de probabilidad  $f_{(x,y)}$ , se puede expresar también en la forma

$$\frac{\partial F_{(x,y)}(x, y)}{\partial x \partial y} = f_{(x,y)}(x, y). \quad (3)$$

De manera semejante que en el tratamiento de los vectores aleatorios discretos, nos ocuparemos primeramente con las distribuciones marginales y nos interesaremos por las características numéricas especiales para los vectores aleatorios continuos; aquí las definiciones y proposiciones son análogas a las correspondientes del epígrafe 6.2.

La distribución marginal de la variable aleatoria  $X$  en el marco del vector aleatorio continuo  $(X, Y)$ , es una distribución continua; en virtud de

$$F_X(x) = \lim_{y \rightarrow \infty} F_{(x,y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{(x,y)}(t, y) dy dt,$$

la densidad de probabilidad  $f_x$  de la variable aleatoria  $X$ , que se denomina en este contexto *densidad de distribución marginal*, está dada por

$$f_x(x) = \int_{-\infty}^{\infty} f_{(x,y)}(x,y) dy. \quad (4)$$

Asimismo, la distribución marginal de  $Y$  es una distribución continua: para la densidad de distribución marginal  $f_y$  se cumple que

$$f_y(y) = \int_{-\infty}^{\infty} f_{(x,y)}(x,y) dx. \quad (5)$$

Ahora señalarémos, sin demostración, una fórmula para el cálculo del valor esperado de una función de un vector aleatorio continuo.

**Teorema 1.** Sea  $(X, Y)$  un vector aleatorio continuo con la densidad de probabilidad  $f_{(x,y)}$  y sea  $g$  una función real continua definida sobre el conjunto de todos los pares de números reales. Si la integral  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{(x,y)}(x,y) dx dy$  converge absolutamente (es decir, si se cumple  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x,y)| f_{(x,y)}(x,y) dx dy < \infty$ ), entonces se cumple que

$$Eg(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{(x,y)}(x,y) dx dy \quad (6)$$

(ver 5.2, teorema 2, y 6.2, teorema 1).

El valor esperado y la varianza de  $X$  y de  $Y$ , en el marco de la distribución conjunta de  $X$  y  $Y$ , se obtienen utilizando las densidades de distribución marginales correspondientes:

$$EX = \int_{-\infty}^{\infty} xf_x(x) dx, \quad EY = \int_{-\infty}^{\infty} yf_y(y) dy, \quad (7)$$

y

$$D^2X = \int_{-\infty}^{\infty} (x - EX)^2 f_x(x) dx, \quad D^2Y = \int_{-\infty}^{\infty} (y - EY)^2 f_y(y) dy, \quad (8)$$

suponiéndose la convergencia absoluta de las integrales que se presentan.

Queremos dedicarnos ahora al cálculo del valor esperado  $E(X+Y)$  en el caso continuo.

**Teorema 2.** Sea  $(X, Y)$  un vector aleatorio continuo. Entonces se cumple que

$$E(X+Y) = EX + EY, \quad (9)$$

suponiéndose la existencia de los valores esperados indicados en el miembro derecho de (9) (ver 6.2, teorema 2).

**Demostración.** La función dada por  $g(x,y) = x+y$  satisface todas las condiciones nombradas en el teorema 1. Por tanto, se cumple (6) y con esto

$$E(X+Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f_{(x,y)}(x,y) dx dy$$

$$\begin{aligned}
 E(X+Y) &= \int_{-\infty}^{\infty} \left( x \int_{-\infty}^{\infty} f_{x,y}(x,y) dy \right) dx + \int_{-\infty}^{\infty} \left( y \int_{-\infty}^{\infty} f_{x,y}(x,y) dx \right) dy \\
 &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\
 &= EX + EY.
 \end{aligned}$$

Por consiguiente, el valor esperado de una suma de variables aleatorias continuas es, como en el caso de variables aleatorias discretas, igual a la suma de los valores esperados. Con esto se cumple también la fórmula

$$D^2(X+Y) = D^2X + D^2Y + 2(EXY - (EX)(EY)) \quad (10)$$

(ver 6.2, teorema 3) para variables aleatorias continuas  $X$  y  $Y$ , pues en la demostración del teorema 3 (6.2) hemos tomado en consideración solo aquellas reglas de cálculo para el valor esperado y la varianza, que son válidas también para el caso continuo.

Apoyándonos en el teorema 1 definiremos, análogamente al procedimiento seguido en el caso discreto, la covarianza y el coeficiente de correlación para el caso continuo.

**Definición 2.** Sea  $(X, Y)$  un vector aleatorio continuo con la densidad de probabilidad  $f_{x,y}$ . Entonces, el número definido por

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY) f_{x,y}(x, y) dx dy \quad (11)$$

se llama *covarianza de  $X$  y  $Y$* ; aquí se supone, junto a la existencia de  $EX$  y  $EY$ , la convergencia absoluta de la integral situada en el miembro derecho de (11).

**Definición 3.** Sea  $(X, Y)$  un vector aleatorio continuo con la densidad de probabilidad  $f_{x,y}$ . Entonces el número definido por

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{D^2X} \sqrt{D^2Y}} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY) f_{x,y}(x, y) dx dy}{\sqrt{\int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx} \sqrt{\int_{-\infty}^{\infty} (y - EY)^2 f_Y(y) dy}} \quad (12)$$

se denomina *coeficiente de correlación de  $X$  y  $Y$* ; aquí se supone la convergencia absoluta de las integrales que aparecen en (12).

Como en la demostración del teorema 4(6.2) no fueron empleadas propiedades especiales de las variables aleatorias discretas, sino solo reglas de cálculo para el valor esperado y la varianza, que también son válidas para variables aleatorias continuas, se cumplen las proposiciones del teorema 4(6.2) para el caso de variables aleatorias continuas.

**Teorema 3.** Sea  $(X, Y)$  un vector aleatorio continuo con el coeficiente de correlación  $\rho(X, Y)$ .

1. Se cumple que  $|\rho(X, Y)| \leq 1$ .
2. Se cumple que  $|\rho(X, Y)| = 1$  si y solo si existen números  $a \neq 0$  y  $b$ , tales que  $Y = aX + b$ .

Cerraremos este epígrafe con el estudio de la llamada *distribución normal bivariada*, que es una distribución de un vector aleatorio continuo bidimensional, muy utilizada en las aplicaciones.

**Definición 4.** Sean  $\mu_1$  y  $\mu_2$  números reales cualesquiera,  $\sigma_1$  y  $\sigma_2$  números positivos arbitrarios y  $\rho$  un número cualquiera con  $|\rho| < 1$ . Un vector aleatorio continuo bidimensional  $(X, Y)$  se denomina *distribuido normalmente* (con los parámetros  $\mu_1$ ,  $\mu_2$ ,  $\sigma_1^2$ ,  $\sigma_2^2$ ,  $\rho$ ), si la densidad de probabilidad  $f_{(X,Y)}$  tiene la forma

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)} \quad (13)$$

$(-\infty < x < \infty, -\infty < y < \infty).$

El teorema siguiente nos aclara la significación de los parámetros de una distribución normal bivariada (ver epígrafe 5.4).

**Teorema 4.** Sea  $(X, Y)$  un vector aleatorio distribuido normalmente con los parámetros  $\mu_1$ ,  $\mu_2$ ,  $\sigma_1^2$ ,  $\sigma_2^2$  y  $\rho$ .

1. La distribución marginal de  $X$  es una distribución  $N(\mu_1, \sigma_1^2)$ .
2. La distribución marginal de  $Y$  es una distribución  $N(\mu_2, \sigma_2^2)$ .
3. Se cumple que  $\text{cov}(X, Y) = \rho\sigma_1\sigma_2$  y  $\rho(X, Y) = \rho$ .

**Demostración.** Para la densidad de distribución marginal  $f_X$ ,  $f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy$ , se obtiene haciendo la sustitución

$$t = \frac{1}{\sqrt{1-\rho^2}} \left( \frac{y-\mu_2}{\sigma_2} - \rho \frac{x-\mu_1}{\sigma_1} \right)$$

y con  $\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$ , la relación

$$f_X(x) = \frac{1}{2\pi\sigma_1} e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

$$= \phi(x; \mu_1, \sigma_1^2),$$

o sea,  $X$  posee una distribución normal con los parámetros  $\mu_1 (=EX)$  y  $\sigma_1^2 (=D^2X)$ . Con esto está claro que  $Y$  posee una distribución normal con los parámetros  $\mu_2 (=EY)$  y  $\sigma_2^2 (=D^2Y)$ .

Para la covarianza

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY)f_{(X,Y)}(x, y) dx dy$$

se obtiene, con las sustituciones  $u = \frac{x - \mu_1}{\sigma_1}$  y  $v = \frac{y - \mu_2}{\sigma_2}$ , la relación

$$\text{cov}(X, Y) = \frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} u e^{-\frac{u^2}{2}} \left( \int_{-\infty}^{\infty} v e^{-\frac{1}{2(1-\rho^2)}(v-\rho u)^2} dv \right) du.$$

Para la integral interna se obtiene, con la sustitución

$$t = \frac{1}{\sqrt{1-\rho^2}} (v - \rho u),$$

con  $\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$  y, además con,  $\int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt = 0$ , el valor  $\rho \mu \sqrt{2\pi}$ : considerando que

$$\int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du = \sqrt{2\pi},$$

se tiene por último que

$$\text{cov}(X, Y) = \frac{\sigma_1 \sigma_2}{2\pi} \rho \sqrt{2\pi} \sqrt{2\pi} = \rho \sigma_1 \sigma_2,$$

y con esto  $\rho(X, Y) = \rho$ .

De esta forma podemos afirmar que las distribuciones marginales de una distribución normal bivariada son también distribuciones normales. Para concluir, observemos que en el caso  $\rho=0$  se cumple la relación

$$f_{(x,y)}(x, y) = \varphi(x; \mu_1, \sigma_1^2) \varphi(y; \mu_2, \sigma_2^2), \quad (14)$$

es decir, que en el caso  $\rho=0$  el producto de las densidades de distribución marginales es igual a la densidad de probabilidad conjunta.

## 6.4 Independencia de variables aleatorias

El concepto independencia de variables aleatorias es de gran importancia en la teoría de probabilidades. Antes de definir la independencia de variables aleatorias recordemos la definición de independencia de sucesos aleatorios: Dos sucesos aleatorios  $A$  y  $B$  se llaman mutuamente independientes, si se cumple que  $P(A \cap B) = P(A)P(B)$  (ver 3.3, definición 1). De manera semejante denominaremos dos variables aleatorias  $X$  y  $Y$  mutuamente independientes, si todo suceso aleatorio  $A$ , que está en relación con la variable aleatoria  $X$ , es independiente de todo suceso  $B$  que está en relación con la variable aleatoria  $Y$ , es decir, si para cualesquiera  $x \in \mathbb{R}$  y  $y \in \mathbb{R}$  los sucesos  $(X < x)$  y  $(Y < y)$  son independientes, y se cumple que  $P(X < x, Y < y) = P(X < x)P(Y < y)$ .

En esto se basa la definición siguiente del concepto independencia de dos variables aleatorias, utilizándose para su formulación la función de distribución conjunta de las variables aleatorias  $X$  y  $Y$ , y las funciones de distribución marginales de  $X$  y  $Y$ .

**Definición 1.** Sea  $(X, Y)$  un vector aleatorio con la función de distribución  $F_{(x,y)}$  y las funciones de distribución marginales  $F_x$  y  $F_y$ . Las variables aleatorias  $X$  y  $Y$  se denominan (mutuamente) *independientes* (también: *estocásticamente independientes*), si se cumple que

$$F_{(x,y)}(x, y) = F_x(x)F_y(y) \quad (1)$$

para todos los números reales  $x$  y  $y$ .

Advertimos que en todos los casos se pueden determinar las funciones de distribución marginales de las variables aleatorias  $X$  y  $Y$  a partir de la función de distribución conjunta de estas variables aleatorias (ver 6.1, definición 2). En caso de independencia de  $X$  y  $Y$ , el recíproco también es posible; se puede calcular la función de distribución conjunta a partir de las funciones de distribución marginales, según (1).

Los dos teoremas siguientes contienen formulaciones equivalentes de la independencia de dos variables aleatorias  $X$  y  $Y$ , para el caso en que  $(X, Y)$  posea una distribución discreta y para el caso continuo respectivamente; estas formulaciones se realizan sobre la base de las probabilidades particulares o de las densidades de probabilidad, pueden comprobarse fácilmente en la situación concreta.

**Teorema 1.** Sea  $(X, Y)$  un vector aleatorio discreto, que toma los valores  $(x_i, y_k)$  con las probabilidades  $p_{ik}$ . Las variables aleatorias  $X$  y  $Y$  son mutuamente independientes si y solo si

$$P(X=x_i, Y=y_k) = P(X=x_i)P(Y=y_k),$$

o sea, si se cumple que

$$p_{ik} = p_i p_{\cdot k}$$

para todo  $i, k$ . (2)

**Demostración.** a) Sean  $X$  y  $Y$  mutuamente independientes. Entonces se cumple (1), y para todo número positivo  $\epsilon$  (ver 6.1 (2))

$$\begin{aligned} & P(x_i \leq X < x_i + \epsilon, y_k \leq Y < y_k + \epsilon) \\ &= F_{(X,Y)}(x_i + \epsilon, y_k + \epsilon) - F_{(X,Y)}(x_i, y_k) - F_{(X,Y)}(x_i, y_k + \epsilon) + F_{(X,Y)}(x_i, y_k) \\ &= F_X(x_i + \epsilon) - F_Y(y_k + \epsilon) - F_X(x_i + \epsilon)F_Y(y_k) - F_X(x_i)F_Y(y_k + \epsilon) + F_X(x_i)F_Y(y_k) \\ &= (F_X(x_i + \epsilon) - F_X(x_i))(F_Y(y_k + \epsilon) - F_Y(y_k)). \end{aligned}$$

Para  $\epsilon \rightarrow 0$  se obtiene de aquí (ver 2.4, teorema 1 y 4.1 (3))

$$P(X=x_i, Y=y_k) = p_{ik} = P(X=x_i)P(Y=y_k) = p_i p_{\cdot k},$$

o sea, se cumple (2).

b) Cúmplase (2) para todo  $i, k$ . Entonces se cumple para números reales cualesquiera  $x$  y  $y$

$$F_{(X,Y)}(x, y) = \sum_{\substack{i: x_i < x \\ k: y_k < y}} p_{ik} = \sum_{\substack{i: x_i < x \\ k: y_k < y}} p_i p_{\cdot k} = \left( \sum_{i: x_i < x} p_i \right) \left( \sum_{k: y_k < y} p_{\cdot k} \right) = F_X(x)F_Y(y),$$

o sea, se cumple (1).

**Teorema 2.** Sea  $(X, Y)$  un vector aleatorio continuo con la densidad de probabilidad  $f_{(X,Y)}$  y las densidades de distribución marginales  $f_X$  y  $f_Y$ . Las variables aleatorias  $X$  y  $Y$  son mutuamente independientes si y solo si se cumple

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y) \quad (3)$$

para todos los números reales  $x$  y  $y$ .

**Demostración.** a) Sean  $X$  y  $Y$  mutuamente independientes. Entonces se cumple (1) y con esto (ver 6.3, (3))

$$f_{(X,Y)}(x, y) = \frac{\delta^2 F_{(X,Y)}(x, y)}{\delta_x \delta_y} = \frac{\delta^2 F_X(x)F_Y(y)}{\delta_x \delta_y} = \frac{dF_X(x)dF_Y(y)}{dx dy} = f_X(x)f_Y(y),$$

o sea, se cumple (3).

b) Cumplese (3) para todo  $x \in \mathbb{R}$  y  $y \in \mathbb{R}$ . Entonces se cumple

$$\begin{aligned} F_{(X,Y)}(x,y) &= \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(u,v) dv du \\ &= \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) dv dy \\ &= \left( \int_{-\infty}^x f_X(u) du \right) \left( \int_{-\infty}^y f_Y(v) dv \right) \\ &= F_X(x) F_Y(y), \end{aligned}$$

o sea, se cumple (1).

En el teorema siguiente se proporcionan consecuencias fácilmente demostrables de la independencia de dos variables aleatorias, que son útiles para el trabajo práctico con variables aleatorias independientes.

**Teorema 3.** Sea  $(X, Y)$  un vector aleatorio discreto (continuo), con

$\sum_{i,k} |x_i y_k| p_{ik} < \infty$  ( $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x y| f_{(X,Y)}(x,y) dx dy < \infty$ ). Entonces se cumple, en el caso de que las variables aleatorias  $X$  y  $Y$  sean independientes:

1.  $E(XY) = (EX)(EY)$ .
2.  $\text{cov}(X, Y) = 0$ .
3.  $\rho(X, Y) = 0$ .
4.  $D(X+Y) = D^2 X + D^2 Y$ .

(En 3 y 4 se supone la existencia y positividad de las varianzas de  $X$  y  $Y$ .)

**Demuestra**ón. Las proposiciones 2, 3 y 4 se obtienen directamente de la proposición 1 (para el caso discreto (ver 6.2 (14), (17) y (15)). Por tanto, es suficiente demostrar la proposición 1.

a) Sea  $(X, Y)$  un vector aleatorio discreto. Entonces se cumple, con el teorema 1 (ver también 6.2 (7) para  $g(x, y) = xy$ ), que

$$\begin{aligned} EXY &= \sum_{i,k} x_i y_k p_{ik} = \sum_{i,k} x_i y_k p_i p_{.k} \\ &= \left( \sum_i x_i p_i \right) \left( \sum_k y_k p_{.k} \right) = (EX)(EY). \end{aligned}$$

b) Sea  $(X, Y)$  continuo. Entonces se cumple, según el teorema 2 (ver 6) para  $g(x, y) = xy$ , que

$$\begin{aligned} EXY &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y f_{(X,Y)}(x,y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y f_X(x) f_Y(y) dx dy \\ &= \left( \int_{-\infty}^{\infty} x f_X(x) dx \right) \left( \int_{-\infty}^{\infty} y f_Y(y) dy \right) = (EX)(EY). \end{aligned}$$

Por consiguiente, de la independencia de las variables aleatorias resulta que el coeficiente de correlación  $\rho(X, Y)$  es igual a cero. El recíproco de esta proposición no se cum-

ple: de  $\rho(X, Y) = 0$  no resulta la independencia de  $X$  y  $Y$  (ver para esto el ejemplo al final de 6.2; se cumple que  $\rho(X, Y)$  es igual a cero, pero, por ejemplo,

$$P(X=1, Y=1) = \frac{1}{3} \neq \frac{1}{3} \cdot \frac{2}{3} = P(X=1)P(Y=1),$$

de modo que  $X$  y  $Y$  no son independientes.

**Definición 2.** Sea  $(X, Y)$  un vector aleatorio (discreto o continuo). Si se cumple que  $\rho(X, Y) = 0$ , las variables aleatorias  $X$  y  $Y$  se denominan *incorelacionadas*.

De gran valor es la proposición siguiente sobre la distribución normal bivariada (ver 6.3, definición 4), que se obtiene directamente del teorema 2 (ver también 6.3 (14)).

**Teorema 4.** Sea  $(X, Y)$  un vector aleatorio que posee una distribución normal. Si las variables aleatorias  $X$  y  $Y$  están incorrelacionadas ( $\rho(X, Y) = \rho = 0$ ), entonces  $X$  y  $Y$  son independientes.

La proposición (4) del teorema 3 se puede extender al caso de un número finito arbitrario de variables aleatorias mutuamente incorrelacionadas, dos a dos.

**Teorema 5.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias mutuamente incorrelacionadas dos a dos ( $\rho(X_j, X_k) = 0$  para  $j \neq k; j, k = 1, 2, \dots, n$ ). Entonces se cumple que

$$D^2(X_1 + X_2 + \dots + X_n) = D^2X_1 + D^2X_2 + \dots + D^2X_n. \quad (4)$$

**Demostración.** Con  $D^2Z = EZ^2 - (EZ)^2$ ,  $\text{cov}(X, Y) = EXY - (EX)(EY)$  y la proposición de que el valor esperado de una suma de variables aleatorias es igual a la suma de los valores esperados de estas variables aleatorias, se obtiene

$$\begin{aligned} D^2\left(\sum_{i=1}^n X_i\right) &= E\left(\sum_{i=1}^n X_i\right)^2 - \left(E\left(\sum_{i=1}^n X_i\right)\right)^2 \\ &= E\left(\sum_{i=1}^n X_i^2 + 2 \sum_{\substack{j,k=1 \\ j < k}}^n X_j X_k\right) - \left(\sum_{i=1}^n EX_i\right)^2 \\ &= \sum_{i=1}^n EX_i^2 + 2 \sum_{\substack{j,k=1 \\ j < k}}^n EX_j X_k - \sum_{i=1}^n (EX_i)^2 - 2 \sum_{\substack{j,k=1 \\ j < k}}^n (EX_j)(EX_k) \\ &= \sum_{i=1}^n (EX_i^2 - (EX_i)^2) + 2 \sum_{\substack{j,k=1 \\ j < k}}^n (EX_j - EX_i - (EX_i)(EX_k)) \\ &= \sum_{i=1}^n D^2X_i + 2 \sum_{\substack{j,k=1 \\ j < k}}^n \text{cov}(X_j, X_k). \end{aligned}$$

Si se cumple ahora que  $\rho(X_j, X_k) = 0$  para  $j \neq k$ , entonces se tiene que  $\text{cov}(X_j, X_k) = 0$  para  $j \neq k$  y, por tanto, se cumple (4).

Queremos aclarar ahora, como ampliación de la definición 1, qué se entiende por independencia de  $n$  variables aleatorias ( $n$ : número natural).

**Definición 3.** Sea  $(X_1, X_2, \dots, X_n)$  un vector aleatorio  $n$ -dimensional, con la función de distribución  $F_n(X_1, X_2, \dots, X_n)$ . Las variables aleatorias  $X_1, X_2, \dots, X_n$  se denominan *com-*

*pletamente independientes* (entre sí) (también: *estocásticamente independientes*), si se cumple que

$$F_{(x_1, x_2, \dots, x_n)}(x_1, x_2, \dots, x_n) = F_{x_1}(x_1) F_{x_2}(x_2) \dots F_{x_n}(x_n) \quad (5)$$

para todos los números reales  $x_1, x_2, \dots, x_n$ ; aquí  $F_{x_i}$  denota la función de distribución marginal de  $X_i$ , ( $i=1, 2, \dots, n$ ).

De la independencia completa de las variables aleatorias  $X_1, X_2, \dots, X_n$  resulta evidentemente la independencia mutua de ellas tomadas dos a dos; el recíproco de esta proposición no se cumple (ver el ejemplo del epígrafe 3.3).

Si  $(X_1, X_2, \dots, X_n)$  es un vector aleatorio discreto o continuo, entonces a la independencia completa de las variables aleatorias  $X_1, X_2, \dots, X_n$  es equivalente una proposición análoga a la fórmula (2) o (3).

En el trabajo con variables aleatorias independientes se necesita a veces la proposición siguiente, muy evidente en cuanto al contenido, pero que no queremos demostrar.

**Teorema 6.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias independientes y  $g_1, g_2, \dots, g_n$  funciones reales continuas definidas sobre el conjunto de los números reales. Entonces,  $g_1(X_1), g_2(X_2), \dots, g_n(X_n)$  son también variables aleatorias independientes.

Concluiremos este epígrafe con la aclaración de qué se entiende por una sucesión de variables aleatorias independientes.

**Definición 4.** Una sucesión infinita  $X_1, X_2, \dots, X_n, \dots$  de variables aleatorias se denomina una *sucesión de variables aleatorias independientes*, si para todo número natural  $n \geq 2$  las variables aleatorias  $X_1, X_2, \dots, X_n$  son completamente independientes entre sí.

## 6.5 Distribución de funciones de variables aleatorias

En este epígrafe queremos determinar, en lo esencial, la distribución de probabilidad de la suma, diferencia, producto y cociente de dos variables aleatorias independientes, para lo cual comenzaremos con proposiciones especiales acerca de la distribución binomial (ver 4.5) y la de Poisson (ver 4.7).

**Teorema 1.** Sean  $X$  y  $Y$  variables aleatorias independientes que poseen una distribución binomial con los parámetros  $n_1$  y  $p$ , y  $n_2$  y  $p$ , respectivamente. Entonces  $Z=X+Y$  posee una distribución binomial con los parámetros  $n_1+n_2$  y  $p$ .

Renunciaremos a la exposición de la demostración, aunque es sencilla; el contenido de la proposición está claro si recordamos que la frecuencia absoluta de la ocurrencia de un suceso aleatorio  $A$  con la probabilidad  $P(A)=p$ , en  $n$  repeticiones independientes del experimento tomado por base, está distribuida binomialmente con los parámetros  $n$  y  $p$  (ver 4.5, en particular, las explicaciones después de la definición 1).

**Teorema 2.** Sean  $X$  y  $Y$  variables aleatorias independientes que poseen una distribución de Poisson con los parámetros  $\lambda$  y  $\mu$ , respectivamente. Entonces  $Z=X+Y$  posee una distribución de Poisson con el parámetro  $\lambda+\mu$ .

Demostración. Los valores de  $Z$  son los números  $0, 1, 2, \dots$ . Se cumple para  $l=0, 1, 2, \dots$

$$\begin{aligned}
P(Z=l) &= P(X+Y=l) = \sum_{j=0}^l P(X=j, Y=l-j) \\
&= \sum_{j=0}^l P(X=j) P(Y=l-j) \\
&= \sum_{j=0}^l p(j; \lambda) p(l-j; \mu) \\
&= \sum_{j=0}^l \frac{\lambda^j}{j!} e^{-\lambda} \frac{\mu^{l-j}}{(l-j)!} e^{-\mu} \\
&= \frac{e^{-(\lambda+\mu)}}{l!} \sum_{j=0}^l \binom{l}{j} \lambda^j \mu^{l-j} \\
&= \frac{(\lambda+\mu)^l}{l!} e^{-(\lambda+\mu)} = p(l; \lambda+\mu),
\end{aligned}$$

o sea,  $Z$  posee una distribución de Poisson con el parámetro  $\lambda+\mu$ ; aquí hemos utilizado el teorema 1(6.4), la definición de la distribución de Poisson (ver 4.7, la definición 1 y la fórmula (2)), la definición del coeficiente binomial y, por último, el teorema del binomio.

Nos ocuparemos ahora del caso de las variables aleatorias continuas. Primeramente deduciremos una fórmula; la llamada *fórmula de descomposición*, para la densidad de probabilidad de dos variables aleatorias no necesariamente independientes.

**Teorema 3.** Sea  $(X, Y)$  un vector aleatorio continuo con la densidad de probabilidad  $f_{(X,Y)}$ . Entonces, la densidad de probabilidad  $f_Z$  de la variable aleatoria  $Z=X+Y$  está dada por

$$f_Z(z) = \int_{-\infty}^z f_{(x,y)}(x, z-x) dx, \quad -\infty < z < \infty. \quad (1)$$

Demostración. Se cumple que

$$F_Z(z) = P(Z < z) = P(X+Y < z) = \int_B \int f_{(x,y)}(x, y) dx dy,$$

siendo la región de integración

$$B = \{(x, y) : x+y < z\} = \{(x, y) : -\infty < x < \infty, -\infty < y < z-x\}.$$

De aquí se obtiene (fig. 43)

$$\begin{aligned}
F_Z(z) &= \int_{-\infty}^z \left( \int_{-\infty}^{z-x} f_{(x,y)}(x, y) dy \right) dx = \int_{-\infty}^z \left( \int_{-\infty}^x f_{(x,y)}(x, t, -x) dt \right) dx \\
&= \int_{-\infty}^z \left( \int_{-\infty}^x f_{(x,y)}(x, t-x) dx \right) dt,
\end{aligned}$$

de lo que resulta

$$P_Z(z) = \int_{-\infty}^z f_{(x,y)}(x, z-x) dx.$$

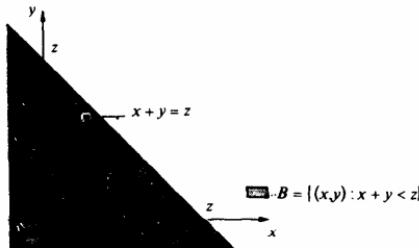


Figura 43

Con la fórmula de descomposición se puede demostrar la siguiente proposición interesante sobre la distribución normal.

**Teorema 4.** Sea  $(X, Y)$  un vector aleatorio que posee una distribución normal (con los parámetros  $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ ). Entonces  $Z = X + Y$  posee una distribución normal (con los parámetros  $\mu_1 + \mu_2$  y  $\sigma_1^2 + \sigma_2^2 + 2\rho \sigma_1 \sigma_2$ ).

No realizaremos la demostración; del teorema 4 inferimos, en particular, que la suma de dos variables aleatorias independientes, que poseen una distribución normal, está también distribuida normalmente. Es notable la validez del recíproco de esta proposición: Si la suma de dos variables aleatorias independientes está distribuida normalmente, entonces los sumandos poseen también una distribución normal. Esta proposición se debe al matemático sueco H. Cramer (nacido en 1893), el cual enriqueció también la estadística matemática con proposiciones importantes.

En el teorema siguiente caracterizaremos la distribución de probabilidad de la suma, diferencia, producto o cociente de dos variables aleatorias continuas independientes.

**Teorema 5.** Sean  $X$  y  $Y$  variables aleatorias continuas independientes, con las densidades de probabilidad  $f_X$  y  $f_Y$ , respectivamente.

1. La variable aleatoria continua  $Z = X + Y$  posee la densidad de probabilidad  $f_Z$ ,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx, -\infty < z < \infty. \quad (2)$$

2. La variable aleatoria continua  $Z = X - Y$  posee la densidad de probabilidad  $f_Z$ ,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(x-z) dx, -\infty < z < \infty. \quad (3)$$

3. La variable aleatoria continua  $Z = X \cdot Y$  posee la densidad de probabilidad  $f_Z$ ,

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y\left(\frac{z}{x}\right) dx, -\infty < z < \infty. \quad (4)$$

4. La variable aleatoria continua  $Z = \frac{X}{Y}$  posee la densidad de probabilidad  $f_Z$ ,

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(xz) f_Y(x) dx, -\infty < z < \infty. \quad (5)$$

**Demostración.** Demostraremos solo la primera proposición; las otras se obtienen en principio de la misma forma.

Para la densidad  $f_Z$  de la suma  $Z$  de dos variables aleatorias continuas  $X$  y  $Y$  se cumple la fórmula de descomposición  $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) dx$ . A causa de la supuesta independencia de las variables aleatorias  $X$  y  $Y$ , se cumple que

$$f_{X,Y}(x, z-x) = f_X(x)f_Y(z-x)$$

(ver 6.4, teorema 2) y con esto

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x) dx.$$

Las proposiciones contenidas en los teoremas siguientes se obtienen aplicando las proposiciones del teorema 5; necesitaremos de estas más adelante en el tratamiento de métodos especiales de la Estadística matemática. En estos teoremas aparecen las distribuciones  $\chi^2$ ,  $t$  y  $F$  (ver 5.6) y se motiva también el concepto grado de libertad que encontramos en estas distribuciones.

**Teorema 6.** Si las variables aleatorias  $X$  y  $Y$  son independientes y poseen una distribución  $\chi^2$  con los grados de libertad  $m_1$  y  $m_2$ , respectivamente; entonces  $Z=X+Y$  posee una distribución  $\chi^2$  con  $m_1+m_2$  grados de libertad.

**Demostración.** Aplicemos la fórmula  $f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x) dx$ . Como  $X$  y  $Y$  poseen una distribución  $\chi^2$ , se cumple (ver 5.6, definición 2) que  $f_X(x)=0$  para  $x \leq 0$  y que  $f_Y(z-x)=0$  para  $z \leq x$ . De aquí se obtiene, por una parte, que  $f_Z(z)=0$  para  $z \leq 0$  y, por otra, que  $f_Z(z) = \int_0^z f_X(x)f_Y(z-x) dx$  para  $z > 0$ .

Si sustituimos aquí las densidades  $f_X$  y  $f_Y$ , obtenemos que

$$\begin{aligned} f_Z(z) &= \frac{1}{2^{\frac{m_1}{2}} \Gamma\left(\frac{m_1}{2}\right) 2^{\frac{m_2}{2}} \Gamma\left(\frac{m_2}{2}\right)} \int_0^z x^{\frac{m_1}{2}-1} e^{-\frac{x}{2}} (z-x)^{\frac{m_2}{2}-1} e^{-\frac{z-x}{2}} dx \\ &= \frac{1}{2^{\frac{m_1+m_2}{2}} \Gamma\left(\frac{m_1}{2}\right) \Gamma\left(\frac{m_2}{2}\right)} z^{\frac{m_1+m_2}{2}-1} e^{-\frac{z}{2}} \int_0^1 t^{\frac{m_1}{2}-1} (1-t)^{\frac{m_2}{2}-1} dt. \end{aligned}$$

Si utilizamos la relación

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \quad (p>0, q>0),$$

que damos sin demostración, obtenemos en total que

$$f_Z(z) = \begin{cases} 0 & \text{para } z \leq 0, \\ \frac{1}{2^{\frac{m_1+m_2}{2}} \Gamma\left(\frac{m_1+m_2}{2}\right)} z^{\frac{m_1+m_2}{2}-1} e^{-\frac{z}{2}} & \text{para } z > 0, \end{cases}$$

o sea, que  $Z$  posee una distribución  $\chi^2$  con  $m_1+m_2$  grados de libertad.

**Corolario 1.** Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes, que poseen una distribución  $N(0,1)$ , entonces  $Z = X_1^2 + X_2^2 + \dots + X_n^2$  posee una distribución  $\chi^2$  con  $n$  grados de libertad.

**Demostración.** Según el teorema 2(5.6), las variables aleatorias  $X_k^2$  ( $k=1, 2, \dots, n$ ) poseen una distribución  $\chi^2$  con un grado de libertad; sobre la base del teorema 6(6.4) estas son, además, independientes. El resto se obtiene entonces del teorema 6 con el principio de inducción completa, debiéndose aún atender a que la independencia de  $X+Y$  y  $Z$  resulta de la independencia (completa) de  $X$ ,  $Y$  y  $Z$ .

**Teorema 7.** Si  $X$  posee una distribución  $N(0,1)$ ,  $Y$  una distribución  $\chi^2$  con  $m$  grados de libertad y  $X$  y  $Y$  son independientes, entonces  $Z = \frac{X}{\sqrt{\frac{Y}{m}}}$  posee una distribución  $t$  con  $m$  grados de libertad.

**Demostración.** De la independencia de  $X$  y  $Y$  resulta la de  $X$  y  $\tilde{Y} = \sqrt{\frac{Y}{m}}$  (ver 6.4, teorema 6). Luego, por la proposición 4 del teorema 5 se cumple que  $f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(xz) f_{\tilde{Y}}(x) dx$ . Calculemos primero la densidad de probabilidad  $f_{\tilde{Y}}$ . Para  $x > 0$  se cumple que

$$F_{\tilde{Y}}(x) = P(\tilde{Y} < x) = P\left(\sqrt{\frac{Y}{m}} < x\right) = P(Y < mx^2) = F_Y(mx^2)$$

$$\text{y con esto (ver 5.1, teorema 1)} \quad f_{\tilde{Y}}(x) = \frac{dF_{\tilde{Y}}(x)}{dx} = f_Y(mx^2) 2mx;$$

para  $x \leq 0$  se cumple  $f_{\tilde{Y}}(x) = 0$ .

De esta forma obtenemos

$$\begin{aligned} f_Z(z) &= \int_0^{\infty} x f_X(xz) f_{\tilde{Y}}(x) 2mx dx \\ &= \frac{2m \cdot m^{\frac{m}{2}-1}}{\sqrt{2\pi} \cdot 2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} \int_0^{\infty} x e^{-\frac{x^2}{2}} x^{m-1} e^{-\frac{mx^2}{2}} x dx \\ &= \frac{m^{\frac{m}{2}}}{\sqrt{\pi} \cdot 2^{\frac{m-1}{2}} \Gamma\left(\frac{m}{2}\right)} \int_0^{\infty} x^m e^{-x^2} \left(\frac{z^2+m}{2}\right) dx \\ &= \frac{m^{\frac{m}{2}}}{\sqrt{\pi} \Gamma\left(\frac{m}{2}\right) (z^2+m)^{\frac{m+1}{2}}} \int_0^{\infty} t^{\frac{m+1}{2}-1} e^{-t} dt. \end{aligned}$$

Con  $\Gamma\left(\frac{m+1}{2}\right) = \int_0^{\infty} t^{\frac{m+1}{2}-1} e^{-t} dt$  (ver 5.6(3)) se obtiene por último

$$f_Z(z) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi m} \Gamma\left(\frac{m}{2}\right)} \cdot \frac{1}{\left(1 + \frac{z^2}{m}\right)^{\frac{m+1}{2}}}$$

o sea,  $Z = \frac{X}{\sqrt{\frac{Y}{m}}}$  posee una distribución  $t$  con  $m$  grados de libertad.

**Teorema 8.** Si las variables aleatorias  $X$  y  $Y$  son independientes y poseen una distribución  $\chi^2$  con  $m_1$  y  $m_2$  grados de libertad, respectivamente; entonces  $Z = \frac{X}{\frac{m_1}{m_2}}$  posee una distribución  $F$  con  $(m_1, m_2)$  grados de libertad.

**Demostración.** De la independencia de  $X$  y  $Y$  resulta la de  $\tilde{X} = \frac{X}{m_1}$  y  $\tilde{Y} = \frac{Y}{m_2}$  (ver 6.4, teorema 6). Luego, con la proposición 4 del teorema 5 se cumple que

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_{\tilde{X}}(xz) f_{\tilde{Y}}(x) dx.$$

En virtud de que  $f_{\tilde{X}}(x) = m_1 f_X(m_1 x)$  y  $f_{\tilde{Y}}(x) = m_2 f_Y(m_2 x)$  (ver 5.1, teorema 2) resulta que

$$f_Z(z) = m_1 m_2 \int_{-\infty}^{\infty} |x| f_X(m_1 xz) f_Y(m_2 x) dx.$$

Como  $X$  y  $Y$  poseen una distribución  $\chi^2$ , se cumple (ver 5.6, definición 2) que  $f_X(m_1 xz) = 0$  para  $xz \leq 0$  y  $f_Y(m_2 x) = 0$  para  $x \leq 0$ .

De aquí se obtiene, por una parte, que  $f_Z(z) = 0$  para  $z \leq 0$  y por otra, que

$$f_Z(z) = m_1 m_2 \int_0^{\infty} x f_X(m_1 xz) f_Y(m_2 x) dx, \text{ para } z > 0.$$

Si sustituimos aquí las densidades  $f_X$  y  $f_Y$  obtenemos

$$\begin{aligned} f_Z(z) &= \frac{m_1 m_2}{2^{\frac{m_1}{2}} \Gamma\left(\frac{m_1}{2}\right) 2^{\frac{m_2}{2}} \Gamma\left(\frac{m_2}{2}\right)} \int_0^{\infty} x (m_1 xz)^{\frac{m_1}{2}-1} e^{-\frac{m_1 xz}{2}} (m_2 x)^{\frac{m_2}{2}-1} e^{-\frac{m_2 x}{2}} dx \\ &= \frac{m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}} z^{\frac{m_1}{2}-1}}{2^{\frac{m_1}{2}} \Gamma\left(\frac{m_1}{2}\right) 2^{\frac{m_2}{2}} \Gamma\left(\frac{m_2}{2}\right)} \int_0^{\infty} x^{\frac{m_1+m_2}{2}-1} e^{-x(\frac{m_1+m_2}{2})} dx \\ &= \frac{m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}}}{\Gamma\left(\frac{m_1}{2}\right) \Gamma\left(\frac{m_2}{2}\right)} \cdot \frac{z^{\frac{m_1}{2}-1}}{(m_1+m_2 z)^{\frac{m_1+m_2}{2}}} \int_0^{\infty} t^{\frac{m_1+m_2}{2}-1} e^{-t} dt. \end{aligned}$$

Con  $\Gamma\left(\frac{m_1+m_2}{2}\right) = \int_0^{\infty} t^{\frac{m_1+m_2}{2}-1} e^{-t} dt$  (ver 5.6 (3)) se obtiene finalmente, en total

$$f_Z(z) = \begin{cases} 0 & \text{para } z \leq 0, \\ \frac{\Gamma\left(\frac{m_1+m_2}{2}\right) m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}}}{\Gamma\left(\frac{m_1}{2}\right) \Gamma\left(\frac{m_2}{2}\right)} \cdot \frac{z^{\frac{m_1}{2}-1}}{(m_1 + m_2 z)^{\frac{m_1+m_2}{2}}} & \text{para } z > 0. \end{cases}$$

O sea,  $Z = \frac{\frac{X}{m_1}}{\frac{Y}{m_2}}$  posee una distribución  $F$  con  $(m_1, m_2)$  grados de libertad.

## 7. Teoremas límites

Los teoremas límites de la teoría de probabilidades ocupan un lugar central en esta disciplina matemática y, en principio, poseen importancia también en la estadística matemática; el contenido de estos teoremas son proposiciones acerca del comportamiento límite de sucesiones de variables aleatorias, siendo de particular interés de acuerdo con las necesidades prácticas, las proposiciones sobre la distribución de la suma de  $n$  variables aleatorias independientes cuando  $n \rightarrow \infty$ .

Los epígrafes 7.1 y 7.2 constituyen una introducción a los teoremas límites de la teoría de probabilidades. Para ello tratamos en el epígrafe 7.1 la llamada *desigualdad de Chebyshev*, que desempeña una importante función como medio auxiliar en la demostración de teoremas límites especiales, y en el epígrafe 7.2 presentamos los tipos de convergencia más importantes utilizados en la teoría de probabilidades para sucesiones de variables aleatorias. Los epígrafes 7.3 y 7.4 están dedicados a la denominada *Ley de los grandes números*. Una ley de los grandes números consiste, hablando sin mucha precisión, en la indicación de condiciones suficientes para que la media aritmética de una sucesión de variables aleatorias tienda hacia una constante, a medida que crece el número de los sumandos. La *Ley de los grandes números de Bernoulli*, tratada en el epígrafe 7.3, facilita una visión más clara y exacta de la relación entre la frecuencia relativa y la probabilidad de un suceso aleatorio; el epígrafe 7.4 proporciona una panorámica sobre las versiones más generales de la Ley de los grandes números.

Los epígrafes 7.5 y 7.6 están dedicados al denominado *teorema central del límite*. Un tal teorema consiste, hablando sin mucha precisión, en la indicación de condiciones suficientes para que la distribución de la suma de una sucesión de variables aleatorias tienda hacia la distribución normal, a medida que crece el número de sumandos. El *teorema integral De Moivre Laplace*, expuesto en el epígrafe 7.5, plantea una proposición semejante a la del teorema central del límite para una sucesión de variables aleatorias distribuidas binomialmente, y constituye la base para una fórmula de aproximación que está destinada al cálculo práctico de probabilidades relacionadas con la distribución binomial (parámetro  $n > > 1$ ). Por último, el epígrafe 7.6 informa acerca de las versiones más generales del teorema central del límite que, en las aplicaciones prácticas, justifican en muchas ocasiones el hecho de considerar distribuida normalmente una variable aleatoria determinada.

## 7.1 Desigualdad de Chebyshev

La función que desempeña la varianza  $D^2X$  de una variable aleatoria  $X$ , como medida para la desviación de los valores de esta variable aleatoria del centro descrito por el valor esperado  $EX$ , se hace muy clara también cuantitativamente en la desigualdad

$$P(|X-EX| \geq k\sqrt{D^2X}) \leq \frac{1}{k^2}, \quad (1)$$

que se cumple para todo número natural  $k$ . Además, esta desigualdad es muy útil en la demostración de las leyes de los grandes números (ver epígrafe 7.3). Dedicaremos la desigualdad (1), que se denomina *desigualdad de Chebyshev* en honor al importante matemático ruso P.L. Chebyshev (1821-1894), como corolario del teorema siguiente.

**Teorema 1.** Sea  $Y$  una variable aleatoria no negativa (o sea, se cumple que  $P(Y \geq 0) = 1$ ) con el valor esperado  $EY$  y  $\delta$ , un número positivo cualquiera. Entonces se cumple que

$$P(Y \geq \delta) \leq \frac{EY}{\delta} \quad (2)$$

o, en una formulación equivalente,

$$P(Y < \delta) \geq 1 - \frac{EY}{\delta}. \quad (3)$$

**Demostración.** Realizaremos la demostración separadamente para variables aleatorias discretas y continuas; el lector debe observar las analogías en el proceder.

a) Sea  $Y$  una variable aleatoria discreta que toma los valores  $y_k \geq 0$ , con las probabilidades  $p_k$ . Entonces se cumple que

$$EY = \sum_k y_k p_k \geq \sum_{k: y_k > \delta} y_k p_k \geq \delta \sum_{k: y_k > \delta} p_k = \delta P(Y \geq \delta),$$

de donde resulta (2) de inmediato.

b) Sea  $Y$  una variable aleatoria continua con la densidad de probabilidad  $f_Y$ . Entonces se cumple, en virtud de que  $P(Y < 0) = 0$ ,

$$EY = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^{\infty} y f_Y(y) dy \geq \int_{\delta}^{\infty} y f_Y(y) dy \geq \delta \int_{\delta}^{\infty} f_Y(y) dy = \delta P(Y \geq \delta),$$

de donde resulta (2) de nuevo.

**Corolario 1.** Sea  $X$  una variable aleatoria con el valor esperado  $EX$  y la varianza  $D^2X$ , y  $\varepsilon$  un número positivo arbitrario. Entonces se cumple la desigualdad de Chebyshev

$$P(|X-EX| \geq \varepsilon) \leq \frac{D^2X}{\varepsilon^2} \quad (4)$$

o, en una formulación equivalente,

$$P(|X-EX| < \varepsilon) \geq 1 - \frac{D^2X}{\varepsilon^2}. \quad (5)$$

**Demostración.** Hagamos  $\delta = \varepsilon^2$  y  $Y = |X - EX|^2$ . Entonces se cumple que  $P(Y \geq 0) = 1, \delta > 0$  y  $EY = E|X - EX|^2 = D^2X$ . Aplicando el teorema 1 obtenemos que  $P(|X - EX|^2 \geq \varepsilon^2) \leq \frac{D^2X}{\varepsilon^2}$ . Consideremos, además, que el suceso  $(|X - EX|^2 \geq \varepsilon^2)$  ocurre si y solo si si lo hace el suceso  $(|X - EX| \geq \varepsilon)$ , con lo cual hemos demostrado (4).

#### Observaciones

1. La desigualdad de Chebyshev solo tiene sentido para aquellas variables aleatorias que poseen una varianza (finita).

2. La forma dada en un inicio de la desigualdad de Chebyshev se obtiene de (4) para  $\varepsilon = k\sqrt{D^2X}$ .

3. Las desigualdades (2) y (3) y las desigualdades (4) y (5) se cumplen, en particular, para  $\delta \leq EY$  y  $\varepsilon \leq \sqrt{D^2X}$ , respectivamente, pero son evidentes en estos casos.

En el caso  $\varepsilon = 3\sqrt{D^2X}$ , la desigualdad (5) expresa que para toda variable aleatoria  $X$  (con varianza finita), la probabilidad de que tome valores cuya distancia del valor esperado sea menor que el triple de la desviación estándar, es por lo menos igual a  $\frac{8}{9}$ ,

$$P(|X - EX| < 3\sqrt{D^2X}) \geq \frac{8}{9} \approx 0,89. \quad (6)$$

Radica en la naturaleza del problema el que una proposición tan general como la desigualdad de Chebyshev, que no requiere más que el valor esperado y la varianza de la distribución de probabilidad de la variable aleatoria considerada, pueda ser muy burda en casos especiales. Por ejemplo, en el caso de que  $X$  posea una distribución normal, se obtiene que  $P(|X - EX| < 3\sqrt{D^2X}) \approx 0,997$  (ver 5.4 (26)). Sin embargo, la desigualdad de Chebyshev no se puede mejorar, como muestra el ejemplo siguiente, sin la adopción de condiciones adicionales sobre la clase de variables aleatorias considerada.

**Ejemplo.** Supongamos que la variable aleatoria  $X$  posee los valores  $-k$ ,  $k$  y  $0$  ( $k$  es aquí un número arbitrario mayor o igual que 1), y se cumple que

$$P(X = -k) = P(X = k) = \frac{1}{2k^2}, \quad P(X = 0) = 1 - \frac{1}{k^2}.$$

Entonces se cumple que  $EX = 0$ ,  $D^2X = EX^2 = k^2 \cdot \frac{1}{2k^2} \cdot 2 = 1$  y con esto

$$P(|X - EX| \geq k\sqrt{D^2X}) = P(|X| \geq k) = P(X = -k) + P(X = k) = \frac{1}{k^2}.$$

Luego, en la desigualdad de Chebyshev está, en este caso, el signo de igualdad.

A continuación indicaremos una generalización de la desigualdad de Chebyshev, la llamada desigualdad de Kolmogorov.

**Teorema 2.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias independientes con varianza (finita) y sea  $\epsilon$  un número positivo arbitrario. Entonces se cumple que

$$P\left(\max_{j \leq k \leq n} \left| \sum_{i=1}^k (X_i - EX_i) \right| \geq \epsilon\right) \leq \frac{\sum_{j=1}^n D^2 X_j}{\epsilon^2} \quad (7)$$

o, en una formulación equivalente,

$$P\left(\bigcap_{k=1}^n \left( \left| \sum_{i=1}^k (X_i - EX_i) \right| < \epsilon \right) \right) \geq 1 - \frac{\sum_{j=1}^n D^2 X_j}{\epsilon^2}. \quad (8)$$

No demostraremos la desigualdad de Kolmogorov; solo observaremos que para  $n=1$  se obtiene la desigualdad de Chebyshev.

## 7.2 Tipos de convergencia en la Teoría de probabilidades

En este epígrafe presentaremos algunas definiciones de convergencia para sucesiones de variables aleatorias. Denotaremos siempre con  $(X_n)$  una sucesión de variables aleatorias y con  $X$ , otra variable aleatoria sobre el mismo espacio de probabilidad  $(\Omega, \mathcal{A}, P)$ .

**Definición 1.** Se dice que una sucesión  $(X_n)$  converge con probabilidad uno (o converge casi seguro) a  $X$ , si se cumple que

$$P\left(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}\right) = 1. \quad (1)$$

Para esto escribimos abreviadamente  $P(\lim_{n \rightarrow \infty} X_n = X) = 1$  y de forma simbólica

$$X_n \xrightarrow{\text{C.S.}} X.$$

Por tanto, la convergencia con probabilidad uno se presenta si el conjunto de todas las  $\omega \in \Omega$ , para las cuales la sucesión numérica  $(X_n(\omega))$  converge al número  $X(\omega)$ , posee la probabilidad uno, es decir, si el suceso  $(\lim_{n \rightarrow \infty} X_n = X)$  es un suceso casi seguro o prácticamente cierto. Por esto, la convergencia casi segura en la Teoría de probabilidades se corresponde, en su esencia, con la convergencia ordinaria de una sucesión de funciones en el Análisis.

El teorema siguiente ofrece una caracterización interesante de la convergencia con probabilidad uno.

**Teorema 1.** Se cumple que  $X_n \xrightarrow{\text{C.S.}} X$  si y solo si para todo número positivo  $\epsilon$  se cumple la relación

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} \{\omega \in \Omega : |X_k(\omega) - X(\omega)| \geq \epsilon\}\right) = 0. \quad (2)$$

**Demostración.** Sea  $\epsilon > 0$  arbitrario. Introduzcamos las notaciones siguientes:

$$A_k(\epsilon) = \{|X_k - X| \geq \epsilon\}, \quad B_n(\epsilon) = \bigcup_{k=n}^{\infty} A_k$$

$$C = \lim_{n \rightarrow \infty} X_n = X, \quad C_n(\epsilon) = C \cap B_n(\epsilon),$$

$$D(\epsilon) = \left( \lim_{n \rightarrow \infty} \sup |X_n - X| \geq \epsilon \right).$$

Entonces se cumple que  $B_{n+1}(\varepsilon) \subseteq B_n(\varepsilon)$ , por consiguiente  $C_{n+1}(\varepsilon) \subseteq C_n(\varepsilon)$  y, por tanto, (ver 2.4, teorema 1)

$$\lim_{n \rightarrow \infty} P(C_n(\varepsilon)) = P\left(\bigcap_{k=1}^{\infty} C_k(\varepsilon)\right).$$

1. Supongamos que se cumple que  $X_n \xrightarrow{\text{c.s.}} X$ , o sea, que  $P(X) = 1$ . Entonces tenemos que  $\bigcap_{n=1}^{\infty} C_n(\varepsilon) = \emptyset$  y, por tanto,  $\lim_{n \rightarrow \infty} P(C_n(\varepsilon)) = 0$ . De  $P(B_n(\varepsilon)) = P(C_n(\varepsilon))$  resulta que  $\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0$ , es decir, se cumple (2).
2. Supongamos que se cumple (2), o sea, que  $\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0$ . Entonces tenemos que  $D(\varepsilon) \subseteq B_n(\varepsilon)$  para  $n=1, 2, \dots$ . Por consiguiente, se cumple que  $P(D(\varepsilon)) = 0$ . De  $\bar{C} \subseteq \bigcup_{k=1}^{\infty} D\left(\frac{1}{k}\right)$  resulta que  $P(\bar{C}) = 0$ , o sea, que es  $P(\bar{C}) = 1$ , lo que es equivalente a  $X_n \xrightarrow{\text{c.s.}} X$ .

**Definición 2.** Se dice que una sucesión  $(X_n)$  converge en probabilidad (o: converge estocásticamente) a  $X$ , si para todo número positivo  $\varepsilon$  se cumple que

$$\lim_{n \rightarrow \infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}) = 1. \quad (3)$$

Para esto escribimos abreviadamente  $\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$  y de forma simbólica:

$$X_n \xrightarrow{P} X.$$

La relación (3) expresa que en la convergencia estocástica de  $(X_n)$  hacia  $X$ , la diferencia de  $X_n$  y  $X$  en al menos  $\varepsilon$ , es decir, el suceso  $(|X_n - X| \geq \varepsilon)$ , posee una probabilidad que converge hacia cero para  $n \rightarrow \infty$ ; aquí  $\varepsilon$  es un número positivo cualquiera. Sin embargo, la relación (3) no dice que para un  $\omega \in \Omega$  fijo exista para todo  $\varepsilon > 0$  un número natural  $n_0$  tal que se cumpla  $|X_n(\omega) - X(\omega)| < \varepsilon$  para todo  $n \geq n_0$ , es decir, que se cumpla

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Entre la convergencia casi segura o prácticamente cierta y la convergencia estocástica existe la relación siguiente.

**Teorema 2.** Si la sucesión  $(X_n)$  converge con probabilidad uno a  $X$ , entonces converge también estocásticamente a  $X$ , es decir, se cumple que

$$X_n \xrightarrow{\text{c.s.}} X \Rightarrow X_n \xrightarrow{P} X. \quad (4)$$

**Demuestra.** Utilicemos las notaciones  $A_n(\varepsilon)$  y  $B_n(\varepsilon)$  introducidas en la demostración del teorema 1. De  $X_n \xrightarrow{\text{c.s.}} X$  resulta, con el teorema 1, que  $\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0$ . En virtud de  $A_n(\varepsilon) \subseteq B_n(\varepsilon)$  se obtiene de aquí directamente que  $\lim_{n \rightarrow \infty} P(A_n(\varepsilon)) = 0$ , es decir, se cumple que  $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$ , lo cual es equivalente a  $\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$  y con esto a  $X_n \xrightarrow{P} X$ .

**Definición 3.** Si las variables aleatorias  $X_n (n=1, 2, \dots)$  y  $X$  poseen una varianza (finita), decimos que la sucesión  $(X_n)$  converge en media cuadrática a  $X$ , si se cumple que

$$\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0. \quad (5)$$

Para esto escribimos simbólicamente  $X_n \xrightarrow{\text{c.m.c.}} X$ .

El contenido de la convergencia en media cuadrática es que  $\lim_{n \rightarrow \infty} D^2(X_n - X) = 0$ , es decir, que la sucesión de las varianzas  $D^2(X_n - X)$  converge hacia la varianza de una variable aleatoria distribuida puntualmente (ver 4.3, teorema 4).

Entre la convergencia en media cuadrática y la convergencia estocástica existe la relación siguiente:

**Teorema 3.** Si la sucesión  $(X_n)$  converge en media cuadrática a  $X$ , entonces converge también estocásticamente a  $X$ , es decir, se cumple que

$$X_n \xrightarrow{\text{c.m.c.}} X \Rightarrow X_n \xrightarrow{P} X. \quad (6)$$

**Demostración.** Sea  $\varepsilon > 0$ . Utilicemos el teorema 1(7.1) con  $\delta = \varepsilon^2$  y  $Y = |X_n - X|^2$  y obtenemos

$$P(|X_n - X| \geq \varepsilon) = P(|X_n - X|^2 \geq \varepsilon^2) \leq \frac{E|X_n - X|^2}{\varepsilon^2} = \frac{E(X_n - X)^2}{\varepsilon^2}.$$

Si se cumple que  $X_n \xrightarrow{\text{c.m.c.}} X$ , es decir, si  $\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0$ , entonces resulta que  $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$  para todo  $\varepsilon > 0$ , es decir, se cumple que  $X_n \xrightarrow{P} X$ .

**Definición 4.** Se dice que la sucesión  $(X_n)$  converge según las funciones de distribución (o: converge en distribución) a  $X$ , si entre las funciones de distribución  $F_{X_n}$  y  $F_X$  se cumple la relación

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (7)$$

en todos los puntos de continuidad  $x$  de  $F_X$ . Para esto escribimos de forma simbólica  $X_n \xrightarrow{\text{e.d.}} X$ .

Advertimos expresamente que la proposición (7) no tiene que cumplirse para todo  $x$ ; esta puede que no se cumpla para aquellos valores de  $x$  en los cuales la función de distribución  $F_X$  de la variable aleatoria  $X$  no es continua. Pero si la función de distribución  $F_X$  es continua (este es por ejemplo el caso si la variable aleatoria  $X$  es continua), entonces la convergencia en distribución de  $(X_n)$  hacia  $X$  es equivalente a la convergencia ordinaria de la sucesión de funciones  $(F_{X_n})$  a la función  $F_X$ .

Entre la convergencia estocástica y la convergencia en distribución existe la relación siguiente:

**Teorema 4.** Si la sucesión  $(X_n)$  converge estocásticamente a  $X$ , entonces converge también en distribución a  $X$ , es decir, se cumple que

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\text{e.d.}} X. \quad (8)$$

**Demostración.** Sea  $\varepsilon > 0$  arbitrario. Hagamos  $A_n = \{|X_n - X| < \varepsilon\}$ .

Entonces se cumple, según la premisa, que  $\lim_{n \rightarrow \infty} P(A_n) = 1$ . Sobre la base de la fórmula de la probabilidad total (ver 3.4, teorema 1) se obtiene para un número real  $x$  cualquiera

$$F_{X_n}(x) = P(X_n < x) = P(X_n < x | A_n)P(A_n) + P(X_n < x | \bar{A}_n)P(\bar{A}_n).$$

Por una parte, resulta de aquí que  $F_{X_n}(x) \leq P(X_n < x | A_n)P(A_n) + P(\bar{A}_n)$ , de donde se obtiene con

$$P(X_n < x | A_n) = \frac{P(X_n < x) \cap (|X_n - X| < \varepsilon)}{P(A_n)}$$

$$\begin{aligned} P(X_n < x | A_n) &= \frac{P((X_n < x) \cap (X < X_n + \varepsilon) \cap (X > X_n - \varepsilon))}{P(A_n)} \\ &\leq \frac{P(X < x + \varepsilon)}{P(A_n)} \end{aligned}$$

y  $\lim_{n \rightarrow \infty} P(\bar{A}_n) = 0$ , la proposición

$$\lim_{n \rightarrow \infty} \sup F_{X_n}(x) \leq F_X(x + \varepsilon).$$

Por otra parte resulta que

$$F_{X_n}(x) \geq P(X_n < x | A_n) P(A_n) = P(|X_n - X| < \varepsilon),$$

de donde se obtiene con

$$P(|X_n - X| < \varepsilon) + P(|X_n - X| \geq \varepsilon) \geq P(X < x - \varepsilon)$$

y

$$\lim_{n \rightarrow \infty} P(\bar{A}_n) = \lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$$

la relación

$$\lim_{n \rightarrow \infty} \inf F_{X_n}(x) \geq F_X(x - \varepsilon).$$

Si  $x$  es un punto de continuidad de  $F_X$ , obtenemos para  $\varepsilon \rightarrow 0$  las desigualdades

$$\lim_{n \rightarrow \infty} \sup F_{X_n}(x) \leq F_X(x) \text{ y } \lim_{n \rightarrow \infty} \inf F_{X_n}(x) \geq F_X(x).$$

Por tanto, se cumple que  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$  en todos los puntos de continuidad de  $F_X$ , es decir, se cumple que  $X_n \xrightarrow{\text{c.d.}} X$ .

Con esto hemos mostrado que la convergencia en distribución es la más débil entre los tipos de convergencia aquí definidos. Si la variable aleatoria  $X$  posee una distribución puntual, o sea, si existe un número  $c$  con  $P(X=c)=1$ , y la sucesión  $(X_n)$  converge en distribución a  $X$ , entonces ella convergerá también estocásticamente a  $X$ . (Para esto escribimos abreviadamente  $X_n \xrightarrow{P} c$  y decimos que la sucesión  $(X_n)$  converge estocásticamente hacia  $c$ .) Se cumple, por consiguiente, el teorema siguiente:

**Teorema 5.** Sea  $X$  una variable aleatoria distribuida puntualmente. Una sucesión  $(X_n)$  converge estocásticamente a  $X$  si y solo si converge en distribución a  $X$ .

**Demostración.** Sea  $X$  una variable aleatoria distribuida puntualmente. Sin restricción de la generalidad podemos suponer que  $P(X=0)=1$ . Sobre la base del teorema 4 solo tenemos que demostrar que la convergencia estocástica resulta, bajo esta condición, de la convergencia en distribución.

Por consiguiente se cumple

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ 1 & \text{para } x > 0, \end{cases}$$

en todos los puntos de continuidad de  $F_X$ , es decir, se cumple que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 0 & \text{para } x < 0, \\ 1 & \text{para } x > 0. \end{cases}$$

Para  $\varepsilon > 0$  arbitrario, se cumple que

$$\begin{aligned} P(|X_n| < \varepsilon) &= P(X_n < \varepsilon) - P(X_n \leq -\varepsilon) \\ &= F_{X_n}(\varepsilon) - F_{X_n}(-\varepsilon + 0), \end{aligned}$$

de donde para  $n \rightarrow \infty$  resulta, sobre la base de las premisas, que

$$\lim_{n \rightarrow \infty} P(|X_n| < \epsilon) = 1 - 0 = 1.$$

Esto significa precisamente que la sucesión  $(X_n)$  converge estocásticamente a 0.

## 7.3 Teoremas de Bernoulli y de Poisson (Ley de los grandes números)

En este epígrafe retrocederemos otra vez a la relación entre la frecuencia relativa y la probabilidad. La Ley de los grandes números de Bernoulli, que se expone a continuación, puede concebirse como una formulación matemática del efecto observado reiteradamente en casos concretos de la estabilización de la frecuencia relativa (ver 2.1).

Designemos con  $A$  un suceso aleatorio que ocurre en el marco de un experimento aleatorio con la probabilidad  $P(A) = p$ ; denotemos con  $f_n(A)$ , al igual que antes (ver epígrafe 4.5), la frecuencia relativa aleatoria de la ocurrencia de  $A$  en una serie de  $n$  repeticiones independientes de este experimento aleatorio.

**Teorema 1.** Para todo número  $\epsilon$  se cumple que

$$\lim_{n \rightarrow \infty} P(|f_n(A) - p| < \epsilon) = 1 \quad (1)$$

o, en una formulación equivalente,

$$\lim_{n \rightarrow \infty} P(|f_n(A) - p| \geq \epsilon) = 0, \quad (2)$$

es decir, la sucesión  $(f_n(A))$  converge estocásticamente hacia  $p$  (*Ley de los grandes números de Bernoulli*, 1712).

**Demostración.** Se cumple que  $E f_n(A) = p$  ( $n = 1, 2, \dots$ ) y  $D f_n(A) = \frac{p(1-p)}{n} \rightarrow 0$  para  $n \rightarrow \infty$  (ver 4.5 (13) y (14)). Aplicando la desigualdad de Chebyshev (ver 7.1, teorema 2, y sustituir  $X$  por  $f_n(A)$ ) se obtiene, para  $\epsilon > 0$  arbitrario, la desigualdad

$$P(|f_n(A) - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \left( \leq \frac{1}{4n\epsilon^2} \right),$$

de donde resulta la proposición (2) del teorema por paso al límite cuando  $n \rightarrow \infty$ .

La Ley de los grandes números de Bernoulli plantea que la probabilidad de que la diferencia entre la frecuencia relativa  $f_n(A)$  de un suceso  $A$  y la probabilidad  $P(A) = p$  de este suceso sea menor que un número positivo  $\epsilon$  cualquiera dado, está arbitrariamente cerca de uno, si el número  $n$  de las repeticiones del experimento aleatorio considerado es suficientemente grande. Esto significa que para un número de experimentos suficientemente grande, la probabilidad de que exista una diferencia insignificante entre la frecuencia relativa y el número  $p$  es aproximadamente igual a uno. En particular, la Ley de los grandes números de Bernoulli muestra que todo suceso aleatorio con probabilidad positiva, por pequeña que esta sea, ocurre al menos una vez en una serie de experimentos suficientemente grande con una probabilidad situada arbitrariamente cerca de uno. De estas explicaciones se deduce por qué se denomina la proposición del teorema 1 como Ley de los grandes números.

Queremos aún deducir una proposición que contiene al teorema 1 como caso particular: la llamada Ley de los grandes números de Poisson. Constituye el punto de partida una serie de  $n$  experimentos aleatorios independientes, en los cuales ocurre un suceso  $A$  con una probabilidad que, en contraposición con el esquema de experimentos de Bernoulli considerado anteriormente, depende del número del experimento aleatorio (esquema de experimentos de Poisson). Designemos con  $p_k$  la probabilidad del suceso  $A$  en el experimento  $k$ . Consideremos la variable aleatoria  $X_k$  tal que

$$X_k = \begin{cases} 1 & \text{en caso de que el suceso } A \text{ ocurra en el experimento,} \\ 0 & \text{en caso de que el suceso } \bar{A} \text{ ocurra en el experimento,} \end{cases} \quad k=1,2,\dots,n.$$

Entonces se cumple que  $P(X_k=1)=p_k$ ,  $P(X_k=0)=1-p_k$ . Por consiguiente se cumplen las ecuaciones

$$EX_k=1 \cdot p_k+0(1-p_k)=p_k$$

y

$$D^2X_k=(1-p_k)^2p_k+(0-p_k)^2(1-p_k)=p_k(1-p_k).$$

Designemos de nuevo con  $f_n(A)$  la frecuencia relativa aleatoria de la ocurrencia de  $A$  en un esquema de experimentos de Poisson.

Entonces se cumple que

$$f_n(A)=\frac{1}{n}(X_1+X_2+\dots+X_n),$$

de donde (ver 4.3, teorema 1 y 6.2, colorario 1)

$$Ef_n(A)=\frac{1}{n}EX_1+\dots+EX_n=\frac{p_1+\dots+p_n}{n}$$

y, en virtud de la independencia de las variables aleatorias  $X_1, X_2, \dots, X_n$  (ver 6.4, teorema 5), resulta

$$\begin{aligned} D^2f_n(A) &= \frac{1}{n^2} D^2(X_1+\dots+X_n) = \frac{D^2X_1+\dots+D^2X_n}{n^2} \\ &= \frac{p_1(1-p_1)+\dots+p_n(1-p_n)}{n^2} \left( \leq \frac{1}{4n} \rightarrow 0 \text{ para } n \rightarrow \infty \right). \end{aligned}$$

De la aplicación de la desigualdad de Chebyshev (ver 7.1, teorema 2 y sustituir  $X$  por  $f_n(A)$ ) se obtiene directamente la proposición del teorema siguiente.

**Teorema 2.** Para todo número positivo  $\varepsilon$  se cumple que

$$\lim_{n \rightarrow \infty} P \left( \left| f_n(A) - \frac{p_1+\dots+p_n}{n} \right| < \varepsilon \right) = 1 \quad (3)$$

o, en una formulación equivalente,

$$\lim_{n \rightarrow \infty} P \left( \left| f_n(A) - \frac{p_1+\dots+p_n}{n} \right| \geq \varepsilon \right) = 0 \quad (4)$$

(*Ley de los grandes números de Poisson*).

Verifiquemos, por una parte, que en el caso de que la probabilidad del suceso  $A$  sea igual en todos los experimentos ( $p_k = p$  para todo  $k$ ), se obtiene de aquí la Ley de los grandes números de Bernoulli; pero observemos también por otra, que una proposición correspondiente a la Ley de los grandes números de Bernoulli se obtiene también con premisas menos limitantes. El epígrafe siguiente trata sobre otras generalizaciones de la Ley de los grandes números de Bernoulli.

## 7.4 Generalización de la Ley de los grandes números

En la deducción de la Ley de los grandes números de Poisson partimos de una sucesión especial  $(X_k)$  de variables aleatorias, consideramos la sucesión de las medias aritméticas  $\frac{1}{n} (X_1 + X_2 + \dots + X_n)$  e investigamos la convergencia de esta sucesión. La proposición del teorema 2(7.3) se puede formular entonces de modo que la sucesión  $(Y_n)$  de las medias aritméticas centradas  $Y_n$ ,

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k - E \left( \frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k) \quad (1)$$

converge estocásticamente a cero. Este hecho es el fundamento de la definición siguiente.

**Definición 1.** Se dice que una sucesión  $(X_k)$  satisface la Ley de los grandes números, si la sucesión  $(Y_n)$  de las medias aritméticas centradas  $Y_n$ ,

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k - E \left( \frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k)$$

converge estocásticamente a cero.

En esta formulación se supone la existencia de los valores esperados que aparecen. Si estos no existen, entonces se dice que la sucesión  $(X_k)$  satisface la Ley de los grandes números si existe una sucesión numérica  $(a_n)$  tal, que la sucesión  $(Y_n)$ ,  $Y_n = \frac{1}{n} \sum_{k=1}^n X_k - a_n$ , converge estocásticamente a cero.

El próximo objetivo consiste en indicar condiciones suficientes para que una sucesión de variables aleatorias satisfaga la Ley de los grandes números.

Algunas proposiciones importantes en esta dirección se deben a nombrados representantes de la escuela rusa de la teoría de probabilidades, fundada por P.L. Chebyshev, la cual representó el centro de la investigación teórica en este campo al inicio de nuestro siglo (en especial se deben a P.L. Chebyshev y su famoso discípulo A.A. Markov (1856-1922), y a los matemáticos soviéticos A. Ja. Kinchine (1894-1959) y A.N. Kolmogorov, el fundador de la teoría axiomática de probabilidades.

**Teorema 1. (Ley de los grandes números de Markov)**

Sea  $(X_k)$  una sucesión de variables aleatorias, que satisfacen la condición

$$\lim_{n \rightarrow \infty} \frac{D^2 \left( \sum_{k=1}^n X_k \right)}{n^2} = 0 \quad (\text{condición de Markov}). \quad (2)$$

Entonces la sucesión  $(X_k)$  satisface la Ley de los grandes números.

**Demostración.** Aplicando la desigualdad de Chebyshev (ver 7.1, corolario 1 y sustituir  $X$  por  $\frac{1}{n} \sum_{k=1}^n X_k$ ) resulta, para  $\varepsilon > 0$  arbitrario,

$$P(|Y_n| \geq \varepsilon) = P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - E\left(\frac{1}{n} \sum_{k=1}^n X_k\right)\right| \geq \varepsilon\right) \leq \frac{D^2 \left(\sum_{k=1}^n X_k\right)}{n^2 \varepsilon^2},$$

de donde se obtiene en virtud de la validez de la condición de Markov (2), que  $\lim_{n \rightarrow \infty} P(|Y_n| \geq \varepsilon) = 0$ . Luego, la sucesión  $(Y_n)$  converge estocásticamente a cero, o sea, la sucesión  $(X_k)$  satisface la Ley de los grandes números.

**Teorema 2.** (Ley de los grandes números de Chebyshev)

Sea  $(X_k)$  una sucesión de variables aleatorias incorrelacionadas dos a dos, cuyas varianzas están acotadas. (Luego, existe un número  $M > 0$ , tal que se cumple  $D^2 X_k \leq M$ , para todo  $k$ .) Entonces la sucesión  $(X_k)$  satisface la Ley de los grandes números.

**Demostración.** Como las variables aleatorias  $X_k$  están incorrelacionadas dos a dos, se cumple (ver 6.4, teorema 5) que

$$D^2 \left( \sum_{k=1}^n X_k \right) = \sum_{k=1}^n D^2 X_k$$

y, por tanto, sobre la base de la premisa,

$$\frac{D^2 \left( \sum_{k=1}^n X_k \right)}{n^2} = \frac{\sum_{k=1}^n D^2 X_k}{n^2} \leq \frac{nM}{n^2} = \frac{M}{n}.$$

De aquí resulta que se cumple la condición de Markov y con esto hemos demostrado la validez de la Ley de los grandes números para la sucesión  $(X_k)$ , en virtud del teorema 1.

Como caso especial de la Ley de los grandes números de Chebyshev se obtiene directamente la Ley de los grandes números de Poisson (ver 7.3, teorema 2; allí se cumple para todo  $k$  que  $D^2 X_k = p_k(1-p_k) \leq \frac{1}{4}$  a causa de que  $0 \leq p_k \leq 1$ ).

En la formulación de otras proposiciones utilizaremos un concepto, que estableceremos en la definición siguiente.

**Definición 2.** Los elementos de un conjunto de variables aleatorias se denominan *distribuidos idénticamente*, si todas las variables aleatorias de este conjunto poseen una misma función de distribución.

En relación con esta definición llamamos la atención de que las variables aleatorias distribuidas idénticamente no tienen que ser iguales; en cambio, las variables aleatorias iguales poseen una distribución idéntica, como es natural. El lector debe aclararse a sí mismo este comportamiento.

**Teorema 3.** Sea  $(X_k)$  una sucesión de variables aleatorias independientes, distribuidas idénticamente, con el valor esperado (común)  $\mu$  y la varianza (común)  $\sigma^2$ . Entonces la sucesión  $(X_k)$  satisface la Ley de los grandes números. En particular, la sucesión

$\left(\frac{1}{n} \sum_{k=1}^n X_k\right)$  de las medias aritméticas de la sucesión  $(X_k)$  converge estocásticamente al valor esperado (común)  $\mu$ .

La proposición de este teorema se obtiene directamente de la Ley de los grandes números de Chebyshev; el lector debe verificar esto. En la parte relativa a la Estadística matemática haremos un empleo provechoso de la proposición del teorema 3. Por último, advertimos que la Ley de los grandes números de Bernoulli (ver 7.3, teorema 1) se obtiene directamente como caso especial de este teorema.

Es de notar que se puede renunciar a la condición de la existencia de la varianza.

**Teorema 4.** (*Ley de los grandes números de Kinchine*). Sea  $(X_k)$  una sucesión de variables aleatorias independientes, distribuidas idénticamente, con el valor esperado (común)  $\mu$ . Entonces, la sucesión  $(X_k)$  satisface la Ley de los grandes números. En particular la sucesión  $\left(\frac{1}{n} \sum_{k=1}^n X_k\right)$  converge estocásticamente a  $\mu$ .

Queremos exponer aún algunas proposiciones sobre la denominada Ley fuerte de los grandes números.

**Definición 3.** Se dice que una sucesión  $(X_k)$  satisface la *Ley fuerte de los grandes números*, si la sucesión  $(Y_n)$ ,

$$Y_n = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k),$$

converge casi seguro a cero, suponiéndose la existencia de los valores esperados  $EX_k$  (Si estos no existen, entonces se dice que la sucesión  $(X_k)$  satisface la Ley fuerte de los grandes números, si existe una

sucesión numérica  $(a_n)$  tal, que la sucesión  $(Y_n)$ ,  $Y_n = \frac{1}{n} \sum_{k=1}^n X_k - a_n$  (converge casi seguro a cero.)

Las definiciones 1 y 3 solo se diferencian en el tipo de la convergencia de la sucesión  $(Y_n)$  hacia cero: en la definición 1 se parte de la convergencia estocástica y la definición 3 se basa en la convergencia con probabilidad uno. Como de la convergencia con probabilidad uno resulta la convergencia estocástica (ver 7.2, teorema 2), una sucesión  $(X_k)$ , para la cual se cumpla la Ley fuerte de los grandes números, satisface también la Ley de los grandes números. (Para una mejor diferenciación, la Ley de los grandes números caracterizada mediante la definición 1, se denomina *Ley débil de los grandes números*.)

Los teoremas siguientes, provenientes de A.N. Kolmogorov, indican condiciones suficientes para la validez de la Ley fuerte de los grandes números.

**Teorema 5.** Sea  $(X_k)$  una sucesión de variables aleatorias independientes que satisface la condición

$$\sum_{k=1}^{\infty} \frac{D^2 X_k}{k^2} < \infty \quad (\text{condición de Kolmogorov}). \quad (3)$$

Entonces la sucesión  $(X_k)$  satisface la Ley fuerte de los grandes números.

La demostración de este teorema se basa fundamentalmente en la desigualdad de Kolmogorov (ver 7.1, teorema 2), pero no la realizaremos; no obstante, observemos que en el teorema 5 se supone la existencia de las varianzas.

Cada una de las condiciones siguientes, impuestas a una sucesión  $(X_k)$  de variables aleatorias, es suficiente para la validez de la condición de Kolmogorov (3) y en unión con la condición de independencia de las variables aleatorias  $X_1, X_2, \dots$ , lo es también para la validez de la Ley fuerte de los grandes números.

1.  $X_1, X_2, \dots$  están distribuidas idénticamente (con el valor esperado  $\mu$  y la varianza  $\sigma^2$ ). (En este caso se obtiene que  $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{c.s.} \mu$ .)

2. Existe  $M > 0$  tal que  $D^2X_k \leq M$  para todo  $k$ .

La última condición mencionada muestra, que en el caso de una sucesión de variables aleatorias independientes, la Ley de los grandes números de Chebyshev (ver teorema 2), -y en particular, la Ley de los grandes números de Poisson (ver 7.3, teorema 2)-, pueden considerarse también como Ley fuerte de los grandes números.

La primera condición nombrada muestra que la Ley de los grandes números formulada en el teorema 3 y, en particular, la Ley de los grandes números de Bernoulli (ver 7.3, teorema 1), puede pasar también como Ley fuerte de los grandes números. La sucesión  $(f_n(A))$  de las frecuencias relativas  $f_n(A)$ , tomadas como variables aleatorias, de la ocurrencia de un suceso aleatorio  $A$  en una serie de  $n$  repeticiones independientes de un mismo experimento aleatorio, para el cual el suceso  $A$  tiene la probabilidad  $P(A) = p$ , converge para  $n \rightarrow \infty$  no solo estocásticamente, sino también con probabilidad uno.\*

Por último, daremos un teorema muy concluyente referente a la validez de la Ley fuerte de los grandes números para una sucesión de variables aleatorias independientes, distribuidas idénticamente.

**Teorema 6.** (Ley de los grandes números de Kolmogorov)

Sea  $(X_k)$  una sucesión de variables aleatorias independientes distribuidas idénticamente.

1. Si existe  $EX_1 = \mu$ , entonces la sucesión  $(X_k)$  satisface la Ley fuerte de los grandes números. En particular, se cumple que  $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{C.S.}} \mu$ .

2. Si la sucesión  $\left( \frac{1}{n} \sum_{k=1}^n X_k \right)$  converge hacia una variable aleatoria  $X$ , entonces  $X$  está distribuida puntualmente, es decir, existe un número  $a$ , tal que  $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{C.S.}} a$ . Además, existe entonces  $EX_1 = a$ .

Renunciaremos a la demostración de este teorema, que es muy difícil; esta se realiza haciendo referencia al lema de Borel-Cantelli (ver 3.3, teorema 1). Advertimos aún que, sobre la base de la primera proposición del teorema 6, la Ley de los grandes números de Kinchine (ver teorema 4) puede considerarse también como Ley fuerte de los grandes números.

## 7.5 Teorema integral de De Moivre-Laplace

Por teorema límite se entiende, en la teoría de probabilidades, en lo esencial, una proposición sobre el comportamiento límite de una sucesión  $(F_{Z_n})$  de funciones de distribución de una sucesión dada  $(Z_n)$  de variables aleatorias. Las leyes de los grandes números, tratadas en los epígrafes 7.3 y 7.4, son ejemplos de teoremas límites semejantes; se indican condiciones suficientes para que dada una sucesión  $(X_n)$ , la sucesión  $(Z_n)$ ,

$$Z_n = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k),$$

converja estocásticamente (o incluso, casi seguro) hacia cero, de donde resulta la convergencia en distribución de la sucesión  $(Z_n)$  hacia cero (ver 7.2, teorema 4).

Muchas veces, y de casos semejantes nos ocuparemos en este y en el próximo epígrafe, los teoremas límites consisten en la indicación de condiciones suficientes para la convergencia de una sucesión de funciones de distribución hacia la función de distribución  $\Phi$  de una variable aleatoria distribuida normalmente con los parámetros  $\mu = 0$  y  $\sigma^2 = 1$ ; con esto se obtienen también caracterizaciones significativas de la distribución normal.

\* Esta proposición fue considerada por primera vez en 1909 por el matemático francés E. Borel (1871-1956); por ello se denomina también Ley de los grandes números de Borel.

En este epígrafe conoceremos el llamado *teorema integral de De Moivre-Laplace* (A. De Moivre, 1730, P.S. Laplace, 1812), que tiene por contenido una proposición semejante para variables aleatorias distribuidas binomialmente.

Sea  $A$  un suceso aleatorio que ocurre en el marco de un experimento aleatorio con la probabilidad  $P(A) = p$ ,  $0 < p < 1$ . Denotemos con  $F_n(A)$ , al igual que antes (ver epígrafe 4.5), el número aleatorio de la ocurrencia de  $A$  en una serie de  $n$  repeticiones independientes de este experimento. Como sabemos, la variable aleatoria discreta  $F_n(A)$  está distribuida binomialmente con los parámetros  $n$  y  $p$ , y se cumplen las relaciones  $EF_n(A) = np$  y  $D^2F_n(A) = np(1-p)$ . Sobre la base de la Ley de los grandes números de Bernoulli (ver 7.3, teorema 1), sabemos que la sucesión  $(Y_n)$ ,

$$Y_n = f_n(A) - p = \frac{F_n(A) - np}{n} = \frac{F_n(A) - EF_n(A)}{n},$$

converge estocásticamente -y según la Ley de los grandes números de Borel (ver 7.4, antes del teorema 6) incluso casi seguro- hacia cero cuando  $n \rightarrow \infty$ . La función de distribución límite es, por consiguiente, la función de distribución de una variable aleatoria distribuida puntualmente, o sea, de una variable aleatoria que posee, la varianza cero. Observemos que

$$D^2Y_n = \frac{1}{n^2} D^2F_n(A) = p \frac{(1-p)}{n}$$

y, por tanto, se cumple que

$$\lim_{n \rightarrow \infty} D^2Y_n = 0.$$

El comportamiento diferente de la función de distribución límite se hace pausible, de esta forma.

Ahora queremos considerar la sucesión  $(Z_n)$  que se obtiene mediante estandarización de la sucesión  $(F_n(A))$ ,

$$Z_n = \frac{F_n(A) - EF_n(A)}{\sqrt{D^2F_n(A)}} = \frac{F_n(A) - np}{\sqrt{np(1-p)}},$$

entre las variables aleatorias  $Z_n$  y las  $Y_n$  consideradas anteriormente, existe la relación

$$Z_n = \sqrt{n} \frac{Y_n}{\sqrt{p(1-p)}},$$

y se cumplen, por tanto, las relaciones  $EZ_n = EY_n = 0$  y  $D^2Z_n = \frac{n}{p(1-p)} D^2Y_n = 1$

$(n=1, 2, \dots)$ . Para la sucesión  $(Z_n)$  se cumple el teorema siguiente:

#### Teorema 1. (Teorema integral de De Moivre-Laplace)

Sea  $(F_n)$  una sucesión de variables aleatorias  $F_n$ , que están distribuidas binomialmente con los parámetros  $n$  y  $p$  ( $0 < p < 1$ ,  $n=1, 2, \dots$ ). Entonces para la sucesión  $(F_{Z_n})$  de las funciones de distribución  $F_{Z_n}$  de las variables aleatorias  $Z_n$ ,

$$Z_n = \frac{H_n - EF_n}{\sqrt{D^2F_n}} = \frac{F_n - np}{\sqrt{np(1-p)}}$$

se cumple para todo  $x$  la relación

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad (1)$$

o sea, la sucesión  $(Z_n)$  converge en distribución hacia una variable aleatoria con distribución  $N(0,1)$ .

Una demostración clara de este teorema exige medios auxiliares que sobrepasan los marcos de este libro. Por eso, nos limitaremos aclarar la significación del teorema 1 y, en particular, la utilización de esta proposición en casos de aplicación.

Si  $X$  es una variable aleatoria distribuida binomialmente con los parámetros  $n$  ( $n > 1$ ) y  $p$  ( $0 < p < 1$ ), entonces el cálculo de las probabilidades

$$P(X=k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

es complicado, como habíamos dicho ya en el epígrafe 4.5. Sin embargo, en este caso ( $n > 1$ ), no nos interesamos tanto por tales probabilidades particulares, que son en su mayoría muy pequeñas, sino por los valores que toma  $X$  de un intervalo cualquiera dado. Aplicando el teorema 1 se obtiene para  $P(a \leq X < b)$

$$\begin{aligned} P(a \leq X < b) &= P\left(\frac{a-np}{\sqrt{np(1-p)}} \leq \frac{X-np}{\sqrt{np(1-p)}} < \frac{b-np}{\sqrt{np(1-p)}}\right) \\ &\sim \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right) \end{aligned} \quad (2)$$

(La expresión señalada representa al mismo tiempo una aproximación para las probabilidades  $P(a \leq X \leq b)$ ,  $P(a < X \leq b)$  y  $P(a \leq X < b)$ .

Una variable aleatoria distribuida binomialmente con los parámetros  $n (> 1)$  y  $p$  ( $0 < p < 1$ ) posee aproximadamente una distribución normal con los parámetros  $\mu = np$  y  $\sigma^2 = np(1-p)$ .

**Ejemplo.** Una fábrica suministra bombillitos en cartones de 1 000 cada uno. Se sabe que la fábrica produce un promedio de bombillitos defectuosos del 3 %. Luego, en un cartón con 1 000 bombillitos es de esperar que alrededor 30 estén defectuosos. Nos interesamos por la probabilidad de que en un cartón se encuentren de 20 a 40 bombillitos defectuosos. Para ello designemos con  $X$  el número (aleatorio) de los bombillos defectuosos en un cartón. La variable aleatoria  $X$  está distribuida binomialmente con los parámetros  $n=1\,000$  y  $p=0,03$ ; se cumple entonces que

$$EX = 1\,000 \cdot 0,03 = 30 \text{ y } D^2X = 1\,000 \cdot 0,03 \cdot (1-0,03) = 29,1.$$

Para la probabilidad buscada se obtiene que

$$P(20 \leq X \leq 40) = \sum_{k=20}^{40} P(X=k) = \sum_{k=20}^{40} \binom{1\,000}{k} 0,03^k (1-0,03)^{1\,000-k}.$$

Con esta fórmula no se puede calcular de forma práctica la probabilidad buscada. Si utilizamos la fórmula de aproximación (2) con  $a=20$ ,  $b=40$ ,  $n=1000$ ,  $p=0,03$  y  $1-p=0,97$ , obtenemos que

$$\begin{aligned} P(20 \leq X \leq 40) &\sim \Phi\left(\frac{40 - 1000 \cdot 0,03}{\sqrt{1000 \cdot 0,03 \cdot 0,97}}\right) - \Phi\left(\frac{20 - 1000 \cdot 0,03}{\sqrt{1000 \cdot 0,03 \cdot 0,97}}\right) \\ &= \Phi\left(\frac{10}{\sqrt{29,1}}\right) - \Phi\left(\frac{-10}{\sqrt{29,1}}\right) \\ &= 2\Phi\left(\frac{10}{\sqrt{29,1}}\right) - 1 \\ &\approx 2\Phi(1,85) - 1 = 2 \cdot 0,97 - 1 = 0,94 = 94 \% \end{aligned}$$

Luego, la probabilidad buscada es aproximadamente de 0,94.

## 7.6 Teorema central del límite

Para la formulación del teorema integral de De Moivre-Laplace partimos de una sucesión de variables aleatorias  $F_n$ , distribuidas binomialmente. Una variable aleatoria  $F_n$  distribuida binomialmente con los parámetros  $n$  y  $p$  se puede representar como suma de  $n$  variables aleatorias discretas  $X_1, X_2, \dots, X_n$ , independientes y distribuidas idénticamente,  $F_n = X_1 + X_2 + \dots + X_n$ , cuya tabla de distribución está dada por

	1	0	
$X_k$ :			
	$p$	$1-p$	

$(k=1,2,\dots,n)$

(ver en 7.3 las explicaciones posteriores a la formulación de la Ley de los grandes números de Bernoulli). Las variables aleatorias  $Z_n = \frac{F_n - EF_n}{\sqrt{D^2 F_n}}$  de la sucesión  $(Z_n)$  considerada en el teorema integral de De Moivre-Laplace, se pueden representar también, debido a que  $EF_n = \sum_{k=1}^n EX_k$  y  $D^2 F_n = \sum_{k=1}^n D^2 X_k$ , en la forma

$$Z_n = \frac{\sum_{k=1}^n (X_k - EX_k)}{\sqrt{\sum_{k=1}^n D^2 X_k}} \quad (1)$$

La proposición del teorema integral de De Moivre-Laplace plantea que la sucesión  $(Z_n)$ , formada según (1) de la sucesión  $(X_k)$  de variables aleatorias independientes, distribuidas idénticamente, converge en distribución hacia una variable aleatoria con distribución  $N(0,1)$ . Este hecho constituye el fundamento de la definición siguiente.

**Definición 1.** Se dice que una sucesión  $(X_k)$  de variables aleatorias independientes satisface al **teorema central del límite**, si la sucesión  $(Z_n)$ ,

$$Z_n = \frac{\sum_{k=1}^n (X_k - EX_k)}{\sqrt{\sum_{k=1}^n D^2 X_k}}. \quad (1)$$

converge en distribución hacia una variable aleatoria con distribución  $N(0,1)$ , es decir, si para la sucesión  $(F_{Z_n})$  de las funciones de distribución de  $Z_n$  se cumple la relación

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, -\infty < x < \infty. \quad (2)$$

Luego, en esta formulación se supone la existencia de los valores esperados y las varianzas que aparecen, así como que  $D^2 X_k > 0$ . Si estas magnitudes no existen, entonces se dice que la sucesión  $(X_k)$  satisface al teorema central del límite, si existen sucesiones numéricas  $(a_n)$  y  $(b_n)$ ,  $b_n \neq 0$ , tales que la sucesión  $(Z_n)$ ,

$$Z_n = \frac{\sum_{k=1}^n X_k - a_n}{b_n}, \quad (3)$$

converge en distribución hacia una variable aleatoria con distribución  $N(0,1)$ .

El próximo objetivo consiste en indicar condiciones suficientes para que una sucesión de variables aleatorias satisfaga al teorema central del límite. Para ello afirmamos primariamente que, sobre la base del teorema integral de De Moivre-Laplace, una sucesión  $(X_k)$  de variables aleatorias independientes, distribuidas idénticamente en dos puntos, satisface al teorema central del límite. A continuación se muestra que se puede renunciar a la condición de la distribución en dos puntos.

**Teorema 1.** Sea  $(X_k)$  una sucesión de variables aleatorias independientes, distribuidas idénticamente y con varianza finita y positiva. Entonces la sucesión  $(X_k)$  satisface al teorema central del límite.

Este teorema se debe a J.W. Lindeberg (1922) y P. Lévy (1925); por eso se denomina también como **Teorema límite de Lindeberg-Lévy**. En la estadística matemática este teorema es de gran significación; en él se plantea que las sumas estandarizadas  $Z_n$  de variables aleatorias  $X_k$  independientes y distribuidas idénticamente, poseen asintóticamente una distribución  $N(0,1)$  y (es decir, cuando el número de los sumandos tiende a  $\infty$ ), si para los sumandos  $X_k$  existe, junto al valor esperado (común)  $\mu$ , la varianza (común)  $\sigma^2$  ( $\sigma^2 < \infty$ ) y esta es positiva ( $\sigma^2 > 0$ ).

Esto significa que las variables aleatorias

$$Z_n = \frac{\sum_{k=1}^n (X_k - EX_k)}{\sqrt{\sum_{k=1}^n D^2 X_k}} = \frac{\sum_{k=1}^n (X_k - \mu)}{\sqrt{n\sigma^2}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} \quad (4)$$

poseen aproximadamente una distribución  $N(0,1)$ , para  $n$  grande, formulado de otra forma.

Si en el teorema 1 se renuncia a la condición de que las variables aleatorias distribuidas idénticamente  $X_1, X_2, \dots$  posean una varianza finita y positiva, o a que las variables aleatorias  $X_1, X_2, \dots$  estén distribuidas idénticamente, entonces una sucesión ~~no satisface~~ por lo general, al teorema central del límite; sin embargo, existen una serie de proposiciones que tratan de la validez del teorema central del límite también en el caso de variables aleatorias no distribuidas idénticamente, por ejemplo, el *teorema límite de Lyapunov* y el *teorema límite de Lindeberg-Feller*.

Primero presentaremos el teorema límite de Lyapunov (A.M. Lyapunov (1857-1918) fue uno de los representantes más significativos de la famosa escuela rusa de teoría de las probabilidades, fundada por P.L. Chebyshov.)

**Teorema 2.** Sea  $(X_k)$  una sucesión de variables aleatorias independientes, que poseen momentos de tercer orden. Si para las sucesiones  $(b_n)$  y  $(c_n)$ , con

$$b_n = \sqrt{\sum_{k=1}^n E|X_k - EX_k|^3} \quad \text{y} \quad c_n = \sqrt{\sum_{k=1}^n D^3 X_k} \quad (5)$$

respectivamente, se satisface la condición

$$\lim_{n \rightarrow \infty} \frac{b_n}{c_n} = 0 \quad (\text{condición de Lyapunov}), \quad (6)$$

la sucesión  $(X_k)$  satisface al teorema central del límite.

La condición de Lyapunov se satisface evidentemente, si, además, las variables aleatorias  $(X_k)$  están distribuidas idénticamente.

Sobre la base del teorema 2, la validez de la condición de Lyapunov es suficiente para el cumplimiento del teorema central del límite, pero no es necesaria. En particular, no es necesario que existan momentos de orden mayor que dos. Lindeberg indicó una condición suficiente para la validez del teorema central del límite, para cuya formulación -a la cual renunciaremos aquí- no se necesitan momentos de orden mayor que dos. De la satisfacción de esta condición -llamada condición de Lindeberg- resulta el cumplimiento de la condición de Lyapunov, en caso de que existan momentos de tercer orden. Además, de la satisfacción de la condición de Lindeberg resulta la proposición

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \frac{D^2 X_k}{\sum_{i=1}^n D^2 X_i} = 0. \quad (7)$$

Esta relación expresa que la varianza de cada sumando  $X_k$  es pequeña en comparación con la varianza de la suma  $X_1 + X_2 + \dots + X_n$ .

Por último, W. Feller demostró (1935) que, suponiendo que (7) se cumpla, para la validez del teorema central del límite es necesaria la satisfacción de la condición de Lindeberg.

Estos teoremas son de gran importancia, tanto en el aspecto teórico -en especial teórico-cognoscitivo como en el aspecto de sus aplicaciones prácticas. De estos teoremas se obtiene con frecuencia la justificación para describir aproximadamente la distribución de una variable aleatoria como una distribución normal. Así, por ejemplo, se puede suponer que una variable aleatoria posee una distribución normal si se obtiene mediante superposición de un número considerable de efectos aleatorios mutuamente independientes, donde cada uno de estos efectos tiene una influencia insignificante sobre la variable aleatoria consi-

derada, en comparación con la suma de los otros efectos (ver (7)). Con esto, el conocimiento de los valores esperados y las varianzas es lo único que se necesita saber acerca de las distribuciones de probabilidad de los efectos aleatorios que intervienen en la superposición. El resultado de una tal superposición se describe muy bien mediante la distribución normal, si el número de los efectos aleatorios es elevado.

Estas notables regularidades en los fenómenos aleatorios, que se expresan en forma cuantitativa en los teoremas centrales del límite y en forma cualitativa, en las leyes de los grandes números, han conducido a realizar y homenajear a la distribución normal; reproducimos en una traducción libre una observación del biólogo y estadístico inglés Sir Francis Galton (1822-1911):

Yo no sabría nombrar algo que pudiera impresionar tanto la fantasía como la forma maravillosa del orden cósmico, que se expresa en la Ley de los grandes números. Si los griegos hubieran conocido esta ley, la hubieran personificado y adorado como divinidad. Con serenidad y completo desconocimiento de si misma ejerce su poder en medio del más salvaje desorden. Mientras más gigantesco es el conjunto y mayor la aparente anarquía, tanto más completa es su fuerza. Ella es la ley superior del caos. Tan pronto una gran masa de elementos sin reglas se ordenan medianamente, se muestra que una imprevista y maravillosa regularidad, sumamente armónica, estaba ya oculta en ellos.

Con esto concluimos nuestras observaciones sobre la Teoría de probabilidades para decirnos a los problemas de la Estadística matemática.

## 8. Estadística descriptiva

Los métodos y procedimientos de la Estadística descriptiva tienen el objetivo de representar y agrupar convenientemente, de forma clara y gráfica, el material de datos obtenidos, para expresar de manera comprensible su esencia. Esto se realiza, por una parte, a través de *listas, tablas y representaciones gráficas* y por otra, mediante el cálculo de las llamadas medidas estadísticas (por ejemplo, medidas de tendencia central). Con esto solo se obtienen proposiciones sobre el material de datos presentado, y se utilizan métodos y procedimientos que son bastante independientes de la Teoría de probabilidades. Sin embargo, el objetivo esencial en la investigación de un material de datos concretos, de una llamada muestra, consiste, en última instancia, en llegar a proposiciones más generales sobre una denominada población. Para esto sirven los métodos y procedimientos de la Estadística matemática (del capítulo 9 al 11), los cuales se basan en la Teoría de probabilidades.

En correspondencia con el objetivo planteado para este libro, nos ocuparemos de forma detallada de la Estadística matemática y solamente abordaremos ligeramente los métodos y procedimientos utilizados en la Estadística descriptiva. Así trataremos en el epígrafe 8.1 los métodos para una característica medible, y en el epígrafe 8.3, los métodos para dos características medibles.

Además, presentaremos algunas medidas estadísticas típicas (epígrafes 8.2 y 8.4), las cuales aparecerán de nuevo, en su mayoría, en los capítulos posteriores relativos a la Estadística matemática.

### 8.1 Métodos para el estudio de una característica medible

La base de una investigación estadística es un conjunto de objetos en el cual una o varias características deben ser investigadas. En este y en el próximo epígrafe partiremos de que se debe investigar una característica medible  $X$ , más general, una característica que se puede describir numéricamente en  $n$  objetos, y designaremos con  $x_1, \dots, x_n$  los valores de medición (números) obtenidos, los cuales no tienen que ser necesariamente diferentes unos de otros.

Se puede tratar, por ejemplo, del número de puntos obtenidos en un trabajo de control por  $n$  estudiantes, o de las medidas del cuerpo de  $n$  estudiantes de la misma edad, o de las temperaturas del mediodía en  $n$  lugares diferentes, o tomando un ejemplo de la técnica, de la diferencia entre el diámetro real y la medida prevista en  $n$  pernos producidos en un taladro automático.

En el marco de la Estadística matemática se considera a  $X$  como una variable aleatoria, y a  $x_1, \dots, x_n$  como valores observados de  $X$  en  $n$  experimentos concretos.

Los números  $x_1, \dots, x_n$  forman una serie de mediciones (de tamaño  $n$ ). La agrupación de los elementos de una serie de mediciones en la sucesión en que van surgiendo, se denomina lista originaria.

Ejemplo 1. La tabla siguiente contiene el resultado de un trabajo de control realizado por 100 estudiantes. Aquí se representó el rendimiento de cada uno de esos estudiantes de acuerdo con una puntuación determinada, pudiéndose alcanzar como máximo 15 puntos.

Tabla 1

7	6	13	7	11	10	13	8	14	10
4	8	3	12	14	8	11	10	2	14
9	8	12	3	9	5	4	9	8	15
12	9	8	10	6	11	7	11	11	12
3	4	13	0	6	3	8	6	7	13
6	13	2	14	4	9	5	9	9	6
9	10	10	9	10	10	10	12	0	12
11	7	5	2	12	1	7	13	6	10
11	9	10	15	11	10	13	8	12	14
8	12	8	11	13	12	10	14	12	9

Como se observa ya en este ejemplo, una lista originaria es bastante incomprensible, y no resulta fácil reconocer en ella lo típico, las particularidades. Por eso se ordenan, generalmente, los valores de medición de la característica y se determina, con ayuda del tarjado la frecuencia absoluta de los diferentes valores. La agrupación de los valores de medición que se realiza de esta forma se denomina *tabla de frecuencia* o *tabla de distribución primaria*.

Ejemplo 2. A continuación se muestra la tabla de frecuencia del material numérico considerado en el ejemplo 1.

Tabla 2

Puntos	Tarjado	Frecuencia	Puntos	Tarjado	Frecuencia
0	II	2	8		10
1	I	1	9	I	11
2	III	3	10	III	13
3		4	11	I	9
4		4	12	III	11
5	III	3	13	III	8
6	II	7	14	I	6
7	I	6	15	II	2

Como se puede apreciar las tablas de frecuencia son más comprensibles y pequeñas que las listas originarias, así como más apropiadas para emitir un juicio sobre la distribución. En ellas no se pierde información con respecto a las listas originarias. Las tablas de frecuencia se pueden ilustrar bien mediante representaciones gráficas.

Ejemplo 3. Ilustraremos la tabla de frecuencia dada en el ejemplo 2 mediante representaciones gráficas (fig. 44).

Una representación gráfica como la de la figura 44a se llama *polígono escalonado o histograma*; la representación gráfica dada en la figura 44b se denomina *polígono de frecuencia* (o abreviadamente: *polígono*). Si lo que se quiere es comparar varias series de mediciones de distintos tamaños (en el marco de un mismo problema), se representa sobre el eje de las ordenadas en lugar de la frecuencia absoluta, la frecuencia relativa.

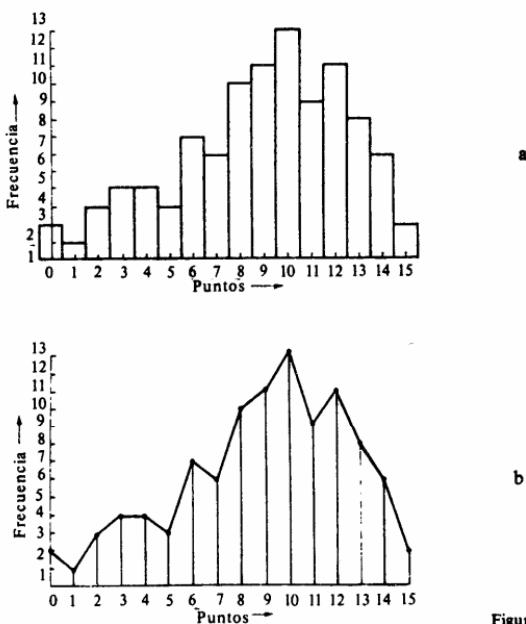


Figura 44

Si se tienen series de mediciones muy grandes, entonces se recomienda realizar una *agrupación o clasificación* de los valores, concentrando algunos consecutivos. Este proceder se basa sobre una partición en clases, es decir, sobre una descomposición en subconjuntos disjuntos, del conjunto de los posibles valores de la característica considerada.

Los conceptos que se relacionan con el de partición en clases, tales como número de clase, amplitud de clase, límites de la clase, medio de la clase, no requieren de más aclaraciones. Todo lo que concierne a la técnica de la formación de clases se encuentra en la bibliografía.

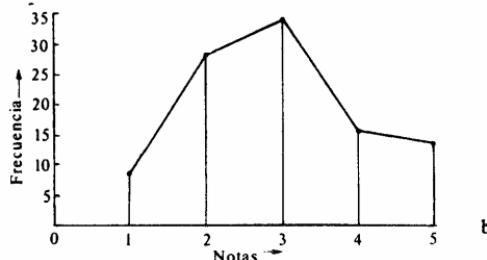
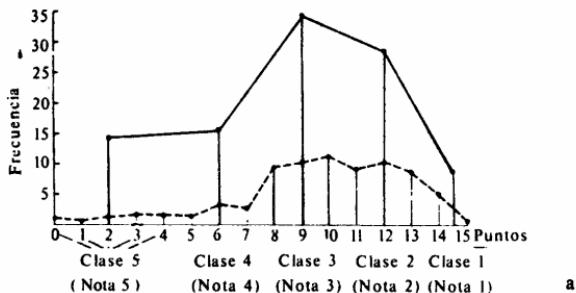


Figura 45

**Ejemplo 4.** Agrupemos el material numérico dado en el ejemplo 1 en correspondencia con la siguiente partición en clases.

Clase 1: 0,1,2,3,4 puntos

Clase 2: 5,6,7 "

Clase 3: 8,9,10 "

Clase 4: 11,12,13 "

Clase 5: 14,15 "

(La evaluación de los rendimientos con las notas 1 hasta 5 constituye la fundamentación para esta partición en clases; de aquí, corresponde a la clase 1 la nota 1, a la clase 2 la nota 2 y así sucesivamente.)

Los resultados se resumen en la tabla siguiente -en una denominada tabla de distribución secundaria- y en la figura 45 se ilustran gráficamente.

Tabla 3		Tarjado	Frecuencia	Frecuencia relativa
Clase	Nota			
1	(5)	JHT JHT IIII	14	0,14
2	(4)	JHT JHT JHT I	16	0,16
3	(3)	JHT JHT JHT JHT JHT IIII	34	0,34
4	(2)	JHT JHT JHT JHT JHT III	28	0,28
5	(1)	JHT III	8	0,08

Observemos que a la par que se gana en claridad mediante una clasificación del material numérico, surge una pérdida de información (con respecto a la lista originaria o a la tabla de distribución primaria).

## 8.2 Medidas estadísticas para el estudio de una característica medible

Para valorar una serie de mediciones se introducen con frecuencia magnitudes, las denominadas medidas estadísticas, que se calculan a partir de los valores de medida. Queremos ocuparnos, en primer lugar, de las medidas de tendencia central, las cuales caracterizan a una serie de medidas mediante un único valor, un valor "promedio", y tratar a continuación las medidas de dispersión empíricas, que ponen de manifiesto la desviación de los valores de medida en la serie de mediciones.

### 8.2.1 Medidas de tendencia central

Entre las medidas de tendencia central la más conocida es la *media aritmética*. La media aritmética  $\bar{x}_n$  de una serie de mediciones  $x_1, \dots, x_n$  está definida de la forma siguiente:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Si el material numérico es dividido en  $k$  clases y el punto medio de la clase  $j$  (marca de clase) se denota por  $u_j$ , y con  $m_j$ , su frecuencia de clase (=número de medidas que se encuentran en la clase  $j$ ), entonces se define la media aritmética de la forma siguiente:

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^k u_j m_j \quad (2)$$

En la práctica, al hallar la media aritmética, en especial cuando se tiene un número grande de medidas, se recomienda el empleo de procedimientos, especialmente concebidos para este caso (por ejemplo, mediante la introducción de un valor medio provisional); no abordaremos esto con más detalle.

Ejemplo. Para el material numérico del ejemplo 1 (8.1) se obtiene  $\bar{x}_n = 8,92$  (utilizando la partición en clases del ejemplo 4 (8.1) se obtiene  $\bar{x}_n = 8,82$ ).

Otras medidas de tendencia central son la mediana empírica  $\tilde{x}_n$ , la moda empírica  $\hat{x}_n$  y la media geométrica  $\tilde{x}_n$ .

Por *mediana empírica*  $\tilde{x}_n$  se entiende, en caso de un número impar  $n$ , el puntaje situado en el medio de una serie de mediciones, ordenadas de mayor a menor; en el caso de un número par  $n$ ,  $\tilde{x}_n$  es igual a la media aritmética de los dos puntajes que se encuentran en el medio de la serie de mediciones, ordenadas de mayor a menor. (Para el ejemplo considerado por nosotros se obtiene  $\tilde{x}_n = 9$ .) La mediana está caracterizada entonces, a grosso modo, porque a cada uno de sus lados se encuentra la mitad de las mediciones.

Por *moda empírica*  $\hat{x}_n$  se entiende aquel puntaje de una serie de mediciones, el cual aparece como mínimo, tantas veces como cualquier otro puntaje en la serie. (Para nuestro

ejemplo se obtiene como moda empírica  $\hat{x}_n = 10$ .) Las modas empíricas de una serie de mediciones son los puntajes de mayor frecuencia en la serie de mediciones considerada.

La medida geométrica  $\hat{x}_n$  de una serie de valores  $x_1, \dots, x_n$  está dada por

$$\hat{x}_n = \sqrt[n]{x_1 \dots x_n};$$

ella está definida solamente para series de mediciones con puntajes positivos. En comparación con la media aritmética está menos influenciada por los valores extremos de la serie de mediciones. En la práctica se utiliza frecuentemente en la Estadística económica (por ejemplo, en la caracterización de un tiempo de crecimiento promedio).

### 8.2.2 Medidas de dispersión

Una primera idea sobre la dispersión de una serie de mediciones nos la puede dar el recorrido  $\delta_n$ , el cual se define como la diferencia del máximo y el mínimo de los puntajes, o sea,

$$\begin{aligned} \delta_n &= x_{\max} - x_{\min} \text{ con} & x_{\max} &= \max \{x_1, \dots, x_n\}, \\ & & x_{\min} &= \min \{x_1, \dots, x_n\}. \end{aligned} \quad (3)$$

El recorrido depende solamente de los valores extremos de una serie de mediciones. no suministra información alguna, por ejemplo, sobre cómo se concentran los valores en torno a la media aritmética en la serie de mediciones. Como medidas adecuadas para esto se tiene la varianza empírica  $s_n^2$ , que se define por

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \quad (4)$$

y la raíz cuadrada positiva de esta  $s_n$ ,

$$s_n = \sqrt{s_n^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}, \quad (5)$$

que se denomina desviación estándar empírica.

(Las razones de por qué no se define  $s_n^2$  como media aritmética de los cuadrados de las desviaciones de los valores de medición de la media aritmética, o sea, como  $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ , se aclararán solo en el marco de las explicaciones sobre la Estadística matemática (ver 10.4.2 b).)

Para el cálculo práctico se utiliza la fórmula (fácilmente deducible de 4)

$$s_n^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2 \right] \quad (6)$$

Si el material numérico se divide en clases, entonces se define la varianza empírica (con las notaciones de 8.2.1) como:

$$s_n^2 = \frac{1}{n-1} \sum_{j=1}^k (u_j - \bar{x}_n)^2 m_j, \quad (7)$$

donde  $\bar{x}_n$  se calcula según (2).

**Ejemplo.** Para el material numérico del ejemplo 1 (8.1) se obtiene según (6), con  $\sum_{i=1}^n x_i^2 = 9216$  y  $\sum_{i=1}^n x_i = 892$ , la varianza empírica  $s_n^2$ ,  $s_n^2 = 12,72$ , de donde resulta para la desviación estándar empírica  $s_n$  el número 3,57. (Utilizando la partición en clases del ejemplo 4 (8.1) se obtiene  $s_n^2 = 13,35$  y de ahí se deriva que  $s_n = 3,65$ .)

Por último queremos llamar la atención sobre el *coeficiente de variación empírica* (o *coeficiente de variabilidad*)  $v_n$  para una serie de mediciones, definido para  $\bar{x}_n \neq 0$  por

$$v_n = \frac{s_n}{\bar{x}_n}. \quad (8)$$

El coeficiente de variación se utiliza para comparar varias series de mediciones con respecto a sus desviaciones estándar empíricas, considerando sus medias aritméticas respectivas y frecuentemente se da en tanto por ciento.

### 8.3 Métodos para el estudio de dos características medibles

En este y en el epígrafe siguiente partiremos de que se van a investigar a la vez dos características medibles  $X$  y  $Y$ , en  $n$  objetos y designaremos con  $(x_1, y_1), \dots, (x_n, y_n)$  los pares de valores de medición que se obtienen (no necesariamente desiguales).

Se puede tratar, por ejemplo, de los números de puntos obtenidos en dos pruebas de Matemática por  $n$  estudiantes, o de la talla y el peso de  $n$  alumnos de la misma edad, o tomando un ejemplo de la economía, del grado de cumplimiento de los planes de producción y de financiamiento en  $n$  fábricas).

En el marco de la Estadística matemática se entiende por  $(X, Y)$  un vector aleatorio (bidimensional), siendo  $(x_1, y_1), \dots, (x_n, y_n)$  los valores observados de  $(X, Y)$  en  $n$  experimentos concretos.

La agrupación de los pares  $(x, y)$  según el orden en el cual van surgiendo, se denomina nuevamente *lista originaria*. Racionalmente, también se pasa en este caso, a la confección de una *tabla de distribución primaria (tabla de frecuencia)*, la cual para cada posible valor  $(x, y)$  de  $(X, Y)$  contiene la frecuencia (absoluta o relativa) de la aparición de este par en el material numérico considerado (ver el ejemplo siguiente), donde dado el caso se realiza una *partición en clases* para las características  $X$  y  $Y$ . Para hacer más comprensible lo anterior sirven las *representaciones gráficas* del material numérico, por ejemplo, mediante puntos en el plano  $x, y$  o en forma de *histogramas* (especiales). No profundizaremos más y terminaremos este corto epígrafe con un ejemplo.

**Ejemplo.** A 100 niños recién nacidos se les midió la talla  $X$  (en cm) y el perímetro de la cabeza  $Y$  (en cm). Obviemos la lista originaria y demos la tabla de frecuencia correspondiente en la cual aparecen redondeados los pares de valores de medición (los cuadros en blanco se interpretan como si tuvieran ceros).

Como se aprecia, aparecen con más frecuencia, entre los 100 recién nacidos investigados niños con una talla entre 48 y 52 cm, y un perímetro de la cabeza, entre 33 y 36 cm. Contrariamente, aparecen muy pocos niños pequeños (o grandes) que presenten un gran (pequeño) perímetro de la cabeza.

Tabla 4

$X \backslash Y$	32	33	34	35	36	37	38	39	
47		1	1	3					5
48	1	6	7						14
49	1	5	10	5					21
50	1	4	9	9	1				24
51			3	6	4	1			14
52			3	1	7	1			12
53		1		1	2	1	1	1	7
54						1	1		2
55									0
56								1	1
	3	17	33	25	14	4	2	2	(100)

## 8.4 Medidas estadísticas para el estudio de dos características medibles

El objetivo de medir las características  $X$  y  $Y$  a la vez, en  $n$  objetos, consiste en ganar claridad sobre si existe relación entre ellas y en qué medida se da esta relación. En este epígrafe queremos introducir dos medidas estadísticas especiales, la denominada *covarianza empírica* y el llamado *coeficiente de correlación empírico*.

Para esto, sean  $(x_1, y_1), \dots, (x_n, y_n)$  los resultados de las mediciones de dos características  $X$  y  $Y$  en  $n$  objetos. Denotemos con  $\bar{x}_n$  y con  $s_{x,n}^2$  la media aritmética y la varianza empírica respectivamente de la serie formada por las componentes  $x$ :  $x_1, \dots, x_n$ . El mismo significado se le asigna a  $\bar{y}_n$  y a  $s_{y,n}^2$  para la serie de valores formada por las componentes  $y$ :  $y_1, \dots, y_n$ . Estas medidas estadísticas no dicen nada, naturalmente, sobre la dependencia mutua de  $X$  y  $Y$ . Para valorar el comportamiento de  $X$  y  $Y$  en este sentido es apropiado el uso de la covarianza empírica  $s_{xy,n}$ , que se define de la forma siguiente:

$$s_{xy,n} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n). \quad (1)$$

Se aprecia claramente, que  $s_{x,y}$  es positiva, cuando a valores grandes de  $x$  se hacen corresponder valores grandes de  $y$  y a valores pequeños de  $x$ , valores pequeños de  $y$ . Además, se reflexiona de forma análoga que la covarianza empírica  $s_{x,y}$  es negativa, cuando se hacen corresponder a valores grandes de  $x$ , pequeños valores de  $y$  y viceversa.

Una medida estadística más potente para la dependencia mutua de  $X$  y  $Y$  se obtiene cuando se relaciona la covarianza empírica con el producto de las desviaciones estándar empíricas  $s_{x,n} = \sqrt{s_{x,x}^2}$  y  $s_{y,n} = \sqrt{s_{y,y}^2}$ , a través del *coeficiente de correlación empírico*, definido por:

$$r_n = \frac{s_{x,y}}{s_{x,n}s_{y,n}} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_n)^2}} \quad (2)$$

Se cumple que  $r_n > 0$  o  $r_n < 0$  si y solo si  $s_{x,y} > 0$  o  $s_{x,y} < 0$  respectivamente. Además se cumple la desigualdad  $|r_n| \leq 1$ , de donde se obtiene que  $|r_n|=1$  si y solo si al representar mediante puntos en el plano  $x,y$  los pares numéricos  $(x_i, y_i)$ , estos se encuentran sobre una misma recta (ver 6.2, teorema 4). El coeficiente de correlación empírica se puede interpretar entonces, como una medida para la tendencia (dirección) e intensidad de la dependencia lineal entre los valores  $x$  y los valores  $y$ .

Para hallar en la práctica el coeficiente de correlación empírica se recomienda utilizar la relación (deducible fácilmente de 2)

$$r_n = \frac{\sum_{i=1}^n x_i y_i - n \bar{x}_n \bar{y}_n}{\sqrt{(n-1)s_{x,n}^2} \sqrt{(n-1)s_{y,n}^2}} \quad (3)$$

y cuando no se han calculado anteriormente  $\bar{x}_n$ ,  $\bar{y}_n$ ,  $s_{x,n}^2$  y  $s_{y,n}^2$ , puede utilizarse la relación

$$r_n = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2} \sqrt{\sum_{i=1}^n y_i^2 - \frac{1}{n} \left( \sum_{i=1}^n y_i \right)^2}} \quad (4)$$

**Ejemplo.** El coeficiente de correlación empírico  $r_n$  para el material numérico del ejemplo del epígrafe 8.3 se obtiene utilizando (4) y con  $n=100$ , de la forma siguiente

$$\begin{aligned} \sum_{i=1}^n x_i &= 5\ 009, & \sum_{i=1}^n x_i^2 &= 251\ 215, \\ \sum_{i=1}^n y_i &= 3\ 460, & \sum_{i=1}^n y_i^2 &= 119\ 908, \\ \sum_{i=1}^n x_i y_i &= 173\ 477, & r_n &= 0.674 \end{aligned}$$

Queremos finalizar las explicaciones sobre la estadística descriptiva con una observación general sobre las propiedades de aplicación de las fórmulas dadas en los epígrafes

8.2 y 8.4. Para la deducción de estas fórmulas hemos partido siempre de que los valores numéricos utilizados son resultados de procesos de mediciones, para los cuales se utilizó una escala de unidades, o con otras palabras, de que los valores de observación utilizados se pueden comparar (en el sentido de mayor que, igual que y menor que), de donde se obtiene que las diferencias de los valores de las mediciones también se pueden interpretar razonablemente.

En especial, en las investigaciones pedagógicas, pero también en los psicológicas y en las sociales, se investigan con frecuencia características que no se pueden medir con una escala de unidades, conocidas como características cualitativas (piense por ejemplo en la característica "resultado de una prueba"; esta característica se puede describir numéricamente, digamos con las notas del 1 al 5, pero la diferencia entre las notas no se puede interpretar razonablemente. Otro ejemplo para esto sería la característica "procedencia social"). En estos casos no se pueden aplicar las fórmulas de manera irreflexiva; no obstante existe una serie de posibilidades de describir numéricamente, por ejemplo, la dependencia mutua de características cualitativas, es decir, de aquellas que no se pueden expresar por medio de una escala de unidades (por ejemplo, mediante el cálculo del llamado coeficiente de correlación del rango o del denominado coeficiente de contingencia).

## 9. Conceptos fundamentales de la Estadística matemática

En este capítulo se hace una introducción a la Estadística matemática. Después de la exposición de las tareas que se plantea esta disciplina (epígrafe 9.1), se realiza en el epígrafe 9.2 la definición de los conceptos *población* y *muestra*. El teorema esencial para todos los procedimientos de la Estadística matemática, el denominado *teorema fundamental de la Estadística matemática*, se explica en el epígrafe 9.3, haciendo referencia directa a la Ley de los grandes números de Bernoulli. Por último, en el epígrafe 9.4 se tratan los llamados estadígrafos, limitándose el tratamiento, en su esencia, a aquellas proposiciones que jugarán un papel importante en la exposición posterior.

### 9.1 Tareas que se plantea la Estadística matemática

Muchos procesos reales se describen convenientemente mediante modelos matemáticos, en los cuales aparecen variables aleatorias y también otros conceptos de la Teoría de probabilidades. Tales modelos matemáticos se denominan *modelos estocásticos*. Las distribuciones de probabilidad de las variables aleatorias que se presentan en la descripción de un proceso real mediante un modelo estocástico, son, con frecuencia, parcial o totalmente desconocidas. Esta es la situación de partida de la Estadística matemática. Sobre la base de observaciones, experimentos y mediciones debe ajustarse el modelo estocástico lo mejor posible al proceso real.

Por ejemplo, en el caso más sencillo se trata de estimar de forma adecuada, sobre la base de los valores observados de una variable aleatoria, parámetros especiales desconocidos de la distribución de probabilidad, por lo demás conocida, de dicha variable aleatoria, digamos, los parámetros  $\mu$  o  $\sigma^2$  de una distribución normal. Otra tarea de la Estadística matemática consiste en someter a prueba, sobre la base de las realizaciones de la variable aleatoria considerada, si nuestra suposición acerca de la distribución de probabilidad que esta posee, digamos, una distribución normal, es correcta en el marco del modelo estocástico.

Estos son ejemplos típicos para dos clases de problemas principales de la Estadística matemática, con los cuales nos ocuparemos en los capítulos 10 y 11.

En el capítulo 10 se exponen los elementos esenciales de la *Teoría de la estimación*, cuya problemática de orden práctico consiste en indicar de forma apropiada valores estimados para parámetros desconocidos de un modelo estocástico. Por parámetros desconocidos debemos entender probabilidades de sucesos aleatorios particulares, características numéricas especiales de una distribución de probabilidad (por ejemplo, el valor esperado, la varianza, el coeficiente de correlación, etc.) y también funciones de distribución. En general, la Teoría de la estimación tiene como propósito indicar valores estimados para tales parámetros desconocidos (lo cual incluye métodos para la construcción de estimadores), el estudio de estimadores con respecto a sus propiedades especiales y, sobre este basamento, la comparación de diversos estimadores para un mismo parámetro. Partiendo de los datos numéricos concretos, las llamadas muestras (ver 9.2), se obtienen valores estimados concretos utilizando los denominados estadígrafos (ver 9.4); luego, estos valores estimados dependen de influjos casuales. En la construcción de estimadores se toma como base frecuentemente, y esto de forma evidente, el principio de utilizar como valores estimados para los parámetros desconocidos, aquellos que atribuyen la mayor probabilidad a los datos concretos de partida (método de máxima verosimilitud, ver 10.3).

En el capítulo 11 se brinda una panorámica de la *Teoría de la docimacia de hipótesis*, cuya problemática de orden práctico consiste en someter a prueba, sobre la base de los datos concretos obtenidos, suposiciones especiales en el marco de un modelo estocástico, que se denominan hipótesis. Semejantes hipótesis pueden referirse a la probabilidad de un suceso aleatorio especial, a parámetros de una distribución de probabilidad, pero también, a la función de distribución de una variable aleatoria. La comprobación de una hipótesis de este tipo mediante una denominada dócima de hipótesis consiste, hablando sin mucha precisión, en averiguar si las magnitudes que se pueden calcular a partir de los datos y que son factibles de comparar con la hipótesis, se diferencian o no sustancialmente de las fijadas por la hipótesis. Las diferencias entre unas y otras magnitudes existirán siempre a causa de las influencias casuales al seleccionar la muestra concreta; por eso una dócima de hipótesis tiene la tarea de comprobar si las diferencias detectadas pueden aclararse mediante estas influencias casuales o por el contrario, indican hacia una hipótesis falsa. Esto último conduciría entonces al rechazo de la hipótesis.

En este punto queremos aún llamar la atención hacia un hecho importante para cualquier aplicación de procedimientos estadísticos, que se refiere al *contenido de verdad de proposiciones estadísticas*. Sobre la base de un procedimiento estadístico, por ejemplo, de una dócima de hipótesis del tipo arriba indicado, no pueden hallarse proposiciones seguras. Otra cosa no es de esperar, ya que siempre se procesa solo un número finito de datos, mientras que las proposiciones que se refieren a una llamada población (ver 9.2) abarcan, por lo general, un conjunto más extenso. La ventaja de la aplicación de procedimientos estadísticos (por ejemplo, en la comprobación de una hipótesis) consiste en que la probabilidad de una decisión errónea (por ejemplo, del rechazo de una hipótesis verdadera) puede calcularse. Abordaremos este aspecto más exactamente en los capítulos 10 y 11.

En la aplicación de procedimientos estadísticos son interesantes los datos, no solo por sí mismos, sino por la forma y modo en que se obtienen. Es de gran importancia conocer, por ejemplo, si los datos se han obtenido mediante observaciones del valor de una variable aleatoria en repeticiones independientes de un experimento aleatorio o si estos experimentos dependían unos de otros. En el siguiente epígrafe nos ocuparemos con problemas fundamentales que se refieren a los métodos de *selección de una muestra*.

## 9.2 Población y muestra

El concepto muestra es de gran significación en los problemas estadísticos y está siempre unido con el concepto población. Queremos explicar estos conceptos con ayuda de ejemplos y más adelante definirlos matemáticamente.

### Ejemplos

1. En una fábrica se producen baterías para linternas. Supongamos que la producción diaria es tan grande, que no es económico comprobar si cada batería funciona correctamente. Sin embargo, para poder tener una impresión de la calidad de las baterías producidas, se extrae un cierto número de baterías, una llamada muestra, y se verifica su funcionamiento; la elección se realiza de modo que cada batería de la producción diaria tenga la misma oportunidad de ser extraída.
2. La efectividad de un medicamento para bajar la presión arterial (hipotensor) se debe investigar. Para ello se probará el medicamento en un número de pacientes que padecen de presión alta. Este conjunto constituye la muestra y el conjunto de todos los hombres que padecen de hipertensión (por ejemplo, en la región de venta del productor) sería la población correspondiente. Luego, una muestra es un subconjunto finito de un conjunto universo  $\Omega$ , que se denomina población en este contexto. Para lograr una conexión con las consideraciones teórico-probabilísticas, supongamos que  $\Omega$  es el conjunto universo de un espacio de probabilidad.

**Definición 1.** Sea  $[\Omega, A, P]$  un espacio de probabilidad. Entonces todo subconjunto no vacío finito  $A$  de  $\Omega$ ,  $A \subseteq \Omega$ , se llama *muestra* (de la población  $\Omega$ ). Si el conjunto  $A$  consta de  $n$  elementos, entonces  $A$  se llama una muestra de tamaño  $n$ , y  $n$  se denomina *tamaño de la muestra*.

En el primer ejemplo indicado,  $\Omega$  es el conjunto de las baterías producidas en un día,  $A$  el conjunto de todos los subconjuntos de  $\Omega$  y  $P(A)$  es igual a la probabilidad de que una batería extraída, de acuerdo con el procedimiento de selección, pertenezca al conjunto  $A \subseteq \Omega$ .

Ahora queremos clasificar los conceptos "muestra con reposición" y "muestra sin reposición", utilizados ya en los epígrafes 4.5 y 4.6. Para ello partiremos del espacio de probabilidad  $[\Omega, A, P]$ , donde  $\Omega$  es un conjunto finito (con  $N$  elementos  $\omega_1, \omega_2, \dots, \omega_N$ ),  $A$  denota al conjunto de todos los subconjuntos de  $\Omega$  y la medida de probabilidad  $P$  está dada por  $P(\{\omega_i\}) = \frac{1}{N}$  ( $i=1, 2, \dots, N$ ). (Una situación semejante se puede producir utilizando un recipiente, denominado comúnmente urna en el cálculo de probabilidades, que contiene  $N$  piezas geométricamente iguales, por ejemplo,  $N$  esferas iguales. Si después de agitar bien las piezas dentro del recipiente, se escoge ciegamente una, cada pieza tendrá igual probabilidad de ser extraída.) Si del conjunto  $\Omega$  tomamos consecutivamente  $n$  elementos, de modo que el elemento recién tomado se reponga antes de la próxima extracción, y que cada pieza tenga de nuevo la misma oportunidad de ser tomada, entonces obtenemos una llamada *muestra con reposición de tamaño n* de la población  $\Omega$ . Una muestra con reposición de tamaño  $n$ , se forma, por tanto, de  $n$  muestras de tamaño 1 (de acuerdo con la definición 1). Por consiguiente, en una muestra con reposición es posible que un mismo elemento  $\omega \in \Omega$  sea extraído varias veces; también el tamaño de la muestra  $n$  puede ser arbitrariamente grande. Si, por el contrario, en cada una de las extracciones no se reponen los elementos, entonces se habla de una *muestra sin reposición de tamaño n* de la pobla-

ción  $\Omega$ . Por tanto, una muestra sin reposición de tamaño  $n$  es una muestra de tamaño  $n$  en el sentido de la definición 1. Por consiguiente, en una muestra sin reposición cada elemento  $\omega \in \Omega$  puede ser extraído a lo sumo una vez, y para el tamaño de la muestra  $n$  se cumple que  $n \leq N$ .

Muchas selecciones de muestras que se hacen con fines económicos, en especial, en el marco del control estadístico de la calidad, y para otras investigaciones científicas, se basan en el modelo de una muestra sin reposición. El objetivo de esta selección consiste, con frecuencia, en obtener información sobre la parte de los elementos de una población que están caracterizados por una determinada propiedad  $P$  (por ejemplo, por una característica cualitativa particular). Para ello se puede describir una muestra de tamaño  $n$  mediante variables aleatorias  $X_1, X_2, \dots, X_n$  de la manera siguiente:

$$X_k = \begin{cases} 1, & \text{si el elemento tomado en la } k\text{-ésima extracción posee la propiedad "P"} \\ 0, & \text{si el elemento tomado en la } k\text{-ésima extracción no posee la propiedad "P"} \end{cases}$$

En una muestra con reposición, las variables aleatorias  $X_1, X_2, \dots, X_n$  son independientes y están distribuidas idénticamente. La variable aleatoria  $S = X_1 + X_2 + \dots + X_n$ , que indica el número (aleatorio) de los elementos con la propiedad "P" en la muestra, está distribuida binomialmente con los parámetros  $n$ =tamaño de la muestra y  $p$ =probabilidad de la propiedad "P" en la población. En una muestra sin reposición, las variables aleatorias  $X_1, X_2, \dots, X_n$  están también distribuidas idénticamente, pero no son independientes entre sí. La variable aleatoria  $S = X_1 + X_2 + \dots + X_n$  posee una distribución hipergeométrica. El resultado concreto de la selección de una muestra, igual si es con o sin reposición, puede describirse por una sucesión finita de los números cero y uno.

En nuestras consideraciones posteriores describiremos las muestras mediante variables aleatorias. Para ello sea  $[\Omega, \mathcal{A}, P]$  un espacio de probabilidad, y sea  $X$  una variable aleatoria sobre este espacio de probabilidad. Para obtener información sobre la distribución de probabilidad de la variable aleatoria  $X$ , por lo general desconocida, se repetirá  $n$  veces un experimento de forma independiente, observándose cada vez un valor concreto, es decir, una realización de la variable aleatoria. Con esto obtendremos los números  $x_1, x_2, \dots, x_n$ , que son realizaciones de la variable aleatoria  $X$ . Si concebimos el número  $x_k$ , o sea, la realización de la variable aleatoria  $X$  en el  $k$ -ésimo experimento, como realización de una variable aleatoria  $X_k$ , entonces las variables aleatorias  $X_1, X_2, \dots, X_n$  son independientes entre sí y están distribuidas idénticamente que  $X$ . Esto constituye el fundamento para la definición siguiente:

**Definición 2.** Sea  $X$  una variable aleatoria con la función de distribución  $F$ . Entonces el vector aleatorio  $(X_1, X_2, \dots, X_n)$ , cuyas componentes  $X_k$  son independientes y están distribuidas idénticamente que  $X$ , se llama una *muestra matemática de tamaño n de la población X con la función de distribución F*. Las variables aleatorias  $X_1, X_2, \dots, X_n$  se denominan en este contexto *variables de la muestra* y a una realización  $(x_1, x_2, \dots, x_n)$  del vector aleatorio  $(X_1, X_2, \dots, X_n)$  se le llama *muestra concreta de tamaño n de la población X con la función de distribución F*.

**Observación.** Anteriormente hemos dicho que por una población se debe entender el conjunto universo de un espacio de probabilidad. Este espacio de probabilidad está caracterizado, en este caso, por el conjunto de todos los  $n$ -uplos de números reales, es decir, por el conjunto  $\mathbb{R}^n$  y por la distribución de probabilidad del vector aleatorio  $(X_1, X_2, \dots, X_n)$ . La distribución de probabilidad del vector aleatorio  $(X_1, X_2, \dots, X_n)$  está

caracterizada por la función de distribución  $F_{(x_1, x_2, \dots, x_n)}$ , que está relacionada con la función de distribución de la variable aleatoria  $X$  (ver 6.4 (1)) según

$$F_{(x_1, x_2, \dots, x_n)}(x_1, x_2, \dots, x_n) = F(x_1) \cdot F(x_2) \cdots F(x_n).$$

Por tanto, la población es en cada caso el conjunto  $\mathbb{R}^n$ ; la función de distribución  $F$  de la variable aleatoria  $X$  contiene la información esencial para las consideraciones teórico-probabilísticas. Esto motiva las denominaciones introducidas en la definición 2.

### 9.3 Teorema fundamental de la Estadística matemática

El teorema fundamental de la Estadística matemática plantea que a través de muestras de tamaño suficientemente grande se puede describir aproximadamente y por tanto, reconocer, la función de distribución de la población considerada, por lo general desconocida. En principio, todos los procedimientos y métodos de la Estadística matemática se basan en este teorema; por esto el representa el eslabón principal entre la Teoría de probabilidades y la Estadística matemática, de donde se infiere también la denominación de este teorema como teorema fundamental de la Estadística matemática.

El punto de partida de nuestras reflexiones será una muestra concreta  $(x_1, x_2, \dots, x_n)$  de tamaño  $n$  de una población  $X$  con la función de distribución  $F$ . Para un número real  $x$  cualquiera dado averigüemos el número  $m_n(x)$  de los elementos de la muestra concreta que son menores que  $x$ , y consideremos para ello la magnitud  $w_n(x) = \frac{m_n(x)}{n}$ , que indica la frecuencia relativa de que los elementos de la muestra se encuentren en el intervalo de  $-\infty$  hasta  $x$ .

Definición 1. La función  $w_n$  definida sobre el eje real por

$$x \rightarrow w_n(x) = \frac{m_n(x)}{n} = \frac{\text{número de los elementos } x_1, x_2, \dots, x_n \text{ menores que } x}{n},$$

cuyos valores son números entre cero y uno, se denomina *función de distribución empírica de la muestra concreta*  $(x_1, x_2, \dots, x_n)$ .

La función de distribución empírica  $w_n$  de una muestra concreta  $(x_1, x_2, \dots, x_n)$  es una función escalonada, continua por la izquierda, que posee saltos en los lugares  $x_i$ ; la altura del salto es igual a  $\frac{1}{n}$ , en caso de que el valor  $x_i$  aparezca en la muestra exactamente una vez, en caso contrario, esta es igual a  $\frac{m_i}{n}$ , donde  $m_i$  denota el número de los elementos de la muestra que son iguales a  $x_i$ . Para  $x \leq \min_{1 \leq i \leq n} x_i$  se cumple que  $w_n(x) = 0$  y para  $x > \max_{1 \leq i \leq n} x_i$  se cumple que  $w_n(x) = 1$ . Estas propiedades muestran que  $w_n$  es una función de distribución (ver en 4.1 la observación después del teorema 1); esto justifica también la denominación introducida en la definición 1. Podemos reconocer en qué sentido esta función  $w_n$  es una aproximación de la función de distribución  $F$  de la población, si tenemos en cuenta la totalidad de todas las posibles muestras concretas, y con esto, la to-

talidad de todas las posibles funciones de distribución empíricas para un tamaño  $n$  fijo de las muestras de la población dada. Escojamos ahora, como punto de partida, una muestra matemática  $(X_1, X_2, \dots, X_n)$  de tamaño  $n$  de la población  $X$  con la función de distribución  $F$ . Para un número real  $x$  arbitrario designe  $M_n(x)$  el número de las variables de la muestra que son menores que  $x$ . Entonces  $M_n(x)$  es una variable aleatoria y la magnitud  $m_n(x)$ , definida anteriormente, puede concebirse como una realización de  $M_n(x)$ . De acuerdo con la forma de proceder seguida en el caso de una muestra concreta, consideremos ahora la variable aleatoria  $W_n(x) = \frac{M_n(x)}{n}$ .

**Definición 2.** La función  $W_n$  definida sobre el eje real por

$$x \rightarrow W_n(x) = \frac{M_n(x)}{n} = \frac{\text{número de las } X_1, X_2, \dots, X_n \text{ menores que } x}{n},$$

cuyos valores son variables aleatorias, se denomina *función de distribución empírica de la muestra matemática  $(X_1, X_2, \dots, X_n)$* .

Por tanto, para todo número  $x \in \mathbb{R}$ ,  $W_n(x)$  es una variable aleatoria; ella indica la frecuencia relativa (aleatoria) de que los elementos  $X_i$  de la muestra matemática  $(X_1, X_2, \dots, X_n)$  se encuentren situados en el intervalo de  $-\infty$  hasta  $x$ . La función  $W_n$ , que asocia a un número real arbitrario  $x$  la variable aleatoria  $W_n(x)$ , es un ejemplo para una denominada *función aleatoria*. El valor  $w_n(x)$  de la función de distribución empírica  $w_n$  de una muestra concreta  $(x_1, x_2, \dots, x_n)$  en el punto  $x$  debe entenderse como una realización de la variable aleatoria  $W_n(x)$ ; en este sentido la función  $w_n$  puede denominarse *realización de la función aleatoria  $W_n$* .

Queremos referirnos ahora a la estrecha relación entre la función de distribución empírica  $W_n$  de una muestra matemática  $(X_1, X_2, \dots, X_n)$  de tamaño  $n$  de una población  $X$  y la función de distribución  $F$  de esta población.

Podemos entender una muestra concreta  $(x_1, x_2, \dots, x_n)$  como resultado de una serie de  $n$  repeticiones independientes de un mismo experimento, consistente en la realización de la variable aleatoria  $X$ . Sea ahora  $x$  un número real arbitrario. El número de veces (concebido como variable aleatoria) de la ocurrencia del suceso  $(X < x)$  - luego, la variable aleatoria  $M_n(x)$  - está distribuida binomialmente con los parámetros  $p = P(X < x) = F(x)$  y  $n =$  tamaño de la muestra. Por consiguiente, se cumplen las relaciones (ver 4.5, teorema 2)

$$EM_n(x) = np = nF(x), \quad D^2M_n(x) = np(1-p) = nF(x)(1-F(x)),$$

de donde, con  $W_n(x) = \frac{M_n(x)}{n}$ , se obtienen las proposiciones

$$EW_n(x) = F(x) \quad (n \in \mathbb{N}) \tag{1}$$

y

$$D^2W_n(x) = \frac{F(x)(1-F(x))}{n} \rightarrow 0 \quad (n \rightarrow \infty). \tag{2}$$

Por tanto, el valor esperado del valor de la función de distribución empírica  $W_n$  de una muestra matemática  $(X_1, X_2, \dots, X_n)$  de tamaño  $n$  de la población  $X$ , en el punto  $x$ , es igual -independientemente del tamaño  $n$  de la muestra- al valor de la función de distribución  $F$  de esta población en el punto  $x$ , y la varianza de la variable aleatoria  $W_n(x)$  converge

hacia cero a medida que crece el tamaño  $n$  de la muestra ( $n \rightarrow \infty$ ). La relación entre la función de distribución empírica  $W_n$  de una muestra y la función de distribución  $F$  de la población considerada, se demuestra aún más claramente en el teorema siguiente, que constituye una forma debilitada del teorema fundamental de la Estadística matemática.

**Teorema 1.** Para todo número positivo  $\epsilon$  y todo número real  $x$  se cumple que

$$\lim_{n \rightarrow \infty} P(|W_n(x) - F(x)| < \epsilon) = 1, \quad (3)$$

O sea, para todo número real  $x$  la sucesión  $(W_n(x))$  converge estocásticamente hacia  $F(x)$ .

**Demuestra.** Sea  $x$  un número real arbitrario. Entonces  $W_n(x)$  es igual a la frecuencia relativa (aleatoria)  $f_n(A)$  del suceso  $A = \{X < x\}$  en una serie de  $n$  repeticiones independientes de un mismo experimento, consistente en la realización de la variable aleatoria  $X$  y  $A$  posee en cada ocasión la probabilidad  $p = P(A) = P(X < x) = F(x)$ . Sobre la base de la Ley de los grandes números de Bernoulli (ver 7.3, teorema 1) se cumple para todo número positivo  $\epsilon$  que

$$\lim_{n \rightarrow \infty} P(|f_n(A) - p| < \epsilon) = 1, \text{ o sea, } \lim_{n \rightarrow \infty} P(|W_n(x) - F(x)| < \epsilon) = 1,$$

lo que queríamos demostrar.

Ya que la Ley de los grandes números de Bernoulli puede considerarse también como ley fuerte de los grandes números (ver 7.4, Ley de los grandes números de Borel), la proposición del teorema 1 puede agudizarse de la forma siguiente:

$$P(\lim_{n \rightarrow \infty} W_n(x) = F(x)) = 1. \quad (4)$$

Esto significa que para todo número real  $x$ , la sucesión  $(W_n(x))$  converge casi seguro hacia  $F(x)$ . El contenido del teorema siguiente es un resultado esencialmente más fuerte, que se debe al matemático soviético V.I. Glivenko (1933).

**Teorema 2 (Teorema de Glivenko).** Se cumple que

$$P(\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |W_n(x) - F(x)| = 0) = 1. \quad (5)$$

No demostraremos este teorema, pero queremos aún aclarar algo. La proposición (4) muestra que se cumple  $P(\lim_{n \rightarrow \infty} |W_n(x) - F(x)| = 0) = 1$  para todo número real  $x$ , o sea, que para todo número real  $x$  la sucesión  $(D_n(x))$ ,  $D_n(x) = |W_n(x) - F(x)|$ , converge casi seguro hacia cero. La proposición (5) significa que esta convergencia es incluso uniforme (en  $x$ ), o sea, que la sucesión  $(D_n)$ ,

$$D_n = \sup_{-\infty < x < \infty} |W_n(x) - F(x)|$$

converge casi seguro hacia cero. La relación, expresada por medio de (5), entre la función de distribución empírica de una muestra matemática y la función de distribución de la población, se denomina teorema fundamental de la Estadística matemática.

Concluyendo este círculo de problemas indicamos sin demostración, una formulación cuantitativa del teorema fundamental de la Estadística matemática.

**Teorema 3 (Teorema de Kolmogorov).** Si la función de distribución  $F$  de la población es continua, entonces se cumple que

$$\lim_{n \rightarrow \infty} P\left(\sqrt{n} \sup_{-\infty < x < \infty} |W_n(x) - F(x)| < y\right) = K(y)$$

con

$$K(y) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k e^{-2ky^2} & \text{para } y > 0, \\ 0 & \text{para } y \leq 0. \end{cases} \quad (6)$$

Para la explicación de este teorema observemos que sobre la base del teorema de Glivenko la sucesión  $(D_n)$ ,  $D_n = \sup_{-\infty < x < +\infty} |W_n(x) - F(x)|$ , converge casi seguro hacia cero, luego, hacia una variable aleatoria distribuida puntualmente. El teorema de Kolmogorov muestra que la sucesión  $(\sqrt{n} D_n)$  converge en distribución hacia una variable aleatoria, cuya función de distribución es la función  $K$ . Notable es, en particular, que esta función de distribución límite  $K$  no depende de  $F$ , bajo la sola condición de que  $F$  sea continua. En esta proposición se basan díctimas de hipótesis para la distribución de una población; los valores necesarios de la función  $K$  pueden encontrarse en tablas de la Estadística matemática.

## 9.4 Estadígrafos

En la aplicación de procedimientos de la Estadística matemática se utilizan con frecuencia magnitudes, que se calculan a partir de una muestra concreta (por ejemplo, la media aritmética o la varianza empírica). Su cálculo se basa, en cada ocasión, sobre una función real  $\phi$  definida sobre un conjunto de  $n$ -úplos de números reales,

$$(x_1, \dots, x_n) (\in \mathbb{R}^n) \rightarrow \phi(x_1, \dots, x_n) (\in \mathbb{R}^1). \quad (1)$$

(Por ejemplo, en el caso de la media aritmética se trata de la función dada por  $\phi(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$ .

De forma general partiremos de una función  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^1$  y consideraremos una variable aleatoria  $X$  definida sobre el espacio de probabilidad  $[\Omega, \mathcal{A}, P]$  y una muestra matemática  $(X_1, \dots, X_n)$  de tamaño  $n$  de la población  $X$ . Entonces se define por

$$[\phi(X_1, \dots, X_n)](\omega) = \phi(X_1(\omega), \dots, X_n(\omega)) \quad (\omega \in \Omega) \quad (2)$$

una función real  $\phi(X_1, \dots, X_n)$  sobre el conjunto  $\Omega$ , que en este contexto se denomina *estadígrafo*, y que consideraremos siempre como una variable aleatoria (sobre  $[\Omega, \mathcal{A}, P]$ ).

A continuación damos algunos ejemplos de estadígrafos que desempeñarán también un papel en las explicaciones posteriores; aquí introduciremos algunas abreviaturas que se utilizarán en lo que sigue.

### Ejemplos

$$1. \phi(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n x_i =: \bar{X}_n.$$

$$2. \phi(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 =: S_n^{*2} \quad (\mu \in \mathbb{R}^1 \text{ fijo}).$$

$$3. \phi(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 =: S_n^2.$$

$$4. \phi(X_1, \dots, X_n) = \max \{X_1, \dots, X_n\}.$$

$$5. \phi(X_1, \dots, X_n) = \min \{X_1, \dots, X_n\}.$$

El conocimiento de la distribución de probabilidad de estadígrafos especiales es de decisiva importancia en la realización de muchos procedimientos de la Estadística matemática; aquí nos interesan tanto las proposiciones acerca de la distribución de un estadígrafo  $\phi(X_1, \dots, X_n)$  para un  $n$  fijo, como aquellas sobre su comportamiento asintótico (o sea,

para  $n \rightarrow \infty$ ). Estos problemas constituyen un interés central de la Estadística matemática. Del gran número de proposiciones que existen la respecto, solo formularemos algunas pocas, y preferentemente aquellas que necesitaremos en el tratamiento de la teoría de la estimación y de la docimasia de hipótesis (capítulos 10 y 11).

**Teorema 1.** Sea  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$  con la función de distribución  $F$ . Para las funciones de distribución  $G$  y  $H$  de los estadígrafos  $\max\{X_1, \dots, X_n\}$  y  $\min\{X_1, \dots, X_n\}$  respectivamente, se cumple que

$$G(x) = [F(x)]^n \quad (-\infty < x < \infty) \quad (3)$$

y

$$H(x) = 1 - [1 - F(x)]^n \quad (-\infty < x < \infty). \quad (4)$$

**Demostración.** Como las variables de la muestra  $X_1, \dots, X_n$  son independientes y están distribuidas idénticamente que  $X$ , se cumple para todo  $x \in \mathbb{R}$  que

$$\begin{aligned} G(x) &= P(\max\{X_1, \dots, X_n\} < x) \\ &= P(X_1 < x, \dots, X_n < x) \\ &= P(X_1 < x) \cdot \dots \cdot P(X_n < x) \\ &= F_{X_1}(x) \cdot \dots \cdot F_{X_n}(x) = [F(x)]^n \end{aligned}$$

y

$$\begin{aligned} H(x) &= P(\min\{X_1, \dots, X_n\} < x) \\ &= 1 - P(\min\{X_1, \dots, X_n\} \geq x) \\ &= 1 - P(X_1 \geq x, \dots, X_n \geq x) \\ &= 1 - P(X_1 \geq x) \cdot \dots \cdot P(X_n \geq x) \\ &= 1 - (1 - F_{X_1}(x)) \cdots (1 - F_{X_n}(x)) \\ &= 1 - [1 - F(x)]^n \end{aligned}$$

Para los teoremas siguientes (2, 3, 4 y 5) se cumple que  $(X_1, \dots, X_n)$  es una muestra matemática de tamaño  $n$  de una población que posee una distribución  $N(\mu, \sigma^2)$ .

**Teorema 2.** El estadígrafo  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  posee una distribución  $N\left(\mu, \frac{\sigma^2}{n}\right)$ .

**Demostración.** Como la suma de variables aleatorias independientes que poseen una distribución normal posee, a su vez, una distribución de dicho tipo (ver en 6.5 observación posterior al teorema 4),  $\sum_{i=1}^n X_i$  posee una distribución  $N(n\mu, n\sigma^2)$  y, por consiguiente,  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  es una variable aleatoria con una distribución  $N\left(\mu, \frac{\sigma^2}{n}\right)$ .

#### Observaciones

- Del teorema 2 resulta directamente que  $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$  es una variable aleatoria con una distribución  $N(0, 1)$ .
- Supongamos acerca de la población  $X$  considerada, que se cumple  $0 < D^2 X < \infty$ . Entonces la sucesión  $\left(\sqrt{n} \frac{\bar{X}_n - EX}{\sqrt{D^2 X}}\right)$  converge en distribución hacia una variable aleatoria

que posee una distribución  $N(0,1)$  (ver 7.6, teorema 1). Luego  $\bar{X}_n$  posee para  $n$  grande aproximadamente una distribución  $N\left(EX, \frac{D^2X}{n}\right)$ .

**Teorema 3.** El estadígrafo  $\frac{n S_n^{*2}}{\sigma^2}$  con  $S_n^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$  posee una distribución  $\chi^2$  con  $n$  grados de libertad.

**Demostración.** Las variables aleatorias  $Y_i = \frac{X_i - \mu}{\sigma}$  ( $i = 1, \dots, n$ ) son independientes y poseen una distribución  $N(0, 1)$ . Luego, según el corolario 1 (6.5)

$$\sum_{i=1}^n Y_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{\sigma^2}{n} \cdot \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 = \frac{n S_n^2}{\sigma^2}$$

posee una distribución  $\chi^2$  con  $n$  grados de libertad.

**Teorema 4.** El estadígrafo  $\frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2$  posee una distribución  $\chi^2$  con  $n-1$  grados de libertad.

Renunciaremos a la demostración de este teorema algo difícil.

**Teorema 5.** El estadígrafo  $\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}}$  posee una distribución  $t$  con  $n-1$  grados de libertad.

La proposición de este teorema se obtiene de los enunciados de los teoremas 2 y 4, de que  $\bar{X}_n$  y  $S_n^2$  son estocásticamente independientes y por último, de la proposición del teorema 7 (6.5).

**Teorema 6.** Sean  $(X_1, \dots, X_m)$  y  $(Y_1, \dots, Y_n)$  dos muestras matemáticas de tamaño  $m$  de una población  $X$  con una distribución  $N(\mu_1, \sigma^2)$  y de tamaño  $N$  de una población  $Y$  con una distribución  $N(\mu_2, \sigma^2)$  respectivamente. Además, sean  $X$  y  $Y$  estocásticamente independientes. Entonces el estadígrafo  $\frac{S_{x,m}^2}{S_{y,n}^2}$ , con

$$S_{x,m}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_n)^2 \text{ y } S_{y,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2,$$

posee una distribución  $F$  con  $(m-1, n-1)$  grados de libertad.

La proposición de este teorema se basa esencialmente en la proposición del teorema 4. De acuerdo con ella,  $\frac{(m-1)S_{x,m}^2}{\sigma^2}$  y  $\frac{(n-1)S_{y,n}^2}{\sigma^2}$  poseen una distribución  $\chi^2$  con  $m-1$  y  $n-1$  grados de libertad respectivamente. Como  $X$  y  $Y$  son independientes, esto se cumple también para  $\frac{(m-1)S_{x,m}^2}{\sigma^2}$  y  $\frac{(n-1)S_{y,n}^2}{\sigma^2}$ . La proposición del teorema 6 se obtiene por último del teorema 8 (6.5).

Daremos algunas otras proposiciones sobre distribuciones de estadígrafos, sin demostración, en los lugares donde las utilicemos.

## 10. Introducción a la Teoría de la estimación

El capítulo siguiente contiene una introducción a la *Teoría de la estimación*, una de las ramas más importantes de la Estadística matemática. Aquí trataremos las llamadas *estimaciones puntuales* (10.2 hasta 10.4), utilizadas con frecuencia en las aplicaciones, y las denominadas *estimaciones por intervalo* (10.5 y 10.6). En relación con un grupo de datos concreto, las estimaciones puntuales conducen a valores aproximados de un parámetro desconocido, mientras que las estimaciones por intervalo conducen a intervalos donde es posible que se encuentre el valor de un parámetro desconocido.

En los epígrafes siguientes se introducen de forma matemáticamente exacta, y se motivan al mismo tiempo, los conceptos básicos estimación puntual y estimación por intervalo, se exponen los métodos y procedimientos generales y se muestra su fundamento teórico-probabilístico y naturalmente, para algunos problemas de estimación frecuente, se indican estimaciones apropiadas, tanto puntuales como por intervalo. La aplicación práctica de estas estimaciones conduce, en lo esencial, a un cálculo de medidas estadísticas y no da lugar a otras complicaciones, de modo que renunciaremos a los ejemplos numéricos.

### 10.1 Tareas que se plantea la Teoría de la estimación

El problema principal de la Teoría de la estimación consiste en indicar métodos para averiguar valores estimados de parámetros desconocidos de un modelo estocástico, sobre la base de muestras.

Nos queremos limitar, en lo esencial, al caso de la estimación de un parámetro desconocido. Este parámetro lo designaremos con  $\gamma$ , al valor verdadero (pero desconocido) del mismo lo denotaremos con  $\gamma_0$ , y al conjunto de sus posibles valores en el marco del problema considerado en cada ocasión, lo designaremos por el símbolo  $\Gamma$ , donde suponemos que  $\Gamma$  es un intervalo sobre el eje real.

Para la formulación matemática del problema fundamental de la Teoría de la estimación partiremos de una población  $X$ , cuya función de distribución  $F$  depende de un pa-

rámetro  $\gamma \in \Gamma$  y consideraremos para ello una muestra matemática  $(X_1, \dots, X_n)$  de tamaño  $n$  de la población  $X$ . La Teoría de la estimación tiene, pues, la tarea de hallar estadígrafos adecuados  $\varphi(x_1, \dots, x_n)$  para la estimación de  $\gamma$  y de investigarlos con respecto a la dependencia de sus correspondientes distribuciones de probabilidad del parámetro  $\gamma$ . Luego, si  $(x_1, \dots, x_n)$  es una muestra concreta de tamaño  $n$  de la población  $X$ , entonces el número  $\varphi(x_1, \dots, x_n)$ , que se concibe como una realización de la variable aleatoria  $\varphi(X_1, \dots, X_n)$ , puede utilizarse como *valor estimado* para  $\gamma_0$ ; el estadígrafo tomado por base  $\varphi(X_1, \dots, X_n)$  se denomina en este contexto un *estimador* (para  $\gamma$ ). Por tanto, un estimador es una variable aleatoria, cuyos valores pertenecen al conjunto  $\Gamma$  de los posibles valores del parámetro; un valor estimado es un número real ( $\in \Gamma$ ).

Para diferenciar las estimaciones que en el caso particular proporcionan números (puntos sobre el eje real), de las llamadas estimaciones por intervalo, que se introducirán más tarde, denominaremos a las primeras *estimaciones puntuales*. Naturalmente, como estimadores puntuales se aspira utilizar estadígrafos que proporcionen una aproximación lo "mejor" posible del parámetro a estimar, sobre la base de sus propiedades teórico probabilísticas.

**Ejemplo.** Supongamos que la población  $X$  posee una distribución normal con la varianza  $D^2 X = \sigma_0^2$  ( $\sigma_0$  conocida, por ejemplo,  $\sigma_0 = 1$ ), y que el valor esperado  $EX$  es desconocido. Por tanto, hacemos  $\gamma = EX$  y  $\Gamma = \mathbb{R}^1$ . Si  $(X_1, \dots, X_n)$  es una muestra matemática de tamaño  $n$  de esta población, entonces el estadígrafo

$$\varphi(X_1, \dots, X_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i =: \hat{\gamma}_n$$

posee el valor esperado  $\gamma(E\hat{\gamma}_n = \gamma)$ , y se cumple que  $D^2 \hat{\gamma}_n = \frac{\sigma_0^2}{n}$ . Sobre la base de la desigualdad de Chebyshev (ver 7.1, corolario 1) se cumple para todo  $\varepsilon > 0$  la relación

$$P(|\hat{\gamma}_n - \gamma| \geq \varepsilon) \leq \frac{\sigma_0^2}{n\varepsilon^2},$$

o sea,  $\lim_{n \rightarrow \infty} P(|\hat{\gamma}_n - \gamma| < \varepsilon) = 1$ .

La sucesión  $(\hat{\gamma}_n)$ ,  $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , converge estocásticamente hacia  $\gamma$ . (Estas proposiciones se cumplen para todo  $\gamma \in \Gamma = \mathbb{R}^1$ , en particular, para el valor "verdadero"  $\gamma_0$ .) Para un tamaño  $n$  de la muestra suficientemente grande se puede esperar que la media aritmética  $\bar{x}_n$  de los elementos de una muestra concreta  $(x_1, \dots, x_n)$  represente un valor estimado pasable para el parámetro desconocido. (Por lo demás, en las reflexiones anteriores no hemos tomado en consideración que la población  $X$  posee una distribución normal; es suficiente saber que la población  $X$  posee una varianza (finita) para todo valor del parámetro.)

Como muestra el ejemplo dado, en la valoración de un estadígrafo como estimador para un parámetro desconocido, es de gran significación el comportamiento asintótico, esto es, el comportamiento para  $n \rightarrow \infty$ . En la aplicación práctica, las proposiciones sobre el comportamiento asintótico son de utilidad solo cuando el tamaño  $n$  de la muestra en cuestión es grande; en realidad, no se puede indicar exactamente qué se debe entender por un tamaño "grande" de la muestra, lo cual depende también estrechamente del problema considerado. Además, se debe llamar la atención de que en vinculación con una estimación

puntual para un parámetro desconocido no se obtienen automáticamente proposiciones acerca de la exactitud de los valores estimados (si, por ejemplo, el estadígrafo utilizado como estimador es una variable aleatoria continua, entonces la probabilidad de que la estimación proporcione el valor verdadero del parámetro es igual a cero). Esto no significa que cuando se halla un valor estimado este no pueda estar situado muy cerca del valor verdadero del parámetro, lo cual es de esperar incluso en el caso en que  $n > 1$ ). Ahora, si se desean proposiciones sobre la exactitud o si el tamaño  $n$  de la muestra es pequeña, planteamos la tarea de construir, sobre la base de una muestra matemática  $(X_1, \dots, X_n)$ , un intervalo  $J(X_1, \dots, X_n)$  que contenga al parámetro desconocido con una probabilidad determinada de antemano (por lo general, cercana a uno). Los puntos extremos de este intervalo dependen de las variables de la muestra  $X_1, \dots, X_n$ , por tanto, son ellos mismos variables aleatorias. Un intervalo  $J(X_1, \dots, X_n)$  aleatorio en este sentido, se denomina *estimador por intervalo de confianza* o *intervalo de confianza*. Para una muestra concreta  $(x_1, \dots, x_n)$  se obtiene, sobre la base de un intervalo de confianza  $J(X_1, \dots, X_n)$ , un intervalo  $J(x_1, \dots, x_n) \subseteq \Gamma$ , denominado *intervalo estimado concreto* para el parámetro desconocido. Las estimaciones por intervalo deben, por una parte, proporcionar intervalos estimados concretos lo más "pequeños" posibles y por otra, deben contener al parámetro desconocido con una probabilidad lo más cercana a uno.

En los epígrafes 10.5 y 10.6 nos ocuparemos, detalladamente, de los estimadores por intervalo; los epígrafes que siguen están dedicados a los estimadores puntuales.

## 10.2 Estimadores puntuales (propiedades)

Como se dijo en el epígrafe 10.1, entenderemos por un estimador puntual -brevemente: estimador-  $\hat{\gamma}_n$  para un parámetro desconocido  $\gamma$ , un estadígrafo  $\phi(X_1, \dots, X_n)$ , cuyos valores pertenecen al conjunto  $\Gamma$  de los posibles valores del parámetro. En este epígrafe definiremos las propiedades de los estimadores puntuales, sobre cuya base podremos hacer una valoración y comparación de estimaciones, en relación con un mismo problema de estimación. Para ello partiremos siempre de la situación bosquejada en el epígrafe 10.1 (Población  $X$ , distribución de probabilidad dependiente de un parámetro  $\gamma \in \Gamma \subseteq \mathbb{R}^1$ ,  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$ ).

**Definición 1.** Un estimador  $\hat{\gamma}_n$  se denomina *estimador insesgado* para  $\gamma$ , si el valor esperado de  $\hat{\gamma}_n$ -calculado bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro- es igual a  $\gamma$  para todo  $\gamma \in \Gamma$ . Para esto escribimos brevemente

$$E_{\gamma} \hat{\gamma}_n = \gamma \quad (\gamma \in \Gamma) \quad (1)$$

La validez de (1) se exige para todo  $\gamma \in \Gamma$ ; con esto se cumple (1) en particular para  $\gamma_0$ , el valor verdadero del parámetro.

**Ejemplo 1.** Supongamos que  $X$  posee una distribución uniforme sobre el intervalo  $[0, b]$ ,  $b > 0$  y que  $b$  sea desconocido. Hagamos  $\gamma = b$  y  $\Gamma = \{\gamma : \gamma > 0\}$ . Además, sea  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$ . Para el estadígrafo

$$\phi(X_1, \dots, X_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

se cumple (ver 5.3 (3)) que

$$E_\gamma \phi(X_1, \dots, X_n) = \frac{1}{n} \cdot n \cdot \frac{\gamma}{2} = \frac{\gamma}{2};$$

para el estimador  $\hat{\gamma}_n = 2\phi(X_1, \dots, X_n) = 2\bar{X}_n$  se obtiene de aquí que

$$E_\gamma \hat{\gamma}_n = 2 \cdot \frac{\gamma}{2} = \gamma \quad (\gamma > 0),$$

o sea,  $\hat{\gamma}_n$  es un estimador insesgado para  $\gamma$ .

En relación con los estimadores sesgados se utiliza el concepto sesgo (error sistemático) que caracterizaremos en la definición siguiente.

**Definición 2** Sea  $\hat{\gamma}_n$  un estimador para  $\gamma$ . Se denomina *sesgo (error sistemático)*<sup>a</sup> de  $\hat{\gamma}_n$  con respecto a  $\gamma$  a

$$b_n(\gamma) = E_\gamma \hat{\gamma}_n - \gamma \quad (\gamma \in \Gamma). \quad (2)$$

Por tanto, para los estimadores insesgados  $\hat{\gamma}_n$  de  $\gamma$  se cumple que  $b_n(\gamma) = 0$  para todo  $\gamma \in \Gamma$ . La variable aleatoria  $\hat{\gamma}_n - E_\gamma \hat{\gamma}_n$  se llama *error aleatorio* de  $\hat{\gamma}_n$  y la variable aleatoria  $\hat{\gamma}_n - \gamma = (\hat{\gamma}_n - E_\gamma \hat{\gamma}_n) + (E_\gamma \hat{\gamma}_n - \gamma)$ , que se obtiene de la suma del sesgo de  $\hat{\gamma}_n$  con respecto a  $\gamma$  y el error aleatorio de  $\hat{\gamma}_n$ , indica la desviación aleatoria del estimador  $\hat{\gamma}_n$  de  $\gamma$ .

**Ejemplo 2.** Consideremos la situación bosquejada en el ejemplo 1 e investiguemos el estadígrafo

$$\hat{\gamma}_n = \max\{X_1, \dots, X_n\}.$$

Para el cálculo de  $E_\gamma \hat{\gamma}_n$  necesitamos la función de distribución o la densidad de  $\hat{\gamma}_n$ , que queremos denotar con  $G_\gamma$  y  $g_\gamma$ , respectivamente, suponiendo que  $\gamma$  es el valor verdadero del parámetro. Se cumple (ver 9.4, teorema 1) que  $G_\gamma(x) = [F_\gamma(x)]^n$ , donde con  $F_\gamma$  denotamos la función de distribución de la población  $X$ , suponiendo que  $\gamma$  es el valor verdadero del parámetro. Con

$$F_\gamma(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ \frac{x}{\gamma} & \text{para } 0 \leq x \leq \gamma, \\ 1 & \text{para } x \geq \gamma. \end{cases}$$

obtenemos

$$G_\gamma(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ \left(\frac{x}{\gamma}\right)^n & \text{para } 0 \leq x \leq \gamma, \\ 1 & \text{para } x \geq \gamma. \end{cases}$$

y con esto,

$$g_\gamma(x) = \begin{cases} 0 & \text{para } x < 0 \text{ y para } x > \gamma, \\ n \frac{x^{n-1}}{\gamma^n} & \text{para } 0 < x < \gamma. \end{cases}$$

Para  $E_{\gamma} \hat{\gamma}_n$  se obtiene entonces que

$$E_{\gamma} \hat{\gamma}_n = \int_{-\infty}^{\infty} x g_{\gamma}(x) dx = \int_0^n n \cdot \frac{x^n}{\gamma^n} dx = \frac{n}{n+1} \gamma$$

y para el sesgo  $b_n(\gamma)$  de  $\hat{\gamma}_n$  con respecto a  $\gamma$ , tenemos que

$$b_n(\gamma) = E_{\gamma} \hat{\gamma}_n - \gamma = \frac{n}{n+1} \gamma - \gamma = -\frac{\gamma}{n+1} \quad (\gamma > 0).$$

Observemos que  $\lim_{n \rightarrow \infty} b_n(\gamma) = 0$  y, por tanto, se cumple que  $\lim_{n \rightarrow \infty} E_{\gamma} \hat{\gamma}_n = \gamma$  para todo  $\gamma$ .

La definición siguiente sirve para caracterizar, de forma general, el comportamiento analizado al final del ejemplo 2.

**Definición 3.** Una sucesión  $(\hat{\gamma}_n)$  de estimadores  $\hat{\gamma}_n$  para  $\gamma$ , se denomina *asintóticamente insesgado*, si se cumple que

$$\lim_{n \rightarrow \infty} E_{\gamma} \hat{\gamma}_n = \gamma \quad (\gamma \in \Gamma). \quad (3)$$

(En caso de que se cumpla (3) para un estimador  $\hat{\gamma}_n$ , se dice también que  $\hat{\gamma}_n$  es asintóticamente insesgado.)

Por lo general, utilizaremos estimadores insesgados, o al menos, asintóticamente insesgados. Como el hecho de que un estimador sea insesgado nada dice acerca de si la distribución de probabilidad del mismo está concentrada o no alrededor del parámetro desconocido, ni del modo en que lo hace, se preferirán especialmente aquellos estimadores que cuando  $n \rightarrow \infty$  se concentren alrededor del parámetro desconocido. Desde el punto de vista matemático expresaremos esta "concentración" por medio de los tipos de convergencia de la Teoría de probabilidades (ver 7.2), en las definiciones siguientes.

**Definición 4.** Una sucesión  $(\hat{\gamma}_n)$  de estimadores para  $\gamma$  se denomina (*débilmente*) *consistente*, si para todo número positivo  $\epsilon$  se cumple que

$$\lim_{n \rightarrow \infty} P_{\gamma}(|\hat{\gamma}_n - \gamma| \geq \epsilon) = 0 \quad (\gamma \in \Gamma); \quad (4)$$

aquí es  $P_{\gamma}(|\hat{\gamma}_n - \gamma| \geq \epsilon)$  la probabilidad del suceso  $(|\hat{\gamma}_n - \gamma| \geq \epsilon)$ , calculada bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro. (En caso de que se cumpla (4) para un estimador  $\hat{\gamma}_n$ , se dice también que  $\hat{\gamma}_n$  es (*débilmente*) consistente.)

Por consiguiente, la consistencia de una sucesión de estimadores significa que existe una convergencia en probabilidad. Las condiciones suficientes para la consistencia, mencionadas en el teorema siguiente, se pueden verificar con frecuencia más fácilmente que (4).

**Teorema 1.** Las condiciones siguientes son, ambas juntas, suficientes para la consistencia (*débil*) de una sucesión  $(\hat{\gamma}_n)$  de estimadores  $\hat{\gamma}_n$  para  $\gamma$ .

1.  $\lim_{n \rightarrow \infty} E_{\gamma} \hat{\gamma}_n = \gamma \quad (\gamma \in \Gamma)$ , es decir, la sucesión  $(\hat{\gamma}_n)$  es asintóticamente insesgada.
2.  $\lim_{n \rightarrow \infty} D_{\gamma} \hat{\gamma}_n = 0 \quad (\gamma \in \Gamma)$ ; aquí  $D_{\gamma} \hat{\gamma}_n$  significa la varianza de  $\hat{\gamma}_n$  calculada bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro.

**Demostración.** Sobre la base del teorema 1(7.1) se cumple para un  $\epsilon$  positivo arbitrario

$$P_{\gamma}(|\hat{\gamma}_n - \gamma| \geq \epsilon) \leq \frac{E_{\gamma} (\hat{\gamma}_n - \gamma)^2}{\epsilon^2}.$$

Ahora, se cumple que

$$\begin{aligned} E_{\gamma}(\hat{\gamma}_n - \gamma)^2 &= E_{\gamma}(\hat{\gamma}_n - E_{\gamma}\hat{\gamma}_n - E_{\gamma}\hat{\gamma}_n - \gamma)^2 \\ &= E_{\gamma}[(\hat{\gamma}_n - E_{\gamma}\hat{\gamma}_n)^2 + 2(\hat{\gamma}_n - E_{\gamma}\hat{\gamma}_n)(E_{\gamma}\hat{\gamma}_n - \gamma) + (E_{\gamma}\hat{\gamma}_n - \gamma)^2] \\ &= E_{\gamma}(\hat{\gamma}_n - E_{\gamma}\hat{\gamma}_n)^2 + 0 + (E_{\gamma}\hat{\gamma}_n - \gamma)^2 \\ &= D_{\gamma}\hat{\gamma}_n + (E_{\gamma}\hat{\gamma}_n - \gamma)^2. \end{aligned}$$

Si las condiciones nombradas en el teorema se satisfacen, entonces resulta de aquí directamente que  $\lim_{n \rightarrow \infty} E_{\gamma}(\hat{\gamma}_n - \gamma)^2 = 0$  y con esto  $\lim_{n \rightarrow \infty} P_{\gamma}(|\hat{\gamma}_n - \gamma| \geq \varepsilon) = 0$ .

**Ejemplo 3.** Consideremos el estimador  $\hat{\gamma}_n = 2\bar{X}_n$ , investigado en el ejemplo 1. Se cumplen las relaciones  $E_{\gamma}\hat{\gamma}_n = \gamma$  y  $D_{\gamma}\hat{\gamma}_n = \frac{4}{n^2} \cdot n \cdot \frac{\gamma^2}{12} = \frac{\gamma^2}{3n}$  (ver 5.3(4)). Según el teorema 1 la sucesión  $(\hat{\gamma}_n)$  es débilmente consistente.

**Ejemplo 4.** Consideremos el estimador  $\hat{\gamma}_n = \max\{X_1, \dots, X_n\}$ , investigado ya en el ejemplo 2. Como fue averiguado allí, se cumple que  $\lim_{n \rightarrow \infty} E_{\gamma}\hat{\gamma}_n = \lim_{n \rightarrow \infty} \frac{n}{n+1} \gamma = \gamma$ . Para  $D_{\gamma}\hat{\gamma}_n$  obtenemos que

$$\begin{aligned} D_{\gamma}\hat{\gamma}_n &= E_{\gamma}\hat{\gamma}_n^2 - (E_{\gamma}\hat{\gamma}_n)^2 = \int_{-\infty}^{\infty} x^2 g_{\gamma}(x) dx - (E_{\gamma}\hat{\gamma}_n)^2 \\ &= \int_0^{\gamma} \frac{x^{n+1}}{\gamma^n} dx - \left( \frac{n}{n+1} \gamma \right)^2 = \frac{n}{n+2} \gamma^2 - \left( \frac{n}{n+1} \gamma \right)^2 \\ &= \frac{n}{(n+1)^2(n+2)} \gamma^2. \end{aligned}$$

Luego, para la sucesión  $(\hat{\gamma}_n)$ ,  $\hat{\gamma}_n = \max\{X_1, \dots, X_n\}$  se satisfacen las condiciones nombradas en el teorema 1, y con esto la sucesión  $(\hat{\gamma}_n)$  es también consistente.

**Definición 5.** Una sucesión  $(\hat{\gamma}_n)$  de estimadores  $\hat{\gamma}_n$  para  $\gamma$  se denomina *fueramente consistente* si se cumple que

$$P_{\gamma}(\lim_{n \rightarrow \infty} \hat{\gamma}_n = \gamma) = 1 \quad (\gamma \in \Gamma). \quad (5)$$

Por consiguiente, la consistencia fuerte de una sucesión de estimadores significa que existe una convergencia con probabilidad uno.

Si para una población  $X$  existe el valor esperado  $EX$ , entonces la sucesión  $(\hat{\gamma}_n)$ ,

$$\hat{\gamma}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

es una sucesión de estimadores fuertemente consistente para  $\gamma = EX$ , sobre la base de la Ley de los grandes números de Kolmogorov (ver 7.4, teorema 6).

Con las definiciones siguientes tendremos distintas posibilidades para comparar diversos estimadores insesgados, por medio de sus varianzas en relación con un mismo problema de estimación. Para ello designe  $\hat{\Gamma}_{\gamma}$  el conjunto de todos los estimadores insesgados para  $\gamma$ , sobre la base de una muestra matemática de tamaño  $n$  con varianza positiva finita; por tanto para  $\hat{\gamma}_n \in \hat{\Gamma}_{\gamma}$  se cumple que  $E_{\gamma}\hat{\gamma}_n = \gamma$  y que  $0 < D_{\gamma}\hat{\gamma}_n < \infty$  para todo  $\gamma \in \Gamma$ .

**Definición 6.** Un estimador  $\hat{\gamma}_n \in \hat{\Gamma}_n$  se llama *mejor* que un estimador  $\tilde{\gamma} \in \hat{\Gamma}_n$ , si se cumple que

$$D_{\gamma}^2 \hat{\gamma}_n \leq D_{\gamma}^2 \tilde{\gamma}_n \quad (\gamma \in \Gamma). \quad (6)$$

La razón  $D_{\gamma}^2 \hat{\gamma}_n : D_{\gamma}^2 \tilde{\gamma}_n$  indica el *grado* en que  $\hat{\gamma}_n$  es mejor que  $\tilde{\gamma}_n$ .

**Ejemplo 5.** Consideremos de nuevo la situación ilustrada en el ejemplo 1 y comparemos los estimadores

$$\hat{\gamma}_n = 2\bar{X}_n = \frac{2}{n} \sum_{i=1}^n X_i \text{ y } \hat{\gamma}_n = \frac{n+1}{n} \max \{X_1, \dots, X_n\}.$$

Se cumplen las proposiciones

$$E_{\gamma} \hat{\gamma}_n = \gamma, \quad D_{\gamma}^2 \hat{\gamma}_n = \frac{\gamma^2}{3n}$$

(ver ejemplos 1 y 3).

$$E_{\gamma} \hat{\gamma}_n = \gamma, \quad D_{\gamma}^2 \hat{\gamma}_n = \frac{\gamma^2}{n(n+2)}$$

(ver ejemplos 2 y 4).

Luego, ambos estimadores son insesgados y poseen una varianza finita para todo  $\gamma > 0$  ( $\hat{\gamma}_n \in \hat{\Gamma}_n$ ,  $\hat{\gamma}_n \in \hat{\Gamma}_n$ ).

En virtud de

$$D_{\gamma}^2 \hat{\gamma}_n = \frac{\gamma^2}{n(n+2)} \leq \frac{\gamma^2}{3n} = D_{\gamma}^2 \tilde{\gamma}_n \quad (\gamma > 0),$$

el estimador  $\hat{\gamma}_n$  es mejor que el estimador  $\tilde{\gamma}_n$ . (Se debe reflexionar otra vez sobre la significación de ambos estimadores, desde el punto de vista del contenido, para este problema de estimación.) El grado en que el estimador  $\hat{\gamma}_n$  es mejor que el estimador  $\tilde{\gamma}_n$  tiene el valor

$$\frac{D_{\gamma}^2 \hat{\gamma}_n}{D_{\gamma}^2 \tilde{\gamma}_n} = \frac{\frac{\gamma^2}{n(n+2)}}{\frac{\gamma^2}{3n}} = \frac{3}{n+2}$$

y es, por tanto, independiente de  $\gamma$ . Para  $n=4$  se obtiene, por ejemplo, que dicho grado es igual a  $\frac{1}{2}$ ; para  $n \rightarrow \infty$  este converge monótonamente hacia cero.

**Definición 7.** Un estimador  $\hat{\gamma}_n^* \in \hat{\Gamma}_n$  se denomina *estimador eficiente*, si para todos los estimadores  $\hat{\gamma}_n \in \hat{\Gamma}_n$  se cumple que

$$D_{\gamma}^2 \hat{\gamma}_n^* \leq D_{\gamma}^2 \hat{\gamma}_n \quad (\gamma \in \Gamma). \quad (7)$$

El grado en que un estimador eficiente  $\hat{\gamma}_n^* \in \hat{\Gamma}_n$  es mejor que  $\hat{\gamma}_n \in \hat{\Gamma}_n$ , es decir,

$$e_n(\gamma) := \frac{D_{\gamma}^2 \hat{\gamma}_n^*}{D_{\gamma}^2 \hat{\gamma}_n} \quad (\gamma \in \Gamma) \quad (8)$$

se llama *eficiencia* de  $\hat{\gamma}_n^*$ .

Luego, un estimador eficiente es el estimador con menor varianza en el conjunto  $\hat{\Gamma}_n$  de estimadores considerado.

Bajo condiciones bastante generales para la distribución de probabilidad de la población considerada, se puede indicar una cota inferior positiva para las varianzas de los estimadores  $\hat{\gamma}_n \in \hat{\Gamma}_n$ . Si se ha encontrado un estimador  $\hat{\gamma}^* \in \hat{\Gamma}_n$ , cuya varianza es igual a esta cota inferior, entonces  $\hat{\gamma}^*$  es evidentemente un estimador eficiente. A continuación trataremos esta problemática de modo más exacto.

Sea  $X$  una variable aleatoria, cuya distribución de probabilidad depende de un parámetro  $\gamma \in \Gamma$ . Supongamos que  $X$  posee, para cada  $\gamma \in \Gamma$ , una distribución continua, y designemos con  $f_\gamma$  la densidad correspondiente. Además, supongamos que la función  $\gamma \mapsto f_\gamma(x)$  ( $\gamma \in \Gamma$ ) es dos veces continuamente diferenciable con respecto a  $\gamma$  para todo  $x \in \mathbb{R}^1$  y que el conjunto  $\{x : f_\gamma(x) > 0\}$  es el mismo para todo  $\gamma \in \Gamma$ .

**Teorema 2.** Para todo estimador  $\hat{\gamma}_n \in \hat{\Gamma}_n$  se cumple, bajo las condiciones de regularidad nombradas, la desigualdad

$$D_{\hat{\gamma}_n}^2 \geq \frac{1}{I_n(\gamma)} \quad (\gamma \in \Gamma) \quad (9)$$

con

$$I_n(\gamma) = n D_\gamma^2 \left( \frac{d \ln f_\gamma(X)}{d\gamma} \right) \quad (10)$$

La desigualdad (9), que proporciona para un estimador  $\hat{\gamma}_n$  dado una proposición acerca de su exactitud, se denomina en la literatura *desigualdad de información* o *desigualdad de Rao-Cramer* (en el ámbito de los países de habla inglesa) o *desigualdad de Fréchet-Darmois* (en los países de lengua francesa). La magnitud dada por la expresión (10) se denomina *información de Fisher*; ella es una medida para la información contenida en la muestra sobre el parámetro que se debe estimar, y depende, en general, tanto de  $\gamma \in \Gamma$  como del tamaño  $n$  de la muestra. En particular, extraemos de la expresión (10) que, bajo las condiciones adicionales halladas, las varianzas de los estimadores  $\hat{\gamma}_n$  de una sucesión de estimadores insegados pueden converger hacia cero a lo sumo en el orden  $\frac{1}{n}$ .

**Ejemplo 6.** Supongamos que  $X$  posee una distribución  $N(\mu, \sigma_0^2)$ ; sea  $\mu$  desconocido y  $\sigma_0^2$  conocido. Hagamos  $\gamma = \mu$  y  $\Gamma = \mathbb{R}^1$ . Entonces se cumple que

$$f_\gamma(x) = \frac{1}{\sqrt{2\pi\sigma_0}} e^{-\frac{(x-\gamma)^2}{2\sigma_0^2}} \quad (-\infty < x < \infty, \gamma \in \mathbb{R}^1),$$

y se satisfacen las condiciones adicionales indicadas anteriormente, para esta población. Para  $I_n(\gamma)$  obtenemos, en virtud de  $D_\gamma^2 X = \sigma_0^2$ ,

$$\begin{aligned} I_n(\gamma) &= n D_\gamma^2 \left( \frac{d \ln f_\gamma(X)}{d\gamma} \right) = n D_\gamma^2 \left( \frac{d}{d\gamma} \left( -\ln \sqrt{2\pi\sigma_0} - \frac{(X-\gamma)^2}{2\sigma_0^2} \right) \right) \\ &= n D_\gamma^2 \left( \frac{X-\gamma}{\sigma_0^2} \right) = n \cdot \frac{1}{\sigma_0^4} D_\gamma^2 X = n \cdot \frac{1}{\sigma_0^4} \sigma_0^2 = \frac{n}{\sigma_0^2} \end{aligned}$$

y con esto se cumple para todos los estimadores insegados  $\hat{\gamma}_n$  para  $\gamma$  que

$$D_{\hat{\gamma}_n}^2 \geq \frac{\sigma_0^2}{n} \quad (\gamma \in \mathbb{R}^1).$$

Para el estimador  $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i$  se cumple que  $E_{\gamma} \hat{\gamma}_n = \gamma$  y que  $D_{\gamma}^2 \hat{\gamma}_n = \frac{\sigma_0^2}{n}$  (ver para ello el ejemplo del epígrafe 10.1). Luego  $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i$  es un estimador eficiente para  $\gamma$ .

Queremos cerrar esta problemática con algunas otras proposiciones interesantes sobre la desigualdad de Rao-Cramer.

**Teorema 3.** Sean satisfechas las condiciones nombradas anteriormente con respecto a la población  $X$ . Entonces se cumplen las proposiciones siguientes:

- Si existe un estimador insesgado  $\hat{\gamma}_n$  con  $D_{\gamma}^2 \hat{\gamma}_n = \frac{1}{I_n(\gamma)}$  ( $\gamma \in \Gamma$ ), entonces  $f_{\gamma}$  posee la representación

$$f_{\gamma}(x) = \exp \{A(\gamma)B(x) + C(\gamma) + D(x)\} \quad (-\infty < x < \infty, \gamma \in \Gamma),$$

o sea,  $f_{\gamma}$  es del llamado tipo exponencial.

- Si  $f_{\gamma}$  es del tipo exponencial, es decir, si se cumple que

$$f_{\gamma}(x) = \exp \{A(\gamma)B(x) + C(\gamma) + D(x)\} \quad (-\infty < x < \infty, \gamma \in \Gamma),$$

y  $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n B(X_i)$  es un estimador insesgado para  $\gamma$ , entonces se cumple que

$$D_{\gamma}^2 \hat{\gamma}_n = \frac{1}{I_n(\gamma)},$$

o sea,  $\hat{\gamma}_n$  es un estimador eficiente para  $\gamma$ .

- Si existe un estimador insesgado  $\hat{\gamma}_n$  con  $D_{\gamma}^2 \hat{\gamma}_n = \frac{1}{I_n(\gamma)}$ , entonces  $\hat{\gamma}_n$  es el único estimador insesgado con esta propiedad.

Ilustraremos este teorema con un ejemplo.

**Ejemplo 7.** La densidad considerada en el ejemplo 6

$$f_{\gamma}(x) = \frac{1}{\sqrt{2\pi} \sigma_0} e^{-\frac{(x-\gamma)^2}{2\sigma_0^2}} = \exp \left( \frac{\gamma x}{\sigma_0^2} - \frac{\gamma^2}{2\sigma_0^2} - \frac{x^2}{2\sigma_0^2} \ln \sqrt{2\pi} \sigma_0 \right)$$

es del tipo exponencial ( $A(\gamma) = \frac{\gamma}{\sigma_0^2}$ ,  $B(x) = x$ ,  $C(\gamma) = -\frac{\gamma^2}{2\sigma_0^2}$ ,  $D(x) = -\ln \sqrt{2\pi} \sigma_0 = -\frac{x^2}{2\sigma_0^2}$ ). Para el estimador

$$\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n B(X_i) = \frac{1}{n} \sum_{i=1}^n X_i$$

se cumple que  $E_{\gamma} \hat{\gamma}_n = \gamma$ . Por tanto, sobre la base de la proposición 2 del teorema 3,  $\hat{\gamma}_n$  es un estimador eficiente para  $\gamma$  (esto lo hemos verificado ya directamente en el ejemplo anterior) y en virtud de la proposición 3,  $\hat{\gamma}_n$  es el único estimador insesgado eficiente para  $\gamma$ .

Muchos de los estimadores utilizados comúnmente poseen, para un tamaño de la muestra suficientemente grande, una distribución aproximadamente normal. Precisaremos este comportamiento en la definición siguiente.

**Definición 8.** Una sucesión  $(\hat{\gamma}_n)$  de estimadores  $\hat{\gamma}_n \in \Gamma$  para  $\gamma$  se dice que está *distribuida normalmente de forma asintótica*, si se cumple que

$$\lim_{n \rightarrow \infty} P_{\gamma} \left( \frac{\hat{\gamma}_n - \gamma}{\sqrt{D_{\gamma}^2 \hat{\gamma}_n}} < x \right) = \Phi(x) \quad (-\infty < x < \infty, \gamma \in \Gamma) \quad (11)$$

(En caso del cumplimiento de (11) para una estimador  $\hat{\gamma}_n$  se dice también que  $\hat{\gamma}_n$  posee una distribución asintóticamente normal.)

Luego, la propiedad caracterizada mediante la definición 8, significa que existe una convergencia en distribución hacia una variable aleatoria  $N(0,1)$ .

**Ejemplo 8.** Sea  $A$  un suceso aleatorio que se presenta en el marco de un experimento aleatorio con la probabilidad  $p$ ;  $p$  sea desconocida ( $0 < p < 1$ ).

Consideremos la variable aleatoria  $X$ .

$$X = \begin{cases} 1, & \text{en caso de la ocurrencia de } A \\ 0, & \text{en caso de la ocurrencia de } \bar{A}. \end{cases}$$

y pongámonos la tarea de estimar el parámetro  $\gamma = p$  sobre la base de una muestra matemática  $(X_1, \dots, X_n)$  de tamaño  $n$  de la población  $X$ . Para ello utilicemos el estimador  $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , que indica la frecuencia relativa aleatoria de la ocurrencia de  $A$  en una serie de  $n$  repeticiones independientes del experimento considerado. Se cumplen las proposiciones

$$E_{\gamma} \hat{\gamma}_n = \gamma \quad \text{y} \quad D_{\gamma}^2 \hat{\gamma}_n = \frac{\gamma(1-\gamma)}{n} \quad (0 < \gamma < 1)$$

(ver 4.5, teorema 4); luego  $(\hat{\gamma}_n)$  es una sucesión de estimadores para  $\gamma = p = P(A)$  débilmente consistente y fuertemente consistente también (ver 7.3, teoremas 1 y 6). Del Teorema integral de De Moivre-Laplace (ver 7.5, teorema 1) se obtiene directamente

$$\lim_{n \rightarrow \infty} P_{\gamma} \left( \frac{\hat{\gamma}_n - \gamma}{\sqrt{D_{\gamma}^2 \hat{\gamma}_n}} < x \right) = \lim_{n \rightarrow \infty} P_{\gamma} \left( \frac{\hat{\gamma}_n - \gamma}{\sqrt{\frac{\gamma(1-\gamma)}{n}}} < x \right) = \Phi(x)$$

$(-\infty < x < \infty, 0 < \gamma < 1)$ , es decir, la sucesión  $(\hat{\gamma}_n)$  posee una distribución asintóticamente normal.

### 10.3 Sobre la construcción de estimadores puntuales

En los ejemplos analizados hasta ahora hemos partido siempre de estimadores puntuales dados y los hemos investigado con respecto a propiedades especiales (por ejemplo, si es insesgado, consistente, eficiente). Ahora se impone naturalmente la pregunta de cómo obtener estimadores puntuales, sobre todo cuando se exigen, además, ciertas propiedades de los mismos (por ejemplo, la consistencia). Para ello han sido desarrollados una serie de métodos, por ejemplo, el llamado *método de máxima verosimilitud* (en la literatura inglesa Maximum-Likelihood-Methode) -que está en estrecha relación con el *método de la suma de los mínimos cuadrados*- y el denominado *método de los momentos*. Aquí trataremos brevemente el método de máxima verosimilitud y después haremos referencia al método de los momentos.

El método de máxima verosimilitud se basa en el principio de estimación siguiente. Como valor estimado para un parámetro desconocido de una distribución de probabilidad se utiliza aquel valor del parámetro para el cual a la muestra concreta le corresponde una probabilidad lo mayor posible. Así se aclara el nombre de este método en la bibliografía inglesa (likelihood- probabilidad, pero más en el sentido del lenguaje usual que en el sentido matemático).

El punto de partida para la exposición de este método es una variable aleatoria  $X$ , cuya distribución de probabilidad depende de un parámetro  $\gamma \in \Gamma$ . En el caso de una variable aleatoria continua  $X$ , designemos con  $f_\gamma(x)$  la densidad de  $X$  en el punto  $x$ , bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro; en el caso discreto sea  $f_\gamma(x) = P_\gamma(X=x)$ . Además, sea  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$ , es decir, un vector aleatorio  $n$ -dimensional, cuyas componentes son independientes y están distribuidas idénticamente que  $X$ . Si  $X$  es continua, entonces  $\prod_{i=1}^n f_\gamma(x_i)$  indica el valor de la densidad de probabilidad del vector aleatorio  $(X_1, \dots, X_n)$  en  $(x_1, \dots, x_n)$ , bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro (ver 6.4, teorema 2); en el caso de una variable aleatoria discreta se cumple que

$$\prod_{i=1}^n f_\gamma(x_i) = P_\gamma(X_1=x_1, \dots, X_n=x_n)$$

(ver 6.4, teorema 1).

**Definición 1.** Si  $(x_1, \dots, x_n)$  es una muestra concreta de tamaño  $n$  de la población  $X$ , entonces la función definida sobre  $\Gamma$  por

$$L(x_1, \dots, x_n; \gamma) = \prod_{i=1}^n f_\gamma(x_i) \quad (\gamma \in \Gamma) \quad (1)$$

se denomina *función de verosimilitud (Likelihood Function) de la muestra concreta  $(x_1, \dots, x_n)$* .

Por tanto, según las explicaciones que se dieron anteriormente,  $L(x_1, \dots, x_n; \gamma)$  indica en el caso discreto la probabilidad de que la muestra matemática  $(X_1, \dots, X_n)$  tome el valor  $(x_1, \dots, x_n)$  (bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro); en el caso continuo,  $L(x_1, \dots, x_n; \gamma)$  indica el valor de la densidad de la muestra matemática  $(X_1, \dots, X_n)$  en  $(x_1, \dots, x_n)$ , bajo la misma suposición.

El principio de estimación sobre el cual se basa el método de máxima verosimilitud consiste en utilizar como valor estimado para el parámetro desconocido un valor tal, que para una muestra concreta  $(x_1, \dots, x_n)$  dada, la función de verosimilitud tome un valor máximo. Para la determinación de un valor estimado semejante se utiliza con frecuencia el cálculo diferencial -supuestas las propiedades de diferenciabilidad correspondientes de la función de verosimilitud que se satisfacen comúnmente en casos de aplicación. Como las funciones  $\gamma \rightarrow L(x_1, \dots, x_n; \gamma)$  y  $\gamma \rightarrow \ln L(x_1, \dots, x_n; \gamma)$  ( $\gamma \in \Gamma$ ) toman valores máximos en los mismos puntos, nos ocuparemos, por conveniencia, no de la ecuación

$$\frac{d}{d\gamma} L(x_1, \dots, x_n; \gamma) = 0,$$

sino de la ecuación (en muchos casos más sencilla)  $\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) = 0$ .

**Definición 2.** Si  $(x_1, \dots, x_n)$  es una muestra concreta de tamaño  $n$  de la población  $X$ , entonces la ecuación

$$\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) = 0 \quad (2)$$

es conocida como *ecuación de verosimilitud* (*Likelihood -Equation*) de la muestra concreta  $(x_1, \dots, x_n)$ .

Si se sustituyen en la solución de esta ecuación los valores  $x_i$  de la muestra concreta por las variables  $X_i$  de la muestra ( $i=1, \dots, n$ ), se obtiene un estimador  $\hat{\gamma}_n = \varphi(X_1, \dots, X_n)$ .

Definición 3. Un estimador  $\hat{\gamma}_n = \varphi(X_1, \dots, X_n)$  que para toda muestra concreta  $(x_1, \dots, x_n)$  es una solución de la ecuación de verosimilitud (o sea, para el que se cumple la relación  $\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) \Big|_{\gamma=\hat{\gamma}_n} = 0$ ) y a la vez, es un punto para el cual la función de verosimilitud tiene un máximo, se denomina estimador máximo verosímil para  $\gamma$  (Maximum Likelihood-Estimation for  $\gamma$ ).

(En nuestra exposición introductoria del método de máxima verosimilitud hemos excluido interrogantes acerca de la existencia de estimadores máximo verosímiles y de su unicidad.)

Ahora queremos demostrar el método de máxima verosimilitud en dos ejemplos.

Ejemplo 1. Supongamos que  $X$  posee una distribución exponencial con el parámetro  $\alpha$  (ver 5.5, definición 1);  $\alpha$  sea desconocido. Luego hagamos  $\gamma = \alpha$ ,  $\gamma > 0$ . Entonces se cumple que

$$f_i(x) = \begin{cases} 0 & \text{para } x \leq 0 \\ \gamma e^{-\gamma x} & \text{para } x > 0. \end{cases}$$

Sea  $(x_1, \dots, x_n)$  una muestra concreta de tamaño  $n$  de la población  $X$ . Para la función de verosimilitud de esta muestra se obtiene que

$$L(x_1, \dots, x_n; \gamma) = \prod_{i=1}^n f_i(x_i) = \prod_{i=1}^n \gamma e^{-\gamma x_i} = \gamma^n e^{-\gamma \sum_{i=1}^n x_i}$$

y de aquí

$$\ln L(x_1, \dots, x_n; \gamma) = n \ln \gamma - \gamma \sum_{i=1}^n x_i$$

Por consiguiente, la ecuación de verosimilitud es

$$\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) = \frac{n}{\gamma} \sum_{i=1}^n x_i = 0.$$

La única solución de esta ecuación es  $\gamma = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i}$ ; en virtud de

$$\frac{d^2}{d\gamma^2} \ln L(x_1, \dots, x_n; \gamma) = -\frac{n}{\gamma^2} < 0$$

se trata del punto de un máximo de la función de verosimilitud.

Por consiguiente, para una muestra concreta se obtiene como valor estimado, según el método de máxima verosimilitud, el recíproco de la media aritmética de los valores de la

muestra. Si sustituimos ahora los valores de la muestra, por las variables correspondientes, obtenemos como estimador máximo verosímil para  $\gamma$

$$\hat{\gamma}_n = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}.$$

Ejemplo 2. Supongamos que  $X$  posee una distribución de Poisson con el parámetro  $\lambda$  (ver 4.7, definición 1);  $\lambda$  sea desconocida. Luego hagamos  $\gamma = \lambda$ ,  $\gamma > 0$ . Entonces se cumple que

$$f(x) = p(x; \gamma) = P(X=x) = \frac{\gamma^x}{x!} e^{-\gamma} \quad (x=0, 1, 2, \dots).$$

Sea  $(x_1, \dots, x_n)$  una muestra concreta de tamaño  $n$  de la población  $X$ . Para la función de verosimilitud de esta muestra se obtiene

$$L(x_1, \dots, x_n; \gamma) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{\gamma^{x_i}}{x_i!} e^{-\gamma} = e^{-n\gamma} \gamma^{\sum x_i} \cdot \frac{1}{\prod_{i=1}^n x_i!}$$

y de aquí

$$\ln L(x_1, \dots, x_n; \gamma) = -n\gamma + \ln \gamma \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i!.$$

Por consiguiente, la ecuación de verosimilitud es

$$\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) = -n + \frac{1}{\gamma} \sum_{i=1}^n x_i = 0$$

La única solución de esta ecuación es  $\gamma = \frac{1}{n} \sum_{i=1}^n x_i$ ; en virtud de

$$\left. \frac{d^2}{d\gamma^2} \ln L(x_1, \dots, x_n; \gamma) \right|_{\gamma = \frac{1}{n} \sum_{i=1}^n x_i} = -n^2 \left( \sum_{i=1}^n x_i \right)^{-1} < 0$$

se trata del punto de un máximo de la función de verosimilitud. Por consiguiente, para una muestra concreta se obtiene como valor estimado, según el método de máxima verosimilitud, la media aritmética de los valores de la muestra. Si sustituimos ahora los valores de la muestra por las variables correspondientes, obtenemos como estimador máximo verosímil para  $\gamma$

$$\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

La significación del método de máxima verosimilitud consiste en que -bajo condiciones bastantes generales- proporciona estimadores con propiedades convenientes. Si existe, por ejemplo, un estimador insesgado y eficiente  $\hat{\gamma}_n$  para  $\gamma$ , este estimador se obtiene de forma unívoca, según el método de máxima verosimilitud, y además, resulta que una sucesión de estimadores semejantes es consistente y posee una distribución asintóticamente normal.

Sin embargo, en el marco de nuestra exposición no podemos tratar estas proposiciones más detenidamente.

Queremos concluir nuestras explicaciones sobre el problema de la construcción de estimadores puntuales con algunas observaciones sobre el método de los momentos.

Sea de nuevo el punto de partida una población  $X$ , cuya distribución de probabilidad depende de un parámetro  $\gamma \in \Gamma$ ; además sea  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$ . Supongamos que  $X$  posee momentos iniciales hasta de orden  $k$ ,  $k \geq 1$  (ver 4.3, definición 3 y 5.2, definición 3). Estos momentos iniciales serán entonces, por lo general, funciones de  $\gamma \in \Gamma$

$$m_j = E_\gamma X^j = f_j(\gamma) \quad (\gamma \in \Gamma); \quad j=1, \dots, k. \quad (3)$$

Ahora queremos suponer que en la relación (3) se puede despejar únicamente  $\gamma$  para  $j=j_0$

$$\gamma = f_{j_0}^{-1}(m_{j_0}). \quad (4)$$

El principio de estimación sobre el cual se basa el método de los momentos consiste en sustituir la variable  $m_{j_0}$  en cada ocasión, por el estadígrafo  $\frac{1}{n} \sum_{i=1}^n X_i^{j_0}$ . De esta forma se obtiene por medio de (4) un estimador  $\hat{\gamma}_{j_0}$  para  $\gamma$ .

$$\hat{\gamma}_{j_0} = f_{j_0}^{-1} \left( \frac{1}{n} \sum_{i=1}^n X_i^{j_0} \right),$$

que se denomina estimador por el método de los momentos.

Ejemplo 3. Supongamos que  $X$  posee una distribución exponencial con el parámetro  $a$ ;  $a$  sea desconocido. Hagamos  $\gamma=a$ ,  $\gamma > 0$ . Entonces se cumple (ver 5.5, teorema 1) que

$$m_1 = E_\gamma X = \frac{1}{\gamma} = f_1(\gamma)$$

y con esto

$$\gamma = \frac{1}{m_1} = f_1^{-1}(m_1).$$

Si sustituimos ahora  $m_1$  por el estadígrafo  $\frac{1}{n} \sum_{i=1}^n X_i$ , obtenemos con esto el estimador

$$\hat{\gamma}_1 = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}$$

para  $\gamma$ . (Por tanto, en este caso se origina el mismo estimador por el método de los momentos que por el método de máxima verosimilitud, ver ejemplo 1.)

(Otro estimador por el método de los momentos -en realidad, más complicado y también menos conveniente en sus propiedades- es el que se obtendría sobre la base de

$$m_2 = E_\gamma X^2 = D_\gamma^2 X + (E_\gamma X)^2 = \frac{1}{\gamma^2} + \frac{1}{\gamma^2} = \frac{2}{\gamma^2} = f_2(\gamma);$$

es decir,

$$\gamma = \sqrt{\frac{2}{m_2}} = f_2^{-1}(m_2)$$

y entonces

$$\hat{\gamma}_n = \sqrt{\frac{2}{\frac{1}{n} \sum_{i=1}^n X_i^2}}$$

La sencillez del método de los momentos habla en muchos casos a favor de su aplicación práctica; no se necesita más que una relación funcional entre el parámetro y un momento inicial que se pueda despejar de forma unívoca, y solo se utilizan estadígrafos del mismo tipo. A decir verdad, desde el punto de vista teórico no se conoce todavía mucho acerca de los estimadores por el método de los momentos. En esencia, se sabe solo que los estadígrafos que sustituyen los momentos iniciales son estimadores de los momentos iniciales insesgados, fuertemente consistentes y con una distribución asintóticamente normal.

## 10.4 Ejemplos importantes de estimadores puntuales

En este epígrafe presentamos algunos estimadores puntuales utilizados con frecuencia en las aplicaciones; en particular, se obtienen aquí estimadores puntuales para los parámetros fundamentales que se presentan en las distribuciones de probabilidad tratadas por nosotros.

### 10.4.1 Estimador puntual para un valor esperado desconocido

El valor esperado  $EX$  de una variable aleatoria  $X$  se debe estimar sobre la base de una muestra matemática  $(X_1, \dots, X_n)$  de tamaño  $n$  de la población  $X$ . Luego, hagamos  $\gamma = EX$  y  $\Gamma = \mathbb{R}^1$ . Como estimador puntual  $\hat{\gamma}$ , para  $\gamma$  utilicemos la media aritmética de las variables de la muestra  $X_1, \dots, X_n$ ,

$$\hat{\gamma}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

El estimador puntual  $\hat{\gamma}_n$  es insesgado,

$$E_{\gamma} \hat{\gamma}_n = E_{\gamma} \left( \frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n E_{\gamma} X_i = \frac{1}{n} \cdot n \cdot \gamma = \gamma \quad (\gamma \in \mathbb{R}^1)$$

con respecto a la población  $X$  solo se supuso que el valor esperado  $EX$  existe.

Además, se cumple bajo la suposición de que  $X$ , independientemente del valor del parámetro, posee una varianza finita ( $D_{\gamma}^2 X < \infty$  para todo  $\gamma \in \mathbb{R}^1$ ) que

$$D_{\gamma}^2 \hat{\gamma}_n = D_{\gamma}^2 \left( \frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n^2} \sum_{i=1}^n D_{\gamma}^2 X_i = \frac{1}{n^2} \cdot n \cdot D_{\gamma}^2 X = \frac{D_{\gamma}^2 X}{n} \rightarrow 0$$

para todo  $\gamma \in \mathbb{R}^1$ . De aquí resulta con el teorema 1 (10.2) la consistencia (débil) de la sucesión  $(\hat{\gamma}_n)$ , una proposición que también se obtiene directamente de las explicaciones sobre la Ley de los grandes números (ver 7.4, teorema 3). (Por lo demás puede renunciarse a la condición  $D_{\gamma}^2 X < \infty$  ( $\gamma \in \mathbb{R}^1$ ), (ver 7.4, teorema 4); además se comprueba que la sucesión  $(\hat{\gamma}_n)$  es fuertemente consistente sobre la base de la Ley de los grandes números de Kolmogorov (ver 7.4, teorema 6).) El estimador puntual  $\hat{\gamma}_n$  posee para  $n$  grande una dis-

tribución aproximadamente  $N\left(\gamma, \frac{D^2X}{n}\right)$  (ver la observación 2 después del teorema 2 (9.4)), y, por consiguiente,  $(\hat{\gamma}_n)$  posee una distribución asintóticamente normal (ver 10.2, definición 8).

En especial obtenemos con (1) estimadores puntuales para el parámetro  $\mu$  de una variable aleatoria con distribución normal y para el parámetro  $\lambda$  de una distribución de Poisson.

#### 10.4.2 Estimadores puntuales para una varianza desconocida

La varianza  $D^2X$  de una variable aleatoria  $X$  se debe estimar sobre la base de una muestra matemática de tamaño  $n$  de la población  $X$ . Luego hagamos  $\gamma = D^2X$  y  $\Gamma = \{\gamma : \gamma > 0\}$ . En lo que sigue diferenciaremos dos casos:

a)  $\mu_0 = EX$  conocido

Como estimador puntual  $\hat{\gamma}_n$  para  $\gamma$  utilizaremos la media aritmética de los cuadrados de las desviaciones de las variables de la muestra  $X_i (i=1, \dots, n)$  del valor esperado (común)  $\mu_0$ ,

$$\hat{\gamma}_n = S_n^* = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2. \quad (2)$$

El estimador  $\hat{\gamma}_n$  es insesgado.

$$E_{\gamma} \hat{\gamma}_n = E_{\gamma} \left( \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2 \right) = \frac{1}{n} \sum_{i=1}^n E_{\gamma} (X_i - \mu_0)^2 = \frac{1}{n} \cdot n \cdot \gamma = \gamma (\gamma \in \Gamma).$$

Además se comprueba que la sucesión  $(\hat{\gamma}_n)$  es fuertemente consistente sobre la base de la Ley de los grandes números de Kolmogorov.

En especial, obtenemos con (2) un estimador puntual para el parámetro  $\sigma^2$  de una variable aleatoria con distribución normal cuando el parámetro  $\mu = \mu_0$  es conocido.

b)  $\mu = EX$  desconocido

En este caso utilizamos el estadígrafo

$$\hat{\gamma}_n = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (3)$$

como estimador puntual para  $\gamma$ .

El estimador (3) es un estimador insesgado para  $\gamma$ . Con esto proporciona (3) un estimador puntual insesgado -y por lo demás también consistente- para el parámetro  $\sigma^2$  de una variable aleatoria con distribución normal, cuando el parámetro  $\mu$  es desconocido.

**Observación.** El estimador puntual dado por (2) no es utilizable aquí, ya que en (2) aparece para el caso considerado un parámetro desconocido. Si se sustituye este por  $\bar{X}_n$ , entonces se obtiene con (2) un estimador no insesgado para  $\gamma$ , pero sí asintóticamente insesgado.

#### 10.4.3 Estimador puntual para una probabilidad desconocida

Como valor estimado para la probabilidad (desconocida)  $p$  de un suceso aleatorio  $A$  utilizamos la frecuencia relativa de la ocurrencia de este suceso en una serie de  $n$  repeticio-

nes independientes de un mismo experimento, en el cual el suceso  $A$  tiene la probabilidad  $p$ . El estimador puntual  $\hat{\gamma}_n$  sobre el cual se basa este procedimiento fue investigado en 10.2 (ejemplo 8); este se mostró como un estimador insesgado para  $p$  y allí se estudió también que la sucesión  $(\hat{\gamma}_n)$  es consistente y posee una distribución asintóticamente normal.

#### 10.4.4 Estimador puntual para una función de distribución desconocida

El problema de la estimación del valor desconocido de la función de distribución  $F$  de una variable aleatoria  $X$  en un punto  $x \in \mathbb{R}$ , o sea, de  $F(x)$ , es equivalente al problema de la estimación de la probabilidad del suceso aleatorio  $(X < x)$ . Si existe una muestra concreta  $(x_1, \dots, x_n)$  de la población  $X$ , entonces se utiliza como valor estimado para  $F(x)$  -de acuerdo al modo de proceder en 10.4.3- el valor de la función de distribución empírica  $W_n$  de la muestra concreta  $(x_1, \dots, x_n)$  (ver 9.3, definición 1) en el punto  $x$ , es decir, el número  $W_n(x)$ . El estimador puntual tomado aquí por base es el valor de la función de distribución empírica  $W_n$  de una muestra matemática  $(X_1, \dots, X_n)$  (ver 9.3, definición 2) de la población  $X$  en el punto  $x$ . Al respecto observemos aún que se puede comprobar que el estimador  $W_n(x)$  es insesgado y que la sucesión  $(W_n(x))$  es consistente mediante la relación (1) y el teorema 1 de 9.3, respectivamente.

#### 10.4.5 Estimador puntual para un coeficiente de correlación desconocido

Sea  $(X, Y)$  un vector aleatorio bidimensional (ver 7.1) con el coeficiente de correlación (desconocido)  $\rho$  (ver 6.2, definición 3 y 6.3, definición 3). El parámetro  $\gamma = \rho$  debe estimarse sobre la base de una muestra matemática  $((X_1, Y_1), \dots, (X_n, Y_n))$  de tamaño  $n$  de la población  $(X, Y)$ - esta es, por tanto, un vector aleatorio  $n$  dimensional, cuyas componentes  $(X_i, Y_i)$  son independientes y están distribuidas idénticamente que  $(X, Y)$ . Para ello se utiliza el estadígrafo.

$$\hat{\gamma}_n = R_n = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}. \quad (4)$$

En el caso de una muestra concreta  $((x_1, y_1), \dots, (x_n, y_n))$  se obtiene como valor estimado, utilizando este estimador puntual para el coeficiente de correlación, el coeficiente de correlación empírica

$$r_n = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_n)^2}} \quad (5)$$

El análisis del estimador puntual  $R_n$  y el tratamiento de problemas referentes a esto (por ejemplo, intervalo de confianza para el coeficiente de correlación, díxicmas de hipótesis

sobre la independencia de variables aleatorias) son tareas parciales del llamado análisis de correlación, de un procedimiento de análisis estadístico, que desempeña un gran papel en los distintos campos de aplicación de la Estadística matemática. En el marco de nuestra introducción no podemos tratar esto de forma más detallada. Solo advertimos (sin demostración) que, en el caso de un vector aleatorio  $(X, Y)$  con distribución normal, se cumplen las proposiciones

$$E_p R_n \approx p \text{ y } D_p^2 R_n \approx \frac{(1-p^2)^2}{n} \quad (n \gg 1).$$

## 10.5 Estimaciones por intervalo de confianza

Nos ocuparemos en este epígrafe de estimaciones por intervalo de confianza, que se utilizan especialmente cuando se desea un grado de exactitud de la estimación de un parámetro desconocido, que no se puede obtener con una estimación puntual (por ejemplo, a causa de un tamaño de la muestra muy pequeño). La situación de partida es, por tanto, la misma que para las estimaciones puntuales: La distribución de probabilidad de una población  $X$  depende de un parámetro  $\gamma \in \Gamma \subseteq \mathbb{R}^1$ ; el valor verdadero -pero desconocido- del parámetro  $\gamma$  se denota con  $\gamma_0$ . Además, sea  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$ . Como se acordó en el epígrafe 10.1, entenderemos por un intervalo de confianza  $J(X_1, \dots, X_n)$  un denominado intervalo aleatorio, es decir, un intervalo cuyos extremos son magnitudes dependientes de las variables de la muestra -luego son variables aleatorias; para toda muestra concreta  $(x_1, \dots, x_n)$ ,  $J(x_1, \dots, x_n)$  es un intervalo comprendido en  $\Gamma$ .

De importancia decisiva para una estimación por intervalo de confianza es la probabilidad de que el intervalo aleatorio  $J(X_1, \dots, X_n)$  contenga al valor verdadero  $\gamma_0$  del parámetro; para este suceso aleatorio escribiremos  $J(X_1, \dots, X_n) \ni \gamma_0$ . Por consiguiente, nos interesa  $P_{\gamma_0} (J(X_1, \dots, X_n) \ni \gamma_0)$ . Pero como no conocemos a  $\gamma_0$ , nos ocuparemos de forma más general con la probabilidad de que el intervalo aleatorio  $J(X_1, \dots, X_n)$  contenga al valor  $\gamma \in \Gamma$ , calculada bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro, o sea, con  $P_{\gamma} (J(X_1, \dots, X_n) \ni \gamma)$  para  $\gamma \in \Gamma$ .

**Definición 1.** Sea  $J(X_1, \dots, X_n)$  un intervalo de confianza. El número

$$\varepsilon = \min_{\gamma \in \Gamma} P_{\gamma} (J(X_1, \dots, X_n) \ni \gamma) \quad (1)$$

se denomina *coeficiente de confiabilidad* del intervalo de confianza  $J(X_1, \dots, X_n)$ .

**Definición 2.** Un intervalo de confianza  $J(X_1, \dots, X_n)$  se denomina un *intervalo de confianza para  $\gamma$  con el nivel de confiabilidad  $1-\alpha$*  ( $0 < \alpha < 1$ , dado) si

$$P_{\gamma} (J(X_1, \dots, X_n) \ni \gamma) \geq 1-\alpha \quad (\gamma \in \Gamma) \quad (2)$$

o sea, si se cumple que  $\varepsilon \geq 1-\alpha$ .

La probabilidad de que el intervalo aleatorio  $J(X_1, \dots, X_n)$  contenga al valor  $\gamma$ , calculada bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro, tiene al menos el valor  $1-\alpha$  para un intervalo de confianza con el nivel de confiabilidad  $1-\alpha$ . Aquí se exige la validez de (2) para todo  $\gamma \in \Gamma$ ; con esto se cumple (2) en particular para  $\gamma_0$ , el valor verdadero del parámetro.

**Ejemplo 1.** Supongamos que la variable aleatoria  $X$  está uniformemente distribuida sobre el intervalo  $[0, b]$ ,  $b > 0$ ;  $b$  sea desconocido. Hagamos  $\gamma = b$  y  $\Gamma = \{\gamma: \gamma > 0\}$ . Queremos indicar para  $\gamma$  un intervalo de confianza con el nivel de confiabilidad  $1 - \alpha$  ( $0 < \alpha < 1$ , fijo). Para ello utilicemos el estimador puntual  $\hat{\gamma}_n = \max\{X_1, \dots, X_n\}$  (ver ejemplo 2 (10.2)). Fijemos el intervalo aleatorio en la forma

$$J(X_1, \dots, X_n) = [\delta_1 \hat{\gamma}_n, \delta_2 \hat{\gamma}_n] \text{ con } 1 \leq \delta_1 < \delta_2.$$

(En principio esto es algo arbitrario, pero razonable.) Ahora determinemos  $\delta_1$  y  $\delta_2$ , de modo que se cumpla la desigualdad  $P_\gamma(J(X_1, \dots, X_n) \ni \gamma) \geq 1 - \alpha$  para todo  $\gamma \in \Gamma$ . Se cumple que

$$P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = P_\gamma(\delta_1 \hat{\gamma}_n \leq \gamma \leq \delta_2 \hat{\gamma}_n) = P_\gamma\left(\frac{\gamma}{\delta_2} \leq \hat{\gamma}_n \leq \frac{\gamma}{\delta_1}\right).$$

Si observamos ahora que la función de distribución  $F_{\hat{\gamma}_n}^\gamma$  de la variable aleatoria  $\hat{\gamma}_n$  calculada bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro, está dada por

$$F_{\hat{\gamma}_n}^\gamma(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ \left(\frac{x}{\gamma}\right)^n & \text{para } 0 \leq x \leq \gamma, \\ 1 & \text{para } x \geq \gamma \end{cases}$$

(ver 9.4, teorema 1), obtenemos que

$$P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = F_{\hat{\gamma}_n}^\gamma\left(\frac{\gamma}{\delta_1}\right) - F_{\hat{\gamma}_n}^\gamma\left(\frac{\gamma}{\delta_2}\right) = \left(\frac{\gamma}{\delta_1 \gamma}\right)^n - \left(\frac{\gamma}{\delta_2 \gamma}\right)^n = \frac{1}{\delta_1^n} - \frac{1}{\delta_2^n}$$

Escojamos, por ejemplo  $\delta_1 = \sqrt[n]{1 - \alpha_1}$  y  $\delta_2 = \sqrt[n]{\alpha_2}$  con  $\alpha_1 \geq 0$ ,  $\alpha_2 > 0$ ,  $\alpha_1 + \alpha_2 = \alpha$ , entonces se cumple que

$$P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha,$$

osea,

$$J(X_1, \dots, X_n) = \left[ \sqrt[n]{\frac{\hat{\gamma}_n}{1 - \alpha_1}}, \sqrt[n]{\frac{\hat{\gamma}_n}{\alpha_2}} \right] (\alpha_1 \geq 0, \alpha_2 > 0, \alpha_1 + \alpha_2 = \alpha)$$

es un intervalo de confianza para  $\gamma$  con un nivel de confiabilidad  $1 - \alpha$ . Para una muestra concreta  $(x_1, \dots, x_n)$  se obtiene por medio de este estimador el intervalo estimado concreto

$$J(x_1, \dots, x_n) = \left\{ \gamma : \frac{x_{\max}}{\sqrt[n]{1 - \alpha}} \leq \gamma \leq \frac{x_{\max}}{\sqrt[n]{\alpha}} \right\}$$

con  $x_{\max} = \max\{x_1, \dots, x_n\}$  (ver fig. 46 a). Para  $\alpha_1 = 0$ ,  $\alpha_2 = \alpha$  se obtiene el intervalo estimado concreto (ver fig. 46 b),

$$J_1(x_1, \dots, x_n) = \left\{ \gamma : x_{\max} \leq \gamma \leq \sqrt[n]{\frac{x_{\max}}{\alpha}} \right\}$$

y para  $\alpha_2 \rightarrow 0$ ,  $\alpha_1 \rightarrow \alpha$  se obtiene el intervalo estimado concreto (ver fig. 46 c)

$$J_2(x_1, \dots, x_n) = \left\{ \gamma : \frac{x_{\max}}{\sqrt[n]{1-\alpha}} \leq \gamma < +\infty \right\}$$

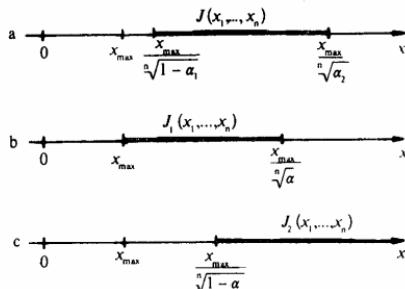


Figura 46

Por medio de un intervalo de confianza con el nivel de confiabilidad  $1-\alpha$  se obtiene para una muestra concreta  $(x_1, \dots, x_n)$  un intervalo  $J(x_1, \dots, x_n) \subseteq \Gamma$  y se decide comúnmente a favor de que " $\gamma_0 \in J(x_1, \dots, x_n)$ ". Esta decisión será verdadera o falsa en el caso concreto; en todo caso ella nada tiene que ver con la causalidad y no se trata tampoco de una proposición que sea correcta con probabilidad  $\geq 1-\alpha$ . A decir verdad, se puede estimar la probabilidad  $\delta$  de una decisión errónea para el principio tomado por base en la decisión concreta ilustrada. Una decisión errónea ocurre siempre y cuando el valor verdadero  $\gamma_0$  del parámetro no pertenezca al intervalo  $J(x_1, \dots, x_n)$ . Luego, se cumple que

$$\delta = P_{\gamma_0}(J(X_1, \dots, X_n) \not\ni \gamma_0). \quad (3)$$

En virtud de (2) resulta que  $\delta \leq \alpha$ , independientemente de qué valor posea  $\gamma_0$  en  $\Gamma$ . Con esto se aclara también el papel del parámetro  $\alpha$ ; con el principio de decisión descrito se necesita calcular como promedio con no más de 100  $\alpha$  % de decisiones erróneas, y de acuerdo con esto -considerando naturalmente el planteamiento de la tarea concreta y en particular, las consecuencias de una decisión errónea- se fijará  $\alpha$ . (Con frecuencia se elige  $\alpha=5\%$ ,  $\alpha=2\%$  o  $\alpha=1\%$ .) Aquí se tiene que reflexionar, en especial, que el hacer  $\alpha$  más pequeño conduce, por lo general, a un intervalo estimado concreto de mayor longitud. (Para  $\alpha=0$  se obtiene forzosamente como intervalo estimado, para todas las muestras concretas  $(x_1, \dots, x_n)$ , el conjunto  $\Gamma$  de todos los posibles valores del parámetro; por tanto, en este caso no se utiliza la información contenida en la muestra acerca del valor verdadero del parámetro.)

En la construcción de un intervalo de confianza con un nivel de confiabilidad  $1-\alpha$  dado, está presente aún -como mostró el ejemplo 1- una cierta arbitrariedad (elección del estimador puntual tomado por base y sustitución para los extremos del intervalo aleatorio).

Por ello nos queremos ocupar un poco más detenidamente de la valoración y -sobre este basamento- de la comparación de intervalos de confianza. Un medio auxiliar esencial para esto es la denominada función de bondad.

**Definición 3.** Sea  $J(X_1, \dots, X_n)$  un intervalo de confianza. Entonces la función  $B$ ,

$$B(\gamma, \gamma') = P_\gamma(J(X_1, \dots, X_n) \ni \gamma'), \quad (4)$$

definida sobre  $\Gamma \times \Gamma$  se denomina *función de bondad* del intervalo de confianza dado.

El valor  $B(\gamma, \gamma')$  de la función de bondad  $B$  en el punto  $(\gamma, \gamma') \in \Gamma \times \Gamma$  indica, por consiguiente, la probabilidad de que el intervalo de confianza considerado contenga al parámetro  $\gamma'$ , calculada bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro. Luego, se cumple siempre que  $0 \leq B(\gamma, \gamma') \leq 1$ . Si  $J(X_1, \dots, X_n)$  es un intervalo de confianza con el nivel de confiabilidad  $1-\alpha$ , entonces se cumple que  $B(\gamma, \gamma') \geq 1-\alpha$  para todo  $\gamma \in \Gamma$ .

Ejemplo 2. Calculemos la función de bondad del intervalo de confianza

$$J(X_1, \dots, X_n) = \left[ \frac{\hat{\gamma}_n}{\sqrt{1-\alpha_1}}, \frac{\hat{\gamma}_n}{\sqrt{\alpha_2}} \right] \quad (\alpha_1 \geq 0, \alpha_2 > 0, \alpha_1 + \alpha_2 = \alpha)$$

con el nivel de confiabilidad  $1-\alpha$ , dado en el ejemplo 1. Para  $\gamma > 0, \gamma' > 0$  se cumple que

$$B(\gamma, \gamma') = P_{\gamma} (J(X_1, \dots, X_n) \ni \gamma') = P_{\gamma} \left( \frac{\hat{\gamma}_n}{\sqrt{1-\alpha_1}} \leq \gamma' \leq \frac{\hat{\gamma}_n}{\sqrt{\alpha_2}} \right)$$

$$= P_{\gamma} \left( \sqrt{\alpha_1} \gamma' \leq \hat{\gamma}_n \leq \sqrt{1-\alpha_1} \gamma' \right) = F_{\hat{\gamma}_n} \left( \sqrt{1-\alpha_1} \gamma' \right) - F_{\hat{\gamma}_n} \left( \sqrt{\alpha_1} \gamma' \right)$$

$$= \begin{cases} \left( \frac{\sqrt{1-\alpha_1} \gamma'}{\gamma} \right)^* - \left( \frac{\sqrt{\alpha_1} \gamma'}{\gamma} \right)^* = 1 - \alpha_1 \left( \frac{\gamma'}{\gamma} \right)^* & \text{para } 0 < \gamma' \leq \frac{\gamma}{\sqrt{1-\alpha_1}}, \\ 1 - \left( \frac{\sqrt{\alpha_1} \gamma'}{\gamma} \right)^* = 1 - \alpha_1 \left( \frac{\gamma'}{\gamma} \right)^* & \text{para } \frac{\gamma}{\sqrt{1-\alpha_1}} \leq \gamma' \leq \frac{\gamma}{\sqrt{\alpha_1}}, \\ 1 - 1 = 0 & \text{para } \gamma' \geq \frac{\gamma}{\sqrt{\alpha_1}}. \end{cases}$$

Para  $\alpha_1=0, \alpha_2=\alpha$ , o sea, para el intervalo de confianza (ver fig. 46 b)

$$J(X_1, \dots, X_n) = \left[ \frac{\hat{\gamma}_n}{\sqrt{\alpha}}, \frac{\hat{\gamma}_n}{\sqrt{\alpha}} \right]$$

con el nivel de confiabilidad  $1-\alpha$ , obtenemos la función de bondad  $B_1$ ,

$$B_1(\gamma, \gamma') = \begin{cases} (1-\alpha) \left( \frac{\gamma'}{\gamma} \right)^* & \text{para } 0 < \gamma' \leq \gamma, \\ 1-\alpha \left( \frac{\gamma'}{\gamma} \right)^* & \text{para } \gamma \leq \gamma' \leq \frac{\gamma}{\sqrt{\alpha}}, \\ 0 & \text{para } \gamma' \geq \frac{\gamma}{\sqrt{\alpha}} \end{cases}$$

Observemos que se cumple que  $B_1(\gamma, \gamma') < B_1(\gamma, \gamma) = 1-\alpha$  para todo  $\gamma > 0, \gamma' > 0$  con  $\gamma \neq \gamma'$ .

La propiedad hallada por último en el ejemplo 2 nos dice que todo valor "falso" del parámetro está contenido en el intervalo de confianza con una probabilidad menor que para el valor verdadero de este, independientemente de qué valor del parámetro es el verdadero. Expresaremos este hecho de forma general en la definición siguiente.

**Definición 4.** Un intervalo de confianza  $J(X_1, \dots, X_n)$  se denomina *admisible*, si para la función de bondad  $B$  se cumple que

$$B(\gamma, \gamma) \geq B(\gamma', \gamma') \quad ((\gamma, \gamma') \in \Gamma \times \Gamma). \quad (5)$$

Por último advertimos que la comparación de intervalos de confianza (en el marco de un mismo problema de estimación) se reduce fundamentalmente a la comparación de las funciones de bondad correspondientes.

**Definición 5.** Sean  $J_1(X_1, \dots, X_n)$  y  $J_2(X_1, \dots, X_n)$  intervalos de confianza (en el marco de un mismo problema de estimación) con las funciones de bondad  $B_1$  y  $B_2$ . El intervalo de confianza  $J_1(X_1, \dots, X_n)$  se llama mejor que el intervalo de confianza  $J_2(X_1, \dots, X_n)$ , si se cumple que

$$B_1(\gamma, \gamma) \leq B_2(\gamma, \gamma) \quad ((\gamma, \gamma) \in \Gamma \times \Gamma, \gamma \neq \gamma'). \quad (6)$$

El motivo para esta definición está claro de acuerdo con lo que precede y a la definición de función de bondad.

**Ejemplo 3.** Como continuación del ejemplo 1 consideremos el intervalo de confianza (ver fig. 46c)

$$J_2(X_1, \dots, X_n) = \left[ \frac{\hat{\gamma}_n}{\sqrt[n]{1-\alpha}}, +\infty \right]$$

con el nivel de confiabilidad  $1-\alpha$ , que se obtiene del intervalo de confianza  $J(X_1, \dots, X_n)$  con el nivel de confiabilidad  $1-\alpha$ , deducido en el ejemplo 1, a través del paso (formal) al límite  $\alpha_i \rightarrow \alpha$ . Para la función de bondad correspondiente se obtiene que

$$B_2(\gamma, \gamma') = \begin{cases} (1-\alpha) \left( \frac{\gamma'}{\gamma} \right)^n & \text{para } 0 \leq \gamma' \frac{\gamma}{\sqrt[n]{1-\alpha}} \\ 1 & \text{para } \gamma' \geq \frac{\gamma}{\sqrt[n]{1-\alpha}} \end{cases}$$

(Observemos al margen que  $J_2(X_1, \dots, X_n)$  no es admisible; por ejemplo, se cumple que  $B_2(\gamma, \gamma') = 1 > B_2(\gamma, \gamma) = 1 - \alpha$  para todo  $(\gamma, \gamma')$  con  $\gamma' \geq \frac{\gamma}{\sqrt[n]{1-\alpha}}$ ,  $\gamma > 0$ ).

Si comparamos esta función de bondad con la función de bondad  $B_1$  del intervalo de confianza  $J_1(X_1, \dots, X_n) = \left[ \hat{\gamma}_n, \frac{\hat{\gamma}_n}{\sqrt[n]{\alpha}} \right]$  considerada en el ejemplo 2, obtenemos que

$$B_1(\gamma, \gamma') \leq B_2(\gamma, \gamma') \quad (\gamma > 0, \gamma' > 0, \gamma \neq \gamma'),$$

es decir, que el intervalo de confianza  $J_1(X_1, \dots, X_n)$  es mejor que el intervalo de confianza  $J_2(X_1, \dots, X_n)$ .

## 10.6 Ejemplos importantes de estimaciones por intervalo de confianza

En este epígrafe indicamos intervalos de confianza con el nivel de confiabilidad  $1-\alpha$  ( $0 < \alpha < 1$ ) para los parámetros de una variable aleatoria con distribución normal, la pro-

babilidad de un suceso aleatorio y la función de distribución de una variable aleatoria. Se recomienda al lector que reflexione acerca de la significación de los extremos del intervalo de confianza (límites de confianza), que motiva con esto la sustitución que se hace en cada ocasión para el intervalo de confianza y que investigue la influencia de  $\alpha$ ,  $n$  y, dado el caso, de otras magnitudes características.

### 10.6.1 Intervalos de confianza para los parámetros de una distribución normal

Sea  $X$  una variable aleatoria con distribución normal y parámetros  $\mu$  y  $\sigma^2$ , y además, sea  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$ . En los extremos de los intervalos indicados a continuación se utilizan los estadígrafos  $\bar{X}_n$ ,  $S_n^2$  y  $S_n^{*2}$  (ver 9.4) y también los percentiles de la distribución normal estandarizada y de las distribuciones  $t$  y  $\chi^2$  (ver 5.6, definiciones 1,2, y 3; tablas 3, 4 y 5); aquí denominamos con  $z_p$  el percentil de orden  $p$  de la distribución normal estandarizada ( $\Phi(z_p) = p$ ). Para indicar intervalos de confianza con el nivel de confiabilidad  $1-\alpha$  para  $\gamma=\mu$ , tenemos que diferenciar si  $\sigma^2$  es conocida o no; de la misma forma, para indicar intervalos de confianza para  $\gamma=\sigma^2$  tenemos que diferenciar si  $\mu$  es conocido o no.

a)  $\gamma=\mu$ ,  $\sigma^2=\sigma_0^2$  (conocida)

**Teorema 1.** Sean  $\alpha_1$  y  $\alpha_2$  números positivos con  $\alpha_1 + \alpha_2 = \alpha$ . Entonces

$$J(X_1, \dots, X_n) = \left[ \bar{X}_n - z_{1-\alpha_1} \cdot \frac{\sigma_0}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha_2} \cdot \frac{\sigma_0}{\sqrt{n}} \right] \quad (1)$$

es un intervalo de confianza para  $\gamma$  con el nivel de confiabilidad  $1-\alpha$ .

**Demostración.** Se debe mostrar que se cumple

$$B(\gamma, \gamma) = P_\gamma (J(X_1, \dots, X_n) \ni \gamma) \geq 1-\alpha \text{ para todo } \gamma \in \mathbb{R}^1:$$

$$\begin{aligned} B(\gamma, \gamma) &= P_\gamma (J(X_1, \dots, X_n) \ni \gamma) = P_\gamma \left( \bar{X}_n - z_{1-\alpha_1} \cdot \frac{\sigma_0}{\sqrt{n}} \leq \gamma \leq \bar{X}_n + z_{1-\alpha_2} \cdot \frac{\sigma_0}{\sqrt{n}} \right) \\ &= P_\gamma \left( -z_{1-\alpha_1} \cdot \frac{\sigma_0}{\sqrt{n}} \leq \frac{\bar{X}_n - \gamma}{\sigma_0} \leq z_{1-\alpha_2} \cdot \frac{\sigma_0}{\sqrt{n}} \right) = \Phi(z_{1-\alpha_2}) - \Phi(-z_{1-\alpha_1}) \\ &= 1 - \alpha_1 - [1 - (1 - \alpha_2)] = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha. \end{aligned}$$

(Aquí fue utilizado el hecho de que para una variable aleatoria con distribución  $N(\gamma, \sigma_0^2)$ , la variable aleatoria  $\frac{\bar{X}_n - \gamma}{\sigma_0}$  posee una distribución  $N(0, 1)$ , ver en 9.4 la primera observación después del teorema 2.)

Observemos que la longitud (en este caso no aleatoria) del intervalo de confianza es igual a  $(z_{1-\alpha_1} + z_{1-\alpha_2}) \frac{\sigma_0}{\sqrt{n}}$ ; ella se hace mínima para  $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ , es decir, para el llamado intervalo de confianza simétrico.

b)  $\gamma = \mu$ ,  $\sigma^2$  (desconocida).

**Teorema 2.** Sean  $a_1$  y  $a_2$  números positivos con  $a_1 + a_2 = \alpha$ . Entonces

$$J(X_1, \dots, X_n) = \left[ \bar{X}_n - t_{n-1; 1-\alpha_1} \sqrt{\frac{S_n^2}{n}}, \bar{X}_n + t_{n-1; 1-\alpha_1} \sqrt{\frac{S_n^2}{n}} \right] \quad (2)$$

es un intervalo de confianza para  $\gamma$  con el nivel de confiabilidad  $1-\alpha$ ; aquí  $t_{n-1; \beta}$  denota al percentil de orden  $\beta$  de la distribución  $t$  con  $n-1$  grados de libertad.

**Observación.** En comparación con el intervalo de confianza (1) indicado en el teorema 1, han sido intercambiados en (2)  $\sigma_0^2$  y los percentiles de la distribución  $N(0, 1)$  por  $S_n^2$  y los percentiles de la distribución  $t$  con  $n-1$  grados de libertad, respectivamente.

Demostración

$$\begin{aligned} B(\gamma, \gamma) &= P_\gamma J(X_1, \dots, X_n) \ni \gamma \\ &= P_\gamma \left( \bar{X}_n - t_{n-1; 1-\alpha_1} \sqrt{\frac{S_n^2}{n}} \leq \gamma \leq \bar{X}_n + t_{n-1; 1-\alpha_1} \sqrt{\frac{S_n^2}{n}} \right) \\ &= P_\gamma \left( -t_{n-1; 1-\alpha_1} \leq \frac{\bar{X}_n - \gamma}{\sqrt{\frac{S_n^2}{n}}} \leq t_{n-1; 1-\alpha_1} \right) \\ &= 1 - \alpha_1 - [1 - (1 - \alpha_1)] = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha. \end{aligned}$$

(Aquí fue utilizado el hecho de que para una variable aleatoria con distribución  $N(\gamma, \sigma^2)$ , la variable aleatoria  $\frac{\bar{X}_n - \gamma}{\sqrt{\frac{S_n^2}{n}}}$  posee una distribución  $t$  con  $n-1$  grados de libertad, ver 9.4, teorema 5.)

Observemos que el valor esperado de la longitud del intervalo de confianza para  $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$  se hace mínimo.

c)  $\gamma = \sigma^2$ ,  $\mu = \mu_0$  (conocida)

**Teorema 3.** Sean  $a_1$  y  $a_2$  números positivos con  $a_1 + a_2 = \alpha$ . Entonces

$$J(X_1, \dots, X_n) = \left[ \frac{n S_n^{*2}}{\chi_{n, 1-a_1}^2}, \frac{n S_n^{*2}}{\chi_{n, a_2}^2} \right] \text{ con } S_n^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2 \quad (3)$$

es un intervalo de confianza para  $\gamma$  con el nivel de confiabilidad  $1-\alpha$ ; aquí  $\chi_n^2$  denota el percentil de orden  $\beta$  de la distribución  $\chi^2$  con  $n$  grados de libertad.

Demostración

$$\begin{aligned} B(\gamma, \gamma) &= P_\gamma J(X_1, \dots, X_n) \ni \gamma = P_\gamma \left( \frac{n S_n^{*2}}{\chi_{n, 1-a_1}^2} \leq \gamma \leq \frac{n S_n^{*2}}{\chi_{n, a_2}^2} \right) \\ &= P_\gamma \left( \chi_{n, a_2}^2 \leq \frac{n S_n^{*2}}{\gamma} \leq \chi_{n, 1-a_1}^2 \right) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha. \end{aligned}$$

(Aqui fue utilizado el hecho que para una variable aleatoria  $X$  con distribución  $N(\mu, \gamma)$ , la variable aleatoria  $\frac{n S_n^2}{\gamma}$  posee una distribución  $\chi^2$  con  $n$  grados de libertad, ver 9.4, teorema 3.)

d)  $\gamma = \sigma^2$ ,  $\mu$  (desconocido)

**Teorema 4.** Sean  $a_1$  y  $a_2$  números positivos con  $a_1 + a_2 = a$ . Entonces

$$J(X_1, \dots, X_n) = \left[ \frac{(n-1)S_n^2}{\chi_{n-1; 1-a_1}^2}, \frac{(n-1)S_n^2}{\chi_{n-1; a_2}^2} \right] \quad (4)$$

es un intervalo de confianza para  $\gamma$  con el nivel de confiabilidad  $1-a$ .

Demostración

$$\begin{aligned} B(\gamma, \gamma) &= P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = P_\gamma \left( \frac{(n-1)S_n^2}{\chi_{n-1; 1-a_1}^2} \leq \gamma \leq \frac{(n-1)S_n^2}{\chi_{n-1; a_2}^2} \right) \\ &= P_\gamma \left( \chi_{n-1; a_2}^2 \leq \frac{(n-1)S_n^2}{\gamma} \leq \chi_{n-1; 1-a_1}^2 \right) = 1 - a_1 - a_2 = 1 - a. \end{aligned}$$

(Aquí fue utilizado el hecho de que para una variable aleatoria con distribución  $N(\mu, \gamma)$ , la variable aleatoria  $\frac{(n-1)S_n^2}{\gamma}$  posee una distribución  $\chi^2$  con  $n-1$  grados de libertad, ver 9.4, teorema 4.)

## 10.6.2 Intervalo de confianza para una probabilidad desconocida

Sea  $A$  un suceso aleatorio que ocurre en el marco de un experimento con la probabilidad  $p$ ;  $p$  sea desconocida ( $0 < p < 1$ ). Consideremos la variable aleatoria  $X$ ,

$$X = \begin{cases} 1, & \text{en caso de que } A \text{ ocurra,} \\ 0, & \text{en caso de que } \bar{A} \text{ ocurra,} \end{cases}$$

y planteémonos la tarea de indicar para el parámetro  $\gamma = p$  un intervalo de confianza con el nivel de confiabilidad  $1-a$ , sobre la base de una muestra matemática de tamaño  $n$  de la población  $X$ . Para ello tomemos por base el estadígrafo

$$M = F_n(A) = \sum_{i=1}^n X_i$$

que proporciona la frecuencia absoluta de  $A$  en  $n$  experimentos. La variable aleatoria  $M$  posee una distribución binomial con los parámetros  $n$  y  $\gamma$ , en el caso que  $\gamma$  sea el valor verdadero del parámetro. Expresamos el intervalo de confianza  $J(X_1, \dots, X_n)$  en la forma

$$J(X_1, \dots, X_n) = [p_1(M), p_2(M)]; \quad (5)$$

luego, los extremos deben ser funciones de la variable aleatoria  $M$ .

**Teorema 5.** El intervalo de confianza (5) es un intervalo de confianza con el nivel de confiabilidad  $1-a$ , si para toda realización  $m$  de  $M$  los extremos  $p_1(m)$  y  $p_2(m)$  del intervalo de confianza concreto  $[p_1(m), p_2(m)]$  están fijados de modo que se cumplen las relaciones

$$\sum_{k=m}^n \binom{n}{k} [p_1(m)]^k [1-p_1(m)]^{n-k} = \frac{a}{2} \quad (6)$$

y

$$\sum_{k=0}^m \binom{n}{k} [p_2(m)]^k [1-p_2(m)]^{n-k} = \frac{a}{2} \quad (7)$$

Renunciaremos a la demostración de esta proposición. Los extremos del intervalo  $p_1(M)$  y  $p_2(M)$  pueden ser tomados de tablas y diagramas para  $\alpha$  especiales ( $\alpha=5\%$ ,  $\alpha=1\%$ ) y  $n$  no muy grandes ( $n \leq 30$ ). Para  $n$  mayores se utilizan fórmulas para el cálculo de los límites de confianza que se obtienen del teorema siguiente.

**Teorema 6.** Para el intervalo de confianza (5) se cumple que

$$\lim_{n \rightarrow \infty} p_\gamma ([(p_1(M), p_2(M))] \ni y) \geq 1 - \alpha \quad (0 < \gamma < 1) \quad (8)$$

(o sea, (5) es un intervalo de confianza con el nivel de confiabilidad  $1 - \alpha$  para  $n \rightarrow \infty$ , si se hace

$$p_1(M) = \frac{2M + z_{1-\frac{\alpha}{2}}^2 - z_{1-\frac{\alpha}{2}}}{2 \left( n + z_{1-\frac{\alpha}{2}}^2 \right)} \sqrt{4n \frac{M}{n} \left( 1 - \frac{M}{n} \right) + z_{1-\frac{\alpha}{2}}^2} \quad (9)$$

y

$$p_2(M) = \frac{2M + z_{1-\frac{\alpha}{2}}^2 + z_{1-\frac{\alpha}{2}}}{2 \left( n + z_{1-\frac{\alpha}{2}}^2 \right)} \sqrt{4n \frac{M}{n} \left( 1 - \frac{M}{n} \right) + z_{1-\frac{\alpha}{2}}^2} \quad (10)$$

y  $z_{1-\frac{\alpha}{2}}$  denota al percentil de orden  $1 - \frac{\alpha}{2}$  de la distribución normal estandarizada.

La demostración de este teorema se basa esencialmente en el Teorema Integral de De Moivre-Laplace (ver 7.5. teorema 1), según el cual se cumple en particular que

$$\lim_{n \rightarrow \infty} P_\gamma \left( \left| \frac{M - n\gamma}{\sqrt{n\gamma(1-\gamma)}} \right| \leq z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

De aquí se obtiene, después de algunos cálculos, los límites de confianza indicados en (9) y (10).

Ilustraremos el teorema 6 con un ejemplo numérico.

**Ejemplo numérico.** Para  $n=200$  y  $m=88$ , se obtiene como valor estimado para la probabilidad desconocida  $\frac{88}{200}=0.44$ . Si escogemos  $\alpha=5\%$ , entonces  $z_{1-\frac{\alpha}{2}}=1.96$ , y obtenemos como límite de confianza inferior el número 0.37, según (9), y como límite de confianza superior el número 0.51, según (10). Como intervalo estimado concreto para la probabilidad desconocida se tiene el intervalo  $[0.37; 0.51]$ . Si escogemos por el contrario  $\alpha=1\%$  entonces obtenemos como intervalo estimado concreto el  $[0.35; 0.53]$ .

Por último queremos advertir que existen medios gráficos auxiliares para el cálculo de los límites de confianza concretos.

### 10.6.3 Intervalo de confianza para una función de distribución desconocida

El problema de la estimación por intervalo de confianza del valor (desconocido) de la función de distribución  $F$  de una variable aleatoria  $X$  en un lugar  $x \in \mathbb{R}^1$ , es equivalente con

el problema de la estimación por intervalo de confianza de la probabilidad del suceso aleatorio ( $X < x$ ). Así, este se puede tratar, en principio, con los métodos expuestos en 10.6.2.

No obstante, queremos explicar otra posibilidad para el tratamiento de este problema. Esta se basa sobre la estrecha relación entre la función de distribución empírica  $W_n$  de una muestra matemática  $(X_1, \dots, X_n)$  de tamaño  $n$  de la población  $X$  y la función de distribución  $F$  de esta población, aclarada en el epígrafe 9.3. Para ello supongamos que  $F$  es continua.

**Teorema 7.** Para el intervalo aleatorio

$$J_x(X_1, \dots, X_n) = \left[ W_n(x) - \frac{y_a}{\sqrt{n}}, W_n(x) + \frac{y_a}{\sqrt{n}} \right], \quad (11)$$

considerado como intervalo de confianza para  $\gamma = F(x)$ , se cumple que

$$\lim_{n \rightarrow \infty} P_\gamma(J_x(X_1, \dots, X_n) \ni \gamma) \geq 1 - \alpha \quad (x \in \mathbb{R}^1, 0 \leq \gamma \leq 1) \quad (12)$$

(o sea, (11) es un intervalo de confianza con el nivel de confiabilidad  $1 - \alpha$  para  $n \rightarrow \infty$ ); aquí  $y_a$  es solución de la ecuación

$$k(y) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2ky} = 1 - \alpha. \quad (13)$$

**Demostración.** Se cumple que

$$\begin{aligned} \lim_{n \rightarrow \infty} P_\gamma(J_x(X_1, \dots, X_n) \ni \gamma) &= \lim_{n \rightarrow \infty} P_\gamma \left( W_n(x) - \frac{y_a}{\sqrt{n}} < \gamma < W_n(x) + \frac{y_a}{\sqrt{n}} \right) \\ &= \lim_{n \rightarrow \infty} P_\gamma \left( \sqrt{n} |W_n(x) - \gamma| < y_a \right) \\ &\geq \lim_{n \rightarrow \infty} P_\gamma \left( \sqrt{n} \sup_{-\infty < x < \infty} |W_n(x) - \gamma| < y_a \right) \\ &= K(y_a) = 1 - \alpha; \end{aligned}$$

aquí hemos utilizado el teorema 3 (9.3) (que a decir verdad no hemos demostrado en este libro).

Para una muestra concreta  $(x_1, \dots, x_n)$  se calcula la función de distribución empírica correspondiente  $w_n$  (ver 9.3, definición 1) y se utiliza -suponiendo un tamaño de la muestra suficientemente grande

$$J_x(x_1, \dots, x_n) = \left[ W_n(x) - \frac{y_a}{\sqrt{n}}, W_n(x) + \frac{y_a}{\sqrt{n}} \right] \quad (14)$$

como intervalo estimado concreto para  $F(x)$ ; el número  $y_a$  puede ser tomado de tablas. La ventaja consiste evidentemente en que se obtienen simultáneamente para todo  $x \in \mathbb{R}^1$  intervalos de confianza concretos para  $F(x)$ . Para la aplicación de esta estimación por intervalo de confianza se pueden utilizar medios gráficos auxiliares.

## 11. Introducción a la teoría de la docimasia de hipótesis

El siguiente capítulo contiene una introducción a la teoría de la docimasia de hipótesis, un campo central ampliamente extendido de la Estadística matemática. Ella ha encontrado una gran aplicación en las más diversas disciplinas científicas. La tarea que se plantea, de forma general, la teoría de la docimasia de hipótesis, consiste en indicar métodos y procedimientos adecuados e investigar, sobre la base de la Teoría de probabilidades, con cuáles de ellos pueden realizarse decisiones objetivas sobre hipótesis -estas son suposiciones en el marco de un modelo estocástico con ayuda de muestras. La ventaja de la utilización de tales procedimientos de decisión estriba también en que permiten valorar cuantitativamente el número de las posibles decisiones erróneas.

Después de la introducción de los conceptos básicos fundamentales de la teoría de la docimasia de hipótesis (epígrafe 11.2), pasaremos a la denominada dócima de significación (epígrafe 11.3) e indicaremos para ella una serie de ejemplos en los epígrafes 11.4 y 11.5 (entre ellos, dócima  $t$ , dócima  $F$  y dócima  $\chi^2$ ). Por último el epígrafe 11.6 contiene un ejemplo de aplicación.

### 11.1 Tareas que se plantea la teoría de la docimasia de hipótesis

Como se bosquejó ya, la tarea fundamental que se plantea la teoría de la docimacia de hipótesis, consiste en indicar y analizar métodos para la verificación de suposiciones acerca de parámetros desconocidos de un modelo estocástico, denominadas hipótesis estadísticas (o brevemente: hipótesis), sobre la base de muestras. La verificación de una hipótesis se realiza con ayuda de una denominada dócima de hipótesis (o brevemente: dócima). Una dócima tiene por objeto producir una decisión acerca de la aceptación o rechazo de una hipótesis, sobre la base de la muestra. Si contamos con una muestra concreta

$(x_1, \dots, x_n)$ , entonces se tomará con ayuda de una dócima la decisión “se rechaza la hipótesis” o la decisión “se acepta la hipótesis”. (Advertimos expresamente que la decisión “se acepta la hipótesis” no significa que ella sea correcta; ver también 9.1.) Luego, una dócima se puede caracterizar en principio por el conjunto de todos los  $(x_1, \dots, x_n)$ , que provocan la decisión “se rechaza la hipótesis”. Este conjunto se denomina *región crítica* o *región de rechazo* (de la hipótesis considerada).

Antes de que nos ocupemos más exactamente en el epígrafe 11.2 de los conceptos básicos mencionados y de otros de la teoría de la docimasia de hipótesis, y en especial, con las exigencias mínimas para establecer de forma adecuada lo que llamamos una región crítica, queremos considerar un ejemplo para ilustrar la problemática y también el procedimiento típico que se utiliza.

**Ejemplo.** Supongamos que la población  $X$  posee una distribución normal con varianza  $D^2X=\sigma_0^2$  ( $\sigma_0$  conocida, por ejemplo,  $\sigma_0=1$ ); el valor esperado  $EX$  sea desconocido. Hagamos  $\gamma=EX$  y designemos con  $\gamma_0$  el valor verdadero (pero desconocido) del parámetro  $\gamma$ . Queremos verificar la hipótesis  $H: \gamma_0=\gamma^*$  con ayuda de una muestra matemática  $(X_1, \dots, X_n)$  de tamaño  $n$  de la población  $X$  ( $\gamma^*$  es un número real dado; puede ser un valor supuesto, pretendido o también dudosos para el parámetro desconocido; con frecuencia tiene el significado de un valor previsto). Para lograr lo anterior consideremos el estadígrafo  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , el cual representa un estimador apropiado para  $\gamma$  (ver 10.4.1). En el caso de que la hipótesis  $H: \gamma_0=\gamma^*$  sea verdadera,  $\bar{X}_n$  posee una distribución  $N\left(\gamma^*, \frac{\sigma_0^2}{n}\right)$  (ver teorema 2(9.4)) y de esto se deriva que  $T = \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sigma_0}$  posee una distribución  $N(0,1)$ . Para una muestra concreta  $(x_1, \dots, x_n)$  se rechazarán la hipótesis  $H: \gamma_0=\gamma^*$  cuando el módulo del valor  $t = \sqrt{n} \frac{\bar{x}_n - \gamma^*}{\sigma_0}$  calculado, se haga muy grande (ver 5.4, fórmulas (17) hasta (19) y figura 35). Para precisar este procedimiento daremos un número pequeño  $\alpha$  ( $0 < \alpha < 1$ , por ejemplo,  $\alpha=0.05$ ) y determinaremos un número  $t^* > 0$ , de modo que se cumpla que  $P_{\gamma^*}(|T| > t^*) = \alpha$ , o sea, de manera que la probabilidad de que se rechace la hipótesis  $H: \gamma_0=\gamma^*$  sea igual al número  $\alpha$  dado- en el caso de que la hipótesis sea cierta. En virtud de que  $P_{\gamma^*}(|T| > t^*) = 1 - P_{\gamma^*}(|T| \leq t^*) = 1 - (2\Phi(t^*) - 1) = 2(1 - \Phi(t^*)) = \alpha$ .

Se obtiene para  $t^*$  el percentil de orden  $1 - \frac{\alpha}{2}$  de la distribución normal estandarizada, o sea,  $t^* = z_{1-\frac{\alpha}{2}}$ ; ver también la figura 47. (Para  $\alpha=0.05$ , se obtiene  $t^*=1.96$ .) Si se cumple la inecuación  $|t| > z_{1-\frac{\alpha}{2}}$  para el valor  $t = \sqrt{n} \frac{\bar{x}_n - \gamma^*}{\sigma_0}$  calculado a partir de una muestra concreta  $(x_1, \dots, x_n)$ , entonces se rechaza la hipótesis  $H: \gamma_0=\gamma^*$ , en caso contrario no se rechaza. Con esto, la región crítica  $K$  de  $H$  está dada por

$$K = \left\{ (x_1, \dots, x_n) : \left| \sqrt{n} \frac{\bar{x}_n - \gamma^*}{\sigma_0} \right| > z_{1-\frac{\alpha}{2}} \right\}$$

y se cumple que:

$$P_{\gamma^*}((X_1, \dots, X_n) \in K) = P_{\gamma^*}(|T| > z_{1-\frac{\alpha}{2}}) = \alpha.$$

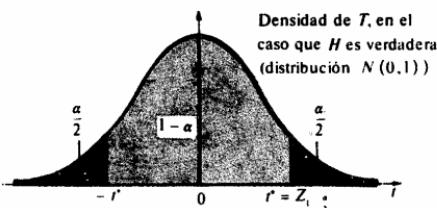


Figura 47

La probabilidad de que se rechace  $H: \gamma_0 = \gamma^*$  es igual a  $\alpha$  -en el caso de que  $H$  sea verdadera. Aquí no hemos reparado en la probabilidad de que la hipótesis  $H: \gamma_0 = \gamma^*$  no se rechace en el caso de que sea falsa-, o sea, no hemos prestado atención a  $P_\gamma(T \leq z_{1-\frac{\alpha}{2}})$  para  $\gamma_0 \neq \gamma^*$ . Por tanto, con el procedimiento indicado comprobamos sólo si la hipótesis  $H$  es compatible con la muestra o si existen diferencias significativas.

## 11.2 Conceptos fundamentales de la teoría de la docimacia de hipótesis

En la formulación matemática general de la tarea que se plantea la teoría de la docimacia de hipótesis partimos de una población  $X$ , cuya función de distribución  $F$  depende de un parámetro  $\gamma \in \Gamma$ . Designemos nuevamente con  $\gamma_0$  el valor verdadero (pero desconocido) del parámetro. Por una hipótesis (estadística) entendemos una proposición de la forma:  $\gamma_0$  es un elemento de un subconjunto no vacío dado  $\Gamma_0$  de  $\Gamma$ . Para ello escribimos abreviadamente  $H: \gamma_0 \in \Gamma_0$ . Si  $\Gamma_0$  contiene un solo elemento,  $\Gamma_0 = \{\gamma^*\}$ , entonces se habla de una **hipótesis simple** y escribimos  $H: \gamma_0 = \gamma^*$ . En el otro caso la hipótesis  $H: \gamma_0 \in \Gamma_0$  se denomina una **hipótesis compuesta**. Si junto a una hipótesis  $H_0$ ;  $\gamma_0 \in \Gamma_0$  se considera otra hipótesis  $H_A$ ;  $\gamma_0 \in \Gamma_1 \subseteq \Gamma \setminus \Gamma_0$ , entonces  $H_0$  se denomina **hipótesis nula** y  $H_A$  **hipótesis alternativa**.

Sea ahora  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$ . Entendemos por una docimia, más exactamente, por una docimia de la hipótesis nula  $H_0$  frente a la hipótesis alternativa  $H_A$ , un procedimiento con el cual es posible una comparación de las hipótesis  $H_0$  y  $H_A$  con respecto a la muestra  $(X_1, \dots, X_n)$  y que conduce para toda muestra concreta  $(x_1, \dots, x_n)$  a una de las decisiones " $H_0$  se rechaza ( $H_A$  se acepta)" o " $H_A$  se rechaza ( $H_0$  se acepta)". En lo sucesivo nos limitaremos fundamentalmente al caso de la hipótesis alternativa  $H_A: \gamma_0 \in \Gamma \setminus \Gamma_0$  y nombraremos sencillamente una docimia de  $H_0: \gamma_0 \in \Gamma_0$  frente a esta hipótesis alternativa una docimia de  $H_0$ . Aquí utilizaremos para las decisiones correspondientes las formulaciones " $H_0$  se rechaza" y " $H_0$  no se rechaza", y evitaremos hablar en este caso de la aceptación de la hipótesis  $H_0$ . Una docimia semejante se describe completamente a través del conjunto  $K$  de todas las muestras concretas  $(x_1, \dots, x_n)$ , para las cuales se toma la decisión " $H_0$  se rechaza", o sea, a través de la **región crítica** o **región de rechazo** de  $H_0$ . Luego, no es necesario diferenciar entre una docimia y la región crítica  $K$  correspondiente; en el futuro hablaremos de la docimia  $K$ , si la docimia posee la región crítica  $K$ . Con esto nada se ha dicho aún sobre el establecimiento adecuado de la región crítica. Antes que nos ocupemos con ciertas exigencias mínimas que se deben ob-

servar en el establecimiento de la región crítica, queremos considerar los posibles errores en el procedimiento de decisión que se realiza en el marco de una dócima:

	$H_0: \gamma_0 \in \Gamma_0$ es verdadera	$H_A: \gamma_0 \in \Gamma \setminus \Gamma_0$ es verdadera
$H_0$ se rechaza	Decisión falsa (error de primer tipo)	Decisión correcta
$H_0$ no se rechaza	Decisión correcta	Decisión falsa (error de segundo tipo)

Un *error de primer tipo* se presenta siempre y cuando la muestra concreta esté situada en la región crítica de  $H_0$  y  $H_0$  sea verdadera. Las probabilidades de cometer errores de primer tipo se pueden estimar (según lo expuesto) mediante

$$\sup_{\gamma \in \Gamma_0} P_\gamma(X_1, \dots, X_n) \in K;$$

en el caso de una hipótesis simple  $H_0: \gamma_0 = \gamma^*$ , la probabilidad de un error de primer tipo es igual a  $P_{\gamma^*}(X_1, \dots, X_n) \in K$ .

Un *error de segundo tipo* se presenta siempre y cuando la muestra concreta no esté situada en la región crítica de  $H_0$  y  $H_A$  sea verdadera las probabilidades de cometer errores de segundo tipo se pueden estimar de forma correspondiente mediante

$$\sup_{\gamma \in \Gamma \setminus \Gamma_0} P_\gamma(X_1, \dots, X_n) \notin K = 1 - \inf_{\gamma \in \Gamma \setminus \Gamma_0} P_\gamma(X_1, \dots, X_n) \in K.$$

Esto nos conduce a valorar una dócima  $K$  de  $H_0$  por medio de la función de potencia definida a continuación

Definición 1. Sea  $K$  una dócima de  $H_0$ . Entonces la función definida sobre  $\Gamma$  por

$$G(\gamma) = P_\gamma((X_1, \dots, X_n) \in K) \quad (\gamma \in \Gamma) \quad (1)$$

se denomina *función de potencia* de la dócima  $K$  (fig. 48).

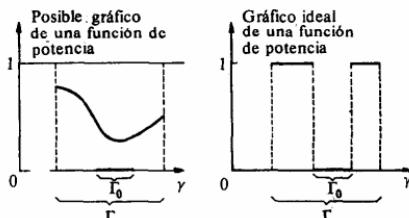


Figura 48

Por tanto, el valor de la función de potencia en el punto  $\gamma (\in \Gamma)$  indica la probabilidad de que la hipótesis  $H_0$  se rechace, calculada bajo la suposición de que  $\gamma$  es el valor verdadero del parámetro. Las probabilidades de cometer errores de primer tipo se describen por medio del gráfico de  $G$  sobre  $\Gamma_0$ , las probabilidades de cometer errores de segundo tipo por medio del gráfico de  $1-G$  sobre  $\Gamma \setminus \Gamma_0$ .

Ejemplo 1. Calculemos la función de potencia  $G$  de la dócima indicada en el epígrafe 11.1 de la hipótesis  $H_0: \gamma_0 = \gamma^*$ , para una población  $X$  con distribución  $N(\gamma_0, \sigma_0^2)$  y con  $\gamma_0$  desconocido y  $\sigma_0^2$  conocido. Para  $\gamma \in \Gamma = \mathbb{R}^1$  se cumple que

$$G(\gamma) = P_\gamma((X_1, \dots, X_n) \in K) = P_\gamma(|T| > z_{1-\frac{\alpha}{2}})$$

$$= 1 - P_{\gamma}(|T| \leq z_{1-\frac{\alpha}{2}}) = 1 - P_{\gamma} \left( -z_{1-\frac{\alpha}{2}} < \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sigma_0} < z_{1-\frac{\alpha}{2}} \right)$$

$$= 1 - P_{\gamma} \left( -z_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} < \sqrt{n} \frac{\bar{X}_n - \gamma}{\sigma_0} < z_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} \right)$$

Observemos ahora que para una variable aleatoria  $X$  con distribución  $N(\gamma, \sigma_0^2)$ , la variable aleatoria  $\sqrt{n} \frac{\bar{X}_n - \gamma}{\sigma_0}$  posee una distribución  $N(0,1)$  (ver en 9.4 la observación 1 después del teorema 2), de modo que con  $\Phi(-x) = 1 - \Phi(x)$  (ver 5.4 (15)) obtenemos (fig. 49)

$$G(\gamma) = 1 - \Phi \left( z_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} \right) + \Phi \left( -z_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} \right)$$

$$= \Phi \left( \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}} \right) + \Phi \left( -\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}} \right)$$

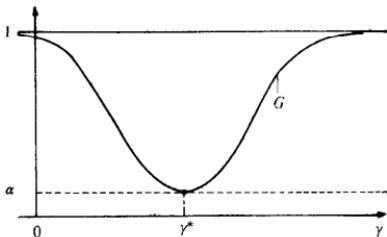


Figura 49

Ahora se intentará establecer la región crítica, de modo que las probabilidades de cometer errores de primer y segundo tipos sean lo más pequeñas posibles. Como no se pueden minimizar ambas al mismo tiempo, se procede por lo general en la determinación de una dócima, de manera que se busca en la clase de todas las dócimas, para las cuales las probabilidades de cometer errores de primer tipo no sobrepasan un número  $\alpha$  dado ( $0 < \alpha < 1$ ), una para la cual las probabilidades de cometer errores de segundo tipo se hagan mínimas. La exigencia de que las probabilidades de cometer errores de primer tipo no sobrepasan una cota  $\alpha$  dada, se considera una exigencia mínima para una dócima.

**Definición 2.** Sea  $\alpha$  ( $0 < \alpha < 1$ ) un número dado. Entonces una dócima  $K$  de  $H_0$ :  $\gamma_0 \in \Gamma_0$  con la función de potencia  $G$  se denomina una *dócima de significación con el nivel de significación  $\alpha$*  (también: *dócima de significación con la seguridad estadística  $1 - \alpha$* ), si se cumple que (fig. 50)

$$G(\gamma) = P_{\gamma}(X_1, \dots, X_n \in K) \leq \alpha \quad (\gamma \in \Gamma_0). \quad (2)$$

**Ejemplo 2.** La dócima indicada en el epígrafe 11.1 de  $H_0: \gamma_0 = \gamma^*$  para una población  $X$  con distribución  $N(\gamma_0, \sigma_0^2)$  y con  $\gamma_0$  desconocido y  $\sigma_0^2$  conocido, es una dócima de significación con el nivel de significación  $\alpha$  (ver también el ejemplo 1; se cumple que

$$G(\gamma^*) = \Phi(-z_{1-\frac{\alpha}{2}}) + \Phi(-z_{1-\frac{\alpha}{2}}) = 2 \left( 1 - \left( 1 - \frac{\alpha}{2} \right) \right) = \alpha$$

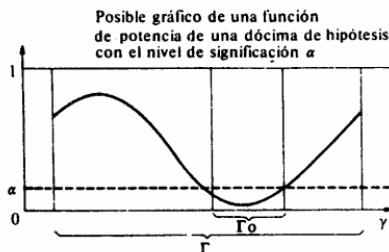


Figura 50

En el epígrafe 11.3 nos ocuparemos aún más detalladamente de las dócimas de significación; los epígrafes 11.4 y 11.5 contienen una serie de ejemplos importantes de tales dócimas.

Nos ocuparemos un poco de la valoración y la comparación de dócimas. El tratamiento de estas tareas se realiza por medio de las funciones de potencia, sobre la base de las definiciones siguientes:

**Definición 3.** Una dócima  $K$  de  $H_0$ :  $\gamma_0 \in \Gamma_0$  con la función de potencia  $G$  se llama *admisible*, si se cumple que

$$\inf_{\gamma \in \Gamma_0} G(\gamma) \geq \sup_{\gamma \in \Gamma_0} G(\gamma). \quad (3)$$

Si  $H_0$  es una hipótesis simple ( $H_0: \gamma_0 = \gamma^*$ ), entonces una dócima de  $H_0$  es, según definición, admisible si se cumple que

$$G(\gamma) \geq G(\gamma^*) \quad (\gamma \in \Gamma). \quad (4)$$

Luego, para una dócima admisible de  $H_0$  la probabilidad de que se rechace  $H_0$  siendo  $H_0$  una hipótesis falsa, no es menor que para el caso en que  $H_0$  sea una hipótesis verdadera, hablando sin mucha precisión.

**Ejemplo 3.** Consideremos de nuevo la dócima expuesta en el epígrafe 11.1 Para la función de potencia de esta dócima se cumple (ver ejemplo 1) que

$$G(\gamma) = \Phi\left(\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right)$$

Se verifica fácilmente que se cumple

$$G(\gamma) > \Phi\left(-z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-z_{1-\frac{\alpha}{2}}\right) = G(\gamma^*) = \alpha$$

para todo  $\gamma \neq \gamma^*$ , es decir, que la dócima tomada por base es admisible (fig. 49).

**Definición 4.** Sean  $K_1$  y  $K_2$  dos dócimas de  $H_0$ :  $\gamma_0 \in \Gamma_0$  con las funciones de potencia  $G_1$  y  $G_2$ , respectivamente. La dócima  $K_1$  se denomina *mejor*, si se cumple que

$$G_1(\gamma) \geq G_2(\gamma) \quad (\gamma \in \Gamma \setminus \Gamma_0). \quad (5)$$

Si  $K_1$  es mejor que  $K_2$ , entonces la probabilidad de que se rechace la hipótesis  $H_0$  para la dócima  $K_1$ , calculada bajo la suposición de que  $\gamma \in \Gamma \setminus \Gamma_0$  es el valor verdadero del parámetro, es para todo  $\gamma$  semejante al menos tan grande como para la dócima  $K_2$ , o -hablando sin mucha precisión- la probabilidad de rechazo de una hipótesis falsa es para  $K_1$  al menos tan grande como para  $K_2$ .

En todas las consideraciones hechas hasta ahora, hemos tomado por base un tamaño de la muestra constante. Radica en la naturaleza de la situación el que se puedan hacer proposiciones, por lo general más confiables, a medida que crece el tamaño  $n$  de la muestra; más confiables en el sentido de una disminución de las probabilidades de cometer errores de primer y segundo tipos. Por ello se investigan su-

cesiones  $(K_n)$  de dócimas -en particular, dócimas de significación con el nivel de significación  $\alpha$  ( $0 < \alpha < 1$ , dado como dato, independiente de  $n$ ) en dependencia de  $n$ : por consiguiente, aquí se cumple para las regiones críticas que  $K_n \subseteq \mathbb{R}^n$  ( $n \in \mathbb{N}$ ).

**Definición 5.** Sea  $(K_n)$  una sucesión de dócimas  $K_n$  de  $H_0$ :  $\gamma_0 \in \Gamma_0$  con la función de potencia  $G_n(n \in \mathbb{N})$ . La sucesión  $(K_n)$  se llama *consistente*, si se cumple que

$$\lim_{n \rightarrow \infty} G_n(\gamma) = 1 \quad (\gamma \in \Gamma \setminus \Gamma_0). \quad (6)$$

Por tanto, para una sucesión consistente  $(K_n)$  la probabilidad de que se rechace  $H_0$ , calculada bajo la suposición de que  $\gamma \in \Gamma \setminus \Gamma_0$  es el valor verdadero del parámetro, converge cuando  $n \rightarrow \infty$  hacia 1, o -hablando sin mucha precisión- la probabilidad de rechazo de una hipótesis falsa tiende a 1.

**Ejemplo 4.** Consideremos la sucesión  $(K_n)$  de dócimas de  $H_0$ :  $\gamma_0 = \gamma^*$  para una población  $X$  con distribución  $N(\gamma_0, \sigma_0^2)$  y con  $\gamma_0$  desconocido y  $\sigma_0^2$  conocida; aquí  $K_n$  es la dócima de significación indicada en el epígrafe 11.1 con el nivel de significación  $\alpha$ . Para la función de potencia  $G_n$  se cumple (ver el ejemplo 1) que

$$\begin{aligned} \lim_{n \rightarrow \infty} G_n(\gamma) &= \lim_{n \rightarrow \infty} \left[ \Phi\left(\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right) \right] \\ &= \begin{cases} 1+0=1 & \text{para } \gamma > \gamma^* \\ 0+1=1 & \text{para } \gamma < \gamma^* \end{cases} = 1 \quad \text{para } \gamma \neq \gamma^*. \end{aligned}$$

o sea, la sucesión  $(K_n)$  es consistente.

### 11.3 Procedimiento general para realizar una dócima de significación

De acuerdo con la definición, se entiende por dócima de significación con el nivel de significación  $\alpha$  ( $0 < \alpha < 1$ , dado) una dócima de  $H_0$ :  $\gamma_0 \in \Gamma_0$  con la región crítica  $K$ , cuya función de potencia  $G$  satisface la condición

$$G(\gamma) = P_\gamma((X_1, \dots, X_n) \in K) \leq \alpha \quad (\gamma \in \Gamma_0) \quad (1)$$

(ver 11.2, definición 2). Luego, en una dócima de significación las probabilidades de cometer errores de primer tipo ( $H_0$  se rechaza, aunque  $H_0$  sea verdadera) no sobrepasan un número prefijado  $\alpha$  -el nivel de significación; errores de segundo tipo ( $H_0$  no se rechaza, aunque  $H_0$  sea falsa) no se toman en consideración. Por ello, las dócimas de significación se utilizan solo cuando, sobre la base de una muestra concreta  $(x_1, \dots, x_n)$  de la población  $X$  considerada, debe valorarse si una hipótesis  $H_0$  sobre la distribución de esta población es compatible con la muestra concreta  $(x_1, \dots, x_n)$ , o si se presentan diferencias significativas (aseguradas estadísticamente). En este último caso se rechaza  $H_0$  sobre la base de la dócima, en el otro nada se puede esgrimir en contra de la hipótesis  $H_0$ . El nivel de significación  $\alpha$  se debe fijar atendiendo al planteamiento concreto del problema y, en particular, a las consecuencias de un error de primer tipo; aquí no se trata propiamente de un interés matemático. (Con frecuencia se eligen en las aplicaciones  $\alpha=5\%$ ,  $\alpha=2\%$  o  $\alpha=1\%$ ).

En la determinación de la región crítica  $K \subseteq \mathbf{R}^n$  se procede por conveniencia, de modo que  $K$  se describa mediante condiciones impuestas a los valores de un estadígrafo apropiado  $T$ . Más exactamente, si  $\varphi$  es una función real definida sobre el conjunto  $\mathbf{R}^n$  y  $T$  denota al estadígrafo  $\varphi(X_1, \dots, X_n)$ ,  $T = \varphi(X_1, \dots, X_n)$ , entonces se elige para el nivel de significación  $\alpha$  prefijado una parte  $K^*$  (lo menor posible) de la imagen de  $T$ , tal que se cumpla que  $P_\gamma((X_1, \dots, X_n) \in K^*) \leq \alpha$  para todo  $\gamma \in \Gamma_0$ . Para la región crítica  $K = \{(x_1, \dots, x_n) : \varphi(x_1, \dots, x_n) \in K^*\}$  se cumple entonces que

$$P_\gamma((X_1, \dots, X_n) \in K) \leq \alpha \text{ para todo } \gamma \in \Gamma_0.$$

es decir,  $K$  es una dócima de significación con el nivel de significación  $\alpha$  (ver el ejemplo del epígrafe 11.1 allí es  $T = \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sigma_0}$  y  $K^* = \left\{ t : |t| > z_{1-\frac{\alpha}{2}} \right\}$ ).

La variable aleatoria  $T$  se llama en este contexto *variable de dócima*. Para fijar la región crítica imagen  $K^*$ , de modo que se cumpla que  $P_\gamma(T \in K^*) \leq \alpha$  ( $\gamma \in \Gamma_0$ ), se tiene que conocer totalmente la distribución de la variable de dócima  $T$  bajo la suposición de que " $H_0$  es verdadera", por lo menos asintóticamente -en el caso de que el tamaño de la muestra  $n$  sea grande (o sea, cuando  $n \rightarrow \infty$ ). Se recomienda utilizar como variables de dócima aquellas variables aleatorias que se deriven de estimaciones puntuales para el parámetro desconocido. Como  $K^*$  determina de forma unívoca la región crítica  $K$  de  $H_0$ , se puede renunciar a la indicación explícita de  $K$  y designar entonces  $K^*$  como región crítica o de rechazo de  $H_0$ .

En la mayoría de los casos  $K^*$  es de la forma  $\{t : t < a\}$ ,  $\{t : t > b\}$  o  $\{t : t < a \text{ o } t > b\}$ .

El procedimiento general para realizar una dócima de significación con el nivel de significación  $\alpha$  prefijado, se puede esquematizar de la manera siguiente (ver también el ejemplo a continuación):

#### 0. Condiciones sobre la población

#### 1. Planteamiento de una hipótesis $H_0$ .

#### 2. Construcción de una variable de dócima o estadígrafo $T$ .

#### 3. Elección de la *región crítica* $K^*$ .

4. Regla de decisión: Para una muestra concreta existente  $(x_1, \dots, x_n)$  se calcula el valor  $t$  del estadígrafo  $T$ . Si se cumple que  $t \in K^*$ , entonces se rechaza a  $H_0$ , en caso contrario ( $t \notin K^*$ ), nada hay que objetar contra  $H_0$  (fig. 51).

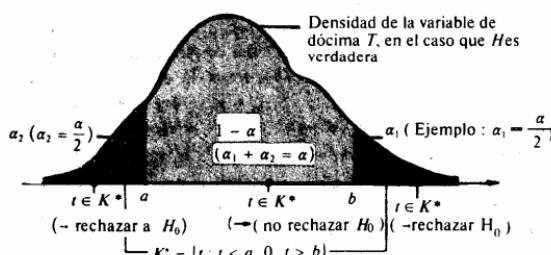


Figura 51

Los pormenores de una dócima, en particular, la elección del nivel de significación y de la región crítica, se deben prefijar necesariamente *antes* de la utilización de una muestra concreta. En caso contrario, es siempre posible -mediante una elección aceptable del nivel de significación y o mediante una fijación ingeniosa de la región crítica- proceder con la hipótesis "según nuestros deseos", por ejemplo, producir un rechazo si este es el deseo del que trabaja. Está claro que para un proceder semejante la aplicación de los métodos de la Estadística matemática pierde todo sentido objetivo.

Consideremos aún un ejemplo; con él queremos también abordar la estrecha relación entre las estimaciones por intervalo de confianza y las dócimas de significación.

### Ejemplos

0. Supongamos que  $X$  posee una distribución uniforme sobre el intervalo  $[0, \gamma_0]$  y que  $\gamma_0 > 0$  es desconocida.

1.  $H_0: \gamma_0 = \gamma^*$  ( $\gamma^*$  número positivo prefijado).

2.  $H_0$  es seguro falsa si para una muestra concreta  $(x_1, \dots, x_n)$  de la población  $X$  se cumple la relación  $\max\{x_1, \dots, x_n\} > \gamma^*$ . Esto sugiere utilizar como variable de dócima la variable aleatoria  $T = \max\{X_1, \dots, X_n\}$ , donde  $(X_1, \dots, X_n)$  es una muestra matemática de tamaño  $n$  de la población  $X$ . Si  $H_0: \gamma_0 = \gamma^*$  es verdadera, entonces la función de distribución  $F_T$  de la variable de dócima  $T$  está dada a través de

$$F_T(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ \left(\frac{x}{\gamma^*}\right)^n & \text{para } 0 \leq x \leq \gamma^*, \\ 1 & \text{para } x \geq \gamma^*, \end{cases}$$

(ver 10.2, ejemplo 2).

3. Establezcamos la región crítica en la forma  $K^* = \{t : t < a \text{ o } t > b\}$  con  $0 \leq a < b \leq \gamma^*$ ; la hipótesis  $H_0$  será rechazada si para una muestra concreta  $(x_1, \dots, x_n)$  se cumple una de las inecuaciones  $\max\{x_1, \dots, x_n\} < a$  o  $\max\{x_1, \dots, x_n\} > b$ . Ahora deben determinarse los números  $a$  y  $b$  de modo que se cumpla que

$$P_T(T \in K^*) = P_T(T < a) + P_T(T > b) = \alpha.$$

Para esto sean primeramente  $\alpha_1$  y  $\alpha_2$  números no negativos cualesquiera con  $\alpha_1 + \alpha_2 = \alpha$ . De

$$P_T(T < a) = F_T(a) = \left(\frac{a}{\gamma^*}\right)^n = \alpha_2 \text{ resulta que } a = \sqrt[n]{\alpha_2 \gamma^*}, \text{ y de}$$

$$P_T(T > b) = 1 - F_T(b) = 1 - \left(\frac{b}{\gamma^*}\right)^n = \alpha_1$$

resulta que  $b = \sqrt[n]{1 - \alpha_1} \gamma^*$ . Para  $K^* = \{t : t < \sqrt[n]{\alpha_2} \gamma^* \text{ o } t > \sqrt[n]{1 - \alpha_1} \gamma^*\}$  se cumple con esto que  $P_T(T \in K^*) = \alpha$

4. Regla de decisión: Si para una muestra concreta  $(x_1, \dots, x_n)$  se cumple una de las inecuaciones  $\max\{x_1, \dots, x_n\} < \sqrt[n]{\alpha_2} \gamma^*$  o  $\max\{x_1, \dots, x_n\} > \sqrt[n]{1 - \alpha_1} \gamma^*$ , entonces  $H_0: \gamma_0 = \gamma^*$  se rechaza; en caso contrario nada hay que objetar contra  $H_0$  sobre la base de esta dócima.

Con esto hemos descrito totalmente una d\'ocima de significaci\'on con el nivel de significaci\'on  $\alpha$  para la hip\'otesis  $H_0: \gamma_0 = \gamma^*$  sobre el par\'ametro  $\gamma_0$  de una variable aleatoria distribuida uniformemente sobre el intervalo  $[0, \gamma_0]$ . Para la ejercitaci\'on de los conceptos introducidos en el ep\'igrafe 11.2 retomaremos \'un \'ultimo ejemplo m\'as adelante.

La funci\'on de potencia  $G$  de esta d\'ocima est\'a dada, como el lector puede comprobar, a trav\'es de

$$G(\gamma) = \begin{cases} 1 & \text{para } 0 < \gamma \leq \sqrt[n]{\alpha_1} \gamma^*, \\ \alpha_1 \left( \frac{\gamma^*}{\gamma} \right)^n & \text{para } \sqrt[n]{\alpha_1} \gamma^* \leq \gamma \leq \sqrt[n]{1-\alpha_1} \gamma^*, \\ 1-(1-\alpha_1) \left( \frac{\gamma^*}{\gamma} \right)^n & \text{para } \sqrt[n]{1-\alpha_1} \gamma^* \leq \gamma. \end{cases}$$

Si consideramos la sucesi\'on  $(K_n^*)$  de d\'ocimas de significaci\'on con el nivel  $\alpha$  y con

$$K_n^* = \{t : t < \sqrt[n]{\alpha_1} \gamma^* \text{ o } t > \sqrt[n]{1-\alpha_1} \gamma^*\}, \alpha_1 + \alpha_2 = \alpha,$$

entonces se cumple para la sucesi\'on  $(G_n)$  de las funciones de potencia correspondientes la relaci\'on  $\lim_{n \rightarrow \infty} G_n(\gamma) = 1$  ( $\gamma \neq \gamma^*$ ), es decir, la sucesi\'on  $(K_n^*)$  es consistente (ver 11.2, definici\'on 5).

Ecojamos especialmente  $\alpha_1=0$  y  $\alpha_2=\alpha$ , entonces obtenemos  $a=\sqrt[n]{\alpha} \gamma^*$  y  $b=\gamma^*$ . Para la regi\'on cr\'itica  $K^*$  de la hip\'otesis  $H_0: \gamma_0 = \gamma^*$  se cumple entonces que  $K^* = \{t : t < \sqrt[n]{\alpha} \gamma^* \text{ o } t > \gamma^*\} =: K_1^*$ ; para la funci\'on de potencia  $G_1$  correspondiente se obtiene que

$$G_1(\gamma) = \begin{cases} 1 & \text{para } 0 < \gamma \leq \sqrt[n]{\alpha} \gamma^*, \\ \alpha \left( \frac{\gamma^*}{\gamma} \right)^n & \text{para } \sqrt[n]{\alpha} \gamma^* \leq \gamma \leq \gamma^*, \\ 1-(1-\alpha) \left( \frac{\gamma^*}{\gamma} \right)^n & \text{para } \gamma^* \leq \gamma. \end{cases}$$

Se verifica f\'cilmente que se cumple  $G_1(\gamma) \geq G_1(\gamma^*) = \alpha$ . La d\'ocima  $K_1^*$  es, por tanto, una d\'ocima admissible (ver 11.2, definici\'on 3). Ecojamos por el contrario  $\alpha_1=\alpha$  y  $\alpha_2=0$ , entonces obtenemos que  $a=0$  y que  $b=\sqrt[n]{1-\alpha} \gamma^*$ . Para la regi\'on cr\'itica  $K^*$  de la hip\'otesis  $H_0: \gamma_0 = \gamma^*$  se cumple entonces que

$K^* = \{t : t < 0 \text{ o } t > \sqrt[n]{1-\alpha} \gamma^*\} =: K_2^*$ ; para la funci\'on de potencia  $G_2$  correspondiente se obtiene que

$$G_2(\gamma) = \begin{cases} 0 & \text{para } 0 < \gamma \leq \sqrt[n]{1-\alpha} \gamma^*, \\ 1-(1-\alpha) \left( \frac{\gamma^*}{\gamma} \right)^n & \text{para } \sqrt[n]{1-\alpha} \gamma^* \leq \gamma. \end{cases}$$

La d\'ocima  $K_2^*$  no es admissible, por ejemplo, se cumple que  $G_2(\sqrt[n]{1-\alpha} \gamma^*) = 0 < G_2(\gamma^*) = \alpha$ . Por lo m\'as, las d\'ocimas  $K_1^*$  y  $K_2^*$  se pueden comparar (en el sentido de la definici\'on 4 (11.2)), y as\'i, la d\'ocima  $K_1^*$  resulta mejor que la d\'ocima  $K_2^*$ , es decir, se cumple que  $G_1(\gamma) \geq G_2(\gamma)$  para todo  $\gamma > 0$ . (El lector debe reflexionar en cada ocasi\'on acerca de la significaci\'on desde el punto de vista del contenido de estas proposiciones.)

Como hab\'iamos anunciado, queremos señalar sobre la base de este ejemplo la estrecha relaci\'on entre las estimaciones por intervalo de confianza y las d\'ocimas de significaci\'on.

El intervalo de confianza  $J(X_1, \dots, X_n)$  con el nivel de confiabilidad  $1-\alpha$ , indicado en el ejemplo 1 (10.5),

$$J(X_1, \dots, X_n) = \left[ \frac{\hat{\gamma}_n}{\sqrt[4]{1-\alpha_1}}, \frac{\hat{\gamma}_n}{\sqrt[4]{\alpha_2}} \right] \text{ con } \hat{\gamma}_n = \max \{X_1, \dots, X_n\},$$

contiene exactamente, para una muestra concreta  $(x_1, \dots, x_n)$ , el valor  $\gamma^*$  para el cual la hipótesis  $H_0: \gamma_0 = \gamma^*$  no se rechaza en la dócima  $K^*$  anterior con el nivel de significación

a. (Esto quiere decir que  $\gamma^* \in J(x_1, \dots, x_n)$ , o sea,  $\frac{t}{\sqrt[4]{1-\alpha_1}} \leq \gamma^* < \frac{t}{\sqrt[4]{\alpha_2}}$ , con

$t = \max \{x_1, \dots, x_n\}$ , es equivalente a  $\sqrt[4]{\alpha_1} \gamma^* \leq t \leq \sqrt[4]{1-\alpha_1} \gamma^*$ , o a  $t \notin K^*$ , y esto es lo mismo que decir que  $H_0: \gamma_0 = \gamma^*$  no se rechaza.

De forma general, si  $J(X_1, \dots, X_n)$  es un intervalo de confianza con el nivel de confiabilidad  $1-\alpha$  ( $0 < \alpha < 1$ ) para un parámetro  $\gamma$ , entonces la regla de decisión siguiente define una dócima de significación de  $H_0: \gamma_0 = \gamma^*$  con el nivel de significación  $\alpha$ : Para una muestra concreta  $(x_1, \dots, x_n)$  se construye el intervalo de confianza concreto  $J(x_1, \dots, x_n)$ . Si se cumple que  $\gamma^* \notin J(x_1, \dots, x_n)$ , se rechaza a  $H_0$ , en el otro caso ( $\gamma^* \in J(x_1, \dots, x_n)$ ) no.

## 11.4 Ejemplos importantes de dócimas paramétricas

Denominaremos dócima paramétrica a aquella destinada a la verificación de una hipótesis sobre un parámetro desconocido de una distribución de probabilidad por lo demás conocida; aquí se utiliza fundamentalmente el conocimiento acerca de la distribución de probabilidad.

A continuación brindamos algunos ejemplos importantes de dócimas paramétricas. En ellos se trata de dócimas de significación (con el nivel de significación  $\alpha$  prefijado,  $0 < \alpha < 1$ ), y se toma por base el esquema general indicado en 11.3. Estas dócimas paramétricas son:

Una dócima para el parámetro  $\mu$  de una población con distribución normal y varianza desconocida (dócima  $t$  simple).

Una dócima para la igualdad de los valores esperados de dos poblaciones independientes con distribuciones normales y varianzas iguales (aunque desconocidas) (dócima  $t$  doble).

Una dócima para el parámetro  $\sigma^2$  de una población con distribución normal y valor esperado desconocido (dócima de varianza  $\chi^2$ ).

Una dócima para la igualdad de las varianzas de dos poblaciones independientes con distribución normal y valores esperados desconocidos (dócima  $F$ ), y por último:

Una dócima para una probabilidad desconocida.

### 11.4.1 Dócima $t$ simple

0. Supongamos que  $X$  es una variable aleatoria con distribución  $N(\gamma_0, \sigma_0^2)$ ;  $\gamma_0$  y  $\sigma_0^2$  sean desconocidas.

1.  $H_0: \gamma_0 = \gamma^*$  ( $\gamma^*$  número real prefijado).

Esta hipótesis es, tomada rigurosamente, una hipótesis compuesta que se debería caracterizar de forma más exacta por  $H_0: (\gamma_0, \sigma^2) \in \{(\gamma^*, \sigma^2): \sigma^2 > 0\}$ . Si  $\sigma_0^2$  es conocida, entonces se trata de una hipótesis simple y se utiliza la dócima indicada en el epígrafe 11.1).

2. Para la construcción de la variable de dócima tomemos por base el estadígrafo

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , que en 10.4.1 se mostró como estimador puntual adecuado para  $\gamma_0$ . La variable  $\bar{X}_n$  posee, en el caso en que  $H_0$  sea verdadera, una distribución  $N\left(\gamma^*, \frac{\sigma_0^2}{n}\right)$

(ver 9.4, teorema 2). Estimemos el parámetro desconocido  $\sigma_0^2$  por medio del estimador puntual  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  (ver 10.4.2 b)) y utilicemos como variable de dócima la

$$\text{variable aleatoria } T = \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sqrt{S_n^2}}$$

que, en el caso en que  $H_0$  sea verdadera, posee una distribución  $t$  con  $n-1$  grados de libertad (ver 9.4, teorema 5) (fig. 52).

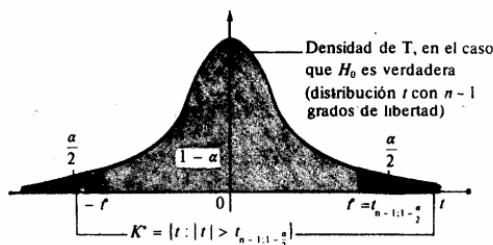


Figura 52

3. Establezcamos la región crítica  $K^*$  en la forma  $K^* = \{t : |t| > t^*\}$  (ver fig. 52) y determinemos  $t^*$  de modo que se cumpla que

$$P_{\gamma^*}(T \in K^*) = P_{\gamma^*}(|T| > t^*) = 1 - P_{\gamma^*}(-t^* \leq T \leq t^*) = \alpha.$$

De aquí se obtiene para  $t^*$  el percentil de orden  $1 - \frac{\alpha}{2}$  de la distribución  $t$  con  $n-1$  grados de libertad ( $t^* = t_{n-1; 1 - \frac{\alpha}{2}}$ ) y con esto la región crítica  $K^* = \{t : |t| > t_{n-1; 1 - \frac{\alpha}{2}}\}$ .

4. Regla de decisión: Para una muestra concreta  $(x_1, \dots, x_n)$  se calcula  $\bar{x}_n$  y  $s_n^2$ , de aquí  $t = \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sqrt{S_n^2}}$ , y se rechaza  $H_0: \gamma_0 = \gamma^*$  si y solo si se cumple que  $t \in K^*$ , es decir,

$$\left| \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sqrt{S_n^2}} \right| > t_{n-1; 1 - \frac{\alpha}{2}}.$$

### 11.4.2 Dócima $t$ doble

0. Sea  $X$  una variable aleatoria con distribución  $N(\mu_1, \sigma_1^2)$  y  $Y$  una variable aleatoria con distribución  $N(\mu_2, \sigma_2^2)$ . Sean  $X$  y  $Y$  variables aleatorias mutuamente independientes; los números  $\mu_1$ ,  $\mu_2$ ,  $\sigma_1^2$  y  $\sigma_2^2$  sean desconocidos y partamos de la condición  $\sigma_1^2 = \sigma_2^2$ . (La última condición se verifica, dado el caso, con la dócima  $F$  que se presenta en 11.4.4.) Además, sean  $(X_1, \dots, X_m)$  y  $(Y_1, \dots, Y_n)$  muestras matemáticas de tamaño  $m$  y  $n$ , respectivamente, de las poblaciones  $X$  y  $Y$  a que corresponden.

1.  $H_0: \mu_1 = \mu_2$
2. Variable de dócima

$$T = \frac{\bar{Y}_n - \bar{X}_m}{\sqrt{(m-1)S_{x,m}^2 + (n-1)S_{y,n}^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}$$

$$\text{con } \bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$S_{x,m}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_m)^2, \quad S_{y,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2$$

La variable de dócima  $T$  posee, en el caso en que  $H_0$  sea verdadera, una distribución  $t$  con  $m+n-2$  grados de libertad.

(Esto puede verificarse sin dificultad considerando la independencia de  $X$  y  $Y$ , utilizando los teoremas 2 y 4 de 9.4 y los teoremas 6 y 7 de 6.5.)

$$3. \text{ Región crítica } K^* = \left\{ t : |t| > t_{m+n-2; 1 - \frac{\alpha}{2}} \right\}$$

4. Regla de decisión: Para muestras concretas  $(x_1, \dots, x_m)$  y  $(y_1, \dots, y_n)$  se calcula  $\bar{x}_m$ ,  $\bar{y}_n$ ,  $s_{x,m}^2$  y  $s_{y,n}^2$  de aquí

$$t = \frac{\bar{Y}_n - \bar{X}_m}{\sqrt{(m-1)s_{x,m}^2 + (n-1)s_{y,n}^2}} \sqrt{\frac{m \cdot n(m+n-2)}{m+n}}$$

y se rechaza  $H_0: \mu_1 = \mu_2$  si y solo si  $t \in K^*$ , es decir, si se cumple que  $|t| > t_{m+n-2; 1 - \frac{\alpha}{2}}$ .

Si los números  $\sigma_1^2$  y  $\sigma_2^2$  son conocidos (no necesariamente iguales), se utiliza entonces la variable de dócima

$$T = \frac{\bar{Y}_n - \bar{X}_m}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

que, en el caso en que  $H_0$  sea verdadera, posee una distribución  $N(0,1)$ , y la región crítica

$$K^* = \left\{ t : |t| > z_{1 - \frac{\alpha}{2}} \right\}$$

La interrogante más general acerca de la verificación de la igualdad de los valores esperados de más de dos variables aleatorias independientes con distribución normal conduce a problemas que pertenecen a la rama del llamado *análisis de varianza*. En el marco de nuestra introducción a la Estadística matemática no podemos adentrarnos en esto.

### 11.4.3 Dócima $\chi^2$

0: Sea  $X$  una variable aleatoria con distribución  $N(\mu_0, \gamma_0)$ ;  $\mu_0$  y  $\gamma_0$  sean desconocidos.

1.  $H_0: \gamma_0 = \gamma^*$  ( $\gamma^*$  número positivo prefijado).

2. Para la construcción de la variable de dócima tomemos por base el estadígrafo

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

que en 10.4.2 b) se mostró como estimador puntual adecuado para  $\gamma_0$ . La variable aleatoria

$$T = \frac{(n-1)S_n^2}{\gamma^*}$$

posee, según el teorema 4(9.4), en el caso en que  $H_0$  sea verdadera, una distribución  $\chi^2$  con  $n-1$  grados de libertad.

3. Establezcamos la región crítica en la forma  $K^* = \{t : t < a \text{ o } t > b\}$  (fig. 53) y determinemos  $a$  y  $b$  de modo que se cumpla que  $P_{\gamma^*}(T < a) = P_{\gamma^*}(T > b) = \frac{\alpha}{2}$ , y por consiguiente, que  $P_{\gamma^*}(T \in K^*) = \alpha$ . De aquí se obtiene para  $a$  y  $b$  los percentiles de orden  $\frac{\alpha}{2}$  y  $1 - \frac{\alpha}{2}$ , respectivamente, de la distribución  $\chi^2$  con  $n-1$  grados de libertad, osea,  $a = \chi^2_{n-1; 1 - \frac{\alpha}{2}}$  y  $b = \chi^2_{n-1; 1 - \frac{\alpha}{2}}$ . Con esto obtenemos la región crítica

$$K^* = \left\{ t : t < \chi^2_{n-1; 1 - \frac{\alpha}{2}} \text{ o } t > \chi^2_{n-1; 1 - \frac{\alpha}{2}} \right\}.$$

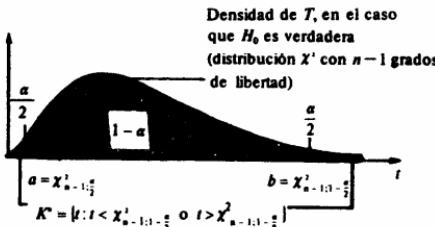


Figura 53

4. Regla de decisión: Para una muestra concreta  $(x_1, \dots, x_n)$  se calcula  $s_n^2$ ; de aquí  $t = \frac{(n-1)s_n^2}{\gamma^*}$ , y se rechaza  $H_0: \gamma_0 = \gamma^*$  si y solo si  $t \in K^*$ , es decir, si se cumple  $t < \chi^2_{n-1; 1 - \frac{\alpha}{2}}$  o  $t > \chi^2_{n-1; 1 - \frac{\alpha}{2}}$ .

### 11.4.4 Dócima $F$

0. Sean  $X$  y  $Y$  variables aleatorias con distribuciones  $N(\mu_1, \sigma_1^2)$  y  $N(\mu_2, \sigma_2^2)$ , respectivamente. Sean las variables aleatorias  $X$  y  $Y$  mutuamente independientes; los números  $\mu_1$ ,  $\mu_2$ ,  $\sigma_1^2$  y  $\sigma_2^2$  sean desconocidos.

Además, sean  $(X_1, \dots, X_m)$  y  $(Y_1, \dots, Y_n)$  muestras matemáticas de tamaño  $m$  y  $n$ , respectivamente, de las poblaciones  $X$  y  $Y$  a que corresponde cada una

$$1. H_0: \sigma_1^2 = \sigma_2^2$$

2. Dócima de prueba:

$$T = \frac{S_{x,m}^2}{S_{y,n}^2} \text{ con } S_{x,m}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_m)^2, S_{y,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2.$$

La variable de dócima  $T$  posee, en el caso en que  $H_0$  sea verdadera, una distribución  $F$  con  $(m-1, n-1)$  grados de libertad (ver 9.4, teorema 6).

3. Región crítica:

$$K^* = \left\{ t : t < F_{m-1, n-1; 1-\frac{\alpha}{2}} \text{ o } t > F_{m-1, n-1; 1-\frac{\alpha}{2}} \right\}; \text{ aquí } F_{m-1, n-1; \beta} \text{ denota al percentil de orden } \beta$$

de la distribución  $F$  con  $(m-1, n-1)$  grados de libertad (fig. 54).

4. Regla de decisión: Para muestras concretas  $(x_1, \dots, x_m)$  y  $(y_1, \dots, y_n)$  se calcula  $s_{x,m}^2$  y  $s_{y,n}^2$  de aquí  $t = \frac{s_{x,m}^2}{s_{y,n}^2}$  y se rechaza  $H_0: \sigma_1^2 = \sigma_2^2$  si y solo si  $t \in K^*$  es decir, si se cumple que  $t < F_{m-1, n-1; 1-\frac{\alpha}{2}}$  o  $t > F_{m-1, n-1; 1-\frac{\alpha}{2}}$ .

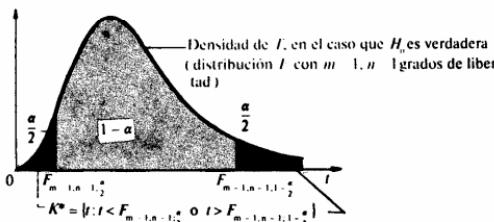


Figura 54

#### 11.4.5 Dócima para una probabilidad desconocida

0. Sea  $A$  un suceso aleatorio que se presenta en el marco de un experimento aleatorio con la probabilidad  $p_0 = P(A); p_0$  sea desconocida. Consideremos la variable aleatoria

$$X = \begin{cases} 1 & \text{en el caso en que } A \text{ ocurra,} \\ 0 & \text{en el caso en que } \bar{A} \text{ ocurra.} \end{cases}$$

Además, sea  $(X_1, \dots, X_n)$  una muestra matemática de tamaño  $n$  de la población  $X$ .

$$1. H_0: p_0 = p^* \quad (p^* \text{ número prefijado entre cero y uno}).$$

2. Variable de dócima

$$T = \frac{M - np^*}{\sqrt{np^*(1-p^*)}} \text{ con } M = \sum_{i=1}^n X_i.$$

(Luego, la variable aleatoria  $M$  indica la frecuencia aleatoria absoluta de  $A$  en  $n$  repeticiones independientes del experimento aleatorio tomado por base y posee con esto, en el caso en que  $H_0$  sea verdadera, una distribución binomial con los parámetros  $n$  y  $p^*$ .) La

variable de dócima  $T$  posee, en el caso en que  $H_0$  sea verdadera, asintóticamente (es decir, cuando  $n \rightarrow \infty$ ) una distribución  $N(0, 1)$ , sobre la base del Teorema Integral de De Moivre-Laplace.

3. Región crítica:  $K^* = \left\{ t : |t| > z_{1 - \frac{\alpha}{2}} \right\}$  (Se cumple que

$$\lim_{n \rightarrow \infty} P_{p^*}(T \in K^*) = \lim_{n \rightarrow \infty} P_{p^*} \left( \left| \frac{M - np^*}{\sqrt{np^*(1-p^*)}} \right| > z_{1 - \frac{\alpha}{2}} \right) \\ = 1 - \lim_{n \rightarrow \infty} P_{p^*} \left( \left| \frac{M - np^*}{\sqrt{np^*(1-p^*)}} \right| \leq z_{1 - \frac{\alpha}{2}} \right) \\ = 1 - (1 - \alpha) = \alpha,$$

o sea,  $K^*$  define para  $n \rightarrow \infty$  una dócima de significación con el nivel de significación  $\alpha$ .)

4. Regla de decisión: Para una muestra concreta  $(x_1, \dots, x_n)$  ( $= n$ -úplo de los números cero y uno) se halla  $m = \sum_{i=1}^n x_i$  ( $=$  número de la ocurrencia de  $A$  en  $n$  experimentos), se calcula

$$t = \frac{m - np^*}{\sqrt{np^*(1-p^*)}},$$

y se rechaza a  $H_0: p_0 = p^*$  si y solo si  $t \in K^*$ , es decir, si se cumple que

$$\left| \frac{m - np^*}{\sqrt{np^*(1-p^*)}} \right| > z_{1 - \frac{\alpha}{2}}.$$

**Observación.** Si  $n$  es tan pequeña que una aplicación del Teorema Integral de De Moivre-Laplace no nos parece justificada, se construye una dócima de significación partiendo directamente de la distribución de la variable de dócima  $M$  (distribución binomial con los parámetros  $n$  y  $p^*$ , en el caso en que  $H_0: p_0 = p^*$  sea verdadera).

## 11.5 Ejemplos importantes de dócimas no paramétricas

Por una *dócima no paramétrica* entendemos aquella destinada a la verificación de una hipótesis acerca de una población, para la cual no se toman en consideración los conocimientos sobre el tipo de distribución de la población considerada.

Como ejemplos importantes de dócimas no paramétricas presentaremos a continuación, utilizando de nuevo el esquema general indicado en 11.3, dos dócimas de ajuste (*dócima de ajuste  $\chi^2$* , *dócima de Kolmogorov*), dos dócimas de homogeneidad (*dócima de homogeneidad  $\chi^2$* , *dócima para dos distribuciones*) y una dócima de independencia (*dócima de independencia  $\chi^2$* ).

Por una *dócima de ajuste* entendemos, de forma general, una dócima para la hipótesis de que la verdadera función de distribución  $F_0$  verdadera (pero desconocida) de una población es igual a una función de distribución  $F^*$  prefijada. Se denomina *dócima de homogeneidad* a una dócima sobre la igualdad de las distribuciones de probabilidad (desconocidas) de varias poblaciones. Por una *dócima de independencia* se entiende aquella que sirve para la verificación de la hipótesis de que dos o más variables aleatorias consideradas sean mutuamente independientes.

### 11.5.1 Dócima de ajuste $\chi^2$

1.  $H_0: F_c = F^*(F^*$  función de distribución prefijada).

2. Construcción de la variable de dócima: Se realiza una partición de la imagen de  $X$  en  $k$  intervalos  $I_j = [\xi_j, \xi_{j+1}], j=1, \dots, k$  -denominados clases- con  $-\infty \leq \xi_1 < \xi_2 < \dots < \xi_k < \xi_{k+1} \leq +\infty$ , siendo  $k \geq 2$  un número natural arbitrario. Para una muestra matemática  $(X_1, \dots, X_n)$  de tamaño  $n$  de la población considerada, denote  $M_j$  la denominada frecuencia de clase (aleatoria) de la clase  $I_j$ , esto es, el número (aleatorio) de las variables de la muestra  $X_i$  que están situadas en  $I_j$ . (Luego se cumple que

$\sum_{j=1}^k M_j = n$ ). La variable aleatoria  $M_j$  está distribuida binomialmente con los parámetros  $n$  y  $p_j = F^*(\xi_{j+1}) - F^*(\xi_j)$ , en el caso en que  $H_0: F_0 = F^*$  sea verdadera ( $j=1, \dots, k$ ):

$$\frac{M_j - np_j}{\sqrt{np_j(1-p_j)}}$$
 posee asintóticamente (es decir, cuando  $n \rightarrow \infty$ ) una distribución  $N(0,1)$

(ver teorema 1(7.5)). Se puede mostrar que la variable aleatoria (utilizada más adelante como variable de dócima)

$$T = \sum_{j=1}^k \frac{(M_j - np_j)^2}{np_j} = \sum_{j=1}^k \frac{M_j^2}{np_j} - n,$$

en el caso en que  $H_0$  sea verdadera, posee asintóticamente (es decir, cuando  $n \rightarrow \infty$ ) una distribución  $\chi^2$  con  $k-1$  grados de libertad. (Renunciaremos a la demostración relativamente difícil de esto.)

3. Si para una muestra concreta  $(x_1, \dots, x_n)$ , las frecuencias de clase  $m_j$ , halladas se diferencian notablemente de los valores  $np_j$ , esperados, dada la validez de  $H_0$ , entonces la variable de dócima  $T$  aceptará valores grandes y se rechazará a  $H_0$ . Por ello establezcamos  $K^*$  en la forma  $K^* = \{t : t > t^*\}$  y fijemos  $t^*$ , de modo tal, que se cumple que

$$\lim_{n \rightarrow \infty} P_{F^*}(T \in K^*) = \lim_{n \rightarrow \infty} P_{F^*}(T > t^*) = \alpha.$$

Como  $T$ , en el caso en que  $H_0: F_0 = F^*$  sea verdadera, posee asintóticamente (es decir, cuando  $n \rightarrow \infty$ ) una distribución  $\chi^2$  con  $k-1$  grados de libertad, se obtiene para  $t^*$  el percentil de orden  $1-\alpha$  de la distribución  $\chi^2$  con  $k-1$  grados de libertad o sea,  $t^* = \chi^2_{k-1, 1-\alpha}$  y con esto  $K^* = \{t : t > \chi^2_{k-1, 1-\alpha}\}$  (fig. 55).

4. Regla de decisión: Para una muestra concreta  $(x_1, \dots, x_n)$  se halla, con respecto a la partición en clases elegida, las frecuencias de clase absolutas  $m_j$  ( $j=1, \dots, k$ ), se calculan las probabilidades  $p_j$  ( $j=1, \dots, k$ ) fijadas por la hipótesis  $H_0$ , y con esto

$$t = \sum_{j=1}^k \frac{m_j^2}{np_j} - n.$$

Si se cumple que  $t \in K^*$ , o sea, si

$$\sum_{j=1}^k \frac{m_j^2}{np_j} - n > \chi_{k-1; 1-\alpha}^2$$

entonces se rechaza a  $H_0: F_0 = F^*$ , en el otro caso no.

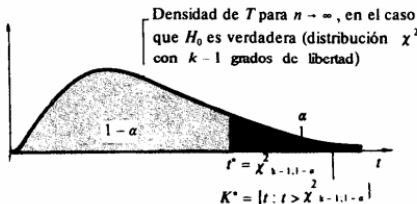


Figura 55

Para la fuerza de la dócima de ajuste  $\chi^2$  es naturalmente importante la elección de la partición en clases. En la práctica se eligen frecuentemente intervalos de igual longitud (en algunos casos con excepción de los intervalos de los extremos). Se ha mostrado conveniente elevar el número de las clases para mayor tamaño  $n$  de la muestra (por ejemplo,  $k \approx \lg n$ ,  $k \approx \sqrt{n}$ ; aquí se recomienda fijar las clases  $I_j$ , de modo que se cumpla que  $np_j \geq 1$  ( $j = 1, \dots, k$ )).

### 11.5.2 Dócima de Kolmogorov

0. Supongamos que la función de distribución  $F_0$  de la población  $X$  es continua.
1.  $H_0: F_0 = F^*$  ( $F^*$  función de distribución continua prefijada).
2. Variable de dócima:  $T = \sqrt{n} \sup_{-\infty < x < \infty} |W_n(x) - F^*(x)|$ ; aquí  $W_n(x)$  denota el valor de la función de distribución empírica de una muestra matemática de tamaño  $n$  de la población  $X$  en el punto  $x$ . La variable de dócima  $T$  posee para  $n \rightarrow \infty$ , en el caso en que  $H_0$  sea verdadera, la función de distribución  $K$  (ver 9.3, teorema 3), dada por

$$K(y) = \begin{cases} 0 & \text{para } y \leq 0, \\ \sum_{k=-\infty}^{\infty} (-1)^k e^{-2ky^2} & \text{para } y > 0. \end{cases}$$

3. Región crítica:  $K^* = \{t: t \geq y_\alpha\}$ , aquí  $y_\alpha$  denota la solución de la ecuación  $K(y) = 1 - \alpha$ . (La probabilidad de que  $T$  tome valores  $\geq y_\alpha$  converge, en el caso en que  $H_0$  sea verdadera, hacia  $\alpha$  para  $n \rightarrow \infty$ .)

4. Regla de decisión: Para una muestra concreta  $(x_1, \dots, x_n)$  se halla la función de distribución empírica concreta  $w_n$  correspondiente, se calcula  $t = \sqrt{n} \sup_{-\infty < x < \infty} |w_n(x) - F^*(x)|$  y se rechaza a  $H_0: F_0 = F^*$  si y sólo si  $t \in K^*$ , es decir, si se cumple que

$$\sqrt{n} \sup_{-\infty < x < \infty} |w_n(x) - F^*(x)| \geq y_\alpha.$$

### 11.5.3 Dócima de homogeneidad $\chi^2$

0. Supongamos que las variables aleatorias  $X$  y  $Y$  son independientes. Denotemos la función de distribución (desconocida) de  $X$  y  $Y$  con  $F_0$  y  $G_0$ , respectivamente.

$$1. H_0: F_0 = G_0$$

2. Construcción de la variable de dócima: Se realiza una partición de la imagen (común) de las variables aleatorias  $X$  y  $Y$  en  $k$  intervalos disjuntos  $I_j (j=1, \dots, k)$ ; aquí  $k (\geq 2)$  es un número natural arbitrario. Si  $M_j$  denota la frecuencia de clase (aleatoria) de la clase  $I_j$ , para una muestra matemática  $(X_1, \dots, X_m)$  de tamaño  $m$  de la población  $X$  y  $N_j$  la de la clase  $I_j$  para una muestra matemática  $(Y_1, \dots, Y_n)$  de tamaño  $n$  de la población  $Y$ , entonces la variable de dócima

$$T = mn \sum_{j=1}^k \frac{1}{M_j + N_j} \left( \frac{M_j}{m} - \frac{N_j}{n} \right)^2$$

posee, en el caso en que  $H_0$  sea verdadera, asintóticamente (es decir, cuando  $m \rightarrow \infty$  y  $n \rightarrow \infty$ ) una distribución  $\chi^2$  con  $k-1$  grados de libertad.

3. Si para muestras concretas  $(x_1, \dots, x_m)$  y  $(y_1, \dots, y_n)$  las frecuencias de clase relativas  $\frac{m_j}{m}$  y  $\frac{n_j}{n}$  ( $j=1, \dots, k$ ) se diferencian notablemente, entonces  $T$  aceptará valores grandes y se rechazará a  $H_0$ . Por ello, fijemos  $K^*$  en la forma  $K^* = \{t : t > \chi^2_{k-1, 1-\alpha}\}$ . (La probabilidad del suceso ( $T \in K^*$ ) converge hacia  $\alpha$  cuando  $m \rightarrow \infty$  y  $n \rightarrow \infty$ , dada la validez de  $H_0$  (fig. 55.).

4. Regla de decisión: Para muestras concretas  $(x_1, \dots, x_m)$  y  $(y_1, \dots, y_n)$  se halla con respecto a la partición en clases elegida las frecuencias de clase absolutas  $m_j$  y  $n_j (j=1, \dots, k)$ , se calcula de aquí

$$t = mn \sum_{j=1}^k \frac{1}{m_j + n_j} \left( \frac{m_j}{m} - \frac{n_j}{n} \right)^2$$

y se rechaza a  $H_0: F_0 = G_0$  si y solo si se cumple que  $t > \chi^2_{k-1, 1-\alpha}$ .

### 11.5.4 Dócima para dos distribuciones

La dócima para dos distribuciones se puede realizar rápidamente, es una denominada dócima rápida, con objeto de verificar la hipótesis  $H_0: F_0 = G_0$  sobre la igualdad de las funciones de distribución desconocidas, supuestas continuas, de dos poblaciones independientes  $X$  y  $Y$ , sobre la base de muestras de igual tamaño de estas poblaciones. En especial se aplica cuando se espera que  $F_0 \neq G_0$ . En principio, la dócima para dos distribuciones es una dócima para una probabilidad desconocida (ver 11.4.5). En el caso en que  $H_0$  sea verdadera, el suceso aleatorio  $A = (X - Y < 0) = (X < Y)$  posee la probabilidad  $\frac{1}{2}$ . Se verifica entonces la hipótesis  $\tilde{H}_0: P(A) = \frac{1}{2}$  (por ejemplo, con la prueba indicada en 11.4.5) y se rechaza a  $H_0$  si  $\tilde{H}_0$  se rechaza.

### 11.5.5 Dócima de independencia $\chi^2$

El punto de partida es una población bidimensional ( $X, Y$ ). En la explicación de la dócima de independencia  $\chi^2$ , que se denomina también *dócima de independencia en tablas de contingencia*, queremos limitarnos para una mayor sencillez al caso de variables aleatorias discretas  $X$  y  $Y$  y aceptar que  $X$  y  $Y$  toman los valores  $1, \dots, r$  y  $1, \dots, s$ , respectivamente.

1.  $H_0$ :  $X$  y  $Y$  son mutuamente independientes (equivalente a esto es la validez de la relación)

$$p_{ik} = P(X=i, Y=k) = P(X=i)P(Y=k) = p_i \cdot p_k$$

para  $i=1, \dots, r$  y  $k=1, \dots, s$  (ver 6.4, teorema 1).

2. Construcción de la variable de dócima. Sea  $((X_1, Y_1), \dots, (X_n, Y_n))$  una muestra matemática de tamaño  $n$  de la población (bidimensional)  $(X, Y)$ . Denotemos con  $N_{ik}$  el número (aleatorio) de las variables de la muestra, cuya primera componente es igual a  $i$  y la segunda a  $k$ . Además, sea

$$N_i = \sum_{k=1}^s N_{ik}, \quad N_k = \sum_{i=1}^r N_{ik}$$

$$(\text{Se cumple entonces que } \sum_{i=1}^r \sum_{k=1}^s N_{ik} = \sum_{i=1}^r N_i = \sum_{k=1}^s N_k = n.)$$

Consideremos la variable aleatoria

$$T = n \sum_{i=1}^r \sum_{k=1}^s \frac{\left( N_{ik} - \frac{N_i N_k}{n} \right)^2}{N_i N_k}.$$

Se puede mostrar que  $T$  posee, en el caso en que  $H_0$  sea verdadera, asintóticamente (es decir, cuando  $n \rightarrow \infty$ ) una distribución  $\chi^2$  con  $(r-1)(s-1)$  grados de libertad.

3. Región crítica:  $K^* = \{t : t > \chi^2_{(r-1)(s-1); 1-\alpha}\}$  (La probabilidad del suceso  $(T \in K^*)$  converge hacia  $\alpha$  cuando  $n \rightarrow \infty$ , dada la validez de  $H_0$ ).

4. Regla de decisión: Para una muestra concreta  $((x_1, y_1), \dots, (x_n, y_n))$  se hallan los números  $n_{ik}$  (=número de los elementos  $(i, k)$  en la muestra),

$$n_i = \sum_{k=1}^s n_{ik}, \quad n_k = \sum_{i=1}^r n_{ik} \quad (i=1, \dots, r; k=1, \dots, s),$$

se calcula de aquí

$$t = n \sum_{i=1}^r \sum_{k=1}^s \frac{\left( n_{ik} - \frac{n_i n_k}{n} \right)^2}{n_i n_k}$$

y se rechaza a  $H_0$  si y solo si se cumple que  $t > \chi^2_{(r-1)(s-1); 1-\alpha}$ .

Para la realización práctica de esta dócima se recomienda la representación de la muestra concreta en una llamada *tabla de contingencia*, que contiene todos los valores numéricos necesarios para la dócima.

$X \backslash Y$	1	2	...	$s$	
1	$n_{11}$	$n_{12}$	...	$n_{1s}$	$n_1$
2	$n_{21}$	$n_{22}$	...	$n_{2s}$	$n_2$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
$r$	$n_{r1}$	$n_{r2}$		$n_{rs}$	$n_r$
	$n_1$	$n_2$		$n_s$	$n$

(En el caso  $r=s=2$  se denomina la tabla de contingencia correspondiente *tabla de cuatro entradas o tablas de  $2 \times 2$* )

## 11.6 Ejemplo de aplicación

En 286 aspirantes para estudiar Matemática fueron investigadas dos características, la calificación  $X$  del examen de ingreso y la nota  $Y$  de la prueba de nivel en la asignatura Matemática. El resultado está agrupado en la tabla de frecuencia siguiente (tabla de contingencia); posteriormente aclararemos la significación de los números indicados en paréntesis y corchetes:

$X \backslash Y$	1	2	3	
1 (especialmente apto)	(26,38) [13,62]	(19,87) [10,87]	(2,74) [2,74]	$49 (=n_1)$
2 (apto)	(77) [11]	(58) [6]	(8) [5]	$143 (=n_2)$
3 (apto condicionalmente)	(14,54) [6,54]	(10,95) [10,95]	(1,51) [0,51]	$27 (=n_3)$
4 (no apto)	(36,08) [18,08]	(27,17) [9,83]	(3,75) [8,25]	$67 (=n_4)$
	$154 (=n_1)$	$116 (=n_2)$	$16 (=n_3)$	$286 (=n)$

Concebimos  $X$  y  $Y$  como variables aleatorias (discretas) y queremos verificar la hipótesis  $H_0$ :  $X$  y  $Y$  son mutuamente independientes, con la prueba de independencia  $\chi^2$  (tratada en 11.5.5) con el nivel de significación  $\alpha=5\%$ . Para nuestro ejemplo se cumple que  $r=4$ ,  $s=3$  y, por tanto,  $(r-1)(s-1)=(4-1)(3-1)=6$ . Como el percentil de orden  $1-\alpha=0,95$  de la distribución  $\chi^2$  con 6 grados de libertad es igual a 2,6, se obtiene para la región crítica,  $K^*=\{t: t>12,6\}$ . Calculemos ahora el valor  $t$ ,

$$t = n \sum_{i=1}^r \sum_{k=1}^s \frac{\left( n_{ik} - \frac{n_i n_k}{n} \right)^2}{\frac{n_i n_k}{n}}$$

$$= \sum_{i=1}^r \sum_{k=1}^s \frac{\left( n_{ik} - \frac{n_i n_k}{n} \right)^2}{\frac{n_i n_k}{n}}$$

de la variable de dómica  $T$  para nuestro ejemplo.

Se cumple que  $n=286$ ,  $r=4$ ,  $s=3$ . Los números  $n_{ik}$ ,  $n_i$  y  $n_k$  se deben tomar directamente de la tabla de contingencia indicada anteriormente. En esta tabla hemos señalado dentro de los paréntesis los números  $\frac{n_i n_k}{n}$  y dentro de los corchetes los números  $|n_{ik} - \frac{n_i n_k}{n}|$  ( $i=1, 2, 3, 4$ ;  $k=1, 2, 3$ ). Con esto se obtiene

$$t = \frac{13,62^2}{26,38} + \frac{10,87^2}{19,87} + \frac{2,74^2}{2,74} + \frac{11^2}{77} + \frac{6^2}{58} + \frac{5^2}{8}$$

$$+ \frac{6,54^2}{14,54} + \frac{7,05^2}{10,95} + \frac{0,51^2}{1,51} + \frac{18,08^2}{36,08} + \frac{9,83^2}{27,17} + \frac{8,25^2}{3,75}$$

$$= 7,03 + 5,94 + 2,74 + 1,57 + 0,62 + 3,12 + 2,94 + 4,54 + 0,17 + 9,06 + 3,56 + 18,16$$

$$= 59,45$$

Por consiguiente, el valor  $t$  está situado en la región crítica y rechazamos la hipótesis  $H_0$  de que la calificación del examen de ingreso para estudiar Matemática y la nota de la prueba de nivel en la asignatura Matemática sean mutuamente independientes. (Al mismo resultado llegaríamos también utilizando el nivel de significación  $\alpha=1\%$ ; se cumple que  $\chi^2_{0,99}=16,8 < 59,45$ .)

## 12. Tablas de algunas distribuciones importantes

Las tablas sobre las distribuciones binomial, de Poisson y normal, dadas en los epígrafes 12.1, 12.2 y 12.3, ofrecen una visión numérica sobre estas distribuciones de probabilidad. Por el contrario, las tablas dadas en los epígrafes 12.4, 12.5, y 12.6 para las distribuciones de prueba de la Estadística matemática (distribuciones  $\chi^2$ ,  $t$  y  $F$ ) contienen solamente algunos percentiles, los cuales deben ser suficientes para la realización práctica de las más importantes estimaciones por intervalo de confianza y dígitos de significación tratadas en este libro. La utilización de las tablas se demostrará con un ejemplo. Se puede encontrar en otra bibliografía tablas más completas para la realización de procedimientos de la Estadística matemática.

### 12.1 Tabla de la distribución binomial

La tabla 1 contiene probabilidades de la distribución binomial,

$$P(X=k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k=0, 1, \dots, n,$$

para  $n=1, 2, \dots, 10, 15, 20$  y algunos  $p \leq 0,50$ . Los lugares vacíos significan aquí  $b(k; n, p) < 0,0005$ .

Para  $p > 0,50$  se utiliza la relación  $b(k; n, p) = b(n-k; n, 1-p)$  (ver 4.5, teorema 1, fórmula (4)).

Para  $n$  grandes y  $p$  pequeños con  $np \leq 20$ , se iguala  $np = \lambda$  y se toma como base la relación  $b(k; n, p) \approx p(k; \lambda)$ , derivada del teorema límite de Poisson (ver 4.7, teorema 3 y fórmula (9)). Para esto se toman los números  $p(k; \lambda)$  de la tabla de la distribución de Poisson (ver 12.2).

Para  $n$  grandes se recomienda la aproximación de la distribución binomial a través de la distribución normal sobre la base del Teorema Integral de De Moivre-Laplace (ver 7.5, teorema 1 y fórmula (2)).

Tabla 1

Ejemplo:  $b(3; 8, 0,30) = 0,254$ 

$n$	$k$	$p=0,01$	$0,02$	$0,05$	$0,10$	$0,15$	$0,20$	$0,25$	$0,30$	$0,40$	$0,50$
1	0	0,990	0,980	0,950	0,900	0,850	0,800	0,750	0,700	0,600	0,500
	1	0,010	0,020	0,050	0,100	0,150	0,200	0,250	0,300	0,400	0,500
2	0	0,980	0,960	0,902	0,810	0,722	0,640	0,562	0,490	0,360	0,250
	1	0,020	0,039	0,095	0,180	0,255	0,320	0,375	0,420	0,480	0,500
	2			0,002	0,010	0,022	0,040	0,062	0,090	0,160	0,250
3	0	0,970	0,941	0,857	0,729	0,614	0,512	0,422	0,343	0,216	0,125
	1	0,029	0,058	0,135	0,243	0,325	0,384	0,422	0,441	0,432	0,375
	2		0,001	0,007	0,027	0,057	0,096	0,141	0,189	0,288	0,375
	3				0,001	0,003	0,008	0,016	0,027	0,064	0,125
4	0	0,961	0,922	0,815	0,656	0,522	0,410	0,316	0,240	0,130	0,062
	1	0,039	0,075	0,171	0,292	0,368	0,410	0,422	0,412	0,346	0,250
	2	0,001	0,002	0,014	0,049	0,098	0,154	0,211	0,265	0,346	0,375
	3				0,004	0,011	0,026	0,047	0,076	0,154	0,250
	4					0,001	0,002	0,004	0,008	0,026	0,062
5	0	0,951	0,904	0,774	0,590	0,444	0,328	0,237	0,168	0,078	0,031
	1	0,048	0,092	0,204	0,328	0,392	0,410	0,396	0,360	0,259	0,156
	2	0,001	0,004	0,021	0,073	0,138	0,205	0,264	0,309	0,346	0,312
	3				0,001	0,008	0,024	0,051	0,088	0,132	0,230
	4					0,002	0,006	0,015	0,028	0,077	0,156
	5						0,001	0,002	0,010	0,031	
6	0	0,941	0,886	0,735	0,531	0,377	0,262	0,178	0,118	0,047	0,016
	1	0,057	0,108	0,232	0,354	0,399	0,393	0,356	0,303	0,187	0,094
	2	0,001	0,006	0,031	0,098	0,176	0,246	0,297	0,324	0,311	0,234
	3				0,002	0,015	0,041	0,082	0,132	0,158	0,276
	4					0,001	0,005	0,015	0,033	0,060	0,138
	5						0,002	0,004	0,010	0,037	0,094
	6							0,001	0,004	0,016	
7	0	0,932	0,868	0,698	0,478	0,321	0,210	0,133	0,082	0,028	0,008
	1	0,066	0,124	0,257	0,372	0,396	0,367	0,311	0,247	0,131	0,055
	2	0,002	0,008	0,041	0,124	0,210	0,275	0,311	0,318	0,261	0,164
	3				0,004	0,023	0,062	0,115	0,173	0,227	0,290
	4					0,003	0,011	0,029	0,058	0,097	0,194
	5						0,001	0,004	0,012	0,025	0,077
	6							0,001	0,004	0,017	0,055
	7								0,002	0,008	

<i>n</i>	<i>k</i>	<i>p</i> =0.01	0.02	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
8	0	0.923	0.851	0.663	0.430	0.272	0.168	0.100	0.058	0.017	0.004
	1	0.075	0.139	0.279	0.383	0.385	0.336	0.267	0.198	0.090	0.031
	2	0.003	0.010	0.051	0.149	0.238	0.294	0.311	0.296	0.209	0.109
	3			0.005	0.033	0.084	0.147	0.208	0.254	0.279	0.219
	4				0.005	0.018	0.046	0.087	0.136	0.232	0.273
	5					0.003	0.009	0.023	0.047	0.124	0.219
	6						0.001	0.004	0.010	0.041	0.109
	7							0.001	0.008	0.031	
	8								0.001	0.004	
9	0	0.914	0.834	0.630	0.387	0.232	0.134	0.075	0.040	0.010	0.002
	1	0.083	0.153	0.299	0.387	0.368	0.302	0.225	0.156	0.060	0.018
	2	0.003	0.013	0.063	0.172	0.260	0.302	0.300	0.267	0.161	0.070
	3			0.001	0.008	0.045	0.107	0.176	0.234	0.267	0.251
	4				0.001	0.007	0.028	0.066	0.117	0.172	0.251
	5					0.001	0.005	0.017	0.039	0.074	0.167
	6						0.001	0.003	0.009	0.021	0.074
	7							0.001	0.004	0.021	0.070
	8								0.004	0.018	
	9									0.002	
10	0	0.904	0.817	0.599	0.349	0.197	0.107	0.056	0.028	0.006	0.001
	1	0.091	0.167	0.315	0.387	0.347	0.268	0.188	0.121	0.040	0.010
	2	0.004	0.015	0.075	0.194	0.276	0.302	0.282	0.233	0.121	0.044
	3		0.001	0.010	0.057	0.130	0.201	0.250	0.267	0.215	0.117
	4			0.001	0.011	0.040	0.088	0.146	0.200	0.251	0.205
	5				0.001	0.008	0.026	0.058	0.103	0.201	0.246
	6					0.001	0.006	0.016	0.037	0.111	0.205
	7						0.001	0.003	0.009	0.042	0.117
	8							0.001	0.011	0.044	
	9								0.002	0.010	
	10									0.001	
15	0	0.860	0.739	0.463	0.206	0.087	0.035	0.013	0.005	0.000	0.000
	1	0.130	0.226	0.366	0.343	0.231	0.132	0.067	0.031	0.005	0.000
	2	0.009	0.032	0.135	0.267	0.286	0.231	0.156	0.092	0.022	0.003
	3		0.003	0.031	0.129	0.218	0.250	0.225	0.170	0.063	0.014
	4			0.005	0.043	0.116	0.188	0.225	0.219	0.127	0.042
	5				0.001	0.010	0.045	0.103	0.165	0.206	0.186
	6					0.002	0.013	0.043	0.092	0.147	0.207
	7						0.003	0.014	0.039	0.081	0.177
	8							0.001	0.003	0.013	0.118
	9								0.001	0.003	0.012
	10									0.001	0.024
	11										0.007
	12										0.002
	13										0.014
	14										0.003
	15										

Tabla 1 (continuación)

<i>n</i>	<i>k</i>	<i>p</i> =0.01	0.02	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
20	0	0.818	0.668	0.358	0.122	0.039	0.012	0.003	0.001	0.000	0.000
	1	0.165	0.272	0.377	0.270	0.137	0.058	0.021	0.007	0.000	0.000
	2	0.016	0.053	0.189	0.285	0.229	0.137	0.067	0.028	0.003	0.000
	3	0.001	0.006	0.060	0.190	0.243	0.205	0.134	0.072	0.012	0.001
	4		0.001	0.013	0.090	0.182	0.218	0.190	0.130	0.035	0.005
	5			0.002	0.032	0.103	0.175	0.202	0.179	0.075	0.015
	6				0.009	0.045	0.109	0.169	0.192	0.124	0.037
	7					0.002	0.016	0.055	0.112	0.164	0.166
	8						0.005	0.022	0.061	0.114	0.180
	9							0.001	0.007	0.027	0.065
	10								0.010	0.031	0.117
	11									0.003	0.012
	12										0.001
	13										0.001
	14										0.005
	15										0.001
	16										0.005
	17										0.001
	18										
	19										
	20										

## 12.2 Tabla de distribución de Poisson

En la tabla 2 se recogen probabilidades de la distribución de Poisson

$$P(X=k) = p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k=0, 1, 2, \dots,$$

para algunas  $\lambda \leq 20$ . Los lugares libres significan que  $p(k : \lambda) < 0,00005$ .

Tabla 2

Ejemplo:  $p(3; 2,0) = 0,1804$

<i>k</i>	$\lambda$							
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383
4		0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077
5				0,0001	0,0002	0,0004	0,0007	0,0012
6						0,0001	0,0002	

$k$	$\lambda$	0.9	1.0	1.5	2.0	2.5	3.0	3.5	4.0
0	0.4066	0.3679	0.2231	0.1353	0.0821	0.0498	0.3020	0.0183	
1	0.3659	0.3679	0.3347	0.2707	0.2052	0.1494	0.1507	0.0733	
2	0.1647	0.1839	0.2510	0.2707	0.2565	0.2240	0.1850	0.1465	
3	0.0494	0.0613	0.1255	0.1804	0.2138	0.2240	0.2158	0.1954	
4	0.0111	0.0153	0.0471	0.0902	0.1336	0.1680	0.1888	0.1954	
5	0.0020	0.0031	0.0141	0.0361	0.0668	0.1008	0.1322	0.1563	
6	0.0003	0.0005	0.0035	0.0120	0.0278	0.0504	0.0771	0.1042	
7		0.0001	0.0008	0.0034	0.0099	0.0216	0.0385	0.0595	
8			0.0001	0.0009	0.0031	0.0081	0.0169	0.0298	
9				0.0002	0.0009	0.0027	0.0066	0.0132	
10					0.0002	0.0008	0.0023	0.0053	
11						0.0002	0.0007	0.0019	
12						0.0001	0.0002	0.0006	
13							0.0001	0.0002	
14								0.0001	
$k$	$\lambda$	4.5	5.0	6.0	7.0	8.0	9.0	10.0	
0	0.0111	0.0067	0.0025	0.0009	0.0003	0.0001			
1	0.0500	0.0337	0.0149	0.0064	0.0027	0.0011	0.0005		
2	0.1125	0.0842	0.0446	0.0223	0.0107	0.0050	0.0023		
3	0.1687	0.1404	0.0892	0.0521	0.0286	0.0150	0.0076		
4	0.1898	0.1755	0.1339	0.0912	0.0573	0.0337	0.0189		
5	0.1708	0.1755	0.1606	0.1277	0.0916	1.0607	0.0378		
6	0.1281	0.1462	0.1606	0.1490	0.1221	0.0911	0.0631		
7	0.0824	0.1044	0.1377	0.1490	0.1396	0.1171	0.0901		
8	0.0463	0.0653	0.1033	0.1304	0.1396	0.1318	0.1126		
9	0.0232	0.0363	0.0688	0.1014	0.1241	0.1318	0.1251		
10	0.0104	0.0181	0.0413	0.0710	0.0993	0.1186	0.1251		
11	0.0043	0.0082	0.0225	0.0452	0.0722	0.0970	0.1137		
12	0.0016	0.0034	0.0113	0.0264	0.0481	0.0728	0.0948		
13	0.0006	0.0013	0.0052	0.0142	0.0215	0.0504	0.0729		
14	0.0002	0.0005	0.0022	0.0071	0.0169	0.0324	0.0521		
15	0.0001	0.0002	0.0009	0.0033	0.0090	0.0194	0.0347		
16			0.0003	0.0014	0.0045	0.0109	0.0217		
17			0.0001	0.0006	0.0021	0.0058	0.0128		
18				0.0002	0.0009	0.0029	0.0071		
19				0.0001	0.0004	0.0014	0.0037		
20					0.0002	0.0006	0.0019		
21					0.0001	0.0003	0.0009		
22						0.0001	0.0004		
23							0.0002		
24							0.0001		

Tabla 2 (continuación)

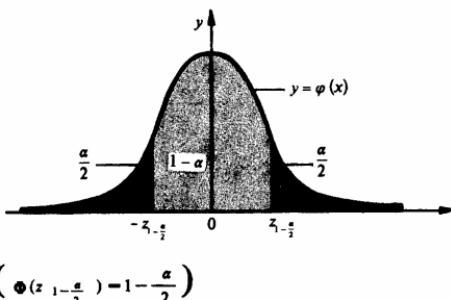
$k$	$\lambda$	12	14	16	18	20
1	0.0001					
2	0.0004	0.0001				
3	0.0018	0.0004	0.0001			
4	0.0053	0.0013	0.0003	0.0001		
5	0.0127	0.0037	0.0010	0.0002		
6	0.0255	0.0087	0.0026	0.0007	0.0002	
7	0.0437	0.0174	0.0060	0.0019	0.0005	
8	0.0655	0.0304	0.0120	0.0042	0.0013	
9	0.0874	0.0473	0.0213	0.0083	0.0029	
10	0.1048	0.0663	0.0341	0.0150	0.0059	
11	0.1144	0.0844	0.0496	0.0245	0.0106	
12	0.1144	0.0984	0.0661	0.0368	0.0176	
13	0.1055	0.1060	0.0814	0.0509	0.0271	
14	0.0905	0.1060	0.0930	0.0655	0.0387	
15	0.0724	0.0989	0.0992	0.0786	0.0517	
16	0.0543	0.0866	0.0992	0.0884	0.0645	
17	0.0383	0.0713	0.0934	0.0936	0.0760	
18	0.0256	0.0554	0.0830	0.0936	0.0844	
19	0.0161	0.0409	0.0699	0.0887	0.0888	
20	0.0097	0.0286	0.0559	0.0798	0.0888	
21	0.0055	0.0191	0.0426	0.0684	0.0888	
22	0.0030	0.0121	0.0310	0.0559	0.0769	
23	0.0016	0.0074	0.0216	0.0438	0.0669	
24	0.0008	0.0043	0.0144	0.0328	0.0557	
25	0.0004	0.0024	0.0092	0.0237	0.0445	
26	0.0002	0.0013	0.0057	0.0164	0.0343	
27	0.0001	0.0007	0.0033	0.0109	0.0254	
28		0.0003	0.0019	0.0070	0.0481	
29		0.0002	0.0011	0.0044	0.0125	
30		0.0001	0.0006	0.0026	0.0084	
31			0.0002	0.0015	0.0053	
32			0.0001	0.0009	0.0034	
33			0.0001	0.0005	0.0020	
34				0.0003	0.0013	
35				0.0001	0.0007	
36					0.0004	
37					0.0002	
38					0.0001	

## 12.3 Tabla de la distribución normal

La tabla 3 da una panorámica sobre la función de distribución  $\Phi$  de la distribución normal estandarizada

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

para  $0 \leq x \leq 3.9$ . Para  $x < 0$  se utiliza la relación  $\Phi(x) = 1 - \Phi(-x)$  (ver 5.4 (15)).



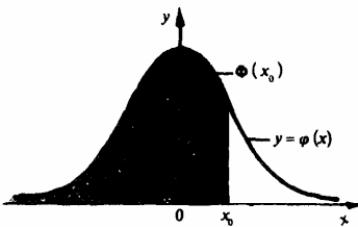
$$\left( \Phi(z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \right)$$

En la tabla siguiente se agrupan algunos percentiles de la distribución normal estandarizada, los cuales se utilizan frecuentemente en la realización práctica de las estimaciones por intervalo de confianza, indicadas en los epígrafes 10.6.1 a) y 10.6.2, y de las dócimas de significación, descritas en los epígrafes 11.1 y 11.4.5.

Tabla 3

Ejemplo:  $\Phi(1,43) = 0,923642$

x	0,00	0,01	0,02	0,03	0,04
0,0	0,500000	0,503989	0,507978	0,511966	0,515953
0,1	0,539828	0,543795	0,547758	0,551717	0,555670
0,2	0,579260	0,583166	0,587064	0,590954	0,594835
0,3	0,617911	0,621720	0,625516	0,629300	0,633072
0,4	0,655422	0,659097	0,662757	0,666402	0,670031
0,5	0,691462	0,694974	0,698468	0,701944	0,705402
0,6	0,725747	0,729069	0,732371	0,735653	0,738914
0,7	0,758036	0,761148	0,764238	0,767305	0,770350
0,8	0,788145	0,791030	0,793892	0,796731	0,799546
0,9	0,815940	0,818589	0,821214	0,823814	0,826391
1,0	0,841345	0,843752	0,846136	0,848495	0,850830
1,1	0,864334	0,866500	0,868643	0,870762	0,872857
1,2	0,884930	0,886861	0,888768	0,890651	0,892512
1,3	0,903200	0,904902	0,906582	0,908241	0,909877
1,4	0,919243	0,920730	0,922196	0,923642	0,925066
1,5	0,933193	0,934478	0,935744	0,936992	0,938220
1,6	0,945201	0,946301	0,947384	0,948449	0,949497
1,7	0,955434	0,956367	0,957284	0,958185	0,959070
1,8	0,964070	0,964852	0,965620	0,966375	0,967116
1,9	0,971283	0,971933	0,972571	0,973197	0,973810
2,0	0,977250	0,977784	0,978308	0,978822	0,979325
2,1	0,982136	0,982571	0,982997	0,983414	0,983823
2,2	0,986097	0,986447	0,986791	0,987126	0,987454
2,3	0,989276	0,989556	0,989830	0,990097	0,990358
2,4	0,991802	0,992024	0,992240	0,992451	0,992656
2,5	0,993790	0,993963	0,994132	0,994297	0,994457
2,6	0,995339	0,995473	0,995604	0,995731	0,995855
2,7	0,996533	0,996636	0,996736	0,996833	0,996928
2,8	0,997445	0,997523	0,997599	0,997673	0,997744
2,9	0,998134	0,998193	0,998250	0,998305	0,998359
	0,0	0,1	0,2	0,3	0,4
3,0	0,998650	0,999032	0,999313	0,999517	0,999663



$x$	0,05	0,06	0,07	0,08	0,09
0,0	0,519938	0,523922	0,527903	0,531881	0,535856
0,1	0,559618	0,563560	0,567495	0,571424	0,575345
0,2	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,708840	0,712260	0,715661	0,719043	0,722405
0,6	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,773373	0,776373	0,779350	0,872305	0,785236
0,8	0,802338	0,805106	0,807850	0,810570	0,813267
0,9	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,874928	0,876976	0,879000	0,881000	0,882977
1,2	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,911492	0,913085	0,914656	0,916207	0,917736
1,4	0,926471	0,927855	0,929219	0,930563	0,931889
1,5	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,950528	0,951543	0,952540	0,953521	0,954486
1,7	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,974412	0,975002	0,975581	0,976148	0,976704
2,0	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,990613	0,990862	0,991106	0,991344	0,991576
2,4	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998411	0,998462	0,998511	0,998559	0,998605
	0,5	0,6	0,7	0,8	0,9
3,0	0,999767	0,999841	0,999892	0,999928	0,999952

## 12.4 Tabla de la distribución $\chi^2$

La tabla 4 contiene algunos porcentiles  $\chi_{m,p}^2$  de la distribución  $\chi^2$  con  $m$  grados de libertad (ver 5.6, definición 2) para  $m=1, 2, \dots, 30, 40, \dots, 100$ , los cuales se utilizan frecuentemente en la realización práctica de las estimaciones por intervalo de confianza, indicadas en los epígrafes 10.6.1 (c) y (d), y de las dócimas de significación descritas en los epígrafes 11.4.3, 11.5.1, 11.5.3 y 11.5.5 (dócima de dispersión  $\chi^2$ , dócima de ajuste  $\chi^2$ , dócima de homogeneidad  $\chi^2$ , dócima de independencia  $\chi^2$ ).

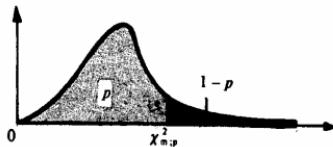


Tabla 4

Ejemplo:  $\chi_{m=p}^2 = 12.59$

$m$	$p=0.99$ ( $1-p=0.01$ )	$0.975$ ( $0.025$ )	$0.95$ ( $0.05$ )	$0.05$ ( $0.95$ )	$0.025$ ( $0.975$ )	$0.01$ ( $0.99$ )
1	6.635	5.024	3.841	0.0039	0.0010	0.0002
2	9.210	7.378	5.991	0.1026	0.0506	0.0201
3	11.34	9.348	7.815	0.3518	0.2158	0.1148
4	13.28	11.14	9.488	0.7107	0.4844	0.2971
5	15.09	12.83	11.07	1.145	0.8312	0.5543
6	16.81	14.45	12.59	1.635	1.237	0.8721
7	18.48	16.01	14.07	2.167	1.690	1.239
8	20.09	17.53	15.51	2.733	2.180	1.646
9	21.67	19.02	16.92	3.325	2.700	2.088
10	23.21	20.48	18.31	3.940	3.247	2.558

<i>m</i>	<i>p</i> =0.99 (1- <i>p</i> =0.01)	0.975 (0.025)	0.95 (0.05)	0.05 (0.95)	0.025 (0.975)	0.01 (0.99)
11	24.72	21.92	19.68	4.575	3.816	3.053
12	26.22	23.34	21.03	5.226	4.404	3.571
13	27.69	24.74	22.36	5.892	5.009	4.107
14	29.14	26.12	23.68	6.571	5.629	4.660
15	30.58	27.49	25.00	7.261	6.262	5.229
16	32.00	28.85	26.30	7.962	6.908	5.812
17	33.41	30.19	27.59	8.672	7.564	6.408
18	34.81	31.53	28.87	9.390	8.231	7.015
19	36.19	32.85	30.14	10.12	8.907	7.633
20	37.57	34.17	31.41	10.85	9.591	8.260
21	38.93	35.48	32.67	11.59	10.28	8.897
22	40.29	36.78	33.92	12.34	10.98	9.542
23	41.64	38.08	35.17	13.09	11.69	10.20
24	42.98	39.36	36.42	13.85	12.40	10.86
25	44.31	40.65	37.65	14.61	13.12	11.52
26	45.64	41.92	38.89	15.38	13.84	12.20
27	46.96	43.19	40.11	16.15	14.57	12.88
28	48.28	44.46	41.34	16.93	15.31	13.56
29	49.59	45.72	42.56	17.71	16.05	14.26
30	50.89	46.98	43.77	18.49	16.79	14.95
40	63.69	59.34	55.76	26.51	24.43	22.16
50	76.15	71.42	67.50	34.76	32.36	29.71
60	88.38	83.30	79.08	43.19	40.48	37.48
70	100.42	95.02	90.53	51.74	48.76	45.44
80	112.33	106.63	101.88	60.39	57.15	53.54
90	124.12	118.14	113.14	69.13	65.65	61.75
100	135.81	129.56	124.34	77.93	74.22	70.06

## 12.5 Tabla de la distribución $t$

La tabla 5 contiene algunos percentiles  $t_{m,p}$  de la distribución  $t$  con  $m$  grados de libertad (ver 5.6, definición 3) para  $m=1, 2, \dots, 30, 40, 60, 120, \infty$ , los cuales se utilizan frecuentemente en la realización práctica de las estimaciones por intervalo de confianza, indicadas en el epígrafe 10.6.1b), y en las dócimas de significación descritas en los epígrafes 11.4.1 y 11.4.2 (dócima  $t$  simple, dócima  $t$  doble).

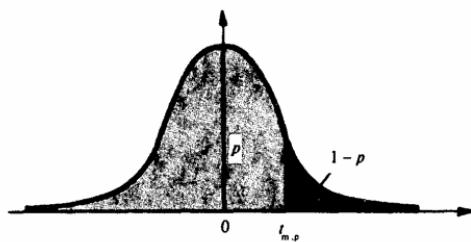


Tabla 5

Ejemplo:  $t_{17;0,975}=2,110$

$m$	$p=0,9$ ( $1-p=0,1$ )	$0,95$ ( $0,05$ )	$0,975$ ( $0,025$ )	$0,99$ ( $0,01$ )	$0,995$ ( $0,005$ )
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845

<i>m</i>	<i>p</i> =0,9 (1− <i>p</i> ,0,1)	0,95 (0,05)	0,975 (0,025)	0,99 (0,01)	0,995 (0,005)
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576

## 12.6 Tabla de la distribución *F*

Las tablas 6a) y 6b) contienen los percentiles  $F_{m_1, m_2, p}$  de la distribución *F* con ( $m_1, m_2$ ) grados de libertad (ver 5.6. definición 4) para  $p=0,95$  y  $p=0,99$ , respectivamente. Estos percentiles se necesitan especialmente para la realización práctica de la dócima de significación descrita en el epígrafe 11.4.4 (dócima *F*) con el nivel de significación  $\alpha=10\%$  o  $\alpha=2\%$ . Además, los números  $F_{m_1, m_2+1, p}$  para  $p=0,95$  y  $p=0,99$  pueden tomarse de las tablas 6a) y 6b) en virtud de la fórmula

$$F_{m_1, m_2+1, p} = \frac{1}{F_{m_1, m_2, p}}.$$

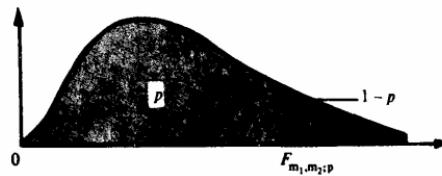
Tabla 6

Ejemplo:  $F_{20,15,0.99}=3.37$ .

$$F_{20,15,0.01} = \frac{1}{F_{15,20,0.99}} = \frac{1}{3.09} = 0.32$$

a)  $p=0.95$  ( $1-p=0.05$ )

$m_1$	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.39	6.59	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.39	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88



$m_1 \backslash m_2$	10	12	15	20	24	30	40	60	120	$\infty$
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
$\infty$	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Tabla 6 (continuación)

b)  $p=0,99$  ( $1-p=0,01$ )

$m_1 \backslash m_2$	1	2	3	4	5	6	7	8	9
1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022
2	98,50	99,90	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40
22	7,95	5,72	4,82	4,31	3,94	3,76	3,59	3,45	3,35
23	7,88	5,66	4,76	3,26	3,71	3,71	3,54	3,41	3,30
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56
$\infty$	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

$m_1$	10	12	15	20	24	30	40	60	120	$\infty$
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
$\infty$	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

### 13. Breve bosquejo de la historia del cálculo de probabilidades

Después que hemos expuesto la construcción matemática, usual hoy día, de la teoría de probabilidades y tratado algunas tareas esenciales que se plantea la estadística matemática, queremos dar en este último capítulo una breve panorámica de la historia del cálculo de probabilidades, con la cual deben ser completadas, perfiladas y clasificadas las observaciones históricas incluidas en los capítulos precedentes.

El cálculo de probabilidades pertenece a las disciplinas matemáticas relativamente jóvenes; ella tiene solo escasamente tres siglos de existencia. Sin embargo, el mundo misterioso de la causalidad interesó a los sabios en el más temprano estadio del pensamiento científico. Así, el concepto probabilidad surgió ya en la filosofía griega antigua. La idea de que las regularidades de la naturaleza se expresan mediante un número enorme de fenómenos aleatorios, se presenta también en los materialistas griegos de la antigüedad. (Esta idea toma cuerpo muy claramente, por ejemplo, en la poesía "De rerum natura" (Sobre la naturaleza de las cosas) de Lukrez (un siglo antes de nuestra era).) Pero el desarrollo hacia una disciplina científica independiente comienza solo en la mitad del siglo XVII. Estimulado por preguntas acerca de las probabilidades de ganancia en juegos de azar, formuladas por un jugador apasionado amigo suyo, el caballero de Méré, el notable matemático francés Blaise Pascal (1623-1662) estableció en el año 1654 un intercambio de correspondencia con el no menos famoso Pierre de Fermat (1601-1665), en la cual fueron desarrollados -yendo más allá del propio motivo- fundamentos importantes del cálculo de probabilidades. Ya desde antes, hubo sabios que se ocuparon con problemas especiales sobre las probabilidades en juegos de azar, como por ejemplo, el monje franciscano Luca de Pacioli (1445-1514) en su libro publicado en 1494 "Summa de Arithmetica, Geometria, Proportioni e Propotionalita", el médico milanés Hieronimo Cardano (1501 hasta 1576) en su obra "Liber de ludo aleae" (Libro sobre los juegos de azar) y también Galileo Galilei (1564-1642). El cálculo de probabilidades fue concebido por primera vez como un medio adecuado para la investigación de fenómenos aleatorios por Pascal y Fermat.

También el físico, matemático y astrónomo holandés Christiaan Huygens (1629-1695) estuvo consciente de la significación de esta nueva dirección matemática. Así escribió él en su libro "De ratiociniis in ludo aleae" (Sobre los cálculos posibles en juegos de azar), publicado en 1658 y en el que se toma como referencia las ideas expresadas por Pascal y Fermat: "... que el lector observa en un estudio atento del objeto, que no se trata solo

de juegos, sino que aquí se desarrollan las bases de una teoría muy interesante y productiva".<sup>1</sup>

Solo que a causa del nivel relativamente bajo de desarrollo de las ciencias naturales fueron en este tiempo los juegos de azar, las interrogantes de la estadística poblacional y las tareas de aseguramiento, los únicos problemas concretos sobre la base de los cuales pudo ser desarrollado el cálculo de probabilidades.

En el libro mencionado de Huygens no aparece, por lo demás, el concepto "probabilidad"; en él siempre se habla de "valor de la esperanza", magnitud que hoy denominamos valor esperado. El concepto probabilidad se definió por primera vez en el libro publicado en 1713 "Ars conjectandi" (El arte del suponer) de Jakob Bernoulli (1654-1705); aquí se entendió por probabilidad "el grado de certeza, que con respecto a la certeza se comporta como la parte al todo"<sup>2</sup>, una definición que tiene más carácter filosófico que matemático.

La obra "Ars conjectandi", que se puede considerar como primer libro de texto del cálculo de probabilidades, contiene, además de un tratamiento completo de todos los problemas sin solución señalados por Huygens, una deducción notablemente exacta (no solo para las condiciones de aquél entonces) de la proposición formulada hoy como Ley de los grandes números de Bernoulli; con ella se da, por consiguiente, una explicación teórica de la estabilización de la frecuencia relativa de este hecho observado una y otra vez y conocido ya antes de Bernoulli. El mérito de Bernoulli no consiste, por tanto, en el descubrimiento de este fenómeno -con referencia a esto, el propio Bernoulli escribió en "Ars conjectandi": "A cada uno le está claro también que no es suficiente para valorar un fenómeno cualquiera hacer una o dos observaciones, sino que es necesario un número grande de ellas. Por esta razón, el hombre más limitado sabe por sí mismo y sin ninguna instrucción anterior (lo cual es asombroso), que cuanto más observaciones se tomen en consideración tanto menor será el peligro de no lograr el objetivo"; el mérito de Jakob Bernoulli consiste sobre todo en la explicación teórica, rigurosamente fundamentada, de esta situación. Para esta época fue característico que hechos empíricos -como por ejemplo, la estabilización de las frecuencias relativas- fueran conocidos, pero que no se buscaran fundamentaciones teóricas para ellos; estos hechos fueron considerados más bien como manifestaciones del orden divino, que no requerían ninguna otra aclaración.

El matemático francés Abraham De Moivre (1667-1754) logró entonces, entre otras cosas, la formulación cuantitativa de la Ley de los grandes números de Bernoulli con la proposición que hemos denominado como Teorema integral de De Moivre-Laplace, y también, relacionado con esto, descubrió la distribución normal (ver el final de 5.4).

La indicación explícita de la llamada definición clásica de probabilidad se encuentra por primera vez en la obra fundamental aparecida en 1812 "Theorie analytique des probabilités" (Teoría analítica de las probabilidades) del importante matemático y físico francés Pierre Simon Laplace (1749-1827). Allí se considera -en completa concordancia con la concepción actual- la definición clásica de probabilidad, no tanto como una definición, sino como una fórmula para el cálculo de probabilidades en casos concretos, para los cuales se satisfacen ciertas condiciones; Laplace escribió: "La probabilidad de un suceso es la razón del número de casos propicios y el de todos los posibles, suponiéndose los distintos casos como igualmente posibles".<sup>3</sup>

<sup>1</sup>B.V.: La cita fue tomada de [15].

<sup>2</sup>B.V.: La cita fue tomada de [6].

<sup>3</sup>La cita fue tomada de [15].

La nombrada obra de Laplace contiene una exposición sistemática de los resultados clásicos del cálculo de probabilidades, se demuestran los teoremas conocidos entonces, en particular la proposición denominada hoy día como Teorema Integral de De Moivre-Laplace; además, Laplace expuso el método de la suma de los mínimos cuadrados desarrollado por él (e independientemente y casi al mismo tiempo por Carl Friedrich Gauss (1777-1855) y por Adrien Marie Legendre (1752-1833)) en relación con problemas del cálculo de errores y de compensación. Él se ocupó también de la aplicación del cálculo de probabilidades a interrogantes de la estadística poblacional y realizó investigaciones estadísticas sobre la base de un amplio material numérico.

Los trabajos de Laplace sobre el cálculo de probabilidades junto con los trabajos del matemático francés Simón Denis Poisson (1781-1840), forman parte importante de los grandes progresos en esta especialidad en las postrimerías del siglo XVIII e inicios del XIX. Poisson realizó una generalización de la Ley de los grandes números de Bernoulli -de él provino también el concepto "Ley de los grandes números" -al caso de experimentos independientes en los cuales la probabilidad de la ocurrencia de un suceso es dependiente del número del experimento. Además, extendió el Teorema integral de De Moivre-Laplace a este caso y descubrió con esto la distribución de probabilidad que lleva su nombre; él aplicó los resultados obtenidos, en particular, a la balística.

Mediante De Moivre, Laplace y Poisson sobrevino un incremento considerable en el desarrollo de métodos analíticos especiales del cálculo de probabilidades, con numerosos resultados hermosos y valiosos; los problemas de las ciencias naturales (por ejemplo, de la balística y la astronomía) y las interrogantes relacionadas con la teoría de los errores de observación sirvieron sobre todo de estímulo para esto.

Es verdad que en aquel tiempo existieron bastantes valoraciones erróneas en cuanto a las posibilidades de aplicación del cálculo de probabilidades, a las cuales dieron lugar sus representantes más prominentes. Así por ejemplo, fue intentado -con intercesión y favorecimiento energético de Laplace y Poisson- abarcar por medio del cálculo de probabilidades el contenido de verdad del veredicto de un jurado llevado a cabo por mayoría de votos. Esto repercutió desventajosamente en el desarrollo del cálculo de probabilidades. Sobre la base de los -forzosamente declarados- fracasos se convirtió en desilusión el entusiasmo existente al principio por el cálculo de probabilidades en los centros científicos de Europa Occidental, surgieron dudas o incluso rechazo; en el mejor de los casos fue concebido el cálculo de probabilidades como objeto de la conversación matemática.

Frente a esto, el desarrollo impetuoso de la física impuso elevadas exigencias a la matemática, en general, y al cálculo de probabilidades, en particular. En este tiempo se desarrolló una fuerte escuela del cálculo de probabilidades en la entonces ciudad de San Petersburgo. Ella fue fundada por Pafnudi Lvovich Chebyshev (1821-1894), quien publicó en total solo cuatro trabajos sobre el cálculo de probabilidades, pero cuya influencia sobre el desarrollo posterior de esta disciplina es considerable. Los méritos de Chebyshev consisten, sobre todo, en que hizo estimaciones acerca de las posibles desviaciones de las regularidades límites y en que elaboró métodos apropiados para describir esto. Además, impuso la exigencia hacia un rigor absoluto en las demostraciones de los teoremas límites e indicó el lugar central correspondiente a los conceptos "variable aleatoria" y "valor esperado" en el sistema de conceptos del cálculo de probabilidades. Famosos representantes de la escuela rusa del cálculo de probabilidades fundada por Chebyshev fueron Andrei Andreevich Markov (1856-1922) y Alexander Mikailovich Liapunov (1857-1918); nos encontramos estos nombres ya, en el tratamiento de las leyes de los grandes números y de los teoremas límites del cálculo de probabilidades.

No obstante la importancia de los resultados logrados al final del siglo pasado y al inicio del nuestro en el cálculo de probabilidades y en su aplicación, este permaneció atrás en comparación con otras teorías, en lo referente al desarrollo de los fundamentos de la teoría matemática. De forma sorprendente, el cálculo de probabilidades no fue alcanzado durante largo tiempo por la enorme transformación de la matemática en el siglo XIX, que estuvo caracterizada por la construcción axiomática de teorías matemáticas, lógicamente compatibles, cerradas en sí y desligadas de la realidad (por ejemplo, la Teoría de Conjuntos, la Topología). Dijimos ya anteriormente (véase para ello la introducción de 2) que en el segundo Congreso Internacional de Matemáticos en París en el año 1900, David Hilbert (1862-1943) mencionó como uno de los problemas matemáticos más importantes la aclaración de los conceptos básicos del cálculo de probabilidades. Con esta tarea se ocuparon muchos matemáticos, entre ellos el matemático austriaco Richard Von Mises (1883-1953), cuya tentativa para la solución de esta tarea provocó vehementes -y por lo demás fructíferas- discusiones y estimuló el interés de muchos matemáticos. Una solución satisfactoria del problema formulado por Hilbert se realizó con la publicación (1933) del famoso matemático soviético Andrei Nikolaevich Kolmogorov (nacido en 1903), quien después de numerosos trabajos preliminares logró emprender una construcción axiomática del cálculo de probabilidades, de acuerdo con el espíritu de la matemática moderna. Aquí fueron representados los sucesos aleatorios mediante conjuntos y la probabilidad se concibió como una función definida sobre estos conjuntos con determinadas propiedades, caracterizadas mediante axiomas. Esta construcción condujo no solo a la aclaración de los fundamentos lógicos del cálculo de probabilidades, sino también permitió, en particular, la utilización de disciplinas matemáticas modernas altamente desarrolladas, por ejemplo, de la Teoría de Conjuntos y del Análisis, en especial, de la Teoría de la Medida y de la Integración. El cálculo de probabilidades se desarrolló desde entonces impetuosamente, tanto respecto a la teoría matemática, como al campo de aplicación de esta teoría.

Hoy en día un gran número de centros de altos rendimientos se ocupan de la Teoría de probabilidades, la Estadística matemática y las numerosas disciplinas especiales surgidas de estas. Una función rectora corresponde a los teóricos soviéticos de las probabilidades cuyos trabajos son de interés y poseen reconocimiento internacional. En los primeros años después de la Revolución de Octubre, se concentró el círculo de los que se ocupaban en la URSS de la Teoría de las probabilidades, sobre todo en Moscú, alrededor de Alexander Jakovlevich Kinchine (1894-1959), uno de los representantes más significativos de la Teoría de probabilidades de nuestro siglo, y de A.N. Kolmogorov; hoy existe una multitud de centros de la Teoría de probabilidades en la URSS, considerados internacionalmente. En la RDA ocupa la Teoría de las probabilidades un lugar fijo en el marco de la formación en universidades e institutos de enseñanza superior y también en la investigación matemática. En el camino hacia este objetivo fue muy provechoso el magisterio de B. V. Gnedenko en el año 1953, en la Universidad de Humboldt, en Berlín, y muchos de los matemáticos de la RDA que hoy investigan en el campo de la Teoría de probabilidades fueron formados en la Unión Soviética o permanecieron allí para realizar estudios.

Desde hace algunos años se hacen mayores esfuerzos -también en marcos internacionales- para incluir el Cálculo de Probabilidades, de forma adecuada, en la formación matemática en las escuelas de enseñanza general.

## Bibliografía

Solo se enumeran títulos sobre Teoría de probabilidades y Estadística matemática en lengua alemana, que han sido publicados o se pueden adquirir en la RDA, sin pretender con ello citar todos los existentes sobre esta temática; las escasas anotaciones complementarias deben auxiliar en la selección de la bibliografía.

- [1] MÜLLER, P.H. (editor y autor coordinador), *Lexikon der Stokastik (Wahrscheinlichkeitstheorie und Mathematische Statistik)*, 2. Auflage, Akademie - Verlag, Berlin, 1975.  
Se explican y se resumen lexicográficamente, en palabras claves, las ideas esenciales de la Teoría de probabilidades, la Estadística matemática y algunas importantes disciplinas especiales que han surgido de éstas.
- [2] MÜLLER, P.H., P. NEUMANN y R. STORM, *Tafeln der Mathematischen Statistik*, 2. Auflage, VEB Fachbuchverlag, Leipzig, 1975.  
Esta colección de tablas contiene un programa básico en tablas, con cuya ayuda pueden tratarse la mayor parte de los problemas prácticos de la Estadística matemática.
- [3] MAIBAUM, G., *Wahrscheinlichkeitsrechnung*, 2. Auflage, Volk und Wissen Volkseigener Verlag, Berlin, 1975.  
Este libro ha sido concebido como texto para las clases facultativas en la escuela media superior ampliada (grados 11 y 12); contiene una exposición detallada del Cálculo de probabilidades en la medida en que esto es preciso para la realización de un curso de esta disciplina, sobre la base de los programas vigentes.
- [4] DONAT, C.D. y G. MAIBAUM, *Wahrscheinlichkeitsrechnung (Fachlichmethodische Hinweise zum Lehrgang Wahrscheinlichkeitsrechnung im Rahmen des fakultativen Unterrichts in der 11. und 12. Klasse)*, Volk und Wissen Volkseigener Verlag, Berlin, 1972.  
El objetivo de este folleto se hace evidente a través del subtítulo. [3] constituye el punto de referencia de las indicaciones metodológicas.
- [5] CLAUS, G. y H. EBNER, *Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen*, Volk und Wissen Volkseigener Verlag, Berlin, 1974.  
Junto a una exposición, realizada conscientemente de manera sencilla, de los fundamentos matemáticos, el libro contiene una serie de procedimientos estadísticos que se aplican de manera creciente en la investigación pedagógica, psicológica y sociológica. Aquí se tratan detalladamente problemas específicos de la aplicación de procedimientos estadísticos a interrogantes de estas ramas. Los numerosos ejemplos de este libro proceden por entero de los dominios de la pedagogía, la psicología y la sociología.
- [6] RÉNYI, A, *Briefe über die Wahrscheinlichkeit*, 2. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin, 1972 (traducción del húngaro).  
En este pequeño libro se explican las cuestiones fundamentales del Cálculo de probabilidades de forma sumamente agradable, desde el punto de vista literario. El lector encuentra, además, detalles interesantes acerca de los inicios del Cálculo de probabilidades.

- Los tres títulos que se mencionan a continuación son colecciones de ejercicios; [7] y [8] contienen, además, breves exposiciones de la materia.
- [7] SWESCHNIKOW, S.A., *Wahrscheinlichkeitsrechnung und Mathematische Statistik in Aufgaben*, BSB B. Teubner Verlagsgesellschaft, Leipzig, 1970 (traducción del ruso).
  - [8] WENTZEL, E.S. y L.A. OWTCHAROW, *Aufgabensammlung zur Wahrscheinlichkeitsrechnung*, Akademie-Verlag, Berlin, 1973 (traducción del ruso).
  - [9] *Wahrscheinlichkeitsrechnung und Mathematische Statistik* (*Übungsaufgaben zur Mathematik, Heft 8*, TU Dresden, Sektion Mathematik). Impreso como manuscrito 1971.

Los siguientes títulos pueden tomarse para ampliar y profundizar el estudio de la Teoría de probabilidades, la Estadística matemática y -como se puede apreciar de los títulos- otras ramas especiales de la estocástica.

- [10] AHRENS, H., *Varianzanalyse*, Akademie-Verlag, Berlin, 1967.
- [11] AHRENS, H. y J. LÄUTER, *Mehrdimensionale Varianzanalyse*, Akademie-Verlag, Berlin, 1974.
- [12] BANDEMER, H. y otros, *Optimale Versuchsplanung*, Akademie-Verlag, Berlin, 1973.
- [13] FABIAN, V., *Statistische Methoden*, 2. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin, 1970 (traducción del checo).
- [14] FISZ, M., *Wahrscheinlichkeitsrechnung und mathematische Statistik*, 7. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin, 1973 (traducción del polaco).
- [15] GNEDENKO, B.W., *Lehrbuch der Wahrscheinlichkeitsrechnung*, 6. Auflage, Akademie-Verlag, Berlin, 1970 (traducción del ruso).
- [16] JAHN, W. y H. VAHLE, *Die Faktorenanalyse und ihre Anwendung*, Verlag Die Wirtschaft, Berlin, 1970.
- [17] NOLLAU, V., *Statistische Analysen*, VEB Fachbuchverlag, Leipzig, 1975.
- [18] PAWLICKI, Z., *Einführung in die mathematische Statistik*, Verlag Die Wirtschaft, Berlin, 1971 (traducción del polaco).
- [19] RAO, C.R., *Lineare statistische Methoden und ihre Anwendungen*, Akademie-Verlag, Berlin, 1973 (traducción del inglés).
- [20] RASCH, D., *Elementare Einführung in die mathematische Statistik*, 2. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin, 1970.
- [21] RÉNYI, A., *Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie*, 3. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin, 1971.
- [22] ROSANOW, J.A., *Wahrscheinlichkeitstheorie*, 2. Auflage, Akademie-Verlag, Berlin, 1972 (traducción del ruso).
- [23] ROSANOW, J.A., *Stochastische Prozesse*, Akademie-Verlag, Berlin, 1975 (traducción del ruso).
- [24] SMIRNOW, N.W. y I.W. DUNIN-BARKOWSKI, *Mathematische Statistik in der Technik*, 3. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin, 1973 (traducción del ruso).
- [25] STORM, R., *Wahrscheinlichkeitsrechnung. Mathematische Statistik. Statische Qualitätskontrolle*, 5. Auflage, VEB Fachbuchverlag, Leipzig, 1974.
- [26] VINCZE, I., *Mathematische Statistik mit industriellen Anwendungen*, Akadémiai Kiadó, Budapest, 1971.
- [27] WEBER, E., *Grundriss der biologischen Statistik*, 7. Auflage, VEB Gustav Fischer Verlag, Jena, 1972.
- [28] WEBER, E., *Einführung in die Faktorenanalyse*, VEB Gustav Fischer Verlag, Jena, 1974.

Por último, llamamos la atención de que [15] contiene un bosquejo de la historia del Cálculo de probabilidades.

*Impreso por el Combinado Poligráfico de Guantánamo  
"Juan Marinello" en el mes de Abril de 1988  
"Año XXX de la Revolución"*