Teorías de la Medida
y de la Probabilidad

Colección manuales uex - 57
(E.E.E.S.)

Agustín
García Nogales

57

TEORÍAS DE LA MEDIDA
Y DE LA PROBABILIDAD

MANUALES UEX

57

(E.E.E.S.)

Espacio
Europeo
Educación
Superior

AGUSTÍN GARCÍA NOGALES

TEORÍAS DE LA MEDIDA
Y DE LA PROBABILIDAD

UNIVERSIDAD DE EXTREMADURA

2008

La publicación del presente manual forma parte de las "Acciones para el Desarrollo del Espacio Europeo de Educación Superior en la Universidad de Extremadura Curso 2007/08" en el marco de la VI Convocatoria de Acciones para la Adaptación de la UEX al Espacio Europeo de Educación Superior (Proyectos Pilotos: modalidad A1) del Vicerrectorado de Calidad y Formación Continua y financiada por la Junta de Extremadura, el Ministerio de Educación y Ciencia y la Universidad de Extremadura.

UNIVERSIDAD DE EXTREMADURA

UNIÓN EUROPEA
Fondo Social Europeo

JUNTA DE EXTREMADURA

Edita

Universidad de Extremadura. Servicio de Publicaciones
C./ Caldereros, 2 - Planta 2ª - 10071 Cáceres (España)
Telf. 927 257 041 - Fax 927 257 046
publicac@unex.es
www.unex.es/publicaciones

ISSN 1135-870-X
ISBN 978-84-691-6411-2
Depósito Legal M-45.182-2008

Edición electrónica: Pedro Cid, S.A.
Teléf.: 914 786 125

A María José y Carlos

A mis padres

Índice general

Prólogo

Este manual sobre las teorías de la Medida y de la Probabilidad ha sido elaborado a partir de los apuntes de clase de una asignatura homónima en estudios de Estadística en la Universidad de Extremadura y pretende, antes que convertirse en una obra exhaustiva y autosuficiente sobre Medida y Probabilidad, ser de utilidad en clase.

Su principal objetivo consiste en presentar la Teoría de la Probabilidad desde un punto de vista matemáticamente riguroso a partir del tronco común que comparte con la Teoría de la Medida, tal y como nos enseñó hace 75 años *Andréi Nikoláyevich Kolmogórov,* matemático ruso que consiguió convencernos de que medir áreas o volúmenes y calcular probabilidades no son problemas tan distintos como pudieran parecer en principio. Permítasenos reproducir aquí parte del Prólogo de Nogales (1998):

"... Las trayectorias paralelas que históricamente siguieron la Teoría de la Medida y el Cálculo de Probabilidades hasta que, en 1933, A.N. Kolmogorov fijó una base axiomática común para ambas teorías, justifica la doble terminología que, para objetos matemáticos idénticos, se usan en las mismas... En palabras de J.L. Doob (véase Doob, J.L. (1989), "Kolmogorov's early work on convergence theory and foundations", Ann. of Prob., Vol. 17, No. 3, p. 815.), "this influential monograph —se refiere a Kolmogorov, A.N. (1933), Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer.— transformed the character of the calculus of probabilities, moving it into mathematics from its previous state as a collection of calculations inspired by a vague nonmathematical context, a context thought to justify the use of a half-defined pseudomathematical concepts." El lector puede encontrar en «Shiryaev, A.N. (1989),

"Kolmogorov-Life and creative activities", Ann. of Prob., Vol. 17, No. 3, p. 866» un estudio detallado del trabajo de Kolmogorov y que, en particular, recoge las dos citas siguientes:

Itô: "Having read Kolmogorov's The Foundations of Probability Theory, I became convinced that probability theory could be developed in terms of measure theory as rigorously as other fields of mathematics."

Kac: (describiendo su colaboración con H. Steinhaus) "Our work began at a time when probability theory was slowly gaining acceptance as a respectable branch of pure mathematics. The turnabout came as a result of a book by the great Soviet mathematician A.N. Kolmogorov on foundations of probability theory, published in 1933."

No cabe duda de que es el trabajo de Kolmogorov en probabilidad el que ha conseguido modificar definitivamente la imagen que los matemáticos tenían del Cálculo de Probabilidades; las consecuencias de ese hecho son evidentes: un número creciente de matemáticos han ido sumando sus esfuerzos en la construcción del Cálculo de Probabilidades hasta el punto de que, hoy en día, es una de las ramas más interesantes, fértiles y elegantes de la matemática..."

Son muchas las obras sobre Medida y Probabilidad en las que se pueden encontrar los resultados que presentamos en este manual en versiones, a menudo, más generales y con todas sus demostraciones; permítasenos citar aquí, por ejemplo, las siguientes referencias, a las que este tratado debe tanto:

- Ash (1972), Real Analysis and Probability, Academic Press (y su sucesor "Ash, R.B., Doleans-Dade, C.A. (2000), Probability & Measure Theory, 2nd ed, Academic Press")
- Bauer, H. (1995), Probability Theory, de Gruyter Studies in Mathematics.
- Billingsley, P. (1995), Probability and Measure, J. Wiley & Sons.
- Cohn. D.L. (1980), Measure Theory, Birkhäuser Verlag.

- Dacunha-Castelle, D., Duflo, M. (1982), Probabilités et Statistiques, 1. Problémes à temps fixe, Masson.
- Nogales, A.G. (2004), Bioestadística Básica, @becedario.
- Nogales, A.G. (1998), Estadística Matemática, Servicio de Publicaciones de la Universidad de Extremadura.

Se asume un bagaje matemático básico de teoría de conjuntos, álgebra lineal, análisis matemático en una y varias variables y topología.

$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ denotarán los conjuntos de los números naturales, enteros, racionales, reales y complejos, respectivamente. Un conjunto se dice numerable si es finito o existe una aplicación biyectiva de $\mathbb{N}$ sobre él.

$\bar{\mathbb{R}} = [-\infty, +\infty]$ denotará la recta real ampliada, en la que asumiremos la aritmética usual: Dados $a \in \mathbb{R}$ y $b \in [-\infty, +\infty]$, se define $a + \infty = \infty + a = \infty$, $a - \infty = -\infty + a = -\infty$, $\infty + \infty = \infty$, $-\infty - \infty = -\infty$, $b \cdot \infty = \infty \cdot b = \infty$ si $b > 0$, $= -\infty$ si $b < 0$, $0 \cdot \infty = \infty \cdot 0 = 0$, $\frac{a}{\infty} = \frac{a}{-\infty} = 0$. No se definen $\infty - \infty$ ni $\frac{\infty}{\infty}$.

Cada capítulo finaliza con una colección de problemas. Algunos de ellos vienen precedidos del símbolo ($\uparrow$); se pretende indicar con ello que, de algún modo, son continuación del problema anterior. Otros, que son interesantes complementos teóricos que pueden obviarse en una primera lectura del manual y que, a menudo, presentan un mayor grado de dificultad, van precedidos del símbolo ($\bigstar$).

Badajoz, verano de 2008

Capítulo 1

Espacios Medibles

En esta sección se definen las nociones de espacio medible y de medida. Presentamos también algunos de los resultados básicos que habrán de ser útiles en el resto de este capítulo.

En lo que sigue Ω será un conjunto no vacío.

DEFINICIÓN 1.1. Sea $\mathcal{A}$ una familia de subconjuntos de Ω. Diremos que $\mathcal{A}$ es un álgebra (resp., σ-álgebra) si contiene a Ω y es cerrado por complementación (es decir, si A está en $\mathcal{A}$, su complementario A^c también está en $\mathcal{A}$) y frente a uniones finitas (resp., numerables). Si $\mathcal{A}$ es una σ-álgebra de partes de Ω, el par $(\Omega, \mathcal{A})$ se llamará un espacio medible, los subconjuntos de Ω que pertenecen a $\mathcal{A}$ se llamarán conjuntos medibles ($\mathcal{A}$-medibles) o sucesos.

***Observación* 1.1.** Longitudes, áreas, volúmenes y probabilidades son ejemplos concretos de medidas en el sentido amplio que estudiaremos en este capítulo. Una medida μ asignará a ciertos subconjuntos A de Ω un número $\mu(A)$ que llamaremos medida de A. Por razones que veremos posteriormente, no siempre es apropiado llamar suceso a cualquier subconjunto de Ω; exigiremos en cualquier caso que la clase de subconjuntos de Ω que sean declarados medibles satisfaga ciertas propiedades: concretamente, exigiremos que sea una σ-álgebra. Consideraciones probabilísticas proporcionan una buena justificación de esa afirmación. Supongamos que Ω es el conjunto de resultados posibles de un cierto experimento aleatorio; a ciertos subconjuntos de Ω se les llamará sucesos y se les asignará una probabilidad. Intuitivamente, si A es un suceso, diremos que el suceso A ha ocurrido en una realización del experimento si el resultado

ω obtenido en esa realización está en A. Sea cual sea el resultado ω del experimento, Ω ocurre; por tanto, Ω debe ser un suceso; le llamaremos suceso seguro. Si pensar en la ocurrencia de A tiene perfecto sentido por ser A un suceso, también tiene perfecto sentido pensar en la no ocurrencia de A, que equivale a la ocurrencia de A^c (complementario de A); es, por tanto, razonable exigir que A^c sea un suceso si A lo es. Si tiene sentido pensar en la ocurrencia de $A \subset \Omega$ y en la ocurrencia de $B \subset \Omega$ por ser A y B sucesos, también tiene sentido pensar en la ocurrencia de $A \cup B$ (si sabemos decidir cuándo ocurre A y cuándo ocurre B, también sabremos decidir cuando ocurre A o B); exigiremos entonces que la unión finita de sucesos sea un suceso. Menos intuitiva es la exigencia de estabilidad frente a uniones numerables; la respuesta más convincente es que, de ese modo, se obtiene una teoría matemática más rica.

***Observación* 1.2.** Los orígenes y primeros pasos de las teorías de la medida y de la probabilidad son diferentes. Fue Kolmogorov quien propuso por primera vez un punto de partida formal común para ambas teorías (que conservan aún sus propios objetivos). Debido a ello, subsiste una doble terminología para objetos matemáticos idénticos que comprobaremos en repetidas ocasiones a lo largo de este manual y que, de hecho, ya hemos comprobado: si en teoría de la medida llamamos conjuntos medibles a los elementos de una σ-álgebra, en probabilidad les llamamos sucesos.

Proposición 1.1. (a) *Si $\mathcal{A}$ es un álgebra (resp., σ-álgebra) de partes de Ω, $\mathcal{A}$ es cerrada frente a la diferencia de conjuntos e intersecciones finitas (resp., numerables) y contiene además al conjunto vacío.*

(b) *Si $\mathcal{A}$ es un álgebra de partes de Ω, será también una σ-álgebra si además es cerrada frente a uniones numerables crecientes o frente a uniones numerables disjuntas.*

Demostración. (a) Trivial.

(b) Si $(A_n)_n$ es una sucesión en $\mathcal{A}$, entonces $A := \cup_n A_n = \cup_n B_n = \cup_n C_n$ donde $B_n = \cup_{i=1}^n A_i$ y $C_n = A_n \setminus \cup_{i=1}^{n-1} A_i$. Siendo $\mathcal{A}$ un álgebra, se verifica que $B_n, C_n \in \mathcal{A}$, para cada n. Si $\mathcal{A}$ se supone estable frente a uniones numerables crecientes (resp., frente a uniones numerables disjuntas) entonces $A = \cup_n B_n \in \mathcal{A}$ (resp., $A = \cup_n C_n \in \mathcal{A}$). $\square$

Antes de ver ejemplos de σ-álgebras, probaremos una proposición que proporciona un método (poco descriptivo, pero frecuentemente utilizado) para construirlas.

Proposición 1.2. *La intersección arbitraria de σ-álgebras de partes de un conjunto Ω es una σ-álgebra en Ω. Así, si $\mathcal{C}$ es una familia de partes de Ω, existe la más pequeña σ-álgebra en Ω que contiene a $\mathcal{C}$; la llamaremos σ-álgebra engendrada por $\mathcal{C}$ y la denotaremos $\sigma(\mathcal{C})$.*

Demostración. La primera afirmación es trivial. La segunda es consecuencia de la primera, pues $\sigma(\mathcal{C})$ es la intersección de todas las σ-álgebras que contienen a $\mathcal{C}$, conjunto de σ-álgebras que es no vacío pues contiene a la σ-álgebra $\mathscr{P}(\Omega)$ de todos los subconjuntos de Ω. $\square$

***Ejemplo* 1.1.** (σ-álgebra discreta) El conjunto $\mathscr{P}(\Omega)$ de todos los subconjuntos de Ω es trivialmente una σ-álgebra en Ω que llamaremos discreta. No obstante, llamaremos espacio medible discreto un conjunto numerable provisto de la σ-álgebra de todas sus partes. En lo que sigue, cualquier conjunto numerable se supondrá provisto de la σ-álgebra discreta, a menos que, expresamente, se indique lo contrario.

***Ejemplo* 1.2.** (σ-álgebra trivial o grosera) $\{\emptyset, \Omega\}$ es la más pequeña σ-álgebra que podemos considerar en Ω. Se llamará σ-álgebra grosera o σ-álgebra trivial.

***Ejemplo* 1.3.** Si A es un subconjunto de Ω, la más pequeña σ-álgebra que contiene a A es $\{\emptyset, A, A^c, \Omega\}$. Si $A_1, \dots, A_n$ es una partición finita de Ω (es decir, subconjuntos dos a dos disjuntos que recubren Ω), entonces la más pequeña σ-álgebra que contiene a los A_i está formada por el conjunto vacío y las uniones finitas de A_i's.

***Ejemplo* 1.4.** (σ-álgebra de Borel) La σ-álgebra de Borel de $\mathbb{R}^n$, que denotaremos por $\mathcal{R}^n$ (o, simplemente, $\mathcal{R}$ si $n = 1$), se define como la σ-álgebra engendrada por los intervalos n-dimensionales de la forma $\{x \in \mathbb{R}^n : x_i < a\}$ para algún $a \in \mathbb{R}$ y algún $1 \leq i \leq n$. La σ-álgebra de Borel de $\mathbb{R}^n$ es también la σ-álgebra engendrada por cada una de las siguientes familias de subconjuntos de $\mathbb{R}^n$: la familia de los abiertos de $\mathbb{R}^n$, la de los compactos, la de los intervalos n-dimensionales cerrados y acotados (ver Problema 1.8). En lo que sigue, $\mathbb{R}^n$ se supondrá provisto de la σ-álgebra de Borel, a menos que, expresamente, se indique lo contrario.

***Ejemplo* 1.5.** $\bar{\mathbb{R}}$ denotará la recta real ampliada y $\bar{\mathcal{R}}$ la σ-álgebra de Borel de $\bar{\mathbb{R}}$ que se define como la σ-álgebra engendrada por los conjuntos de la forma $[-\infty, x]$, $x \in \bar{\mathbb{R}}$. Análogamente se define la σ-álgebra $\bar{\mathcal{R}}^n$ de Borel en $\bar{\mathbb{R}}^n$. En lo que sigue, $\bar{\mathbb{R}}^n$ se supondrá provisto de la σ-álgebra de Borel, a menos que, expresamente, se indique lo contrario.

***Ejemplo* 1.6.** En general, la σ-álgebra de Borel en un espacio topológico es la engendrada por sus abiertos; los sucesos de esta σ-álgebra se llaman conjuntos de Borel o borelianos.

***Ejemplo* 1.7.** (σ-álgebra producto) Sean $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$, espacios medibles y hagamos $\Omega = \Omega_1 \times \Omega_2$. Un rectángulo medible en Ω es un conjunto de la forma $A_1 \times A_2$ con $A_i \in \mathcal{A}_i$, $i = 1, 2$. La familia de los rectángulos medibles no es una σ-álgebra, salvo en casos triviales (ver Problema 1.6); la σ-álgebra engendrada por ellos se llamará σ-álgebra producto y se denotará $\mathcal{A}_1 \times \mathcal{A}_2$. Análogamente se define el producto de una cantidad finita de espacios medibles. Si $(\Omega_i, \mathcal{A}_i)_{i \in I}$ es una familia de espacios medibles y $\Omega = \prod_{i \in I} \Omega_i$, llamaremos σ-álgebra producto en Ω, y la denotaremos $\prod_{i \in I} \mathcal{A}_i$, la σ-álgebra engendrada por los rectángulos medibles (def.: un rectángulo medible en Ω es un producto de la forma $\prod_{i \in I} A_i$, donde $A_i \in \mathcal{A}_i$, $\forall i$ y $A_i = \Omega_i$ si $i \in I \setminus I_0$ para un cierto subconjunto finito I_0 de I). La σ-álgebra de Borel en $\mathbb{R}^n$ coincide con el producto de n copias de la de Borel en $\mathbb{R}$; la notación $\mathcal{R}^n$ utilizada para ella no puede dar lugar a confusión.

***Ejemplo* 1.8.** (σ-álgebra inducida en un subespacio) Si $(\Omega, \mathcal{A})$ es un espacio medible y B es un subconjunto no vacío de Ω entonces $\mathcal{A}_B := \{A \cap B \colon A \in \mathcal{A}\}$ es una σ-álgebra en B. Si $B \in \mathcal{A}$ los elementos de $\mathcal{A}_B$ son $\mathcal{A}$-medibles. De forma general, todo subconjunto B de una espacio medible $(\Omega, \mathcal{A})$ se supondrá provisto, a menos que se indique lo contrario, de la σ-álgebra $\mathcal{A}_B$ (ver Problema 1.9).

***Ejemplo* 1.9.** (Sub-σ-álgebra) Si $(\Omega, \mathcal{A})$ es un espacio medible y $\mathcal{B}$ es una σ-álgebra en Ω tal que $\mathcal{B} \subset \mathcal{A}$, diremos que $\mathcal{B}$ es una sub-σ-álgebra de $\mathcal{A}$. Por ejemplo, $\{\emptyset, [0,1], [0,1]^c, \mathbb{R}\}$ es una sub-σ-álgebra de $\mathcal{R}$, y ésta lo es de $\mathscr{P}(\mathbb{R})$ (ver Problema 10.26).

El teorema siguiente es un resultado de Dynkin que se utiliza para simplificar el trabajo de probar si una cierta familia de subconjuntos de Ω es una σ-álgebra.([1])

DEFINICIÓN 1.2. Sea $\mathcal{C}$ una familia de subconjuntos de Ω.

(a) Diremos que $\mathcal{C}$ es un π-sistema si es estable frente a la intersección finita.

(b) Diremos que $\mathcal{C}$ es un d-sistema (o una clase de Dynkin) si $\Omega \in \mathcal{C}$, es estable frente a la unión numerable creciente y frente a diferencias propias (esto último quiere decir que se verifica la implicación $[A, B \in \mathcal{C}, A \subset B] \Longrightarrow [B \setminus A \in \mathcal{C}]$).

Igual que ocurría con σ-álgebras, la intersección de una colección arbitraria de d-sistemas es un d-sistema y la familia de todos los subconjuntos de Ω es un d-sistema;

[1] Con ese mismo objetivo se suele usar en la literatura el llamado teorema de la clase monótona; véase, por ejemplo, Ash (1972). No obstante, suele ser más sencillo de manejar el resultado de Dynkin.

por tanto, dada una familia $\mathcal{C}$ de partes de Ω, la intersección de todos los d-sistemas que contienen a $\mathcal{C}$ es un d-sistema y es el más pequeño d-sistema que contiene a $\mathcal{C}$; se llama d-sistema engendrado por $\mathcal{C}$.

TEOREMA 1.1. (Teorema de Dynkin) *Si $\mathcal{C}$ es un π-sistema en Ω y $\mathcal{A}$ un d-sistema que contiene a $\mathcal{C}$, entonces $\mathcal{A}$ contiene a la σ-álgebra engendrada por $\mathcal{C}$. En particular, el d-sistema engendrado por un π-sistema coincide con la σ-álgebra engendrada por el π-sistema.*

Demostración. Sea $\mathcal{D}$ el d-sistema engendrado por $\mathcal{C}$. Puesto que las σ-álgebras son d-sistemas, se verifica que $\mathcal{D} \subset \sigma(\mathcal{C})$. Para probar la otra inclusión, bastará probar que $\mathcal{D}$ es una σ-álgebra. Comenzaremos probando que $\mathcal{D}$ es estable por intersección finita. Consideremos la familia $\mathcal{D}_1 := \{A \in \mathcal{D}\colon A \cap C \in \mathcal{D}$ para cada $C \in \mathcal{C}\}$. Puesto que $\mathcal{C}$ es estable por intersección finita, $\mathcal{C} \subset \mathcal{D}_1$; además $\mathcal{D}_1$ es un d-sistema ($\Omega \in \mathcal{D}_1$ pues $\mathcal{C} \subset \mathcal{D}$, y las igualdades $(A \setminus B) \cap C = (A \cap C) \setminus (B \cap C)$ y $(\cup_n A_n) \cap C = \cup_n (A_n \cap C)$ implican que $\mathcal{D}_1$ es cerrada frente a la formación de diferencias propias y frente a uniones numerables crecientes). Entonces $\mathcal{D} \subset \mathcal{D}_1$. Pero por definición de $\mathcal{D}_1$ se tiene que $\mathcal{D} \supset \mathcal{D}_1$; luego $\mathcal{D} = \mathcal{D}_1$.

Hagamos ahora $\mathcal{D}_2 := \{B \in \mathcal{D}\colon A \cap B \in \mathcal{D}$ para cada $A \in \mathcal{D}\}$. Puesto que $\mathcal{D}_1 = \mathcal{D}$, se tiene que $\mathcal{C} \subset \mathcal{D}_2$. Argumentos análogos a los precedentes prueban que $\mathcal{D}_2$ es un d-sistema y que $\mathcal{D}_2 = \mathcal{D}$. Por tanto, $\mathcal{D}$ es estable por intersección finita.

De la definición de d-sistema se deduce que $\mathcal{D}$ es estable por paso al complementario (si $A \in \mathcal{D}$ entonces $A^c = \Omega \setminus A \in \mathcal{D}$); puesto que hemos visto que es estable por intersección finita, $\mathcal{D}$ es un álgebra. Pero un álgebra estable frente a uniones crecientes es una σ-álgebra. $\square$

PROBLEMAS DEL CAPÍTULO 1

Problema 1.1. Si A y B son sucesos disjuntos, determinar $A \setminus B$, $A \triangle B$ y $A^c \cap B^c$.

Problema 1.2. a) Calcular $\cap_n \left]-\frac{1}{n}, \frac{1}{n}\right[$ y $\cap_n \left[\frac{1}{n}, 1 - \frac{1}{n}\right]$.

b) Para cada $n \in \mathbb{N}$, sea $A_n = \left[-\frac{1}{n}, 1\right]$. Determinar $\limsup_n A_n := \cap_n \cup_{k \geq n} A_k$ y $\liminf_n A_n := \cup_n \cap_{k \geq n} A_k$.

c) Repetir el apartado anterior para $A_n = \left[-\frac{1}{n}, 1\right]$ si n es par, $= \left[-1, \frac{1}{n}\right]$ si n es impar.

Problema 1.3. Sea $\Omega = \{1, 2, 3, 4\}$. Describir explícitamente los conjuntos $\mathscr{P}(\Omega)$ y $A = \{(i, j) \in \Omega^2 : i + j \leq 4\}$.

Problema 1.4. Si $(A_n)_n$ es una sucesión de subconjuntos de Ω y $A = \cup_n A_n$, entonces $A = \cup_n B_n$, donde $B_n = \cup_{k=1}^n A_k$, y $B_1 \subset B_2 \subset \dots$. También $A = \cup_n C_n$, donde $C_n = A_n \setminus \cup_{k=1}^{n-1} A_k$, y los C_n son dos a dos disjuntos.

Problema 1.5. Si $(A_n)_n$ es una sucesión de subconjuntos de Ω dos a dos disjuntos y $A = \cup_n A_n$, entonces $I_A = \sum_n I_{A_n}$.

Problema 1.6. En cada una de las situaciones siguientes decidir si la familia $\mathcal{A}$ de subconjuntos de Ω es una σ-álgebra, un álgebra, un π-sistema o un d-sistema.

1. La familia $\mathcal{A}$ de los subconjuntos finitos de un conjunto Ω.
2. $\Omega = \mathbb{R}$ y $\mathcal{A}$ es la familia de sus subconjuntos numerables.
3. $\Omega = \{1, 2, 3, 4\}$ y $\mathcal{A} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}$.
4. La familia $\mathcal{A}$ de los intervalos abiertos de $\Omega = \mathbb{R}$.
5. La familia de los rectángulos medibles en un espacio medible producto.

Problema 1.7. ¿Cuál es la más pequeña σ-álgebra en Ω que contiene a ciertos subconjuntos $A_1, \dots, A_n$ de Ω?

Problema 1.8. a) La σ-álgebra de Borel de $\mathbb{R}$ es también la σ-álgebra engendrada por cada una de las siguientes familias de intervalos: la familia de los intervalos de la forma $]-\infty, x]$, $x \in \mathbb{R}$; la de los intervalos de la forma $]x, \infty[$, $x \in \mathbb{R}$; la de los intervalos de la forma $[x, \infty[$, $x \in \mathbb{R}$; la de los intervalos de la forma $]-\infty, r]$, $r \in \mathbb{Q}$.

b) La σ-álgebra de Borel de $\mathbb{R}^n$ es también la σ-álgebra engendrada por cada una de las siguientes familias de subconjuntos de $\mathbb{R}^n$: la familia de los abiertos de $\mathbb{R}^n$, la de los compactos, la de los intervalos n-dimensionales cerrados y acotados, la de los semiespacios de la forma $\{x \in \mathbb{R}^n : x_i < a\}$ $(1 \leq i \leq n,\ a \in \mathbb{Q})$.

c) Probar que la σ-álgebra de Borel en $\mathbb{R}^n$ coincide con el producto de n copias de la σ-álgebra de Borel en $\mathbb{R}$.

Problema 1.9. Sean $(\Omega, \mathcal{A})$ un espacio medible y B un subconjunto no vacío de Ω. Probar que $\mathcal{A}_B := \{A \cap B \colon A \in \mathcal{A}\}$ es una σ-álgebra de partes de B y que $\mathcal{A}_B \subset \mathcal{A}$, si $B \in \mathcal{A}$. Probar que, si $\mathcal{A} = \sigma(\mathcal{C})$, entonces $\mathcal{A}_B = \sigma_B(\mathcal{C}_B)$, donde $\mathcal{C}_B = \{C \cap B \colon C \in \mathcal{C}\}$ y $\sigma_B(\mathcal{C}_B)$ es la más pequeña σ-álgebra de partes de B que contiene a $\mathcal{C}_B$ (ver Ejemplo 1.8).

Problema 1.10. Sean $\Omega = \{1, 2, 3\}$ y $\mathcal{A} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}$. Notar que $\mathcal{A}$ es una σ-álgebra en Ω. Describir explícitamente el espacio medible producto $(\Omega^2, \mathcal{A}^2)$.

Problema 1.11. a) Sea Ω un conjunto infinito no numerable y $\mathcal{C} = \{A \in \mathcal{P}(\Omega) : A \text{ o } A^c \text{ es numerable}\}$. Demostrar que $\mathcal{C}$ es una σ-álgebra.

b) Describir la σ-álgebra engendrada por los conjuntos unitarios en $\mathbb{R}$.

Problema 1.12. Probar que una familia $\mathcal{A}$ de partes de Ω es una σ-álgebra si y sólo si es un d-sistema y un π-sistema.

Capítulo 2

Funciones Medibles. Variables Aleatorias

La siguiente definición introduce los morfismos en la categoría de los espacios medibles: las funciones medibles (en lenguaje de teoría de la medida) o variables aleatorias (en lenguaje probabilístico).

DEFINICIÓN 2.1. (Variable aleatoria) Sean $(\Omega, \mathcal{A})$ y $(\Omega', \mathcal{A}')$ espacios medibles y $X : \Omega \longrightarrow \Omega'$ una aplicación. Diremos que X es una variable aleatoria (v.a.), una función medible (o una función $(\mathcal{A}, \mathcal{A}')$-medible, si se desea más precisión) si

$$X^{-1}(A') \in \mathcal{A}, \quad \forall A' \in \mathcal{A}'.$$

Si $(\Omega', \mathcal{A}') = (\mathbb{R}^n, \mathcal{R}^n)$ (o $(\bar{\mathbb{R}}^n, \bar{\mathcal{R}}^n)$), diremos que X es una v.a. n-dimensional; si $n = 1$, diremos que X es una v.a. real (v.a.r.) Cuando no hay duda de qué σ-álgebra $\mathcal{A}$ estamos considerando en Ω, llamaremos también funciones Borel-medibles a las v.a. n-dimensionales.

***Observación* 2.1.** Supondremos en lo que sigue que la notación $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ lleva implícito que X es $(\mathcal{A}, \mathcal{A}')$-medible.

Proposición 2.1. (a) *Si $\mathcal{C}$ es una familia de partes de Ω' que engendra la σ-álgebra $\mathcal{A}'$, para comprobar que X es una v.a., es suficiente probar que $X^{-1}(C) \in \mathcal{A}$, $\forall C \in \mathcal{C}$; ello prueba que una función real X en $(\Omega, \mathcal{A})$ es Borel-medible si, y sólo si, $\{\omega \in \Omega \colon X(\omega) \leq x\} \in \mathcal{A}$, $\forall x \in \mathbb{R}$.*

(b) *Sean Ω un conjunto, $(\Omega', \mathcal{A}')$ un espacio medible y $X : \Omega \longrightarrow \Omega'$ una aplicación. $X^{-1}(\mathcal{A}')$ es una σ-álgebra en Ω que llamaremos σ-álgebra inducida por X, pues es la más pequeña σ-álgebra en Ω que hace a X medible. En las hipótesis de la definición anterior, para que X sea $(\mathcal{A}, \mathcal{A}')$-medible es necesario y suficiente que $X^{-1}(\mathcal{A}') \subset \mathcal{A}$.*

Demostración. (a) Sea $\mathcal{B} := \{A' \in \mathcal{A}' \colon X^{-1}A' \in \mathcal{A}\}$. Por hipótesis $\mathcal{C} \subset \mathcal{B}$. Se trata de probar que $\mathcal{A}' \subset \mathcal{B}$. Puesto que $\mathcal{A}' = \sigma(\mathcal{C})$ será suficiente probar que $\mathcal{B}$ es una σ-álgebra. Pero $\Omega = X^{-1}\Omega' \in \mathcal{A}$; si $A' \in \mathcal{B}$ entonces $A'^c \in \mathcal{B}$ pues $X^{-1}(A'^c) = X^{-1}(A')^c \in \mathcal{A}$; y si $(A'_n)_n$ es una sucesión en $\mathcal{B}$, entonces $\cup_n A'_n \in \mathcal{B}$ pues $X^{-1}(\cup_n A'_n) = \cup_n X^{-1}(A'_n) \in \mathcal{B}$.

(b) La conmutatividad de la imagen inversa con el paso al complementario y la unión numerable utilizadas en la parte (a) prueban ahora que $X^{-1}(\mathcal{A}')$ es una σ-álgebra que obviamente hace medible a X; por definición de medibilidad, cualquier sub-σ-álgebra de $\mathcal{A}$ que haga X medible debe contener a $X^{-1}(\mathcal{A}')$ que, por tanto, es la más pequeña σ-álgebra que hace X medible. $\square$

***Ejemplo* 2.1.** Toda función constante $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ es $(\mathcal{A}, \mathcal{A}')$-medible.

***Ejemplo* 2.2.** (Indicador de un conjunto) Si $A \subset \Omega$, llamaremos indicador de A a la aplicación

$$I_A : \omega \in \Omega \longrightarrow I_A(\omega) = \begin{cases} = 1 & \text{si } \omega \in A \\ = 0 & \text{si } \omega \notin A \end{cases}.$$

Es evidente que I_A es una v.a. si, y sólo si, $A \in \mathcal{A}$. Diremos que A es el soporte de I_A.

***Ejemplo* 2.3.** (Función simple) Una aplicación $X : \Omega \longrightarrow \mathbb{R}$ se dice simple si toma un número finito de valores. Sean $x_1, \dots, x_n \in \mathbb{R}$ los valores que toma X y $A_i = \{X = x_i\}$ ($\{X = x_i\}$ es una forma breve de denotar el conjunto $X^{-1}(x_i) = \{\omega \in \Omega \colon X(\omega) = x_i\}$), es claro que $X = \sum_i x_i I_{A_i}$ y que X es medible si, y sólo si, los A_i son medibles.

La siguiente proposición proporciona otros ejemplos de v.a.

Proposición 2.2. (a) *La composición de funciones medibles es medible.*

(b) *Las funciones crecientes de $\mathbb{R}$ en $\mathbb{R}$ son Borel-medibles.*

(c) *Las aplicaciones continuas entre espacios topológicos son medibles si aquellos se suponen provistos de sus correspondientes σ-álgebras de Borel.*

(d) *Las proyecciones en un espacio medible producto (finito o no) son v.a. La σ-álgebra producto es la más pequeña que hace medibles las proyecciones. Se verifica también que una aplicación a valores en un producto es medible si, y sólo si, se obtienen funciones medibles al componer con las proyecciones. En particular, una aplicación $X = (X_1, \dots, X_n)$ a valores en $\mathbb{R}^n$ es una v.a. si, y sólo si, las componentes X_i son v.a.r.*

Demostración. (a) En efecto, si $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ e $Y : (\Omega', \mathcal{A}') \to (\Omega'', \mathcal{A}'')$, entonces, para cada $A'' \in \mathcal{A}''$, $(Y \circ X)^{-1}(A'') = X^{-1}(Y^{-1}(\mathcal{A}'')) \in X^{-1}(\mathcal{A}') \subset \mathcal{A}$.

(b) La Proposición 2.1 reduce el problema a probar que la contraimagen de un intervalo de la forma $]-\infty, y]$ es un boreliano, lo cual es trivial (es, de hecho, un intervalo).

(c) Puesto que la σ-álgebra de Borel de un espacio topológico está generada por sus conjuntos abiertos, basta probar -de nuevo la Proposición 2.1- que la contraimagen de un abierto por una aplicación continua es un boreliano; pero esa contraimagen es, incluso, un abierto, por definición de continuidad.

(d) Sean I un conjunto de índices y π_j, $j \in I$, la proyección j-ésima del espacio medible producto $\left(\prod_{i\in I} \Omega_i, \prod_{i\in I} \mathcal{A}_i\right)$ sobre el factor j-ésimo $(\Omega_j, \mathcal{A}_j)$. Si $A_j \in \mathcal{A}_j$, $X_j^{-1}(A_j)$ coincide con el rectángulo medible $\prod_{i\in I} A_i \in \prod_{i\in I} \mathcal{A}_i$, donde $A_i = \Omega_i$ si $i \neq j$.

Por otra parte, si $\mathcal{B}$ es una σ-álgebra en $\prod_i \Omega_i$ que hace medibles a las proyecciones, entonces, para cada $i \in I$ y cada $A_i \in \mathcal{A}_i$, se verifica que $\pi_i^{-1}(A_i) \in \mathcal{B}$ y, por tanto, para cada subconjunto finito I_0 de I y cada elección de sucesos $A_i \in \mathcal{A}_i$, $i \in I_0$, se verifica que $\cap_{i\in I_0} \pi_i^{-1}(A_i) \in \mathcal{B}$, lo que prueba que $\prod_i \mathcal{A}_i \subset \mathcal{B}$, pues $\prod_i \mathcal{A}_i$ está engendrada por los conjuntos de la forma $\cap_{i\in I_0} \pi_i^{-1}(A_i)$.

Finalmente, la composición de una función medible a valores en un espacio producto con una de las proyecciones es también una función medible, por el apartado (a); recíprocamente, si $(\Omega, \mathcal{A})$ y $(\Omega_i, \mathcal{A}_i)_{i\in I}$ son espacios medibles y $X : \Omega \longrightarrow \prod_i \Omega_i$ es una aplicación tal que $\pi_i \circ X$ es $(\mathcal{A}, \mathcal{A}_i)$-medible, para cada $i \in I$, entonces, dado un rectángulo medible $\prod_i A_i \in \prod_i \mathcal{A}_i$, existe I_0 subconjunto finito de I tal que $A_i = \Omega_i$ si $i \in I \setminus I_0$, y

$$X^{-1}(\textstyle\prod_i A_i) = X^{-1}\left(\cap_{i\in I_0}\pi_i^{-1}(A_i)\right) = \cap_{i\in I_0}(\pi_i \circ X)^{-1}(A_i) \in \mathcal{A}$$

□

Probamos a continuación algunas propiedades de estabilidad de la clase de las funciones medibles.

Proposición 2.3. (a) *Sea $(X_n)_n$ una sucesión de v.a. $\bar{\mathbb{R}}$-valoradas convergente puntualmente a una función $X : \Omega \to \bar{\mathbb{R}}$. Entonces X es también una v.a.*

(b) *El supremo y el ínfimo de una sucesión de funciones $\bar{\mathbb{R}}$-valoradas Borel-medibles son también Borel-medibles.*

(c) *Si X es una v.a.r., se definen la parte positiva X^+ de X y la parte negativa X^- de X mediante: $X^+ = \text{máx}(X, 0)$ y $X^- = \text{máx}(-X, 0)$. Tanto X^+ como X^- son v.a.r. y se verifica que $X = X^+ - X^-$ y $|X| = X^+ + X^-$.*

Demostración. (a) Nótese que, dado $x \in \mathbb{R}$,

$$\{\omega\colon X(\omega) > x\} = \cup_{n=1}^\infty \cap_{k=n}^\infty \{\omega\colon X_k(\omega) > x + 1/n\} \in \mathcal{A}.$$

(b) Sea $(X_n)_n$ una sucesión de funciones Borel-medibles; dado $x \in \mathbb{R}$ se verifica que $\{\omega\colon \sup_n X_n(\omega) \le x\} = \cap_n\{\omega\colon X_n(\omega) \le x\} \in \mathcal{A}$ y $\{\omega\colon \inf_n X_n(\omega) \ge x\} = \cap_n\{\omega\colon X_n(\omega) \ge x\} \in \mathcal{A}$.

(c) La primera afirmación es consecuencia de (b). La segunda es trivial. □

Un método a menudo útil a la hora de probar resultados sobre integración consiste en probar en primer lugar el resultado para indicadores para extenderlo después por linealidad a funciones simples medibles y, en fin, por continuidad a v.a.r. Ese proceso de extensión es posible gracias al teorema siguiente, que utilizaremos también para completar las propiedades de estabilidad de la clase de funciones reales medibles.

TEOREMA 2.1. (a) *Sea $X : (\Omega, \mathcal{A}) \longrightarrow [0, +\infty]$ una v.a. Existe entonces una sucesión creciente $(X_n)_n$ de funciones reales simples medibles no negativas que converge puntualmente a X.*

(b) *Sea $X : (\Omega, \mathcal{A}) \longrightarrow \bar{\mathbb{R}}$ una v.a. Existe entonces una sucesión $(X_n)_n$ de funciones reales simples medibles $\mathbb{R}$-valoradas que converge puntualmente a X y tales que $|X_n| \leq |X|$, para cada $n \in \mathbb{N}$.*

Demostración. (a) Se define

$$X_n(\omega) = \begin{cases} = \frac{k-1}{2^n} & \text{si } \frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n}, \quad k = 1, 2, \dots, n2^n. \\ = n & \text{si } X(\omega) \geq n \end{cases}$$

Es claro que X_n es una función simple medible no negativa. Veamos que, para cada ω, la sucesión $(X_n(\omega))_n$ es creciente. Si existe $k \in \{1, 2, \dots, n2^n\}$ tal que $\frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n}$, entonces $\frac{2k-2}{2^{n+1}} \leq X(\omega) < \frac{2k}{2^{n+1}}$ y $X_n(\omega) := \frac{k-1}{2^n} \leq X_{n+1}(\omega) \in \{\frac{2k-2}{2^{n+1}}, \frac{2k-1}{2^{n+1}}\}$. Si $X(\omega) \geq n$ entonces $X_n(\omega) := n \leq X_{n+1}(\omega)$ pues $X_{n+1}(\omega) = n + 1$ si $X(\omega) \geq n + 1$ o $X_{n+1}(\omega) \in \{\frac{(2n+k)2^n}{2^{n+1}} : k = 0, 1, \dots, 2^{n+1} - 1\}$ si $n = \frac{2n2^n}{2^{n+1}} \leq X(\omega) < n + 1 = \frac{2(n+1)2^n}{2^{n+1}}$. Por último, X_n converge puntualmente a X pues, dado $\omega \in \Omega$ tal que $X(\omega) < \infty$, existe $n' \in \mathbb{N}$ tal que $X(\omega) < n$ si $n > n'$. Por definición de X_n, se verifica que, si $n > n'$, entonces $X(\omega)$ dista de $X_n(\omega)$ menos que 2^{-n} y, por tanto, $X(\omega) = \lim_n X_n(\omega)$. El caso $X(\omega) = \infty$ es trivial.

(b) Sean (S_n) e (S'_n) sucesiones crecientes de funciones simples medibles no negativas puntualmente convergentes a X^+ y X^-, resp. Basta tomar $X_n = S_n - S'_n$, pues S_n es una función simple medible que converge puntualmente a X y tal que $|X_n| \leq S_n + S'_n \leq X^+ + X^- = |X|$. $\square$

TEOREMA 2.2. *Si X_1, X_2 son funciones Borel-medibles de Ω en $\bar{\mathbb{R}}$, también son Borel-medibles $X_1 + X_2$, $X_1 - X_2$, $X_1 \cdot X_2$ y X_1/X_2 (suponiendo que están bien definidas, es decir, $X_1 + X_2$ no es nunca de la forma $+\infty - \infty$ y X_1/X_2 no es nunca de la forma ∞/∞ o $a/0$).*

Demostración. Sean s_n, t_n sucesiones de funciones simples medibles $\mathbb{R}$-valoradas puntualmente convergentes a X_1 y X_2. Entonces s_n+t_n converge puntualmente a X_1+X_2, $s_n t_n I_{\{X_1\neq 0\}} I_{\{X_2\neq 0\}}$ converge puntualmente a X_1X_2 y

$$\frac{s_n}{t_n+\frac{1}{n}I_{\{t_n=0\}}}$$

converge a X_1/X_2. Puesto que todas esas son sucesiones de funciones simples medibles, el resultado se sigue de la Proposición 2.3. □

PROBLEMAS DEL CAPÍTULO 2

Problema 2.1. Sean $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$, espacios medibles, $A_1 \in \mathcal{A}_1$ y $X : A_1 \longrightarrow \Omega_2$ y $X' : A_1^c \longrightarrow \Omega_2$ funciones medibles para las σ-álgebras que $\mathcal{A}_1$ induce en A_1 y A_1^c. Entonces, la aplicación $S : \Omega_1 \longrightarrow \Omega_2$ que coincide con X en A_1 y con X' en A_1^c es medible.[1]

Problema 2.2. a) Sean $X, Y : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{R})$ dos v.a.r. Probar que $\{X = Y\} \in \mathcal{A}$, $\{X \neq Y\} \in \mathcal{A}$, $\{X < Y\} \in \mathcal{A}$ y $\{X > Y\} \in \mathcal{A}$.

b) En cada uno de los apartados siguientes, decidir si el conjunto $A \subset \mathbb{R}^n$ es un boreliano: b.1) $A = \{(x, y) \in \mathbb{R}^2 \colon y = x^2\}$. b.2) $A = \{x \in \mathbb{R}^n \colon \sum_{i=1}^n x_i^2 < 1\}$. b.3) $A = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$. b.4) $A = \cup_{k=1}^\infty]k - 1/k, k + 1/k[\subset \mathbb{R}$.

Problema 2.3. Sean Ω' un conjunto, $(\Omega, \mathcal{A})$ un espacio medible y $X : \Omega \longrightarrow \Omega'$ una aplicación. Probar que la familia $\{B \subset \Omega' \colon X^{-1}(B) \in \mathcal{A}\}$ es la mayor σ-álgebra en Ω' que hace medible a X.

Problema 2.4. Sea $(X_n)_n$ una sucesión de funciones $\bar{\mathbb{R}}$-valoradas y Borel-medibles. Probar que $\{\omega \in \Omega \colon \text{lím}_n X_n(\omega) \text{ existe y es finito}\}$ es un conjuntos medible.

Problema 2.5. Calcular $\sup_n X_n$, siendo

$$X_n : x \in [0, 1] \longrightarrow X_n(x) = \begin{cases} = 2^n x & \text{si } 0 \le x \le 2^{-n} \\ = 1 & \text{si } 2^{-n} < x \le 1 \end{cases}$$

Problema 2.6. Sean $X : \Omega \longrightarrow \Omega'$ una aplicación, $(A_i)_{i \in I}$ una familia de subconjuntos de Ω y $(A'_i)_{i \in I}$ una familia de subconjuntos de Ω'. Probar las siguientes proposiciones:

(a) $X^{-1}(\cup_i A'_i) = \cup_i X^{-1}(A'_i)$.

(b) $X^{-1}(\cap_i A'_i) = \cap_i X^{-1}(A'_i)$.

(c) $X(\cup_i A_i) = \cup_i X(A_i)$.

(d) $X(\cap_i A_i) \subset \cap_i X(A_i)$. La igualdad no siempre se da en este caso.

(e) Si $A' \subset \Omega'$ entonces $X^{-1}(A'^c) = X^{-1}(A')^c$.

Problema 2.7. Consideremos las aplicaciones $X : (x, y) \in \mathbb{R}^2 \longrightarrow X(x, y) := x + y \in \mathbb{R}$ e $Y : (x, y) \in \mathbb{R}^2 \longrightarrow X(x, y) := x^2 + y^2 \in \mathbb{R}$. Determinar $X^{-1}([0, 1])$ e $Y^{-1}(]0, 1])$.

[1] A_1 y A_1^c constituyen una partición medible de Ω, es decir, son sucesos disjuntos que recubren Ω. El resultado se generaliza sin dificultad al caso de una partición medible numerable de Ω

Problema 2.8. Para cada $n \in \mathbb{N}$, consideremos la aplicación $X_n = I_{[\frac{1}{n},1-\frac{1}{n}]}$. Determinar $\lim_n X_n$ (si existe).

Problema 2.9. a) Probar que si $(A_n)_n$ es una sucesión creciente de conjuntos entonces I_{A_n} es una sucesión creciente de funciones y $\lim_n A_n = I_{\cup_n A_n}$.

b) Probar que si $(A_n)_n$ es una sucesión decreciente de conjuntos entonces $\lim_n A_n = I_{\cap_n A_n}$.

Problema 2.10. a) Sean $\Omega = \{1, 2, 3, 4\}$ y $\mathcal{A}$ la σ-álgebra engendrada por los conjuntos $\{1, 2\}$, $\{3\}$ y $\{4\}$. Probar que, si $X : \Omega \to \mathbb{R}$ es $\mathcal{A}$-medible, entonces $X(1) = X(2)$.

b) Sea $A_1, A_2 \dots$ una partición numerable de un conjunto Ω y $\mathcal{A}$ la σ-álgebra engendrada por los A_n. Probar que $\mathcal{A} = \{\cup_{n\in I} A_n \colon I \subset \{1, 2, \dots\}\}$. ¿Cómo son las funciones reales $\mathcal{A}$-medibles? Ejemplo concreto: Sea $\Omega = [0, 3]$. Describir explícitamente la σ-álgebra engendrada por los intervalos $[0, 1]$, $]1, 2[$ y $[2, 3]$. ¿Cómo son las funciones reales medibles para esa σ-álgebra?

c) Un átomo en un espacio medible $(\Omega, \mathcal{A})$ es un suceso $A \in \mathcal{A}$ tal que, si $B \in \mathcal{A}$ y $B \subset A$, entonces $B = \emptyset$ o $B = A$ (ver una definición de átomo relativo a una probabilidad en el Problema 12.11). Probar que, si A es un átomo de $\mathcal{A}$ y $X : \Omega \to \mathbb{R}$ es $\mathcal{A}$-medible, entonces X es constante en A.

Problema 2.11. Sea $f : \Omega \to \Omega$ una aplicación biyectiva. Demostrar que la clase $S = \{A \in \mathcal{P}(\Omega) : f(A) = A\}$ de los subconjuntos de Ω invariantes por f es una σ-álgebra.

Problema 2.12. (★) Sea $f : \Omega \to \Omega'$, y sea $\mathcal{C}$ una clase de subconjuntos de Ω'. Probar que:

$$\sigma(f^{-1}(\mathcal{C})) = f^{-1}(\sigma(\mathcal{C}))$$

donde $f^{-1}(\mathcal{C}) = \{f^{-1}(A) : A \in \mathcal{C}\}$.

Problema 2.13. (★) ($\bar{\mathbb{R}}^n$ es Borel equivalente a $\bar{\mathbb{R}}$) Definición: Dos espacios medibles se dicen Borel-equivalentes si existe una biyección bimedible entre ellos.

1. Si $(\Omega, \mathcal{A})$ y $(\Omega', \mathcal{A}')$ son Borel-equivalentes y $n \in \mathbb{N}$, entonces $(\Omega^n, \mathcal{A}^n)$ y $(\Omega'^n, \mathcal{A}'^n)$ son Borel-equivalentes.

2. Dos espacios topológicos homeomorfos son Borel-equivalentes si se suponen provistos de sus σ-álgebras de Borel.

3. La aplicación

$$f : x \in \bar{\mathbb{R}} \mapsto f(x) = \begin{cases} \frac{1}{2}\left(\frac{x}{1+|x|}\right) + \frac{1}{2} & \text{si } x \in \mathbb{R} \\ 0 & \text{si } x = -\infty \\ 1 & \text{si } x = +\infty \end{cases}$$

es un homeomorfismo entre $\bar{\mathbb{R}}$ y $[0, 1]$.

4. Si $n \in \mathbb{N}$, $\{0,1\}^{\mathbb{N}}$ es homeomorfo a $\left(\{0,1\}^{\mathbb{N}}\right)^n$ cuando $\{0,1\}$ se supone provisto de la topología discreta y los productos de la correspondiente topología producto.

5. Sea $A = \cup_{n\geq 1}\{k2^{-n}/1 \leq k \leq 2^n - 1\}$. Probar que los puntos de A, y sólo ellos, admiten dos representaciones en base 2 de la forma
$$0'\underline{b_1}\,\underline{b_2}\,\underline{b_3}\,\dots$$
con $b = (b_i) \in \{0,1\}^{\mathbb{N}}$.

6. Hagamos
$$S = \cup_{n\geq 1}\left[\{b \in \{0,1\}^{\mathbb{N}}/b_i = 0, \forall i \geq n\}\right.$$
$$\left.\cup\{b \in \{0,1\}^{\mathbb{N}}/b_i = 1, \forall i \geq n\}\right] \setminus \{(0,0,\dots),(1,1,\dots)\}.$$
A y S son numerables. Sean $a^1, a^2, \dots$ y $s^1, s^2, \dots$ enumeraciones de A y S, resp. Se define una aplicación $T : \{0,1\}^{\mathbb{N}} \mapsto [0,1]$ mediante $T(b) = \sum_{i=1}^{\infty} b_i/2^i$ si $b \notin S$, $T(s^i) = a^i, i \in \mathbb{N}$. Probar que T es biyectiva y que T y T^{-1} son medibles.([2])

7. Probar ya que $\bar{\mathbb{R}}^n$ es Borel-equivalente a $\bar{\mathbb{R}}$.

[2] Esta aplicación biyectiva y bimedible, además de identificar como espacios medibles el intervalo $[0,1]$ y el espacio medible producto $\{0,1\}^{\mathbb{N}}$ nos permitirá -ver Problema 8.31- obtener una descripción alternativa de la medida de Lebesgue en $[0,1]$. es decir, de la distribución uniforme continua en ese intervalo.

Capítulo 3

Espacios de Medida. Espacios de Probabilidad

DEFINICIÓN 3.1. (Medida) Sea $(\Omega, \mathcal{A})$ un espacio medible. Una medida en $(\Omega, \mathcal{A})$ es una función de conjuntos $\mu : \mathcal{A} \longrightarrow [0, +\infty]$ numerablemente aditiva (es decir, si (A_n) es una sucesión disjunta en $\mathcal{A}$, entonces $\mu(\cup_n A_n) = \sum_n \mu(A_n)$) y nula en el vacío (es decir, $\mu(\emptyset) = 0$). La terna $(\Omega, \mathcal{A}, \mu)$ se llamará espacio de medida. Una medida μ se dice finita si $\mu(\Omega) < +\infty$; se dice σ-finita si Ω puede ser recubierto por una colección numerable de conjuntos medibles de medida finita. Una medida μ se dice una probabilidad si $\mu(\Omega) = 1$; la terna $(\Omega, \mathcal{A}, \mu)$ se llamará entonces espacio de probabilidad.

***Observación* 3.1.** El concepto de probabilidad nos viene bien para justificar la definición de medida. Parece razonable suponer que la probabilidad del conjunto imposible (o la medida del conjunto vacío) sea cero. Parece natural también exigir que la probabilidad de la unión de dos sucesos disjuntos (resp., la medida de la unión de dos conjuntos medibles disjuntos) sea la suma de las probabilidades (resp., medidas) de ambos. Menos natural puede parecer el requisito de que eso siga siendo cierto para la unión numerable disjunta, axioma que asumiremos por el hecho de que así se obtiene una teoría matemática de la medida y la integración más satisfactoria.

Otras propiedades igualmente razonables para la medida o la probabilidad son consecuencia de la definición anterior. En adelante $(\Omega, \mathcal{A}, \mu)$ será un espacio de medida.

Proposición 3.1. (a) (Monotonía) *Si A y B son sucesos tales que $A \subset B$, entonces $\mu(A) \le \mu(B)$, y, si además $\mu(A) < +\infty$, entonces $\mu(B \setminus A) = \mu(B) - \mu(A)$.*

(b) (Continuidad hacia arriba) *Si $(A_n)_n$ es una sucesión creciente en $\mathcal{A}$, entonces $\mu(\cup_n A_n) = \text{lím}_n\, \mu(A_n)$.*

(c) (Continuidad hacia abajo) *Si $(A_n)_n$ es una sucesión decreciente en $\mathcal{A}$ y alguno de los A_n tiene medida finita, entonces $\mu(\cap_n A_n) = \text{lím}_n\, \mu(A_n)$.*

(d) (Subaditividad numerable o σ-subaditividad) *Si $(A_n)_n$ es una sucesión en $\mathcal{A}$, entonces $\mu(\cup_n A_n) \le \sum_n \mu(A_n)$.*

Demostración. (a) $\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A)$. Si $\mu(A)$ es finito, se sigue de la igualdad anterior que $\mu(B \setminus A) = \mu(B) - \mu(A)$.

(b) Sea $A = \cup_n A_n$. Si $\mu(A_n) = +\infty$ para algún n, entonces $\mu(A_k) = \infty$ para cada $k \ge n$ y el resultado es obvio. Si todos los A_n tienen medida finita podemos escribir

$$A = A_1 \cup (A_2 \setminus A_1) \cup \cdots \cup (A_n \setminus A_{n-1}) \cup \ldots,$$

con lo cual

$$\mu(A) = \mu(A_1) + \mu(A_2) - \mu(A_1) + \cdots + \mu(A_n) - \mu(A_{n-1}) + \cdots = \lim_n \mu(A_n).$$

(c) Podemos suponer sin pérdida de generalidad que A_1 tiene medida finita. Si $(A_n)_n$ es una sucesión decreciente a $A := \cap_n A_n$, entonces $(A_1 \setminus A_n)_n$ es una sucesión creciente a $A_1 \setminus A$. Por (2) se verifica que $\mu(A_1 \setminus A_n)$ converge a $\mu(A_1 \setminus A)$. El resultado se sigue de (a).

(d) Puesto que $A := \cup_{n=1}^\infty A_n = \cup_{n=1}^\infty (A_n \setminus \cup_{i=1}^{n-1} A_i)$ se deduce que

$$\mu(A) = \sum_n \mu(A_n \setminus \cup_{i=1}^{n-1} A_i) \le \sum_n \mu(A_n).$$

$\square$

***Ejemplo* 3.1.** (Medida de Dirac) Si $(\Omega, \mathcal{A})$ es un espacio medible y $\omega \in \Omega$, la función de conjuntos ε_ω definida para $A \in \mathcal{A}$ por $\varepsilon_\omega(A) = 1$ si $\omega \in A$, $= 0$ si no, es una probabilidad que llamaremos probabilidad de Dirac (o distribución de probabilidad degenerada) en el punto ω.

***Ejemplo* 3.2.** (Medida cardinal) Si, para $A \in \mathcal{A}$, denotamos por $\mu(A)$ el cardinal de A ($\mu(A) = +\infty$ si A es un conjunto infinito) obtenemos una medida que llamaremos medida cardinal; esta medida es σ-finita si Ω es numerable y $\mathcal{A}$ es el conjunto de todas las partes de Ω (es decir, si $(\Omega, \mathcal{A})$ es un espacio discreto).

***Ejemplo* 3.3.** Si $\Omega = \{x_1, x_2, \dots\}$ es un conjunto numerable y $p_1, p_2, \dots$ son números no negativos, tomando como $\mathcal{A}$ la σ-álgebra de las partes de Ω, obtenemos una medida μ en $\mathcal{A}$ mediante: $\mu(A) = \sum_{i \geq 1} p_i I_A(x_i)$; esta medida es una probabilidad si $\sum_i p_i = 1$.

DEFINICIÓN 3.2. (Propiedades que ocurren casi seguro) Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida. Diremos que una propiedad de puntos de Ω ocurre casi seguro respecto a μ (o μ-c.s.) si el conjunto de los puntos de Ω que no verifican dicha propiedad está contenido en un conjunto medible de medida nula.

***Observación* 3.2.** La medida μ se dice concentrada en un suceso $A \in \mathcal{A}$ si $\mu(A^c) = 0$.

***Observación* 3.3.** Si A es un suceso tal que $\mu(A) > 0$, entonces $\mu^A : B \in \mathcal{A} \mapsto \mu^A(B) := \mu(B \cap A) \in [0, +\infty]$ es una medida concentrada en el suceso A, que llamaremos restricción al suceso A de la medida μ.

Describimos a continuación un método clásico de construcción de medidas, basado en el concepto de medida exterior, que utilizaremos para construir la medida de Lebesgue en $\mathbb{R}^n$.

DEFINICIÓN 3.3. (Medida exterior) Una medida exterior en un conjunto no vacío Ω es una función de conjuntos $\mu^*: \mathscr{P}(\Omega) \to [0, +\infty]$ nula en el vacío, monótona y numerablemente subaditiva.

DEFINICIÓN 3.4. (Medida exterior de Lebesgue en $\mathbb{R}^n$) Comenzaremos con el caso $n = 1$. Si $A \subset \mathbb{R}$, denotaremos por $\mathcal{C}_A$ el conjunto de todas las sucesiones $(]a_i, b_i[)_{i \in \mathbb{N}}$ de intervalos abiertos acotados tales que $A \subset \cup_i]a_i, b_i[$. Se define también la medida exterior de Lebesgue $\lambda^*(A)$ como el ínfimo de las sumas $\sum_i (b_i - a_i)$ cuando $(]a_i, b_i[)_i$ recorre $\mathcal{C}_A$. Se prueba que λ^* es una medida exterior que asigna a cada intervalo de la recta su longitud. En el caso general, si $A \subset \mathbb{R}^n$, denotaremos por $\mathcal{C}_A$ el conjunto de todas las sucesiones $(R_i)_{i \in \mathbb{N}}$ de intervalos abiertos n-dimensionales acotados (es decir, R_i es un conjunto de la forma $I_1 \times \cdots \times I_n$ donde los I_k son intervalos abiertos acotados de $\mathbb{R}$) tales que $A \subset \cup_i R_i$. Se define la medida exterior de Lebesgue $\lambda_n^*(A)$ como el ínfimo de las sumas $\sum_i \mathrm{vol}(R_i)$ cuando $(R_i)_i$ recorre $\mathcal{C}_A$, donde $\mathrm{vol}(R_i)$ denota el volumen de R_i, es decir, el producto de las longitudes de los intervalos I_k si $R_i = I_1 \times \cdots \times I_n$.

Proposición 3.2. *La medida exterior de Lebesgue λ_n^* en $\mathbb{R}^n$ es una medida exterior que asigna a cada intervalo n-dimensional su volumen.*

Demostración. $\lambda_n^*(\emptyset) = 0$ pues $\emptyset \subset] - \epsilon^{\frac{1}{n}}/2, \epsilon^{\frac{1}{n}}/2[^n$ y $\mathrm{vol}(] - \epsilon^{\frac{1}{n}}/2, \epsilon^{\frac{1}{n}}/2[^n) = \epsilon$, para cada $\epsilon > 0$.

λ_n^* es trivialmente monótona pues, para cada $A \subset B \subset \mathbb{R}^n$, $\{\sum_{i=1}^\infty \mathrm{vol}(R_i) : B \subset \cup_{i=1}^\infty R_i\} \subset \{\sum_{i=1}^\infty \mathrm{vol}(R_i) : A \subset \cup_{i=1}^\infty R_i\}$ y, por tanto, $\lambda_n^*(A) \le \lambda_n^*(B)$. Por otra parte, λ_n^* es numerablemente subaditiva pues, para cada sucesión $(B_k)_{k\in\mathbb{N}}$ de subconjuntos de $\mathbb{R}^n$, cada $\epsilon > 0$ y cada $k \in \mathbb{N}$, existe una sucesión $(R_{ik})_{i=1}^\infty$ de intervalos n dimensionales abiertos y acotados tal que $B \subset \cup_{i=1}^\infty R_{ik}$ y $\lambda_n^*(B_k) \ge \sum_{i=1}^\infty \mathrm{vol}(R_{ik}) - \epsilon/2^k$; por tanto, $(R_{ik})_{i,k}$ es una colección numerable de intervalos n-dimensionales abiertos y acotados y

$$\sum_k \lambda_n^*(B_k) \ge \sum_{k,i} \mathrm{vol}(R_{ik}) - \epsilon \ge \lambda_n^*(\cup_k B_k) - \epsilon.$$

Veamos ahora que la medida exterior de Lebesgue asigna a cada intervalo n-dimensional I su volumen. Consideremos en primer lugar el caso $n = 1$ e $I = [a, b]$. Puesto que, para cada $\epsilon > 0$, $I \subset]a - \frac{\epsilon}{2}, b + \frac{\epsilon}{2}[$, se tiene que $\lambda_n^*(I) \le (b - a) + \epsilon$; por tanto, $\lambda_n^*(I) \le b - a$. Por otra parte, si $(]a_i, b_i[)_i$ es una sucesión de intervalos abiertos que recubren el intervalo compacto I, existe $k \in \mathbb{N}$ tal que $]a, b[\subset \cup_{i=1}^k]a_i, b_i[$. Se deduce de ahí (ver Problema 3.14) que $b - a \le \sum_{i=1}^k (b_i - a_i) \le \sum_{i=1}^\infty (b_i - a_i)$; luego $b - a \le \lambda_n^*(I)$ y, en definitiva, $\lambda_n^*([a, b]) = b - a$. Si I es un intervalo acotado arbitrario de extremos a y b, entonces, para cada $\epsilon > 0$, $[a+\epsilon, b-\epsilon] \subset I \subset [a-\epsilon, b+\epsilon]$ y, por tanto, $(b - a) - 2\epsilon \le \lambda_n^*(I) \le (b - a) + 2\epsilon$, que prueba que $\lambda_n^*(I) = b - a$. Un intervalo no acotado de la recta tiene medida exterior de Lebesgue infinita pues contiene intervalos compactos de longitud arbitrariamente grande.

Supongamos ahora que $n > 1$ y que $I = \prod_{i=1}^n [a_i, b_i]$ es un intervalo compacto n-dimensional. Los detalles son similares al caso $n = 1$ (ver Problema 3.14). Notemos solamente que, si $I \subset \cup_{j=1}^k I_j$, donde $I_j = \prod_{i=1}^n]a_{ji}, b_{ji}[$, podemos descomponer I como unión de intervalos n-dimensionales compactos K_s cuyos interiores son disjuntos y cada uno de ellos está contenido en alguno de los intervalos I_j. Entonces $\mathrm{vol}(I) = \sum_s \mathrm{vol}(K_s) \leq \sum_j \mathrm{vol}(I_j)$, y de aquí se deduce el resultado como en el caso real. □

En lo que sigue μ^* será una medida exterior en Ω.

DEFINICIÓN 3.5. (Conjunto μ^*-medible) Un subconjunto B de Ω se dice μ^*-medible si, para cada $A \subset \Omega$,

$$\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c).$$

Denotaremos por $\mathcal{A}_{\mu^*}$ la familia de subconjuntos de Ω que son μ^*-medibles. Un conjunto Lebesgue-medible en $\mathbb{R}^n$ es un conjunto λ_n^*-medible.

TEOREMA 3.1. *$\mathcal{A}_{\mu^*}$ es una σ-álgebra de partes de Ω que contiene a los conjuntos $B \subset \Omega$ tales que $\mu^*(B) = 0$ o $\mu^*(B^c) = 0$. Además, la restricción de μ^* a $\mathcal{A}_{\mu^*}$ es una medida.*

Demostración. Siendo μ^* subaditiva, siempre ocurre que $\mu^*(A) \leq \mu^*(A\cap B)+\mu^*(A\cap B^c)$; así pues, para probar que B es μ^*-medible, basta probar que $\mu^*(A) \geq \mu^*(A \cap B) + \mu^*(A \cap B^c)$, para cada $A \subset \Omega$ tal que $\mu^*(A) < +\infty$, ya que esa desigualdad es trivial si $\mu^*(A) = +\infty$. Por otra parte, si $\mu^*(B) = 0$ o $\mu^*(B^c) = 0$, la monotonía de μ^* prueba inmediatamente que $\mu^*(A) \geq \mu^*(A \cap B) + \mu^*(A \cap B^c)$, es decir, que B es μ^*-medible.

Para probar que $\mathcal{A}_{\mu^*}$ es una σ-álgebra, probaremos que es un álgebra estable frente a la unión numerable disjunta (eso basta, de acuerdo con la Proposición 1.1). $\emptyset, \Omega \in \mathcal{A}_{\mu^*}$ pues $\mu^*(\emptyset) = 0$. Por otra parte, en la definición de que B sea μ^*-medible los papeles de B y B^c son intercambiables y, por tanto, si B es μ^*-medible, B^c también lo es.

Supongamos ahora que B_1 y B_2 son μ^*-medibles y probemos que también lo es $B_1 \cup B_2$. Siendo B_1 μ^*-medible

$$\begin{aligned}\mu^*(A \cap (B_1 \cup B_2)) &= \mu^*(A \cap (B_1 \cup B_2) \cap B_1) + \mu^*(A \cap (B_1 \cup B_2) \cap B_1^c) \\ &= \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c \cap B_2).\end{aligned}$$

Puesto que $(B_1 \cup B_2)^c = B_1^c \cap B_2^c$ y teniendo en cuenta primero que B_2 es μ^*-medible, y luego que lo es B_1, se tiene que

$$\begin{aligned}&\mu^*(A \cap (B_1 \cup B_2)) + \mu^*(A \cap (B_1 \cup B_2)^c) \\ &\qquad = \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c \cap B_2) + \mu^*(A \cap B_1^c \cap B_2^c) \\ &\qquad = \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c) = \mu^*(A).\end{aligned}$$

Queda, pues, probado que $\mathcal{A}_{\mu^*}$ es un álgebra.

Sea ahora (B_n) una sucesión disjunta de conjuntos μ^*-medibles. Probemos por inducción que, para cada A y cada $n \in \mathbb{N}$,

$$\mu^*(A) = \sum_{i=1}^{n} \mu^*(A \cap B_i) + \mu^*(A \cap (\cap_{i=1}^{n} B_i^c)). \tag{3.1}$$

El caso $n = 1$ se obtiene de la medibilidad de B_1. Supuesta cierta para n esa expresión, teniendo en cuenta que B_{n+1} es medible y que los B_i son disjuntos dos a dos, se verifica que

$$\begin{aligned}\mu^*(A \cap (\cap_{i=1}^{n} B_i^c)) &= \mu^*(A \cap (\cap_{i=1}^{n} B_i^c) \cap B_{n+1}) + \mu^*(A \cap (\cap_{i=1}^{n} B_i^c) \cap B_{n+1}^c) \\ &= \mu^*(A \cap B_{n+1}) + \mu^*(A \cap (\cap_{i=1}^{n+1} B_i^c)).\end{aligned}$$

Eso prueba la igualdad (3.1). Se deduce de ella y de la monotonía de μ^* que

$$\mu^*(A) \geq \sum_{i=1}^{n} \mu^*(A \cap B_i) + \mu^*(A \cap (\cap_{i=1}^{\infty} B_i^c)).$$

Notando que $\cap_{i=1}^{\infty} B_i^c = (\cup_{i=1}^{\infty} B_i)^c$ y tomando límite cuando $n \to \infty$, se tiene que

$$\mu^*(A) \geq \sum_{i=1}^{\infty} \mu^*(A \cap B_i) + \mu^*(A \cap (\cup_{i=1}^{\infty} B_i)^c). \tag{3.2}$$

De esto y de la subaditividad numerable de μ^* se deduce que

$$\begin{aligned}\mu^*(A) &\geq \sum_{i=1}^{\infty} \mu^*(A \cap B_i) + \mu^*(A \cap (\cup_{i=1}^{\infty} B_i)^c) \\ &\geq \mu^*(A \cap (\cup_{i=1}^{\infty} B_i)) + \mu^*(A \cap (\cup_{i=1}^{\infty} B_i)^c) \geq \mu^*(A)\end{aligned}$$

que prueba que $\cup_{i=1}^{\infty} B_i$ es medible y, en definitiva, que $\mathcal{A}_{\mu^*}$ es una σ-álgebra.

Para probar que la restricción de μ^* a $\mathcal{A}_{\mu^*}$ es una medida, basta comprobar la aditividad numerable. Si (B_n) es una sucesión disjunta en $\mathcal{A}_{\mu^*}$, reemplazando A por $\cup_{i=1}^{\infty} B_i$ en la desigualdad (3.2), se obtiene

$$\mu^*(\cup_{i=1}^{\infty} B_i) \geq \sum_{i=1}^{\infty} \mu^*(B_i) + 0;$$

puesto que la desigualdad contraria es cierta por la subaditividad numerable, queda probado el teorema. □

DEFINICIÓN 3.6. (Medida de Lebesgue en $\mathbb{R}^n$) $\mathcal{A}_{\lambda_n^*}$ es una σ-álgebra de partes de $\mathbb{R}^n$ cuyos elementos se llaman conjuntos Lebesgue medibles. La restricción de λ_n^* a $\mathcal{A}_{\lambda_n^*}$ se denotará por λ_n (λ, en el caso $n = 1$) y se llamará medida de Lebesgue en $\mathbb{R}^n$.

Proposición 3.3. (a) *Los borelianos de $\mathbb{R}^n$ son conjuntos Lebesgue-medibles.*

(b) *La medida de Lebesgue es invariante por traslaciones.*

Demostración. (a) Puesto que $\mathcal{R}^n$ está generada por los intervalos n-dimensionales de la forma $I = \prod_{i=1}^{n} I_i$, donde $I_j =]-\infty, a]$ e $I_i = \mathbb{R}$, si $i \neq j$, para algún $1 \leq j \leq n$ y algún $a \in \mathbb{R}$, basta probar que, para cada $A \subset \mathbb{R}$ tal que $\lambda_n^*(A) < \infty$, se verifica que

$$\lambda_n^*(A) \geq \lambda_n^*(A \cap I) + \lambda_n^*(A \cap I^c)$$

para un tal intervalo (ver el comienzo de la demostración del Teorema 3.1). Dado $\epsilon > 0$, podemos elegir una sucesión (R_i) de intervalos n-dimensionales abiertos y acotados que recubren A y satisfacen $\sum_{i=1}^{\infty} \mathrm{vol}(R_i) < \lambda_n^*(A) + \epsilon$. Para cada i, $R_i \cap I$ y $R_i \cap I^c$ son intervalos n-dimesionales disjuntos y acotados (posiblemente vacíos) y, por tanto, podemos tomar intervalos abiertos n dimensionales R_i' y R_i'' tales que

$$R_i \cap I \subset R_i', \quad R_i \cap I^c \subset R_i'' \quad \text{y} \quad \mathrm{vol}(R_i') + \mathrm{vol}(R_i'') \leq \mathrm{vol}(R_i) + \epsilon/2^i.$$

Las sucesiones (R_i') y (R_i'') recubren $A\cap I$ y $A\cap I^c$, resp., y, por tanto,

$$\lambda_n^*(A\cap I)+\lambda_n^*(A\cap I^c) \leq \sum_{i=1}^\infty \mathrm{vol}(R_i')+\sum_{i=1}^\infty \mathrm{vol}(R_i'') \leq (\sum_{i=1}^\infty \mathrm{vol}(R_i))+\epsilon \leq \lambda_n^*(A)+2\epsilon$$

lo que acaba la prueba.

(b) Se deduce inmediatamente de la definición de λ_n^* y de que el volumen de un intervalo n-dimensional es invariante por traslaciones. □

***Ejemplo* 3.4.** (Ejemplo de Vitali de un conjunto no Lebesgue-medible) No todo subconjunto de $\mathbb{R}$ es Lebesgue-medible. Vitali, haciendo uso del axioma de elección, propuso un ejemplo: Se define una relación de equivalencia $\sim$ en $\mathbb{R}$ del siguiente modo: $x \sim y \iff x-y \in \mathbb{Q}$. Las clases de equivalencia son de la forma $\mathbb{Q}+x$ y son densas en $\mathbb{R}$. El axioma de elección asegura la existencia de un subconjunto E de $]0,1[$ que contiene uno y sólo un elemento de cada clase de equivalencia. Sea $(r_n)_n$ una enumeración de los racionales del intervalo $]-1,1[$ y hagamos $E_n = r_n + E$. Entonces los E_n son dos a dos disjuntos, están contenidos en el intervalo $]-1,2[$ y recubren el intervalo $]0,1[$. Si E fuese Lebesgue-medible, puesto que $\lambda(\cup_n E_n) = \sum_n \lambda(E_n)$ y que los E_n tienen la misma medida de Lebesgue que E, llegaríamos a una contradicción con lo dicho anteriormente (distinguir los casos $\lambda(E)=0$ y $\lambda(E)>0$).

***Observación* 3.4.** (Compleción de un espacio de medida) Una medida μ en una σ-álgebra $\mathcal{A}$ se dice completa si, dado $A\in\mathcal{A}$ tal que $\mu(A)=0$, se verifica que $B\in\mathcal{A}$ para cada subconjunto B de A. La medida μ^* en $\mathcal{A}_{\mu^*}$ definida anteriormente es completa. Un espacio de medida $(\Omega,\mathcal{A},\mu)$ se completa del siguiente modo: se define $\mathcal{A}_\mu$ como la clase de los conjuntos de la forma $A\cup N$ cuando A recorre $\mathcal{A}$ y N la clase de los subconjuntos de los sucesos μ-nulos. Se prueba que $\mathcal{A}_\mu$ es una σ-álgebra. μ se extiende a $\mathcal{A}_\mu$ definiendo $\mu(A\cup N)=\mu(A)$ si $A\in\mathcal{A}$ y N es un subconjunto de un suceso μ-nulo. Se prueba que la definición es correcta y que μ es una medida completa en $\mathcal{A}_\mu$. El espacio de medida $(\Omega,\mathcal{A}_\mu,\mu)$ se llama compleción de $(\Omega,\mathcal{A},\mu)$. La medida de Lebesgue en la σ-álgebra de los conjuntos Lebesgue-medibles de $\mathbb{R}^n$ es la compleción de la medida de Lebesgue sobre la σ-álgebra de Borel. El lector puede encontrar la demostración de estas afirmaciones en Cohn (1980, p.36 y ss).

Para concluir este tema, presentaremos el teorema de extensión de Carathéodory que necesitaremos para construir medidas de Lebesgue-Stieljes y para extender el teorema de la medida producto al caso de una cantidad numerable de factores.

TEOREMA 3.2. (Teorema de extensión de Carathéodory) *Sea μ_0 una medida en un álgebra $\mathcal{A}_0$ de subconjuntos de Ω (es decir, $\mu_0 : \mathcal{A}_0 \to [0,+\infty]$ es una función de* *conjuntos nula en el vacío y numerablemente aditiva, en el sentido de que* $\mu_0(\cup_n A_n) = \sum_n \mu_0(A_n)$ *para cada sucesión* (A_n) *en* $\mathcal{A}_0$ *tal que* $\cup_n A_n \in \mathcal{A}_0$*) y supongamos que* μ_0 *es* σ*-finita en* $\mathcal{A}_0$ *(es decir,* Ω *puede descomponerse como* $\cup_n A_n$*, con* $A_n \in \mathcal{A}_0$ *y* $\mu_0(A_n) < +\infty$ *para cada* n*). Entonces* μ_0 *admite una única extensión a una medida* μ *en* $\sigma(\mathcal{A}_0)$.

Demostración. (Primera etapa) Se define

$$\mu^* : B \subset \Omega \mapsto \mu^*(B) := \text{ínf}\{\textstyle\sum_n \mu_0(A_n) \colon B \subset \cup_n A_n,\ A_n \in \mathcal{A}_0, \forall n\}.$$

Es claro que μ^* es una función de conjuntos nula en el vacío y monótona. Veamos que también es numerablemente subaditiva: sean (B_n) una sucesión de subconjuntos de Ω y, para cada $\epsilon > 0$ y cada $n \in \mathbb{N}$, $(A_{nk})_k$ una sucesión en $\mathcal{A}_0$ que recubre a B_n y tal que

$$\sum_k \mu_0(A_{nk}) \leq \mu^*(B_n) + \epsilon/2^n.$$

Entonces

$$\cup_n \cup_k A_{nk} \supset \cup_n B_n$$

y

$$\sum_n \sum_k \mu_0(A_{nk}) \leq \big(\sum_n \mu^*(B_n)\big) + \epsilon;$$

siendo $\epsilon > 0$ arbitrario y $(A_{nk})_{nk}$ un recubrimiento numerable en $\mathcal{A}_0$ de $\cup_n B_n$, se verifica que

$$\mu^*(\cup_n B_n) \leq \sum_n \mu^*(B_n).$$

Por tanto, μ^* es una medida exterior.

(Segunda etapa) Veamos ahora que μ^* coincide con μ_0 sobre $\mathcal{A}_0$. Reemplazando $\cup_n A_n$ por $\cup_n A'_n$, donde $A'_n = A_n \setminus \cup_{k=1}^{n-1} A_k$, se prueba fácilmente que

$$\mu^*(B) := \text{ínf}\{\textstyle\sum_n \mu_0(A_n) \colon B \subset \cup_n A_n,\ A_n \in \mathcal{A}_0, \forall n;\ A_n \cap A_m = \emptyset,\ \text{si } n \neq m\}.$$

Dados $A, A_1, A_2, \dots \in \mathcal{A}_0$ tales que $A \subset \cup_n A_n$, se verifica que, si los A_n son dos a dos disjuntos, $A = \cup_n (A_n \cap A)$ y $\mu_0(A) = \sum_n \mu_0(A \cap A_n) \leq \sum_n \mu_0(A_n)$, lo que prueba que $\mu_0(A) \leq \mu^*(A)$. Por otra parte, por definición de μ^*, para cada $A \in \mathcal{A}_0$, $\mu^*(A) \leq \mu_0(A) + \mu_0(\emptyset) + \mu_0(\emptyset) + \cdots = \mu_0(A)$. Luego $\mu^*(A) = \mu_0(A)$ si $A \in \mathcal{A}_0$.

(Tercera etapa) Probaremos ahora que $\mathcal{A}_0 \subset \mathcal{A}_{\mu^*}$. Dados $A \in \mathcal{A}_0$ y $B \subset \Omega$, para cada $\epsilon > 0$, existe una sucesión disjunta (A_n) en $\mathcal{A}_0$ tal que $B \subset \cup_n A_n$ y

$$\sum_n \mu_0(A_n) \leq \mu^*(B) + \epsilon.$$

Puesto que μ_0 es finitamente aditiva,

$$\mu^*(B) + \epsilon \geq \sum_n \mu_0(A_n) = \sum_n [\mu_0(A_n \cap A) + \mu_0(A_n \cap A^c)] \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

donde la última desigualdad es cierta por definición de μ^* y porque $\cup_n(A_n \cap A) \supset B \cap A$ y $\cup_n(A_n \cap A^c) \supset B \cap A^c$. Siendo $\epsilon > 0$ arbitrario, queda probado que $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$, es decir, que $A \in \mathcal{A}_{\mu^*}$. Por tanto, $\mathcal{A}_0 \subset \mathcal{A}_{\mu^*}$ y, siendo $\mathcal{A}_{\mu^*}$ una σ-álgebra, $\sigma(\mathcal{A}_0) \subset \mathcal{A}_{\mu^*}$.

(Cuarta y última etapa) La restricción μ de la medida exterior μ^* a la σ-álgebra $\mathcal{A} := \sigma(\mathcal{A}_0)$ es una medida que coincide con μ_0 en $\mathcal{A}_0$. Para probar la unicidad de la extensión, supongamos que μ' es una segunda extensión a $\mathcal{A}$ de μ_0; siendo μ_0 σ-finita, podemos suponer sin pérdida de generalidad que $\mu(\Omega) = \mu'(\Omega) < \infty$; un procedimiento similar al utilizado en la resolución del Problema 3.8 acaba la prueba. □

***Observación* 3.5.** (Medidas de Lebesgue-Stieljes en $\mathbb{R}$) Una medida de Lebesgue-Stieljes en $\mathbb{R}$ es una medida μ en $\mathcal{R}$ que asigna medida finita a cada intervalo acotado de la recta. Una función de distribución en $\mathbb{R}$ es una aplicación $F : \mathbb{R} \to \mathbb{R}$ que es creciente y continua por la derecha en cada punto. Ocurre que la fórmula $\mu(]a, b]) = F(b) - F(a)$ establece una correspondencia biunívoca entre las medidas de Lebesgue-Stieljes y las funciones de distribución, si identificamos dos funciones de distribución que difieren en una constante. Concretamente: si μ es una medida de Lebesgue-Stieljes y definimos F fijando arbitrariamente $F(0)$ y exigiendo que $F(b)-F(a) = \mu(]a, b])$ (p.ej., $F(x) = F(0)+\mu(]0, x])$ si $x > 0$ y $F(x) = F(0)-\mu(]x, 0])$ si $x < 0$), entonces F es una función de distribución; recíprocamente, como consecuencia del teorema de Carathéodory, dada una función de distribución F, existe una única medida de Lebesgue-Stieljes μ tal que $\mu(]a, b]) = F(b) - F(a)$. La medida de Lebesgue-Stieljes correspondiente a la función de distribución $F(x) = x$ es la medida de Lebesgue en $\mathbb{R}$. En particular, $\mu(]-\infty, a[) = F(a-) = \lim_{x \to a-} F(x)$, $\mu(\{a\}) = F(a) - F(a-)$ y $\mu([a, b]) = F(b) - F(a-)$. Una medida de Lebesgue-Stieljes μ en $\mathcal{R}$ se dice continua si $\mu(\{x\}) = 0$, para cada $x \in \mathbb{R}$; ello es equivalente a que su función de distribución sea continua. El lector interesado puede encontrar la demostración de todas las afirmaciones de esta observación en Ash (1972, p. 22 y ss). Ver también la observación 5.2.

PROBLEMAS DEL CAPÍTULO 3

Problema 3.1. Probar que en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ se verifican las siguientes relaciones, donde $A, B, A_1, A_2, \ldots$ son sucesos:

1. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
2. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.
3. $P(\cup_{n=1}^{\infty} A_n) = 1 - P(\cap_{n=1}^{\infty} A_n^c)$.
4. (Principio de inclusión-exclusión)

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) + \sum_{i_1<i_2<i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \ldots + (-1)^{n+1} P(\cap_{i=1}^{n} A_i)$$

Problema 3.2. Sea Ω un conjunto infinito numerable y sea $\mathcal{F} = \mathcal{P}(\Omega)$. Definimos sobre $(\Omega, \mathcal{F})$, la siguiente aplicación:

$$\mu(A) = \begin{cases} 0 & \text{si } A \text{ es finito} \\ +\infty & \text{si } A \text{ es infinito} \end{cases}$$

1. Probar que μ es finitamente aditiva pero no numerablemente aditiva.
2. Probar que Ω es el límite de una sucesión creciente de conjuntos A_n con $\mu(A_n) = 0$, $\forall n \in \mathbb{N}$, pero $\mu(\Omega) = \infty$.

Problema 3.3. Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $\{A_1, \ldots, A_n\}$ una partición medible de Ω (es decir, $A_1, \ldots, A_n$ son sucesos dos a dos disjuntos que recubren Ω). Recordemos que, dados dos sucesos A y B, se define la probabilidad condicionada de B supuesto que ha ocurrido el suceso A mediante:

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

si $P(A) > 0$.

(a) Probar que la aplicación $P_A : B \in \mathcal{B} \mapsto P_A(B) := P(B|A) \in [0, 1]$ es una probabilidad (es, de hecho, la restricción P^A de P al suceso A dividida por $P(A)$).

(b) (Teorema de la probabilidad total) Probar que

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

si los sucesos A_i tienen probabilidad no nula.

(c) (Teorema de Bayes) Para $1 \leq k \leq n$,

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

(d) Una población tiene un 40 % de varones y un 60 % de mujeres. Se sabe que un 10 % de los hombre llevan pelo largo y un 40 % de las mujeres lo llevan corto. Se presenta una persona con pelo largo: ¿cuál es la probabilidad de que sea una mujer?

Problema 3.4. (Orígenes del cálculo de probabilidades)

(a) El Caballero de Méré se sorprendía de que al lanzar dos dados se obtenga con más frecuencia 11 que 12 al sumar los resultados, a pesar de que uno y otro se obtienen solamente con una combinación (5+6 y 6+6). ¿Que piensas de ello?

(b) (★) Dos jugadores igualmente hábiles juegan una serie de partidas, habiendo depositado previamente cada uno de ellos a monedas en la mesa. El jugador que gane 4 partidas antes que su contrincante se llevará la cantidad total depositada. Pero la serie de partidas ha debido interrumpirse en un instante en que un jugador había ganado 2 partidas y perdido una. ¿Cómo debe repartirse la cantidad depositada? ([1])

Problema 3.5. Describir el espacio de probabilidad correspondiente al experimento aleatorio que consiste en tres lanzamientos independientes al aire de un dado perfecto y calcular la probabilidad de que la suma de los resultados sea menor o igual que 5.

Problema 3.6. Calcular la medida de Lebesgue del conjunto

$$\{(x, y) \in \mathbb{R}^2 \colon y = x,\ x \geq 0\}.$$

Problema 3.7. Sea $(A_n)_n$ una sucesión decreciente de sucesos en un espacio de medida $(\Omega, \mathcal{A}, \mu)$. Probar que, en general, $\mu(\cap_n A_n)$ no coincide con $\lim_n \mu(A_n)$ si $\mu(A_n) = \infty$, para cada $n \geq 1$.

Problema 3.8. Sean $\mathcal{C}$ un π-sistema en Ω y P_1 y P_2 dos probabilidades en $\sigma(\mathcal{C})$ que coinciden en $\mathcal{C}$. Probar que $P_1 = P_2$. Consecuencia: dos distribuciones de probabilidad en $\mathbb{R}$ con la misma función de distribución coinciden en todo $\mathcal{R}$.

Problema 3.9. Si μ es una medida σ-finita, entonces Ω puede ser recubierto por una sucesión $(A_n)_n$ de conjuntos medibles de medida finita que puede suponerse creciente o disjunta, si conviene.

[1] En el siglo XVII, Antoine de Gambaud (Caballero de Méré y jugador profesional) planteó estos dos problemas -que, por cierto, ya habían sido considerados por Cardano un siglo antes- a Blaise Pascal, y se acostumbra a fijar el origen del Cálculo de Probabilidades en la correspondencia mantenida entre Pascal y Fermat para la resolución de éstos y otros problemas similares.

Problema 3.10. (★) (Lema de Borel-Cantelli) Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Si $(A_n)_n$ es una sucesión de sucesos tal que $\sum_n P(A_n) < \infty$ entonces

$$P(\limsup_n A_n) = 0.$$

Problema 3.11. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Si $(A_n)_n$ es una sucesión de sucesos, se verifica que

$$P(\liminf_n A_n) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P(\limsup_n A_n).$$

Nota: Si reemplazamos P por una medida μ, la primera desigualdad sigue siendo cierta y la última también es cierta si $\mu(\cup_{k\geq n} A_k) < \infty$ para algún n.

Problema 3.12. Sea F la función de distribución sobre $\mathbb{R}$ dada por:

$$F(x) = \begin{cases} 0 & \text{si } x < -1 \\ 1+x & \text{si } -1 \leq x < 0 \\ 2+x^2 & \text{si } 0 \leq x < 2 \\ 9 & \text{si } x \geq 2 \end{cases}$$

Si μ es la medida de Lebesgue-Stieltjes correspondiente a F, calcular la medida de cada uno de los siguientes conjuntos: a) $\{2\}$, b) $[-1/2, 3)$, c) $]-1, 0] \cup]1, 2[$, d) $[0, 1/2] \cup]1, 2[$, e) $\{x : |x| + 2x^2 > 1\}$.

Problema 3.13. Encontrar la probabilidad de que la suma de dos números no negativos elegidos al azar (ambos menores o igual a 1) sea mayor que 1 y su producto sea menor que 1/4.

Problema 3.14. (ver Proposición 3.2 y su demostración) (a) Probar que si $[a,b] \subset \cup_{i=1}^k]a_i, b_i[$, entonces $b - a \leq \sum_{i=1}^k (b_i - a_i)$.

(b) Completar la demostración de que la medida exterior de Lebesgue asigna a cada intervalo n-dimensional su volumen en el caso $n > 1$.

> *Indicación:* (a) Sean $c_1 < \ldots < c_d$ tales que $\{a_1, \ldots, a_k\} \cup \{b_1, \ldots, b_k\} = \{c_1 \ldots, c_d\}$. Es claro que, para cada $1 \leq i \leq d-1$, existe $j_i \in \{1, \ldots, k\}$ tal que $]c_i, c_{i+1}[\subset]a_{j_i}, b_{j_i}[$. Por tanto,
>
> $$b - a \leq \textstyle\sum_{i=1}^{d-1} (c_{i+1} - c_i) \leq \sum_{i=1}^{d-1} (b_{j_i} - a_{j_i}) \leq \sum_{i=1}^k (b_i - a_i).$$
>
> (b) Con las notaciones del final de la demostración de la Proposición 3.2, para $1 \leq i \leq n$, sean $c_{1i} < \ldots < c_{d_i i}$ tales que $\{c_{1,i}, \ldots, c_{d_i,i}\} = \{a_{1i}, \ldots, a_{ki}\} \cup \{b_{1i}, \ldots, b_{ki}\}$. Bastaría tomar los K_s como los intervalos n-dimensionales que son de la forma $I \cap \prod_{i=1}^n [c_{j_i,i}, c_{j_i+1,i}]$, para $j_i \in \{1, \ldots, d_i - 1\}$, $1 \leq i \leq n$. En realidad, lo que se ha hecho es prolongar las "hipercaras" $(n-1)$-dimensionales de los k intervalos n-dimensionales I_j que recubren I para obtener una malla de intervalos n-dimensionales acotados cuya unión es $\prod_{i=1}^n [c_{1,i}, c_{d_i,i}]$; los $(K_s)_s$ son aquellos intervalos n-dimensionales acotados de la malla que cortan a I.

Capítulo 4

Integral. Esperanza

Definimos a continuación la integral de una v.a.r. respecto a una medida μ.

DEFINICIÓN 4.1. (Integral) Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida.

(a) Si $A \in \mathcal{A}$, se llama integral de I_A respecto a μ, y se denota $\int I_A \, d\mu$, el número real $\mu(A)$.

(b) Sea $f = \sum_1^n a_i I_{A_i}$ una función simple donde los A_i son sucesos dos a dos disjuntos y los a_i son números reales no negativos. Se define la integral de f respecto a μ, y se denota por $\int f d\mu$, como el número real $\sum_i a_i \mu(A_i)$.

(c) Si $f : \Omega \longrightarrow \bar{\mathbb{R}}$ es Borel-medible y no negativa, se define la integral de f respecto a μ por $\int f d\mu = \sup\{\int s d\mu \colon s \text{ simple y } 0 \leq s \leq f\}$.

(d) Si $f : \Omega \longrightarrow \bar{\mathbb{R}}$ es Borel medible, diremos que f es μ-integrable si las integrales de sus partes positiva y negativa son finitas; diremos que existe la integral de f respecto a μ si al menos una de ellas es finita. En ambos casos se define $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$.

(e) Si $f = (f_1, \ldots, f_n)$ es una función medible a valores en $\bar{\mathbb{R}}^n$, diremos que existe la integral de f respecto a μ (resp., diremos que f es μ-integrable) si existen (resp., existen y son finitas) las integrales $\int f_i d\mu$, $1 \leq i \leq n$, y, en ese caso, se define

$$\int f \, d\mu = \left(\int f_1 \, d\mu, \ldots, \int f_n \, d\mu \right).$$

(f) Si f es Borel-medible y $A \in \mathcal{A}$, denotaremos $\int_A f d\mu = \int (f I_A) d\mu$, si esta última integral existe.

***Observación* 4.1.** En la parte (b) de la definición anterior debemos verificar que la integral de la función simple considerada no depende de la representación de f como combinación lineal finita de indicadores con soportes dos a dos disjuntos, es decir, habrá que probar que si $f = \sum_{i=1}^n a_i I_{A_i} = \sum_{j=1}^m b_j I_{B_j}$ entonces $\sum_i a_i \mu(A_i) = \sum_j b_j \mu(B_j)$; pero eso se prueba sin dificultad notando que ambas expresiones coinciden con $\sum_i \sum_j z_{ij} \mu(A_i \cap B_j)$, donde $z_{ij} = a_i = b_j$ si $A_i \cap B_j \neq \emptyset$.

***Observación* 4.2.** Usaremos distintas notaciones para la integral de una función medible respecto a una medida. Así, las notaciones $\int f(\omega) d\mu(\omega)$, $\int f(\omega)\mu(d\omega)$ y $\int_\Omega f(\omega) d\mu(\omega)$ significan lo mismo que $\int f d\mu$.

***Observación* 4.3.** Si μ es una probabilidad, entonces $\int f d\mu$ se suele denotar en la forma $E_\mu(f)$ (o, simplemente, $E(f)$ o Ef, si no hay confusión) y se llama esperanza o media de f.

***Observación* 4.4.** Si $f : \mathbb{R} \longrightarrow \mathbb{R}$ es una v.a.r., su integral respecto a la medida de Lebesgue (si existe) se denota usualmente por $\int f(x)\, dx$, o $\int f$, o $\int_{\mathbb{R}} f(x)\, dx$ o también $\int_{-\infty}^{+\infty} f(x)\, dx$.

***Observación* 4.5.** El conjunto $\mathbb{C}$ de los números complejos se identifica con $\mathbb{R}^2$ en la forma usual. Los borelianos de $\mathbb{C}$ son exactamente los mismos que los de $\mathbb{R}^2$. Así, una función compleja f es medible si y sólo si lo son su partes real e imaginaria, y la integral de f se define como $\int Ref d\mu + i \int Imf d\mu$.

El resultado siguiente establece algunas propiedades de la integral. En lo que sigue $(\Omega, \mathcal{A}, \mu)$ será un espacio de medida.

Proposición 4.1. *Sean X e Y funciones medibles $\bar{\mathbb{R}}$-valoradas en Ω.*

(a) *Si $c \in \mathbb{R}$ y existe $\int X d\mu$, entonces existe también $\int cX d\mu$ y coincide con $c \int X d\mu$.*

(b) *Si $X \leq Y$ y sus integrales respecto a μ existen, entonces $\int X d\mu \leq \int Y d\mu$.*

(c) *Si existe $\int X d\mu$, entonces $|\int X d\mu| \leq \int |X| d\mu$.*

(d) *Si $X \geq 0$ y $A \in \mathcal{A}$, entonces $\int_A X d\mu = \sup\{\int_A s d\mu \colon 0 \leq s \leq X,\ s \text{ simple}\}$.*

(e) *Si existe (resp., es finita) $\int X\, d\mu$, entonces existe (resp., es finita) $\int_A X\, d\mu$ para cada $A \in \mathcal{A}$.*

Demostración. (a) El resultado es trivialmente cierto si X es simple. Si X es no negativa y $c > 0$, entonces

$$\begin{aligned}\int cX d\mu &= \sup\left\{\int s d\mu \colon 0 \le s \le cX, \ \ s \text{ simple}\right\} \\ &= c \sup\left\{\int \frac{s}{c} d\mu \colon 0 \le \frac{s}{c} \le X, \ \ \frac{s}{c} \text{ simple}\right\} = c\int X d\mu.\end{aligned}$$

En general, si $X = X^+ - X^-$ y $c > 0$, entonces $(cX)^+ = cX^+$ y $(cX)^- = cX^-$; por tanto, $\int cX d\mu = \int cX^+ d\mu - \int cX^- d\mu = c\int X d\mu$. Si $c < 0$, entonces $(cX)^+ = -cX^-$, $(cX)^- = -cX^+$; por tanto,

$$\int cX d\mu = -c\int X^- d\mu + c\int X^+ d\mu = c\int X d\mu.$$

(b) Si X e Y son no negativas y $0 \le s \le X$, s simple, entonces $0 \le s \le Y$; por tanto, $\int X d\mu \le \int Y d\mu$. En general, $X \le Y$ implica que $X^+ \le Y^+$ y $X^- \ge Y^-$. De eso y de que ambas integrales existen se deduce que

$$\int Y d\mu = \int Y^+ d\mu - \int Y^- d\mu \ge \int X^+ d\mu - \int X^- d\mu = \int X d\mu.$$

(c) Se tiene que $-|X| \le X \le |X|$ y, por (a) y (b), $-\int |X| d\mu \le \int X d\mu \le \int |X| d\mu$, de donde se sigue el resultado.

(d) Si s es una función simple medible y $s \le X$, entonces $\int_A s d\mu \le \int_A X d\mu$ por (b); entonces

$$\int_A X d\mu \ge \sup\left\{\int_A s d\mu \colon 0 \le s \le X, \ s \text{ simple}\right\}.$$

Si t es una función simple medible tal que $0 \le t \le XI_A$, entonces $t = tI_A \le X$ y $\int t d\mu \le \sup\{\int sI_A d\mu \colon 0 \le s \le X, \ s \text{ simple}\}$. Tomando supremo en t se deduce que

$$\int_A X d\mu \le \sup\left\{\int_A s d\mu \colon 0 \le s \le X, \ s \text{ simple}\right\}.$$

(e) Se sigue de (b) y de que $(XI_A)^+ = X^+ I_A \le X^+$ y $(XI_A)^- = X^- I_A \le X^-$. □

Los objetivos fundamentales de este capítulo son los teoremas de convergencia de Lebesgue y el teorema de aditividad. La proposición siguiente, parte fundamental del Teorema 6.1, resultará de utilidad en su demostración.

Proposición 4.2. *Sea h una función Borel-medible cuya integral respecto a μ existe. Hagamos $\lambda(A) = \int_A h d\mu$, para $A \in \mathcal{A}$. Entonces λ es numerablemente aditiva en $\mathcal{A}$. Si $h \geq 0$, λ es una medida.*

Demostración. Supongamos en primer lugar que h es una función simple medible no negativa: $h = \sum_{i=1}^n x_i I_{A_i}$. Entonces $\lambda(A) = \int_A h d\mu = \sum_{i=1}^n x_i \mu(A \cap A_i)$. Puesto que μ es numerablemente aditiva, también lo es λ.

Supongamos ahora que h es una función medible no negativa y sean $(A_n)_n$ una sucesión disjunta en $\mathcal{A}$ y $A = \cup_n A_n$. Si s es una función simple medible tal que $0 \leq s \leq h$ entonces, puesto que el resultado es cierto para funciones simples, $\int_A s d\mu = \sum_{i=1}^\infty \int_{A_n} s d\mu$ y esto es $\leq \sum_{i=1}^\infty \int_{A_n} h d\mu$, por definición de integral. Tomando el supremo en s queda probado que $\lambda(A) \leq \sum_{i=1}^\infty \lambda(A_n)$. Para probar la desigualdad contraria supondremos que todos los A_n verifican que $\lambda(A_n) < \infty$ (si alguno de ellos fuera infinito, puesto que $\lambda(A_n) \leq \lambda(A)$ por la parte (b) de la proposición anterior, tendríamos que $\lambda(A) = \infty$ y habríamos acabado). De las partes (d) y (b) de la proposición anterior y de que el máximo de un número finito de funciones simples es una función simple, se deduce que, dado $\epsilon > 0$, existe una función simple $0 \leq s \leq h$ tal que

$$\int_{A_i} s d\mu \geq \int_{A_i} h d\mu - \frac{\epsilon}{n}, \quad i = 1, \dots, n.$$

Entonces, de acuerdo con lo probado para funciones simples,

$$\lambda(\cup_{i=1}^n A_i) = \int_{\cup_{i=1}^n A_i} h d\mu \geq \int_{\cup_{i=1}^n A_i} s d\mu = \sum_{i=1}^n \int_{A_i} s d\mu.$$

Por tanto,

$$\lambda(\cup_{i=1}^n A_i) \geq \sum_{i=1}^n \int_{A_i} h d\mu - \epsilon = \sum_{i=1}^n \lambda(A_i) - \epsilon.$$

Puesto que $\lambda(A) \geq \lambda(\cup_{i=1}^n A_i)$ y que $\epsilon > 0$ y n son arbitrarios, se tiene que $\lambda(A) \geq \sum_{i=1}^\infty \lambda(A_i)$.

Por último, sea $h = h^+ - h^-$ una función Borel-medible arbitraria. Entonces $\lambda(A) = \int_A h^+ d\mu - \int_A h^- d\mu$. Puesto que $\int h^+ d\mu < \infty$ o $\int h^- d\mu < \infty$, queda probado el resultado. $\square$

TEOREMA 4.1. (Teorema de la convergencia monótona de Lebesgue) *Sean $f, f_1, f_2, \dots$ funciones medibles $[0, +\infty]$-valoradas. Si $f_1(\omega) \leq f_2(\omega) \leq \dots$ y $f(\omega) = \lim_n f_n(\omega)$, $\forall \omega \in \Omega$, entonces $\int f d\mu = \lim_n \int f_n d\mu$.*

Demostración. Puesto que $(\int f_n d\mu)_n$ es una sucesión creciente y que $\int f_n d\mu \leq \int f d\mu$, se verifica que $k := \lim_n \int f_n d\mu \leq \int f d\mu$. Fijemos ahora $0 < b < 1$ y una función simple medible no negativa $s \leq f$. Sea $B_n := \{f_n \geq bs\}$. Entonces $(B_n)_n$ es una sucesión creciente de sucesos cuya unión es Ω pues f_n converge a f y s toma valores en $\mathbb{R}$. Pero, por un resultado anterior, $k \geq \int f_n d\mu \geq \int_{B_n} f_n d\mu$, y $\int_{B_n} f_n d\mu \geq b \int_{B_n} s d\mu$. Por el resultado anterior, $\int_{B_n} s d\mu \to \int s d\mu$. Siendo eso cierto para cada $0 < b < 1$, se tiene que $k \geq \int s d\mu$. Tomando el supremo en s, queda probado que $k \geq \int f d\mu$. □

TEOREMA 4.2. (Teorema de aditividad) *Sean f y g funciones Borel-medibles en Ω tales que $f + g$ está bien definida. Si existen $\int f d\mu$ y $\int g d\mu$ y $\int f d\mu + \int g d\mu$ está bien definida (es decir, no es de la forma $+\infty - \infty$ o $-\infty + \infty$), entonces*

$$\int (f+g) d\mu = \int f d\mu + \int g d\mu.$$

En particular, si f y g son integrables, también lo es $f + g$.

Demostración. Si f y g son funciones reales simples medibles no negativas, el resultado se prueba fácilmente por la definición de la integral. Supongamos ahora que f y g son funciones medibles no negativas, y sean $(t_n)_n$ y $(u_n)_n$ sucesiones crecientes de funciones reales simples medibles no negativas puntualmente convergentes a f y g, resp. Entonces $(s_n)_n := (t_n)_n + (u_n)_n$ es una sucesión creciente de funciones reales simples medibles no negativas convergente a $f + g$. Pero $\int s_n d\mu = \int t_n d\mu + \int u_n d\mu$; por el teorema de la convergencia monótona, $\int (f+g) d\mu = \int f d\mu + \int g d\mu$.

Supongamos ahora que $f \geq 0$, $g \leq 0$ y $h := f + g \geq 0$ (g es, por tanto, finita). Entonces $f = h + (-g)$ y $\int f d\mu = \int h d\mu - \int g d\mu$. Si $\int g d\mu$ es finita, entonces $\int (f+g) d\mu = \int f d\mu + \int g d\mu$; si $\int g d\mu = -\infty$, siendo $h \geq 0$, tendríamos que $\int f d\mu \geq -\int g d\mu = \infty$, en contra de que $\int f d\mu + \int g$ está bien definida. Análogamente, si

$f \geq 0$, $g \leq 0$ y $h \leq 0$, se obtiene que $\int h d\mu = \int f d\mu + \int g$ reemplazando todas las funciones por sus opuestas ($-g \geq 0$, $-f \leq 0$ y $-h = -g - f \geq 0$).

En el caso general, hagamos

$$\begin{aligned}
&E_1 = \{f, g \geq 0\},\\
&E_2 = \{f \geq 0,\ \ g < 0,\ h \geq 0\},\quad E_3 = \{f \geq 0,\ \ g < 0,\ h < 0\},\\
&E_4 = \{f < 0,\ \ g \geq 0,\ h \geq 0\},\quad E_5 = \{f < 0,\ \ g \geq 0,\ h < 0\},\\
&E_6 = \{f, g < 0\}.
\end{aligned}$$

Los argumentos usados arriba prueban que $\int_{E_i} h d\mu = \int_{E_i} f d\mu + \int_{E_i} g d\mu$, $1 \leq i \leq 6$. Pero $\int f d\mu = \sum_{i=1}^6 \int_{E_i} f d\mu$ y $\int g d\mu = \sum_{i=1}^6 \int_{E_i} g d\mu$ por el Teorema 4.2; luego $\int f d\mu + \int g d\mu = \sum_{i=1}^6 \int_{E_i} h d\mu$, y esto coincide con $\int h d\mu$ si podemos probar que $\int h d\mu$ existe, es decir, que $\int h^+ d\mu$ y $\int h^- d\mu$ no son ambas infinitas. Si ése fuese el caso, tendríamos que $\int_{E_i} h^+ d\mu = \int_{E_j} h^- d\mu = \infty$ para algún par $i, j \in \{1, \ldots, 6\}$. Por tanto, $\int_{E_i} h d\mu = \infty$ y $\int_{E_j} h d\mu = -\infty$. Pero entonces $\int_{E_i} f d\mu = \infty$ o $\int_{E_i} g d\mu = \infty$ y, por tanto, $\int_\Omega f d\mu = \infty$ o $\int_\Omega g d\mu = \infty$. Análogamente, $\int_\Omega f d\mu = -\infty$ o $\int_\Omega g d\mu = -\infty$, que es una contradicción. □

Corolario 4.1. *Sean $f, g, f_1, f_2, \ldots$ funciones reales medibles.*

(a) (Teorema de Beppo-Levy) *Si $f_1, f_2, \ldots$ son no negativas, entonces*

$$\int \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

(b) *f es integrable si y sólo si lo es $|f|$.*

(c) *Si $|f| \leq g$ y g es integrable, entonces f es integrable.*

(d) *Si $f = 0$ μ-c.s. entonces $\int f d\mu = 0$.*

(e) *Si $f = g$ μ-c.s. y $\int g d\mu$ existe, entonces existe $\int f d\mu$ y $\int f d\mu = \int g d\mu$.*

(f) *Si f es μ-integrable, entonces f es finita μ-c.s.*

(g) *Si $f \geq 0$ y $\int f d\mu = 0$ entonces $f = 0$ μ-c.s.*

Demostración. (a) $\sum_{i=1}^n f_i$ es una sucesión creciente de funciones reales medibles no negativas que converge a $\sum_{i=1}^\infty f_i$ y el resultado se sigue del teorema de aditividad y del teorema de la convergencia monótona (Teorema 4.1).

(b) Puesto que $|f| = f^+ + f^-$, el resultado se sigue de la definición de integral y del teorema anterior.

(c) La monotonía de la integral prueba que $|f|$ es integrable, y el resultado se deduce de (b).

(d) Si $f = \sum_{i=1}^n x_i I_{A_i}$ es una función simple, entonces $x_i \neq 0$ implica $\mu(A_i) = 0$, por hipótesis, y por tanto, $\int f d\mu = 0$. Si $f \geq 0$ y $0 \leq s \leq f$, s simple, entonces $s = 0$ μ-c.s. y $\int s d\mu = 0$; luego $\int f d\mu = 0$. Si $f = f^+ - f^-$, entonces f^+ y f^- son cero μ-c.s. por ser menores o iguales que $|f|$, y esto acaba la prueba.

(e) Sean $A = \{f = g\}$ y $B = A^c$. Entonces $f = fI_A + fI_B$ y $g = gI_A + gI_B = fI_A + gI_B$. Puesto que $fI_B = gI_B = 0$ salvo en B, que tiene medida nula, el resultado se deduce de la proposición anterior y del teorema de aditividad.

(f) Sea $A = \{|f| = \infty\}$. Si $\mu(A) > 0$ entonces $\int |f| d\mu \geq \int_A |f| d\mu = \infty\mu(A) = \infty$, que es una contradicción.

(g) Sean $B = \{f > 0\}$ y $B_n = \{f \geq 1/n\}$. B_n es una sucesión creciente a B. Se verifica que $0 \leq fI_{B_n} \leq fI_B = f$. Por la monotonía de la integral, $\int_{B_n} f d\mu = 0$. Pero $\int_{B_n} f d\mu \geq \mu(B_n)/n$. Se deduce que $\mu(B_n) = 0$ para cada n y, por tanto, $\mu(B) = 0$. □

***Observación* 4.6.** Consecuencia del corolario anterior es que, en resultados de integración, podemos reemplazar cualquier función medible por otra que coincida con ella c.s. sin que ello afecte al resultado, o reemplazar condiciones que ocurren en todo punto por otras que se verifiquen c.s. Por ejemplo, el teorema de la convergencia monótona podría enunciarse en los siguientes términos: "Sean $f, f_1, f_2, \dots$ funciones reales medibles tales que las relaciones $f_1 \leq f_2 \leq f_3 \leq \dots$, $f_1, f_2, \dots \geq 0$ y $f = \lim_n f_n$ se verifican μ-c.s. Entonces $\int f d\mu = \lim_n \int f_n d\mu$."

El siguiente resultado extiende sensiblemente el teorema de la convergencia monótona:

TEOREMA 4.3. (Teorema de la convergencia monótona extendido) *Sean $f, g, f_1, f_2, \dots$ funciones Borel-medibles.*

(a) *Si f_n es una sucesión creciente convergente a f puntualmente y $f_n \geq g$ para cada n y $\int g d\mu > -\infty$, entonces $\int f d\mu = \lim_n \int f_n d\mu$.*

(b) *Si f_n es una sucesión decreciente convergente a f puntualmente y $f_n \leq g$ para cada n y $\int g d\mu < +\infty$, entonces $\int f d\mu = \text{lím}_n \int f_n d\mu$.*

Demostración. (a) Si $\int g d\mu = +\infty$, entonces $\int f_n d\mu = +\infty$ para cada n y $\int f d\mu = +\infty$, y el resultado es trivial. Si $\int g d\mu < +\infty$, entonces g es finita μ-c.s.; cambiemos g por 0 donde sea infinita. Entonces $(f_n - g)_n$ es una sucesión creciente de funciones medibles no negativas que converge puntualmente c.s. a $f - g$; por el teorema de la convergencia monótona (Teorema 4.1), $\int (f_n - g) d\mu$ converge a $\int (f - g) d\mu$ y el resultado se deduce del teorema de aditividad; este teorema se aplica pues, siendo $\int g d\mu > -\infty$, las integrales $\int f_n d\mu$ y $\int f d\mu$ existen y son $> -\infty$; además, $\int g d\mu$ es finita y, por tanto, $\int f_n d\mu - \int g d\mu$ y $\int f d\mu - \int g d\mu$ están bien definidas.

(b) $-f_n \geq -g$, $\int -g d\mu > -\infty$, y $-f_n$ es una sucesión creciente convergente a $-f$. Por la parte (a), $-\int f_n d\mu$ converge a $-\int f d\mu$ y de ahí se sigue el resultado. □

TEOREMA 4.4. (Lema de Fatou) *Sean $f, f_1, f_2, \dots$ funciones $\bar{\mathbb{R}}$-valoradas y Borel-medibles en Ω.*

(a) *Si $f_n \geq f$, para cada n, y $\int f d\mu > -\infty$, entonces*

$$\liminf_n \int f_n d\mu \geq \int \liminf f_n d\mu.$$

(b) *Si $f_n \leq f$, para cada n, y $\int f d\mu < +\infty$, entonces*

$$\limsup_n \int f_n d\mu \leq \int \limsup f_n d\mu.$$

Demostración. (a) Sean $g_n = \inf_{k \geq n} f_k$ y $g = \liminf f_n$. Entonces $g_n \geq f$ para cada n, $\int f d\mu > -\infty$ y (g_n) es una sucesión creciente convergente a g. Del teorema anterior se deduce que $\int g_n d\mu$ converge crecientemente a $\int \liminf_n f_n d\mu$. Pero $g_n \leq f_n$; por tanto,

$$\int \liminf f_n d\mu = \int \lim_n g_n d\mu = \lim \int g_n d\mu = \liminf \int g_n d\mu \leq \liminf \int f_n d\mu.$$

(b) Podemos escribir

$$\begin{aligned} \int \limsup_n f_n d\mu &= -\int \liminf_n (-f_n) d\mu \\ &\geq -\liminf_n \int (-f_n) d\mu = \limsup_n \int f_n d\mu. \end{aligned}$$

□

TEOREMA 4.5. (Teorema de la convergencia dominada de Lebesgue) *Sean $g, f, f_1, f_2, \ldots$ funciones $\bar{\mathbb{R}}$-valoradas y Borel-medibles en Ω. Supongamos que g es μ-integrable y que las relaciones $|f_n| \le g$ $(n = 1, 2, \ldots)$ y $f = \lim_n f_n$ ocurren μ-c.s. en Ω. Entonces $f, f_1, f_2, \ldots$ son también μ-integrables y $\int f d\mu = \lim_n \int f_n d\mu$.*

Demostración. Puesto que $f, f_1, f_2, \ldots$ están dominadas por una función μ-integrable, son μ-integrables. Se sigue del lema de Fatou que

$$\begin{aligned}\int_\Omega (\liminf_n f_n) d\mu &\le \liminf_n \int_\Omega f_n d\mu \le \limsup_n \int_\Omega f_n d\mu \\ &\le \int_\Omega (\limsup_n f_n) d\mu\end{aligned}$$

Por hipótesis, $\liminf_n f_n = \limsup_n f_n = f$ c.s. y, por tanto, todas las desigualdades precedentes son igualdades e iguales a $\int_\Omega f d\mu$. □

Sabemos que dos funciones reales medibles que coinciden c.s. tienen la misma integral; el recíproco es evidentemente falso. No obstante, si dos funciones reales medibles tienen la misma integral sobre cualquier conjunto medible, entonces coinciden c.s. Eso es lo que prueba, entre otras cosas, el siguiente resultado.

TEOREMA 4.6. *Supongamos que μ es σ-finita sobre $\mathcal{A}$ y que f y g son funciones reales $\mathcal{A}$-medibles cuyas integrales respecto a μ existen.*

(a) *Si $\int_A f d\mu \le \int_A g d\mu$ para cada suceso $A \in \mathcal{A}$, entonces $f \le g$ μ-c.s.*

(b) *Si $\int_A f d\mu = \int_A g d\mu$ para cada suceso $A \in \mathcal{A}$, entonces $f = g$ μ-c.s.*

Demostración. Es claro que basta probar (a) en el caso de que μ sea finita. Para cada $n \in \mathbb{N}$, hagamos

$$A_n = \{f \ge g + 1/n;\ |g| \le n\}.$$

Entonces

$$\int_{A_n} g d\mu \ge \int_{A_n} f d\mu \ge \int_{A_n} g d\mu + \frac{1}{n}\mu(A_n).$$

Pero

$$\left|\int_{A_n} g d\mu\right| \leq \int_{A_n} |g| d\mu \leq n\mu(A_n) < \infty,$$

y, por tanto, podemos restar $\int_{A_n} g d\mu$ para obtener que $\mu(A_n)/n \leq 0$; luego $\mu(A_n) = 0$. Entonces $\mu(\cup_n A_n) = 0$, es decir, $\mu(\{f > g, g \text{ finita}\}) = 0$. Por tanto, $f \leq g$ c.s. en $\{g \text{ finita}\}$. Obviamente, $f \leq g$ cuando $g = \infty$. Por último, haciendo $C_n = \{g = -\infty; f \geq -n\}$, se obtiene

$$-\infty\mu(C_n) = \int_{C_n} g d\mu \geq \int_{C_n} f d\mu \geq -n\mu(C_n);$$

se sigue de ello que $\mu(C_n) = 0$. Por tanto, $\mu(\cup_n C_n) = 0$, es decir,

$$\mu(\{f > g; g = -\infty\}) = 0.$$

Entonces $f \leq g$ c.s. en $\{g = -\infty\}$. □

***Observación* 4.7.** En el caso de que f y g sean μ-integrables, la demostración del teorema anterior es mucho más simple, y no es necesario exigir que μ sea σ-finita. Baste notar que si $B = \{f > g\}$ entonces $\int_B f d\mu \leq \int_B g d\mu \leq \int_B f d\mu$; las tres integrales son iguales y, por tanto, $\int_\Omega (f - g) I_B d\mu = 0$; siendo $f - g > 0$ en B se deduce que $\mu(B) = 0$. Puesto que $f \leq g$ en B^c, se deduce que $f \leq g$ c.s.

PROBLEMAS DEL CAPÍTULO 4

Problema 4.1. Sean $\sum_{i=1}^m a_i I_{A_i} = \sum_{j=1}^n b_j I_{B_j}$ dos representaciones de una misma función simple medible no negativa $f : (\Omega, \mathcal{A}, \mu) \to [0, +\infty[$. Probar que la integral de f respecto a μ no depende de la representación elegida, es decir, $\sum_{i=1}^m a_i \mu(A_i) = \sum_{j=1}^n b_j \mu(B_j)$.

> *Indicación:* Probar que ambas expresiones coinciden con $\sum_{i=1}^m \sum_{j=1}^n z_{ij}\mu(A_i \cap B_j)$, donde $z_{ij} = a_i = b_j$ si $A_i \cap B_j \neq \emptyset$.

Problema 4.2. Sean $(\Omega, \mathcal{A})$ un espacio medible, ω_0 un punto de Ω tal que $\{\omega_0\} \in \mathcal{A}$ y $f(\Omega, \mathcal{A}) \longrightarrow \bar{\mathbb{R}}$ una función medible. Probar que

$$\int f d\varepsilon_{\omega_0} = f(\omega_0).$$

Problema 4.3. Sea μ la medida cardinal en $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$. Consideremos la función $f(n) = 2^{-n}$, $n \in \mathbb{N}$. Calcular $\int f d\mu$.

Problema 4.4. Decidir si existen y calcular, en su caso, las siguientes integrales: $\int I_{\mathbb{Q}}(x)dx$, $\int_0^1 (e^x - 1)dx$, $\int_{\mathbb{Q}} x^2 dx$, $\int_{\mathbb{R}} |x| dx$, $\int_{]0,1[\cup]2,3[} x dx$, $\int_{\{0\}} e^x dx$, $\int_{\{0\}} e^x d\varepsilon_0(x)$.

Problema 4.5. Sean $p \in]0, 1[$, $\Omega = \{0, 1, \dots, n\}$, $\mathcal{A} = \mathscr{P}(\Omega)$ y P la probabilidad en $(\Omega, \mathcal{A})$ definida por

$$P(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n.$$

Calcular la esperanza de la aplicación identidad en Ω.

Problema 4.6. (Esperanza condicional respecto a un suceso) Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y A un suceso tal que $P(A) > 0$. Si Y es una v.a.r. con media finita, probar que

$$E_{P_A}(Y) = \frac{1}{P(A)} \int_A Y dP,$$

donde P_A es la probabilidad definida en el Problema 3.3. Se suele denotar $E(Y|A) := E_{P_A}(Y)$ y se llama media condicional de Y dado el suceso A. *Observación:* $E(Y|A)$ se interpreta como el valor medio de la variable Y en una larga serie de realizaciones del experimento aleatorio si solo tenemos en cuenta aquellas realizaciones en que ha ocurrido el suceso A.

Problema 4.7. (★) (Espacios $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$) Sean μ una medida en $(\Omega, \mathcal{A})$, X una v.a.r. en $(\Omega, \mathcal{A})$ y $1 \leq p < \infty$. Notar que X es μ-integrable si, y sólo si, $|X|$ es μ-integrable. X se dice p-integrable respecto a μ y se denota $X \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu; \mathbb{R})$ si $|X|^p$ es μ-integrable.

1. Probar que, si $X \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu; \mathbb{R})$ y μ es finita, entonces

$$X \in \mathcal{L}^q(\Omega, \mathcal{A}, \mu; \mathbb{R}), \forall 1 \leq q < p.$$

   Probar que eso no es cierto, en general, si μ no es finita.

2. (Desigualdad de Hölder) Si $0 < \alpha < 1$ y $a, b \geq 0$, entonces

$$a^\alpha b^{1-\alpha} \leq \alpha a + (1-\alpha) b.$$

3. (Desigualdad de Hölder) Si $1 < p < \infty$ y

$$f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu; \mathbb{R}) \text{ y } g \in \mathcal{L}^{p'}(\Omega, \mathcal{A}, \mu; \mathbb{R}),$$

   con $1/p + 1/p' = 1$, entonces $fg \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu; \mathbb{R})$ y

$$\int_\Omega |f(\omega)g(\omega)| d\mu(\omega) \leq \left(\int_\Omega |f(\omega)|^p d\mu(\omega) \right)^{1/p} \left(\int_\Omega |g(\omega)|^{p'} d\mu(\omega) \right)^{1/p'}.$$

   Observación. Si $p = 2$, entonces $p' = 2$ y la desigualdad de Hölder suele conocerse como desigualdad de Cauchy-Schwarz. En particular, si X e Y son dos v.a.r. de cuadrado integrables en un espacio de probabilidad, se verifica que $E(|XY|) \leq E(|X|^2)^{\frac{1}{2}} E(|Y|^2)^{\frac{1}{2}}$.

4. (Desigualdad de Minkowsky) Si $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu; \mathbb{R})$, denotaremos

$$\|f\|_p = \left(\int_\Omega |f(\omega)|^p d\mu(\omega) \right)^{1/p}.$$

   Probar que

$$\|f + g\|_p = \|f\|_p + \|g\|_p.$$

   En particular, si $f, g \in \mathcal{L}^p$, entonces $f + g \in \mathcal{L}^p$.

Problema 4.8. Consideremos un espacio de medida $(\Omega, \mathcal{A}, \mu)$. Sea $\{f_n\}_{n\geq 1}$ una sucesión de funciones medibles-borel tal que $|f_n| \leq a_n$, con $\sum_n a_n$ convergente, y sea g una función integrable. Probar que se verifica:

$$\int_\Omega \left(\sum_{n=1}^\infty f_n g \right) d\mu = \sum_{n=1}^\infty \int_\Omega f_n g d\mu.$$

Problema 4.9. (Media) Probar que, si X es una v.a.r. de cuadrado integrable en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y $\mu = E(X)$, entonces para cada $a \in \mathbb{R}$,

$$E((X-\mu)^2) \leq E((X-a)^2)$$

En particular, si $x_1, \dots, x_n \in \mathbb{R}$ y $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, entonces, para cada $a \in \mathbb{R}$,

$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq \sum_{i=1}^n (x_i - a)^2$$

Problema 4.10. (Mediana) Probar que, si X es una v.a.r. integrable en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y m es una mediana de X (es decir, un número real que verifica que $P(X < m) \leq \frac{1}{2} \leq P(X \leq m)$), entonces para cada $a \in \mathbb{R}$,

$$E(|X-\mu|) \leq E(|X-a|)$$

En particular, si $x_1, \dots, x_n \in \mathbb{R}$ y m es la mediana de ese conjunto de datos (es decir, m es cualquier número entre los dos valores centrales una vez ordenados los datos de menor a mayor en el caso de que n sea par, y el valor central en la sucesión ordenada si n es impar), entonces, para cada $a \in \mathbb{R}$,

$$\sum_{i=1}^n |x_i - m| \leq \sum_{i=1}^n |x_i - a|.$$

Indicación: Notar que, si $a > m$, entonces $|X-a| - |X-m| = m - a$ si $X \geq a$, $= m + a - 2X$ si $m < X < a$, $= a - m$ si $X \leq m$, y, por tanto,

$$\begin{aligned} E(|X-a|) - E(|X-m|) \geq & (m-a)P(X \geq a) + (m-a)P(m < X < a) \\ & + (a-m)P(X \leq m) \\ \geq & (a-m)[P(X \leq m) - P(X > m)] \geq 0. \end{aligned}$$

Proceder de forma análoga en el caso $a < m$.

Capítulo 5

Suma de Medidas. Medida Imagen. Distribuciones de Probabilidad

En esta sección se presentan diversos procedimientos para construir nuevas medidas a partir de otras conocidas que son especialmente útiles en el desarrollo del Cálculo de Probabilidades y de la Estadística Matemática. Presentaremos también resultados que nos indican cómo se integra respecto a esas nuevas medidas si sabemos integrar respecto a las originales.

El primero de esos procedimientos es la suma de medidas.

TEOREMA 5.1. *Sean $(\Omega, \mathcal{A})$ un espacio medible, (a_n) una sucesión de números reales positivos y (μ_n) una sucesión de medidas en $(\Omega, \mathcal{A})$. La función de conjuntos $\mu = \sum_n a_n \mu_n$ es una medida en $\mathcal{A}$, y, para cada v.a.r. no negativa f se verifica que*

$$\int f d\mu = \sum_n a_n \int f d\mu_n.$$

Demostración. Es claro que μ es una función de conjuntos, no negativa y nula en el vacío. Además, si (A_m) es una sucesión disjunta en $\mathcal{A}$, entonces

$$\mu(\cup_m A_m) = \sum_n a_n \sum_m \mu_n(A_m) = \sum_m \sum_n a_n \mu_n(A_m).$$

Por otra parte, la igualdad

$$\int f d\mu = \sum_n a_n \int f d\mu_n$$

es cierta para indicadores por definición de μ, y para funciones simples medibles no negativas por linealidad de la integral. En el caso no negativo podemos escribir f como límite puntual de una sucesión (f_m) creciente de funciones simples, medibles y no negativas. Entonces, tras una aplicación juiciosa del teorema de la convergencia monótona (Teorema 4.1), se tiene que

$$\int f d\mu = \lim_m \int f_m d\mu = \lim_m \sum_n a_n \int f_m d\mu_n = \sum_n a_n \lim_m \int f_m d\mu_n = \sum_n a_n \int f d\mu_n.$$

El caso integrable se reduce a éste descomponiendo previamente f como diferencia de sus partes positiva y negativa. □

Otro procedimiento para construir nuevas medidas consiste en transportar una medida por una v.a. A lo largo de toda esta sección, $(\Omega, \mathcal{A})$ será un espacio medible.

Proposición 5.1. (Medida imagen. Distribución de probabilidad de una v.a.) *Sean $(\Omega, \mathcal{A}, \mu)$ un espacio de medida, $(\Omega', \mathcal{A}')$ un espacio medible y $X : \Omega \longrightarrow \Omega'$ una función medible. La función de conjuntos μ^X definida en $\mathcal{A}'$ por $\mu^X(A') = \mu(X^{-1}(A'))$ es una medida en $\mathcal{A}'$ que llamaremos medida imagen de μ por X. $(\Omega', \mathcal{A}', \mu^X)$ se llamará espacio de medida imagen de X. Si P es una probabilidad en $(\Omega, \mathcal{A})$ y $X : \Omega \longrightarrow \Omega'$ es una v.a., la medida imagen de P por X, que denotamos por P^X, es una probabilidad en $\mathcal{A}'$ que llamaremos distribución de probabilidad de X (respecto a P).*

Demostración. Es claro que μ^X es una función de conjuntos no negativa. Puesto que la contraimagen del vacío es el vacío se verifica que $\mu^X(\emptyset) = 0$. Además, la contraimagen de la unión es la unión de las contraimágenes, y las contraimágenes de conjuntos disjuntos son disjuntas. Por tanto, μ^X es numerablemente aditiva. □

***Observación* 5.1.** Hagamos notar aquí que todo el valor probabilístico de una v.a. queda determinado por su distribución; dicho de otro modo, dos v.a. idénticamente distribuidas son probabilísticamente equivalentes.

***Observación* 5.2.** Dada una v.a.r. X en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se define su función de distribución como la aplicación $F : x \in \mathbb{R} \longrightarrow F(x) = P(X \leq x) = P^X(]-\infty, x]) \in [0,1]$. Siendo P^Xmonótona, es claro que F es creciente; por otra parte, F es continua por la derecha en cada punto $x \in \mathbb{R}$ (si (x_n) es una sucesión decreciente convergente a x, entonces $]0,x] = \cap_n]0, x_n]$ y el resultado en consecuencia inmediata de la Proposición 3.1 (c)). F es, pues la función de distribución asociada a la medida de Lebesgue-Stieljes P^X (ver la Observación 3.5). Es sencillo asimismo probar que $0 = F(-\infty) := \text{lím}_{x\to-\infty F(x)}$ y $1 = F(+\infty) := \text{lím}_{x\to+\infty F(x)}$. Es claro que F queda unívocamente determinada por su distribución P^X; como aplicación del teorema de Dynkin (Teorema 1.1) se prueba también que dos v.a.r. con la misma función de distribución tienen la misma distribución.

El teorema siguiente establece la relación entre las integrales respecto a una medida y su imagen por una cierta función medible.

TEOREMA 5.2. (Teorema de la medida imagen) *Conservamos las hipótesis y notaciones de la proposición anterior. Si $g : \Omega' \longrightarrow \bar{\mathbb{R}}$ es una función Borel-medible, entonces $\int_{\Omega'} g d\mu^X = \int_\Omega (g \circ X) d\mu$, en el sentido de que, si una de esas dos integrales existe, entonces la otra también existe y ambas coinciden.*

Demostración. Si g es un indicador, digamos $g = I_{A'}$ con $A' \in \mathcal{A}'$, entonces $\int_{\Omega'} g d\mu^X = \mu^X(A') = \mu(X^{-1}(A')) = \int I_{X^{-1}(A')} d\mu = \int I_{A'} \circ X d\mu$, pues $I_{X^{-1}(A')} = I_{A'} \circ X$. El resultado se extiende por linealidad a funciones simples medibles no negativas. El teorema de la convergencia monótona (Teorema 4.1) nos permite extender el resultado a funciones simples medibles no negativas. En el caso general descomponemos g como diferencia de sus partes positiva y negativa y el teorema de aditividad de la integral finaliza la prueba. □

***Ejemplo* 5.1.** Si X es una v.a.r. en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, entonces

$$E(X) = \int_\Omega X dP = \int_\mathbb{R} x dP^X(x),$$

y, si X es de cuadrado integrable,

$$\text{Var}(X) = E[(X - EX)^2] = \int_\Omega [X(\omega) - EX]^2 dP(\omega) = \int_\mathbb{R} (x - EX)^2 dP^X(x).$$

***Ejemplo* 5.2.** Reemplazando g por $gI_{A'}$ en el enunciado se obtiene

$$\int_{A'} g d\mu^T = \int_{T^{-1}(A')} (g \circ T) d\mu, \quad \forall A' \in \mathcal{A}'.$$

***Ejemplo* 5.3.** Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, $X : (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$ y $f : (\Omega', \mathcal{A}') \to \mathbb{R}$ una v.a.r. Entonces $E_{P^X}(f) = E_P(f \circ X)$.

DEFINICIÓN 5.1. (Momentos de la distribución de una v.a.r.) Sean $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ una v.a.r. y $k > 0$.

(a) A $E(X^k)$ (si existe) se le llama momento de orden k de (la distribución de) X. La media es el momento de orden 1 de X.

(b) $E(|X|^k)$ se llamará momento absoluto de orden k de X.

(c) $E[(X - E(X))^k]$ (si existe) se llamará momento central de orden k de X. La varianza de X es el momento central de orden 2 de X.

Proposición 5.2. (a) *Si $k > 0$ y X tiene momento finito de orden k, entonces tiene también momento finito de orden j, para cada $0 < j \leq k$.*

(b) *Si X tiene momento finito de orden $n - 1$ y existe el momento de orden n, entonces*

$$E[(X - E(X))^n] = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} E(X^k) E(X)^{n-k}.$$

Demostración. (a)

$$E(|X|^j) = \int_{\{|X|<1\}} |X|^j dP + \int_{\{|X|\geq 1\}} |X|^j dP \leq P(|X| < 1) + \int_{\{|X|\geq 1\}} |X|^k dP < \infty$$

(b) Trivial. □

Proposición 5.3. (a) (Desigualdad de Markov) *Si X es una v.a.r. y $k, \alpha > 0$, entonces*

$$P(|X| \geq \alpha) \leq \alpha^{-k} E(|X|^k).$$

(b) (Desigualdad de Chebyshev) *Sean X una v.a.r. con media finita μ y varianza finita σ^2, y $k > 0$. Entonces*

$$P(|X - \mu| \geq k\sigma) \leq k^{-2}.$$

Demostración. (a) $\alpha^k P(|X| \geq \alpha) = \int_{\{|X|\geq\alpha\}} \alpha^k dP \leq \int_{\{|X|\geq\alpha\}} |X|^k dP \leq E(|X|^k)$.

(b) Es consecuencia inmediata de (a). $\square$

DEFINICIÓN 5.2. (Covarianza, coeficiente de correlación) Si $X, Y \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu)$ se define $\mathrm{Cov}(X,Y) = E[(X - EX)(Y - EY)]$. El coeficiente de correlación entre X e Y se define como $\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$.

***Observación* 5.3.** La desigualdad de Cauchy-Schwarz prueba que la covarianza está bien definida. Es sencillo probar que $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$. Por otra parte, esa misma desigualdad prueba que $-1 \leq \rho(X,Y) \leq 1$.

PROBLEMAS DEL CAPÍTULO 5

Problema 5.1. Haciendo $T(x) = -x$, $x \in \mathbb{R}$, probar que una función medible $g : \mathbb{R} \longrightarrow \mathbb{R}$ es integrable (respecto a la medida de Lebesgue) si, y sólo si, la función $h(x) = g(-x)$ lo es. Análogamente, si $c \in \mathbb{R}$, g es integrable si, y sólo si, $g(x+c)$ lo es. Se verifica además que

$$\int g(x)dx = \int g(-x)dx = \int g(x+c)dx.$$

Problema 5.2. Determinar la distribución de una v.a. respecto a una medida de Dirac.

Problema 5.3. Sean $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{A} = \mathscr{P}(\Omega)$ y $P = \sum_{i=1}^{6} p_i \varepsilon_i$ (donde $p_i \geq 0$ y $\sum_i p_i = 1$). Consideremos la v.a.r T definida por $T(1) = T(2) = T(3) = -1$, $T(4) = T(5) = 0$ y $T(6) = 1$. Determinar P^T, $E(T)$ y $E_{P^T}(f)$, donde $f(x) = x^2+x+1$.

Problema 5.4. Sea $P = \frac{1}{6}\sum_{k=1}^{6} \varepsilon_k$. Probar que P es una probabilidad en $\mathcal{R}$. Calcular

$$P(\mathbb{N}),\ \ P([-2,2]\setminus]-\infty,1]),\ \ P(\sqrt{2}+\mathbb{N}) \text{ y } P(\mathbb{Z}\triangle[-3,3]).$$

Problema 5.5. (★) Sean X una variable aleatoria real en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y y $g : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ una función creciente no negativa tal que $E(g(X)) < \infty$. Supongamos que g es una función par y no decreciente en $[0,\infty)$. Entonces para cada $\eta > 0$,

$$P(|X| \geq \eta) \leq \frac{E(|g(X)|)}{g(\eta)}.$$

Nota: Este resultado generaliza la desigualdad de Chebyshev.

Problema 5.6. Sea $\lambda = 1/2$ y consideremos en $(\mathbb{N}_0, \mathscr{P}(\mathbb{N}_0))$ la probabilidad $P = e^{-\lambda}\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}\varepsilon_n$. Consideremos las funciones $f(n) = n$ y $g(n) = n!$, $n \in \mathbb{N}_0$. Calcular $E(f)$ y $E(g)$.

Problema 5.7. Supongamos que el número de veces por semana que un individuo aparca en doble fila, sigue una distribución de Poisson de parámetro λ. Si la probabilidad que tiene de ser denunciado por cada aparcamiento indebido es p (y admitimos que no recibe denuncias injustificadas) determinar:

1. La probabilidad de que haya aparcado en doble fila exactamente k veces una semana en la que recibió n denuncias.

2. La probabilidad de que en una semana comete k infracciones no denunciadas.

3. La probabilidad de que la j–ésima denuncia se produzca en la k–ésima infracción (en una semana cualquiera).

Problema 5.8. En cierta votación entre dos candidatos, el candidato A recibe a votos y el candidato B recibe b votos ($a > b$). Supongamos que n de los $N = a + b$ votos totales son irregulares (no válidos). El candidato B cuestiona entonces el resultado de la votación, reclamando que si esos n votos se hubiesen sacado al azar del total y hubieran sido anulados, la votación podría variar. ¿Cuál sería la probabilidad de que se produzca una variación en el resultado?

Problema 5.9. (Distribución uniforme discreta) La distribución uniforme discreta en el conjunto $\{1, \dots, n\}$ es la distribución $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$. Determinar su media y su varianza.([1])

Problema 5.10. (Combinatoria) Recordemos que, si $n \in \mathbb{N}$ y $k \in \{0, 1, \dots, n\}$ se define el número combinatorio n sobre k mediante:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

donde $n! = n \cdot (n-1) \dots 3 \cdot 2 \cdot 1$ (se lee n factorial y es el número de permutaciones en el conjunto $\{1, \dots, n\}$, es decir, el número de aplicaciones biyectivas de $\{1, \dots, n\}$ en sí mismo). Por convenio, 0!=1!=1 y $\binom{n}{0} = 1$. El número combinatorio $\binom{n}{k}$ es el número de combinaciones de n elementos tomados de k en k, siendo dos combinaciones diferentes cuando una de ellas contiene un elemento que no contiene la otra, es decir, $\binom{n}{k}$ es el número de subconjuntos de $\{1, \dots, n\}$ que tienen k elementos. Es obvio que $\binom{n}{k} = \binom{n}{n-k}$. Si $k, n \in \mathbb{N}$, los elementos del conjunto producto $\{1, \dots, n\}^k$ se llaman también variaciones con repetición de n elemento tomados de k en k; hay un total de n^k variaciones con repetición de n elementos tomados de k en k. Si, además, $k \le n$, las aplicaciones inyectivas de $\{1, \dots, k\}$ en $\{1, \dots, n\}$ se llaman variaciones ordinarias (o sin repetición) de n elementos tomados de k en k. Su número es $n \cdot (n-1) \cdot \dots \cdot (n-k+1) = n!/(n-k)!$ ([2]) En una baraja española de 40 cartas agrupadas en cuatro "palos" (oros, copas, espadas y bastos) con 10 cartas en cada

[1] La distribución uniforme discreta corresponde a la definición clásica de probabilidad de Laplace en el caso de un número finito de sucesos elementales -digamos, 1,. . . ,n- equiprobables -en cuyo caso cada uno tiene probabilidad $\frac{1}{n}$-. En ese contexto, la probabilidad de un suceso es el cociente entre el *número de casos favorables al suceso* -es decir, el número de sucesos elementales que componen el suceso- y el *número de casos posibles* -es decir, n-.

[2] El número de aplicaciones biyectivas (o permutaciones) del conjunto $\{1, 2, 3, 4\}$ en sí mismo es igual a $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$. Un ejemplo de permutación en ese conjunto es $(3, 1, 4, 2)$, forma abreviada esta de describir la aplicación biyectiva que lleva el 1 en el 3 (que aparece en primer lugar en la permutación), el 2 en el 1 (que aparece en segundo lugar en la permutación), el 3 en el 4 y el 4 en el 2. Esa permutación se puede escribir también en la forma 3142, identificándose así con un número de cuatro cifras. Podemos, por tanto, afirmar que con las cifras 1, 2, 3 y 4 se pueden escribir $4! = 24$ números de cuatro cifras diferentes. Como vemos, en cada permutación en un conjunto se utiliza cada elemento de este una y sólo una vez, y una permutación se diferencia de otra en el orden en que aparecen colocados los elementos. Por otra parte, el conjunto $\{1, 2, 3, 4\}$ tiene $\binom{4}{2} = 6$ subconjuntos con 2 elementos: $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$ y $\{3, 4\}$, que también se escriben en la forma

palo numeradas del 1 al 10, calcular, *utilizando la definición clásica de probabilidad de Laplace -no de otro modo-,* las siguientes probabilidades:

1. La probabilidad de que al barajar los 10 oros, éstos salgan en orden creciente.
2. La probabilidad de sacar cuatro ases en 4 extracciones de cartas con reemplazamiento (es decir, tras cada extracción, se toma nota de la carta obtenida y ésta es devuelta a la baraja) si se baraja después de cada extracción.
3. La misma probabilidad del apartado anterior en el caso de que el experimento se haga sin reemplazamiento.
4. La probabilidad de que en 3 extracciones sin reemplazamiento se obtenga el as de oros.

Problema 5.11. a) Se lanzan cuatro dados equilibrados. ¿Cuál es la probabilidad de obtener, al menos, un 6?

b) Se reúnen n personas en un local. ¿Cuál es la probabilidad de que al menos dos de ellas cumplan años el mismo día?

Problema 5.12. (Distribución uniforme continua) Si $a < b$, la distribución uniforme en el intervalo $[a, b]$ queda definida por la función de densidad

$$\frac{1}{b-a} I_{[a,b]}.$$

Determinar su media y su varianza. *Observación:* En general, si B es un boreliano de $\mathbb{R}^k$ con medida de Lebesgue $\lambda_k(B)$ finita y > 0, la distribución uniforme (multivariante, si $k > 1$) sobre B queda definida por la función de densidad $\lambda_k(b)^{-1} I_B$.

Problema 5.13. Un móvil se desplaza en un plano cartesiano realizando dos posibles desplazamientos: con probabilidad p se desplaza del punto (x, y) al punto $(x, y + 1)$ y con probabilidad $1 - p$ se desplaza del punto (x, y) al punto $(x + 1, y)$, siendo $0 < p < 1$. Supuesto que inicialmente se encuentra en el punto $(0, 0)$, determinar justificando la respuesta:

12, 13, 14, 23, 24 y 34. Se dice que hay 6 combinaciones de 4 elementos tomados de 2 en 2. Por ejemplo, hay 6 formas distintas de asignar dos pelotas idénticas a 4 niños si cada niño recibe a lo sumo una pelota. Así, en una combinación en un conjunto no pueden aparecer elementos repetidos, y el orden no influye en la combinación, de tal suerte que dos combinaciones son diferentes cuando una de ellas contiene un elemento que no contiene la otra. En una variación con repetición se usan k de los números $1, \ldots, n$ pudiendo aparecer un mismo número más de una vez e influyendo el orden de aparición, mientras que en las variaciones ordinarias de orden $k \leq n$ se usan k de los números $1, \ldots, n$ sin que puedan repetirse pero manteniendo la influencia del orden de aparición -dicho de otro modo, las variaciones ordinarias serían los elementos de $\{1, \ldots, n\}^k$ con coordenadas dos a dos distintas-.

1. La probabilidad de que alcance el punto (a, b).

2. La probabilidad de que alcance el segmento determinado por los puntos $(n, 0)$ y (n, n).

Problema 5.14. Se elige de forma aleatoria un punto en el círculo unidad. Calcular la distribución de probabilidad de la ordenada del punto.

Capítulo 6

Medidas Definidas por Densidades. Teorema de Cambio de Variables

El tercer procedimiento para construir nuevas medidas a partir de otras conocidas hace uso del concepto de densidad. Las distribuciones de probabilidad más utilizadas en Estadística Matemática quedan determinadas por su densidad respecto a una cierta medida: la medida cardinal en el caso discreto y la medida de Lebesgue en $\mathbb{R}$ o en $\mathbb{R}^n$ en el caso continuo. Los resultados siguientes permiten con frecuencia obtener resultados generales que se aplican a las tres situaciones mencionadas (evitando así la triplicación de las demostraciones).

DEFINICIÓN 6.1. Sean λ y μ medidas en $\mathcal{A}$. Diremos que λ es absolutamente continua respecto a μ o que λ está dominada por μ (o que μ domina a λ, o que μ es una medida dominante de λ), y lo denotaremos por $\lambda \ll \mu$, si $\lambda(A) = 0$ cada vez que $A \in \mathcal{A}$ y $\mu(A) = 0$.

La siguiente proposición-definición introduce la definición de densidad o derivada de Radon-Nikodym.

TEOREMA 6.1. (Medidas definidas por densidades) *Sean μ una medida en $\mathcal{A}$ y $h : (\Omega, \mathcal{A}) \longrightarrow [0, +\infty]$ una función medible. La función de conjuntos $\lambda : A \in \mathcal{A} \longrightarrow$*

$\lambda(A) = \int_A h d\mu$ es una medida absolutamente continua respecto a μ. λ suele llamarse integral indefinida de h respecto a μ; a h la llamaremos derivada de Radon-Nikodym o densidad de λ respecto a μ y denotaremos $h = d\lambda/d\mu$. Además, si $f : (\Omega, \mathcal{A}) \longrightarrow \bar{\mathbb{R}}$ es una función medible, entonces

$$\int f\, d\lambda = \int f \cdot h\, d\mu,$$

en el sentido de que una de esas integrales existe si, y sólo si, existe la otra, y en ese caso coinciden.

Demostración. La primera parte ha sido probada en el Teorema 4.2. La segunda es cierta en el caso de que f sea un indicador, por definición; a partir de ahí, el resultado se extiende a funciones simples medibles no negativas por linealidad de la integral, y a funciones medibles no negativas haciendo uso del Teorema 2.1 y del teorema de la convergencia monótona de Lebesgue, y de ahí al caso general mediante la habitual descomposición de $f$como diferencia de sus partes positiva y negativa. □

***Ejemplo* 6.1.** (Integral respecto a una medida de Dirac, es decir, esperanza de una v.a.r. respecto a la distribución de probabilidad degenerada en un punto) Sean $(\Omega, \mathcal{A})$ un espacio medible, $f : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}$ una v.a.r. y ω_0 un punto de Ω tal que $\{\omega_0\} \in \mathcal{A}$. Puesto que $f = f(\omega_0)\ \varepsilon_{\omega_0} - c.s.$, se verifica que

$$\int f d\varepsilon_{\omega_0} = \int f(\omega_0) d\varepsilon_{\omega_0} = f(\omega_0).$$

***Ejemplo* 6.2.** (Integral respecto a la medida cardinal y esperanza respecto a una distribución de probabilidad discreta) Consideremos el espacio discreto $(\Omega, \mathcal{A}) = (\mathbb{N}, \mathscr{P}(\mathbb{N}))$, y denotemos por μ la medida cardinal sobre $\mathcal{A}$, es decir, $\mu = \sum_{n=1}^{\infty} \varepsilon_n$. Cualquier función real f en $\mathbb{N}$ es $\mathcal{A}$-medible. Si f es no negativa, se verifica que

$$\int f d\mu = \int f d(\textstyle\sum_n \varepsilon_n) = \sum_n \int f d\varepsilon_n = \sum_n f(n).$$

En el caso de una función $\overline{\mathbb{R}}$-valorada f, que f sea μ-integrable equivale a que lo sea $|f|$, y, por tanto, f es integrable respecto a la medida cardinal si, y sólo si, la serie $\sum_n f(n)$ es absolutamente convergente. Análogamente se obtiene la integral respecto a la medida cardinal en un conjunto numerable provisto de la σ-álgebra de todos sus subconjuntos.

***Ejemplo* 6.3.** (Esperanza respecto a una distribución de probabilidad discreta) Sean $\{x_1, x_2, \dots\}$ un subconjunto numerable de $\mathbb{R}$, $p_1, p_2, \dots \geq 0$ y $P = \sum_{n\geq 1} p_n \varepsilon_{x_n}$. P será una probabilidad si $\sum_n p_n = 1$. Si $X : \mathbb{R} \to \mathbb{R}$ es una v.a.r. no negativa, entonces

$$E_P(X) = \sum_n p_n X(x_n). \quad \text{(I)}$$

Si X toma valores en $\overline{\mathbb{R}}$, que X tenga media finita equivale a que la serie $\sum_n p_n X(x_n)$ sea absolutamente convergente, en cuyo caso también se verifica (I). En particular, la media de la distribución P se obtiene cuando X es la identidad

$$\sum_n p_n x_n,$$

y no es más que una suma ponderada de los valores x_i siendo el "peso"de x_i la probabilidad p_i con que P toma el valor x_i. ([1])

***Ejemplo* 6.4.** (Integrales de Riemann y de Lebesgue) Sean $X : [a, b] \to \mathbb{R}$ una aplicación acotada y $a = a_0 < a_1 < \dots < a_n = b$ una partición del intervalo $[a, b]$, que denotaremos por $\pi :=< a_0, a_1, \dots, a_n >$. Denotemos $m_i = \inf_{x\in[a_{i-1},a_i]} X(x)$ y $M_i = \sup_{x\in[a_{i-1},a_i]} X(x)$, $1 \leq i \leq n$. Se definen las sumas inferior y superior de Riemann de X relativas a la partición π mediante

$$l(X, \pi) := \sum_{i=1}^n m_i(a_i - a_{i-1}), \quad u(X, \pi) := \sum_{i=1}^n M_i(a_i - a_{i-1}),$$

respectivamente. Si $\pi' :=< a = b_0, b_1, \dots, b_m = b >$ es otra partición de $[a, b]$ y $\{a_0, a_1, \dots, a_n\} \subset \{b_0, b_1, \dots, b_m\}$, diremos que π' es más fina que π. En ese caso, es sencillo probar que

$$l(X, \pi) \leq l(X, \pi') \leq u(X, \pi') \leq u(X, \pi).$$

X se dice Riemann-integrable en $[a, b]$ si $\sup_\pi l(X, \pi) = \inf_\pi u(X, \pi)$; en ese caso, ese valor común se denota $\int_a^b X$ y se llama integral de Riemann de X en el intervalo $[a, b]$. Se prueba fácilmente (ver Problema 6.5) que una función acotada $X : [a, b] \to \mathbb{R}$ es

[1] La expresión “peso” queda justificada por la siguiente interpretación física de esa esperanza: Consideremos un cuerpo de masa 1 formado por n masas puntuales $p_1, p_2, \dots$ colocadas, respectivamente, en los puntos de abscisa $x_1, x_2, \dots$. El centro de gravedad del cuerpo -es decir, el punto sobre el que debemos apoyarlo para que quede en equilibrio- se sitúa en el punto $\sum_n p_n x_n$. Análogamente, si $f : \mathbb{R} \longrightarrow [0, +\infty[$ es la función de densidad de una distribución de probabilidad en $\mathbb{R}$, podemos imaginar una barra metálica de longitud infinita y grosor nulo situada sobre el eje de abscisas y de masa 1 desigualmente repartida a lo largo de la barra: es la función de densidad la que nos indica cómo se reparte la masa total a lo largo de la barra, cómo de “densa” es ésta en cada segmento de barra (el área que sobre el segmento hay bajo la gráfica de la función de densidad).

Riemann-integrable si, y sólo si, para cada $\epsilon > 0$, existe una partición π tal que $u(X,\pi) - l(X,\pi) < \epsilon$. Puesto que una función real continua en un intervalo compacto de la recta es uniformemente continua, se prueba fácilmente que toda función continua en $[a,b]$ es Riemann-integrable. Se prueba que una función acotada X en $[a,b]$ es Riemann-integrable si, y sólo si, el conjunto de puntos de discontinuidad de X tiene medida de Lebesgue nula, y que, en ese caso, la función es, también, Lebesgue-integrable y que las integrales de Riemann y de Lebesgue de X en $[a,b]$ coinciden. Podemos usar, pues los resultado clásicos de cálculo de integrales definidas para calcular integrales de Lebesgue.([2]) Pero la integral de Lebesgue nos permite integrar más funciones: por ejemplo, el indicador de $\mathbb{Q} \cap [0,1]$ no es Riemann-integrable en $[0,1]$ (no es continua en ningún punto de ese intervalo), pero tiene integral de Lebesgue nula, pues coincide c.s. con la función idénticamente nula.

Algunos autores denominan teorema de cambio de variables el que nosotros hemos denominado de la medida imagen; aquí reservaremos ese nombre para el teorema siguiente, que no demostraremos (puede verse la demostración en Cohn (1980)).

TEOREMA 6.2. (Cambio de variables) *Sean U y V abiertos de $\mathbb{R}^n$ y $T : U \longrightarrow V$ una biyección tal que T y T^{-1} son diferenciables. Denotemos por λ la medida de Lebesgue en $\mathbb{R}^n$ y sea*

$$J_T(x) = \det\left(\frac{\partial T_i}{\partial x_j}(x)\right)_{1\le i,j\le n}$$

el jacobiano de T en el punto $x \in U$. Para cada boreliano B de U se verifica

$$\lambda(T(B)) = \int_B |J_T(x)| d\lambda(x).$$

Para cada función Borel medible $f : V \longrightarrow \mathbb{R}$ se verifica que

$$\int_V f d\lambda = \int_U (f \circ T)(x)|J_T(x)|d\lambda(x)$$

en el sentido de que, si una de esas integrales existe, entonces ambas existen y coinciden.

[2] Recordemos la regla de Barrow: Sean X una función acotada y Riemann-integrable en $[a,b]$ y f una función derivable en $[a,b]$ tal que $X = f'$ (es decir, f es una primitiva de X); entonces, $\int_a^b X(x)dx = f(b) - f(a)$. Recordemos también que, si X es continua en un punto c, la aplicación $f(x) := \int_a^x X$ es derivable en el punto c y $f'(c) = X(c)$.

Como consecuencia inmediata del teorema de cambio de variables (Teorema 6.2) se tiene el siguiente resultado, que describe la densidad de una v.a. n-dimensional transformada por un difeomorfismo (es decir, una aplicación biyectiva que es de clase C^1 y tiene inversa de clase C^1) de otra v.a. n-dimensional con densidad.

Corolario 6.1. (Transformación de variables) *Sean U y V abiertos de $\mathbb{R}^n$, $T : U \to V$ una biyección tal que T y T^{-1} son de clase C^1, y $X : (\Omega, \mathcal{A}, P) \to U$ una v.a. n-dimensional con densidad f_X (respecto a la medida de Lebesgue λ en $\mathbb{R}^n$). Entonces la v.a. $Y := T \circ X$ tiene densidad*

$$f_Y(y) = \frac{dP^{T\circ X}}{d\lambda}(y) = f(T^{-1}(y))|J_{T^{-1}}(y)|, \quad y \in V.$$

Demostración. Si $B \in \mathcal{R}^n(V)$, entonces

$$P^{T\circ X}(B) = P(T \circ X \in B) = P(X \in T^{-1}B) = \int_{T^{-1}B} f(x)dx = (*).$$

Haciendo el cambio $x = T^{-1}(y)$, se deduce del teorema de cambio de variables (Teorema 6.2) que

$$(*) = \int_B f(T^{-1}(y))|J_{T^{-1}}(y)|dy,$$

lo que acaba la prueba. □

***Observación* 6.1.** Con las notaciones e hipótesis del teorema anterior, se tiene que, para cada boreliano B de V,

$$\lambda^T(B) = \int_{T^{-1}B} dx = \int_B |J_{T^{-1}}(y)|dy,$$

lo que prueba que la medida imagen λ^T tiene densidad $|J_{T^{-1}}|$ respecto a λ.

PROBLEMAS DEL CAPÍTULO 6

Problema 6.1. En los apartados siguientes, decidir si existe la derivada de Radon-Nikodym de λ respecto a μ y, en su caso, calcularla.

1. $\Omega = \{1,2,3\}$, $\mathcal{A} = \mathscr{P}(\Omega)$, $\lambda = \frac{1}{2}(\varepsilon_1 + \varepsilon_2)$ y $\mu = \frac{1}{2}(\varepsilon_1 + \varepsilon_3)$.
2. $\Omega = \{1,2,3\}$, $\mathcal{A} = \mathscr{P}(\Omega)$, $\lambda = \frac{1}{2}(\varepsilon_1 + \varepsilon_2)$ y $\mu = \frac{1}{3}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)$.
3. $\Omega = [0,1]$, $\mathcal{A} = \mathcal{R}([0,1])$, $\lambda(A) = \int_A 3x^2 dx$ si $A \in \mathcal{A}$ y μ la distribución uniforme en $[0,1]$.
4. $\Omega = [0,1]$, $\mathcal{A} = \mathcal{R}([0,1])$, $\lambda(A) = \int_A 3x^2 dx$ y $\mu(A) = \int_A e^x dx$, si $A \in \mathcal{A}$.

Problema 6.2. Sean Z una variable aleatoria con función de densidad:

$$f(z) = \begin{cases} z+1 & \text{si} \quad z \in (-1,0) \\ -z+1 & \text{si} \quad z \in [0,1) \\ 0 & \text{en otro caso} \end{cases}$$

Determinar, la función de densidad de las variables aleatorias S y T donde

$$S = Z^2 + 3 \ \text{ y } \ T = e^{Z+2}.$$

Problema 6.3. Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} 0 & \text{si} \quad x \le 0 \\ e^{-x} & \text{si} \quad x > 0 \end{cases}$$

1. Determinar la función de probabilidad de la variable aleatoria $Y = [X]$ (parte entera de X)
2. Determinar la función de densidad de $Z = \dfrac{X}{1+X}$

Problema 6.4. (Distribución de Cauchy) Probar que, si $\alpha > 0$,

$$c_\alpha(x) := \frac{\alpha}{\pi(\alpha^2 + x^2)}$$

es una función continua no negativa tal que $\int_{\mathbb{R}} c_\alpha = 1$. Es la función de densidad de una distribución de probabilidad en $\mathbb{R}$ que llamaremos distribución de Cauchy de parámetro α. Probar que esa distribución no tiene media finita.

Problema 6.5. (★) (Integrales de Riemann y de Lebesgue) En este problema se pretenden justificar las afirmaciones del ejemplo 4 de la página 69. Con las definiciones y notaciones allí utilizadas, probar:

1. $l(X,\pi) \leq l(X,\pi') \leq u(X,\pi') \leq u(X,\pi)$ si π' es una partición de $[a,b]$ más fina que π.

   Indicación: Proceder por inducción, añadiendo los puntos extra uno a uno a la partición π.

2. Una función acotada $X : [a,b] \to \mathbb{R}$ es Riemann-integrable si, y sólo si, para cada $\epsilon > 0$, existe una partición π tal que $u(X,\pi) - l(X,\pi) < \epsilon$.

3. Toda función continua en $[a,b]$ es Riemann-integrable.

   Indicación: Utílicese que toda función continua en un compacto es uniformemente continua.

4. Una función acotada X en $[a,b]$ es Riemann-integrable si, y sólo si, el conjunto de puntos de discontinuidad de X tiene medida de Lebesgue nula. En ese caso, las integrales de Riemann y de Lebesgue de X en $[a,b]$ coinciden.

   Indicación: ($\Longrightarrow$) Para cada n, elegir una partición π_n (más fina que π_{n-1}) tal que $u(X,\pi_n) - l(X,\pi_n) < 1/n$. Definir funciones simples s_n y t_n en $[a,b]$ que coinciden con X en el punto a y valen $m_i := \text{ínf}_{]a_{i-1},a_i]} X$ y $M_i := \sup_{]a_{i-1},a_i]} X$, resp., si $\pi_n =< a_0, a_1, \dots, a_{k_n} >$. Entonces $s_n \leq X \leq t_n$ e $\int_a^b s_n = l(X,\pi_n)$ e $\int_a^b t_n = u(X,\pi_n)$. (s_n) es una sucesión creciente a una función s y (t_n) es una sucesión decreciente a una función t. Usar el teorema de la convergencia dominada de Lebesgue para probar que $\int_a^b (t-s) d\lambda = 0$; siendo $t \geq s$, debe ser $t = s$ c.s. en $[a,b]$. Si $s(x) = t(x)$ y x no está en ninguna de las particiones π_n, X es continua en x: luego el conjunto de puntos de discontinuidad de X tiene medida de Lebesgue nula. Además $s \leq X \leq t$ implica que $X = s$ c.s. y, por tanto, X es Lebesgue-medible y Lebesgue-integrable y su integral de Lebesgue coincide con la de s, es decir, con su integral de Riemann. ($\Longleftarrow$) Sea X continua c.s. Sea π_n una partición que divide $[a,b]$ en 2^n intervalos de igual longitud y construyamos a partir de ella funciones s_n y t_n como antes. Entonces $X(x) = \text{lím}_n s_n(x) = \text{lím}_n t_n(x)$ en cada punto x de continuidad de x y, por tanto, c.s. Por el teorema de la convergencia dominada, $\text{lím}_n(u(X,\pi_n) - l(X,\pi_n)) = 0$, que prueba que X es Riemann-integrable.

5. (Regla de Barrow) Sean X una función acotada y Riemann-integrable en $[a,b]$ y f una función derivable en $[a,b]$ tal que $X = f'$ (es decir, f es una primitiva de X); entonces, $\int_a^b X(x)dx = f(b) - f(a)$.

   Indicación: Sean $\pi =< a_0, a_1, \dots, a_n >$ una partición de $[a,b]$, $m_i = \text{ínf}_{[a_{i-1},a_i]} X$ y $M_i = \sup_{[a_{i-1},a_i]} X$. Por el teorema de los incrementos finitos, para cada i, existe $t_i \in]a_{i-1}, a_i[$ tal que

   $$f(a_i) - f(a_{i-1}) = f'(t_i)(a_i - a_{i-1}) = X(t_i)(a_i - a_{i-1}).$$

Entonces

$$\begin{aligned} l(X,\pi) &= \sum_{i=1}^n m_i(a_i - a_{i-1}) \le \sum_{i=1}^n X(t_i)(a_i - a_{i-1}) \\ &\le \sum_{i=1}^n M_i(a_i - a_{i-1}) = u(X,\pi). \end{aligned}$$

Pero

$$\sum_{i=1}^n X(t_i)(a_i - a_{i-1}) = \sum_{i=1}^n (f(a_i) - f(a_{i-1})) = f(b) - f(a).$$

Luego, para cada partición π, $l(X,\pi) \le f(b) - f(a) \le u(X,\pi)$. Puesto que, si X es Riemann-integrable, también $l(X,\pi) \le \int_a^b X \le u(X,\pi)$, se deduce que, dado $\epsilon > 0$, si π es suficientemente fina,

$$\left| \int_a^b X - [f(b) - f(a)] \right| \le u(X,\pi) - l(X,\pi) < \epsilon.$$

6. Si X es acotada en $[a,b]$ y continua en un punto $c \in]a,b[$, la aplicación $f(x) := \int_a^x X$ es derivable en el punto c y $f'(c) = X(c)$.

Capítulo 7

Medida Producto. Medidas de Transición

Se introducen a continuación las nociones de medida de transición (y probabilidad de transición) y de medida producto. Ambas nociones juegan un papel fundamental en esta obra (p.ej., en los conceptos de independencia y de distribución condicional). La noción de medida de transición permite presentar los teoremas clásicos de la medida producto y de Fubini con mayor generalidad sin un gran esfuerzo adicional.

En lo que sigue $(\Omega_1, \mathcal{A}_1)$ y $(\Omega_2, \mathcal{A}_2)$ serán espacios medibles. La siguiente definición construye a partir de ellos un nuevo espacio medible. Ω denotará el conjunto producto $\Omega_1 \times \Omega_2$ que supondremos provisto de la σ-álgebra producto $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$. Si μ_i es una medida en $\mathcal{A}_i$, $i = 1, 2$, estamos ahora interesados en construir una medida μ en $\mathcal{A}$ tal que $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ (ésa sería una medida razonable en el producto en el sentido de que, para ella, "el área de un rectángulo es la longitud de su base por la de su altura"). Como veremos, podemos conseguir mucho más que esto.

DEFINICIÓN 7.1. (Medida de transición) Una medida de transición es una aplicación $S : \Omega_1 \times \mathcal{A}_2 \longrightarrow [0, +\infty]$ tal que

(i) Para cada $A_2 \in \mathcal{A}_2$, $S_{A_2}(\omega_1) = S(\omega_1, A_2)$ es una función Borel-medible de ω_1.

(ii) Para cada $\omega_1 \in \Omega_1$, $S_{\omega_1}(A_2) = S(\omega_1, A_2)$ es una medida en $\mathcal{A}_2$.

Diremos que la medida de transición S es uniformemente σ-finita si existen sucesiones (B_n) en $\mathcal{A}_2$ y (k_n) en $\mathbb{R}$ tales que $S(\omega_1, B_n) \le k_n$, para cada $\omega_1 \in \Omega_1$ y cada entero n. S se dice una probabilidad de transición si, además de (i), verifica que las funciones de conjunto $S(\omega_1, \cdot)$ son probabilidades en $\mathcal{A}_2$.

TEOREMA 7.1. (Teorema de la medida producto generalizado) *Consideremos dos espacios medibles $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$, y sea μ_1 una medida σ-finita en $\mathcal{A}_1$. Sea S una medida de transición uniformemente σ-finita en $\Omega_1 \times \mathcal{A}_2$. Existe entonces una única medida μ en $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$ tal que*

$$\mu(A_1 \times A_2) = \int_{A_1} S(\omega_1, A_2) d\mu_1(\omega_1),$$

para cada $A_1 \in \mathcal{A}_1$ y cada $A_2 \in \mathcal{A}_2$. Concretamente, si $A \in \mathcal{A}$, entonces

$$\mu(A) = \int_{\Omega_1} S(\omega_1, A(\omega_1)) d\mu_1(\omega_1),$$

donde $A(\omega_1) = \{\omega_2 \in \Omega_2 \colon (\omega_1, \omega_2) \in A\}$. Además, μ es σ-finita en $\mathcal{A}$. Por último, si S es una probabilidad de transición y μ_1 es una probabilidad en $\mathcal{A}_1$, entonces μ es una probabilidad en $\mathcal{A}$.

Demostración. Primer caso: Supongamos en primer lugar que las medidas $S(\omega_1, \cdot)$ son finitas.

(i) Probemos que si $A \in \mathcal{A}$, entonces $A(\omega_1) \in \mathcal{A}_2$, $\forall \omega_1 \in \Omega_1$: consideremos la familia

$$\mathcal{C} := \{A \in \mathcal{A} \colon A(\omega_1) \in \mathcal{A}_2\}.$$

$\mathcal{C}$ es una σ-álgebra (pues $(\cup_n A_n)(\omega_1) = \cup_n A_n(\omega_1)$ y $A^c(\omega_1) = A(\omega_1)^c$) que contiene a los rectángulos medibles (pues, para cada $A_1 \in \mathcal{A}_1$ y cada $A_2 \in \mathcal{A}_2$, $(A_1 \times A_2)(\omega_1) = A_2$ si $\omega_1 \in A_1$ y $(A_1 \times A_2)(\omega_1) = \emptyset$ si $\omega_1 \notin A_1$). Por tanto, $\mathcal{C} = \mathcal{A}$.

(ii) Probemos ahora que, si $A \in \mathcal{A}$, $S(\omega_1, A(\omega_1))$ es una función $\mathcal{A}_1$-medible: consideremos ahora la familia

$$\mathcal{C} := \{A \in \mathcal{A} \colon S(\omega_1, A(\omega_1)) \text{ es una función } \mathcal{A}_1\text{-medible}\}.$$

Basta probar que $\mathcal{C} = \mathcal{A}$. Para ello probaremos que $\mathcal{C}$ es un d-sistema que contiene al π-sistema formado por los rectángulos medibles; el teorema de Dynkin (Teorema

1.1) probará que $\mathcal{C}$ es una σ-álgebra y, por tanto, que $\mathcal{C} = \mathcal{A}$. Pero, dados $A_1 \in \mathcal{A}_1$ y $A_2 \in \mathcal{A}_2$, si $A = A_1 \times A_2$, entonces

$$S(\omega_1, A(\omega_1)) = S(\omega_1, A_2) I_{A_1}(\omega_1),$$

que es una función $\mathcal{A}_1$-medible. Además $\mathcal{C}$ es un d-sistema pues contiene a $\Omega_1 \times \Omega_2$ por ser un rectángulo medible, es estable frente a diferencias propias (pues si $A, B \in \mathcal{C}$ y $A \subset B$ entonces $S(\omega_1, (B \setminus A)(\omega_1)) = S(\omega_1, B(\omega_1)) - S(\omega_1, A(\omega_1))$, que es $\mathcal{A}_1$-medible, lo que prueba que $B \setminus A \in \mathcal{C}$) y es también estable frente a la unión numerable creciente (pues, si $(A_n)_n$ es una sucesión creciente en $\mathcal{C}$ entonces $S(\omega_1, (\cup_n A_n)(\omega_1)) = \text{lím}_n\, S(\omega_1, A_n(\omega_1))$, que es $\mathcal{A}_1$-medible, lo que prueba que $\cup_n A_n \in \mathcal{C}$).

(iii) Se define una función de conjuntos en la σ-álgebra producto $\mathcal{A}$ por

$$\mu(A) = \int_{\Omega_1} S(\omega_1, A(\omega_1)) \mu_1(d\omega_1).$$

μ está bien definida por (ii). Veamos que μ es una medida en $\mathcal{A}$ que verifica que

$$\mu(A_1 \times A_2) = \int_{A_1} S(\omega_1, A_2) d\mu_1(\omega_1),$$

para cada $A_1 \in \mathcal{A}_1$ y cada $A_2 \in \mathcal{A}_2$. Es claro que μ es una función de conjuntos no negativa y nula en el vacío. Además, si $(A^{(n)})_n$ es una sucesión disjunta en $\mathcal{A}$, se verifica que

$$\begin{aligned}
\mu(\cup_n A^{(n)}) &= \int_{\Omega_1} S(\omega_1, \cup_n A^{(n)}(\omega_1)) d\mu_1(\omega_1) \\
&= \sum_n \int_{\Omega_1} S(\omega_1, A^{(n)}(\omega_1)) d\mu_1(\omega_1) \\
&= \sum_n \mu(A^{(n)}).
\end{aligned}$$

Eso prueba que μ es una medida. Por otra parte

$$\begin{aligned}
\mu(A_1 \times A_2) &= \int_{\Omega_1} S(\omega_1, (A_1 \times A_2)(\omega_1)) d\mu_1(\omega_1) \\
&= \int_{\Omega_1} S(\omega_1, A_2) I_{A_1}(\omega_1) d\mu_1(\omega_1) \\
&= \int_{A_1} S(\omega_1, A_2) d\mu_1(\omega_1).
\end{aligned}$$

Segundo caso: En el caso general, la medida de transición S es uniformemente σ-finita y existen una sucesión $(k_n)_n$ en $\mathbb{R}$ y una sucesión disjunta $(A_2^{(n)})_n$ en $\mathcal{A}_2$ tales que $S(\omega_1, A_2^{(n)}) \leq k_n$ para cada $\omega_1 \in \Omega_1$ y cada $n \in \mathbb{N}$. Haciendo

$$S_n(\omega_1, A_2) := S(\omega_1, A_2 \cap A_2^{(n)})$$

se obtienen medidas de transición finitas S_n a las que podemos aplicar el primer caso para obtener medidas ν_n en $\mathcal{A}$ tales que

$$\nu_n(A) = \int_{\Omega_1} S_n(\omega_1, A(\omega_1))\mu_1(\omega_1)$$

y

$$\nu_n(A_1 \times A_2) = \int_{A_1} S_n(\omega_1, A_2) d\mu_1(\omega_1).$$

Se define ahora $\mu = \sum_n \nu_n$. Es fácil probar que μ conviene.

Unicidad: Supongamos ahora que λ es otra medida en $\mathcal{A}$ tal que

$$\lambda(A_1 \times A_2) = \int_{A_1} S(\omega_1, A_2) d\mu_1(\omega_1).$$

Entonces λ y μ coinciden en el π-sistema formado por los rectángulos medibles. Supongamos en primer lugar que μ es finita. Entonces $\mathcal{D} := \{A \in \mathcal{A} \colon \lambda(A) = \mu(A)\}$ es un d-sistema que contiene a los rectángulos medibles y, por tanto, $\mathcal{D} = \mathcal{A}$. En el caso general, podemos tomar una sucesión creciente $A_n = A_{1n} \times A_{2n}$ de rectángulos medibles que recubren $\Omega_1 \times \Omega_2$ y con medidas λ y μ finitas. Las restricciones λ_n y μ_n de λ y μ a A_n son medidas finitas que coinciden según hemos probado anteriormente. Eso acaba la prueba, pues para cada $A \in \mathcal{A}$, $\mu(A) = \text{lím}_n \mu_n(A) = \text{lím}_n \lambda_n(A) = \lambda(A)$. $\square$

El teorema de Fubini aclara cómo se integra respecto a la medida producto que acabamos de construir y reduce el cálculo de integrales en los espacios de medida construidos anteriormente al cálculo de integrales (iteradas) en los espacios de medida de partida.

TEOREMA 7.2. (Teorema de Fubini generalizado) *Mantenemos las notaciones e hipótesis del teorema de la medida producto generalizado. Sea $f : \Omega_1 \times \Omega_2 \longrightarrow \bar{\mathbb{R}}$ una función Borel-medible respecto a la σ-álgebra producto.*

(a) *Si f es no negativa, entonces existe $\int_{\Omega_2} f(\omega_1,\omega_2)S(\omega_1,d\omega_2)$ y define una función medible de ω_1. Además*

$$\int f d\mu = \int_{\Omega_1}\left(\int_{\Omega_2} f(\omega_1,\omega_2)S(\omega_1,d\omega_2)\right)\mu_1(d\omega_1). \tag{7.1}$$

(b) *Si $\int_\Omega f d\mu$ existe (resp., es finita), entonces $\int_{\Omega_2} f(\omega_1,\omega_2)S(\omega_1,d\omega_2)$ existe (resp., es finita) para casi todo ω_1, es decir, μ_1-c.s., y define una función medible de ω_1 si se define como cero en el conjunto excepcional. Además, (7.1) se verifica también en este caso.*

Demostración. (a) Notemos en primer lugar que, para cada $\omega_1 \in \Omega_1$, la aplicación $f(\omega_1,\cdot):\Omega_2 \longrightarrow \bar{\mathbb{R}}$ es $\mathcal{A}_2$-medible, pues si $B \in \bar{\mathcal{R}}$, entonces

$$\{\omega_2 \in \Omega_2 \colon f(\omega_1,\omega_2) \in B\} = \{\omega_2 \colon (\omega_1,\omega_2) \in f^{-1}(B)\} = f^{-1}(B)(\omega_1) \in \mathcal{A}_2,$$

por el teorema anterior. Entonces, la integral

$$\int_{\Omega_2} f(\omega_1,\omega_2)S(\omega_1,d\omega_2)$$

existe. Para probar que es $\mathcal{A}_1$-medible, comenzaremos suponiendo que f es un indicador: $f = I_A$. En ese caso

$$\int_{\Omega_2} I_A(\omega_1,\omega_2)S(\omega_1,d\omega_2) = \int_{\Omega_2} I_{A(\omega_1)}(\omega_2)S(\omega_1,d\omega_2) = S(\omega_1,A(\omega_1)),$$

que es $\mathcal{A}_1$-medible por el teorema anterior. Además,

$$\int_\Omega I_A d\mu = \int_{\Omega_1} S(\omega_1,A(\omega_1))\mu_1(d\omega_1) = \int_{\Omega_1}\int_{\Omega_2} I_A(\omega_1,\omega_2)S(\omega_1,d\omega_2)\mu_1(d\omega_1).$$

Eso prueba el teorema de Fubini para indicadores. Si f es una función simple medible no negativa, el teorema es también cierto por linealidad de la integral.

Supongamos ahora que f es una función medible $[0,+\infty]$-valorada. Existe entonces una sucesión creciente $(f_n)_n$ de funciones simples medibles no negativas que converge puntualmente a f. Por el teorema de la convergencia monótona (Teorema 4.1),

$$\int_{\Omega_2} f(\omega_1,\omega_2)S(\omega_1,d\omega_2) = \lim_n \int_{\Omega_2} f_n(\omega_1,\omega_2)S(\omega_1,d\omega_2),$$

y define una función medible de ω_1 como límite puntual de funciones medibles. Además, aplicando repetidamente el teorema de la convergencia monótona (Teorema 4.1),

$$\begin{aligned}\int_\Omega f d\mu &= \lim_n \int_\Omega f_n d\mu = \lim_n \int_{\Omega_1}\int_{\Omega_2} f_n(\omega_1,\omega_2)S(\omega_1,d\omega_2)\mu_1(d\omega_1)\\ &= \int_{\Omega_1}\int_{\Omega_2} f(\omega_1,\omega_2)S(\omega_1,d\omega_2)\mu_1(d\omega_1).\end{aligned}$$

(b) Si existe $\int f d\mu$, o bien su parte positiva o bien su parte negativa tiene integral finita. Supongamos que $\int f^- d\mu < \infty$. De acuerdo con (a),

$$\int_{\Omega_1}\int_{\Omega_2} f^-(\omega_1,\omega_2)S(\omega_1,d\omega_2)\mu_1(d\omega_1) = \int f^- d\mu < \infty;$$

por tanto, la integral $\int_{\Omega_2} f^-(\omega_1,\omega_2)S(\omega_1,d\omega_2)$ define una función de ω_1 que es medible y μ_1-integrable; en particular, es finita μ_1-c.s. Por tanto, para casi todo ω_1 podemos escribir

$$\int_{\Omega_2} f(\omega_1,\omega_2)S(\omega_1,d\omega_2) = \int_{\Omega_2} f^+(\omega_1,\omega_2)S(\omega_1,d\omega_2) - \int_{\Omega_2} f^-(\omega_1,\omega_2)S(\omega_1,d\omega_2).$$

Terminamos la demostración definiendo $\int_{\Omega_2} f(\omega_1,\omega_2)S(\omega_1,d\omega_2) = 0$ en el conjunto excepcional e integrando respecto a μ_1. □

El corolario siguiente es un caso particular del teorema anterior: si μ_2 es una medida en $\mathcal{A}_2$, basta considerar la medida de transición $S(\omega_1,A_2) = \mu_2(A_2)$.

Corolario 7.1. (Teoremas clásicos de la medida producto y de Fubini) *Consideremos dos espacios de medida σ-finitos $(\Omega_i,\mathcal{A}_i,\mu_i)$, $i = 1,2$. Entonces la función de conjuntos*

$$\mu(A) = \int_{\Omega_1} \mu_2(A_{\omega_1})\mu_1(d\omega_1) = \int_{\Omega_2} \mu_1(A^{\omega_2})\mu_2(d\omega_2)$$

es la única medida en $\mathcal{A}_1\times\mathcal{A}_2$ tal que $\mu(A_1\times A_2) = \mu_1(A_1)\mu_2(A_2)$ para cada $A_1\in\mathcal{A}_1$ y cada $A_2\in\mathcal{A}_2$, donde $A_{\omega_1} := \{\omega_2\in\Omega_1\colon (\omega_1,\omega_2)\in A\}$ y $A^{\omega_2} := \{\omega_1\in\Omega_1\colon (\omega_1,\omega_2)\in A\}$. Además μ es σ-finita, y es una probabilidad si μ_1 y μ_2 lo son. μ se llamara medida *producto de* μ_1 *y* μ_2 *y se denotará* $\mu_1 \times \mu_2$*;* $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2, \mu_1 \times \mu_2)$ *se llamará espacio de medida producto de los espacios de medida considerados.*

Si $f : \Omega_1 \times \Omega_2 \longrightarrow \bar{\mathbb{R}}$ *es Borel-medible y si existe* $\int f d\mu$*, entonces*

$$\int f d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} f d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left(\int_{\Omega_1} f d\mu_1 \right) d\mu_2$$

(donde, como antes, las integrales interiores existen c.s. y se definen como cero allí donde no existan).

***Observación* 7.1.** El teorema clásico de la medida producto se generaliza sin dificultad al caso de un número finito de factores. El espacio de medida producto de n copias de una espacio de medida σ-finito $(\Omega, \mathcal{A}, \mu)$ se denotará por $(\Omega^n, \mathcal{A}^n, \mu^n)$ o, también, por $(\Omega, \mathcal{A}, \mu)^n$.

***Observación* 7.2.** Ya sabemos que $\mathcal{R}^2 = \mathcal{R} \times \mathcal{R}$. Se deduce del teorema anterior que, si λ_n es la medida de Lebesgue en $\mathbb{R}^n$, $n = 1, 2$, entonces $\lambda_2 = \lambda_1 \times \lambda_1$. En general, $\lambda_n = \lambda_1^n$ (ver Problema 7.3).

El teorema clásico de Fubini también se generaliza de forma natural al caso de un número finito arbitrario de factores: la integral múltiple (es decir, respecto al producto de n medidas σ-finitas) se reduce al cálculo de n integrales iteradas.

El teorema clásico de la medida producto n-dimensional formaliza el concepto de n realizaciones independientes de un experimento aleatorio. Desde un punto de vista teórico, es conveniente considerar un número finito arbitrariamente grande de realizaciones independientes de ese experimento; para ello necesitaremos un teorema de la medida producto infinito dimensional, aunque sólo en un contexto probabilístico.

Sean $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2, \ldots$, una sucesión de espacios medibles. Denotaremos por Ω el conjunto producto $\prod_{i=1}^{\infty} \Omega_i$. Un subconjunto C de Ω se llamará cilindro medible si existen un entero n y $C_n \in \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ tal que $C = C_n \times \Omega_{n+1} \times \Omega_{n+2} \times \ldots$ (se dice por ello que C es un cilindro n-dimensional). La σ-álgebra producto $\mathcal{A}$ en Ω se define ahora como la σ-álgebra engendrada por los cilindros medibles. Se verifica entonces el siguiente:

Teorema 7.3. *Sean* $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1, 2, \dots$, *espacios de probabilidad. Si* Ω *es el conjunto producto de los* Ω_i *y* $\mathcal{A}$ *la* σ*-álgebra producto de las* $\mathcal{A}_i$, *entonces existe una única probabilidad* P *en* $\mathcal{A}$ *tal que, para cada entero* n *y cada elección de sucesos* $A_i \in \mathcal{A}_i$, $1 \le i \le n$, *se verifica que*

$$P(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \dots) = P_1(A_1) \cdots P_n(A_n).$$

Además, la integral (si existe) respecto a P *de una función* $\mathcal{A}$*-medible que sólo depende de las* n *primeras coordenadas coincide con su integral respecto al producto finito* $P_1 \times \cdots \times P_n$.

Demostración. Denotemos $Q_n = P_1 \times \cdots \times P_n$. Un cilindro medible n-dimensional en Ω es un conjunto de la forma

$$C_n = B_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots$$

donde $B_n \in \prod_{i=1}^n \mathcal{A}_i$. Para un tal cilindro medible se define

$$P(C_n) = Q_n(B_n).$$

Probemos en primer lugar que la definición es correcta, es decir, que si C_n admite también la expresión

$$C_n = B'_m \times \Omega_{m+1} \times \Omega_{m+2} \times \cdots$$

con $m \le n$, entonces $Q_n(B_n) = Q_m(B'_m)$. Pero esos se deduce sin problemas de que

$$(\omega_1, \dots, \omega_n) \in B_n \iff (\omega_1, \dots, \omega_m) \in B'_m$$

y de que

$$Q_n(B_n) = \int_{\Omega_1} \cdots \int_{\Omega_n} I_{B_n}(\omega_1, \dots, \omega_n) dP_n(\omega_n) \cdots dP_1(\omega_1).$$

Es claro que P es finitamente aditiva en el álgebra de los cilindros medibles. Si probamos que P es continua hacia abajo en el vacío, quedará probado que es numerablemente aditiva en ese álgebra y el teorema de extensión de Carathéodory asegurará la existencia de una única extensión de P a una probabilidad en el espacio producto.

Sea, pues, $(C_n)_n$ una sucesión decreciente al vacío de cilindros medibles. Puesto que todo cilindro n-dimensional es también $n+k$-dimensional, para cada $k \geq 0$, podemos suponer sin pérdida de generalidad que C_n es un cilindro n-dimensional que escribiremos en la forma $B_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots$ (añadiendo, si fuera preciso, cilindros a la sucesión de forma que esta siga siendo decreciente al vacío). Si $\lim_n P(C_n) > 0$, entonces, para cada $n > 1$,

$$P(B_n) = \int_{\Omega_1} g_n^{(1)}(\omega_1) dP_1(\omega_1)$$

donde

$$g_n^{(1)}(\omega_1) = \int_{\Omega_2} \cdots \int_{\Omega_n} I_{C_n}(\omega_1, \omega_2, \ldots, \omega_n) dP_n(\omega_n) \ldots dP_2(\omega_2).$$

Puesto que $C_{n+1} \subset C_n$, se tiene que $C_{n+1} \subset C_n \times \Omega_{n+1}$ y, por tanto,

$$I_{C_{n+1}}(\omega_1, \ldots, \omega_{n+1}) \leq I_{C_n}(\omega_1, \ldots, \omega_n),$$

lo que prueba que $g_n^{(1)}$ es una sucesión decreciente. Sea $h_1 = \lim_n g_n^{(1)}$. El teorema de la convergencia monótona extendido prueba que

$$\lim_n P(C_n) = \int_{\Omega_1} h_1 dP_1.$$

Puesto que $\lim_n P(C_n) > 0$, existe ω_1' tal que $h(\omega_1') > 0$; además, $\omega_1' \in B_1$ pues, en otro caso, $I_{C_n}(\omega_1', \omega_2, \ldots, \omega_n) = 0$, para todo $\omega_2, \ldots, \omega_n$ y todo n, y se tendría que $g_n^{(1)}(\omega_1') = 0$, para todo n, y $h(\omega_1') = 0$, contra la elección de ω_1'. Ahora, para $n > 2$,

$$g_n^{(1)}(\omega_1') = \int_{\Omega_2} g_n^{(2)}(\omega_2) dP_2(\omega_2)$$

donde

$$g_n^{(2)}(\omega_2) = \int_{\Omega_3} \cdots \int_{\Omega_n} I_{C_n}(\omega_1', \omega_2, \ldots, \omega_n) dP_n(\omega_n) \ldots dP_3(\omega_3).$$

Como arriba, $(g_n^{(2)})$ es una sucesión decreciente a un límite h_2, y $(g_n^{(1)}(\omega_1'))$ converge a $\int_{\Omega_2} h_2(\omega_2) dP_2(\omega_2)$. Como $\lim_n g_n^{(1)}(\omega_1') = h(\omega_1') > 0$, existe ω_2' tal que $h_2(\omega_2') > 0$ y, como antes, $(\omega_1', \omega_2') \in B_2$. Por inducción construimos un punto $(\omega_1', \omega_2', \ldots) \in \prod_{i=1}^\infty \Omega_i$ tal que $(\omega_1', \ldots, \omega_n') \in B_n$, para cada n. Luego $(\omega_1', \omega_2', \ldots) \in \cap_n C_n = \emptyset$, lo que es una contradicción. Eso prueba la existencia de P; la unicidad se deduce del teorema de Dynkin en la forma habitual. □

PROBLEMAS DEL CAPÍTULO 7

Problema 7.1. En el proceso de fabricación de un cierto objeto, se seleccionan al azar n objetos con el fin de decidir si se acepta o se rechaza la producción total en función del número de objetos defectuosos presentes en la muestra. Supongamos que la proporción de objetos defectuosos es un número $p \in]0,1[$ y que decidimos rechazar la producción total con probabilidad k/n si hay k objetos defectuoso entre los n elegidos. Sabemos que la distribución del número de objetos defectuosos entre n elegidos al azar es la distribución binomial $b_n(p)$. Sean, pues, $\Omega_1 = \{0,1,\dots,n\}$, $\mathcal{A}_1 = \mathscr{P}(\Omega_1)$ y P_1 la distribución binomial $b_n(p)$ de parámetro p en Ω_1. Sea $(\Omega_2, \mathcal{A}_2) = (\{0,1\}, \mathscr{P}(\{0,1\}))$ y consideremos la probabilidad de transición S de $(\Omega_1, \mathcal{A}_1)$ en $(\Omega_2, \mathcal{A}_2)$ definida por

$$S(k,\{1\}) = \frac{k}{n} = 1 - S(k,\{0\}).$$

Esta probabilidad de transición describe la estrategia que seguiremos a la hora de tomar la decisión 0 de aceptar la producción total o 1 de rechazarla. Supongamos conocida una función $f : \{0,1,\dots,n\} \times \{0,1\} \longrightarrow [0,+\infty[$ que evalúa el coste que ocasiona la toma de una determinada decisión del siguiente modo:

$$f(k,1) = c, \quad f(k,0) = c'k, \qquad k = 0,1,\dots,n,$$

donde c y c' son constantes positivas conocidas. Calcular la media de f, valor que podemos utilizar para evaluar el coste medio del proceso de producción.

Problema 7.2. Sea μ la medida cardinal en $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$. Probar que μ^2 es la medida cardinal en $(\mathbb{N}^2, \mathscr{P}(\mathbb{N}^2))$. Calcular $\int_{\mathbb{N}^2} f d\mu^2$ si $f(m,n) = 2^{-(m+n)}$.

Problema 7.3. Sea λ_n la medida de Lebesgue en $(\mathbb{R}^n, \mathcal{R}^n)$.

1. Probar que $\lambda_n = \lambda_1^n$.
2. Calcular $\int_{[0,1]^n} f(x_1,\dots,x_n) d\lambda_n(x_1,\dots,x_n)$ si $f(x_1,\dots,x_n) = x_1 \cdot \ldots \cdot x_n$.
3. Calcular $\int_{[-1,1]^2} (x+y) d\lambda_2(x,y)$.

Problema 7.4. Probar que $\int_0^\infty e^{-x^2} dx = \sqrt{\pi}/2$.

> *Indicación:* Aplicar el teorema de cambio de variables -cambio a coordenadas polares- utilizando la transformación $(r,\theta) \in]0,\infty[\times]0,2\pi[\longrightarrow (r\cos\theta, r\sin\theta)$ para calcular $\iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dxdy$. Utilizar el teorema de Fubini (Teorema 7.2) para obtener de ahí que $\int_0^\infty e^{-x^2} dx = \sqrt{\pi}/2$.

Problema 7.5. ($\uparrow$) Concluir que $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}dx = \sqrt{2\pi}$, con lo cual la aplicación $f : \mathbb{R} \longrightarrow \mathbb{R}$ definida por $f(x) = (\sqrt{2\pi})^{-1}e^{-\frac{1}{2}x^2}$ es la densidad de una probabilidad P en $(\mathbb{R}, \mathcal{R})$. Ello nos permite dar la siguiente definición: Sea X una v.a.r. en $(\Omega, \mathcal{A}, P)$; diremos que X es una v.a.r. con distribución $N(0, 1)$ (*normal* de media cero y varianza 1) si su distribución de probabilidad P^X en $(\mathbb{R}, \mathcal{R})$ admite como densidad (respecto a la medida de Lebesgue) la aplicación f que acabamos de definir.

Problema 7.6. ($\uparrow$) Si X es una v.a.r. con distribución $N(0, 1)$ probar que $E(X) = 0$.

Problema 7.7. ($\uparrow$) Sean $\alpha > 0$ y $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1}e^{-x}dx$ (es la llamada función gamma de Euler). Probar que $\Gamma(\alpha)$ está bien definida, es decir, que la integral que la define existe y es finita.

> *Indicación:* La integral existe pues el integrando es una función medible no negativa; acotar el integrando por $e^{-x/2}$ en un cierto entorno de $+\infty$ y por una función integrable en un entorno de cero (si $\alpha \geq 1$ no hay problemas en este último caso, y si $0 < \alpha < 1$ se podrá acotar $x^{\alpha-1}e^{-x} \leq x^{\alpha-1}$ para x próximo a cero).

Problema 7.8. ($\uparrow$) $\Gamma(1) = 1$.

Problema 7.9. ($\uparrow$) Si $\alpha > 1$ entonces $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

> *Indicación:* Integrar por partes.

Problema 7.10. ($\uparrow$) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

> *Indicación:* $\Gamma(\frac{1}{2}) = \int_0^\infty x^{-\frac{1}{2}}e^{-x}dx$; aplicar el teorema de cambio de variables utilizando la transformación $z \in]0, +\infty[\longrightarrow z^2 \in]0, +\infty[$.

Problema 7.11. ($\uparrow$) Si X tiene distribución $N(0, 1)$ probar que $\mathrm{Var}(X) = 1$.

> *Indicación:* Notar que
>
> $$\begin{aligned}\mathrm{Var}(X) = E[(X - E(X))^2] &= \int_{-\infty}^{+\infty} x^2(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}dx \\ &= 2(2\pi)^{-\frac{1}{2}}\int_0^\infty x^2e^{-\frac{1}{2}x^2}dx;\end{aligned}$$
>
> hacer ahora el cambio de variables $y = x^2$ y utilizar los dos problemas anteriores para calcular $\Gamma(3/2)$.

Problema 7.12. ($\uparrow$) Si $\mu \in \mathbb{R}$ y $\sigma > 0$ probar que $[\sigma\sqrt{2\pi}]^{-1}\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = 1$. Def.: Si X es una v.a.r. cuya distribución queda determinada por la densidad $[\sigma\sqrt{2\pi}]^{-1}e^{-\frac{(x-\mu)}{2\sigma^2}}$, diremos que X es una v.a. $N(\mu, \sigma^2)$.

Indicación: Efectuar el cambio de variables $y = (x - \mu)/\sigma$.

Problema 7.13. ($\uparrow$) Probar que, si X es una v.a. con distribución $N(\mu, \sigma^2)$, entonces $E(X) = \mu$ y $\mathrm{Var}(X) = \sigma^2$.

Problema 7.14. ($\uparrow$) Si $\alpha, \beta > 0$ entonces $\int_0^\infty y^{\alpha-1}e^{-y/\beta}dx = \Gamma(\alpha)\beta^\alpha$. Def.: Una v.a.r. X se dice que tiene distribución *gamma* de parámetros α, β ($\alpha > 0$, $\beta > 0$), y se escribirá $X \sim G(\alpha, \beta)$, si su distribución admite como densidad la aplicación

$$[\Gamma(\alpha)\beta^\alpha]^{-1}y^{\alpha-1}e^{-y/\beta}I_{]0,+\infty[}(y).$$

Indicación: En la definición de $\Gamma(\alpha)$ hacer el cambio $x = y/\beta$.

Problema 7.15. ($\uparrow$) Probar que, si $X \sim G(\alpha, \beta)$, entonces $E(X) = \alpha\beta$ y $\mathrm{Var}(X) = \alpha\beta^2$.

Problema 7.16. ($\uparrow$) Def.: Si $n \in \mathbb{N}$ y X tiene distribución $G(n/2, 2)$ diremos que X tiene distribución *ji-cuadrado* con n grados de libertad y escribiremos $X \sim \chi^2(n)$. Su densidad es

$$[\Gamma(n/2)2^{n/2}]^{-1}e^{-x/2}x^{\frac{n}{2}-1}I_{]0,+\infty[}(x);$$

su media es $E(X) = n$ y su varianza $\mathrm{Var}(X) = 2n$.

Problema 7.17. ($\uparrow$) Si $X \sim N(0, 1)$ entonces X^2 tiene distribución $\chi^2(1)$. En particular, si $X \sim N(\mu, \sigma^2)$ entonces $\sigma^{-2}(X - \mu)^2 \sim \chi^2(1)$.

Problema 7.18. Sea (X, Y) una variable bidimensional, con función de densidad:

$$f(x, y) = \begin{cases} \alpha^2 e^{-\alpha(x+y)} & \text{si } x, y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

siendo $\alpha > 0$. Calcular la distribución de probabilidad de la variable aleatoria $Z = X - Y$.

Capítulo 8

Independencia. Convolución

Un concepto típicamente probabilístico está especialmente relacionado con la idea de medida producto: independencia.

Definición 8.1. (Independencia) Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad.

(a) Dos sucesos $A, B \in \mathcal{A}$ se dicen independientes si

$$P(A \cap B) = P(A)P(B).$$

(b) Los sucesos de una familia $(A_i)_{i \in I}$ de $\mathcal{A}$ se dicen independientes si para cada subfamilia finita $\{A_{i_1}, \ldots, A_{i_n}\}$ se verifica $P(\cap_{k=1}^n A_{i_k}) = \prod_{k=1}^n P(A_{i_k})$.

(c) Una familia $(\mathcal{A}_i)_{i \in I}$ de sub-σ-álgebras de $\mathcal{A}$ se dicen independientes si cada familia $(A_i)_{i \in I}$ tal que $A_i \in \mathcal{A}_i$, $\forall i \in I$, está formada por sucesos independientes.

(d) Sean $(\Omega_i, \mathcal{A}_i)_{i \in I}$ una familia de espacios medibles y, para cada $i \in I$, sea $T_i : \Omega \longrightarrow \Omega_i$ una v.a. Diremos que las v.a. T_i son independientes si lo son las sub-σ-álgebras $T^{-1}(\mathcal{A}_i)$.

Definición 8.2. Sean $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega_i, \mathcal{A}_i)$, $i = 1, 2, \ldots$, v.a. La distribución P^X de la v.a.

$$X : \omega \in (\Omega, \mathcal{A}, P) \longrightarrow X(\omega) = (X_1(\omega), X_2(\omega), \ldots) \in \left(\prod_{i \geq 1} \Omega_i, \prod_{i \geq 1} \mathcal{A}_i\right)$$

se llama distribución conjunta de las v.a. X_i. En esa situación, la distribución de cada una de las X_i recibe el nombre de distribución marginal.

TEOREMA 8.1. (Caracterización de independencia) *Con las notaciones de la definición anterior, las v.a. X_i son independientes si, y sólo si, $P^X = \prod_{i\geq 1} P^{X_i}$.*

Demostración. Si las v.a. X_i son independientes, entonces, dados $n \in \mathbb{N}$ y sucesos $A_i \in \mathcal{A}_i$, $1 \leq i \leq n$, se tiene que, haciendo $A_i = \Omega_i$ si $i > n$,

$$P^X(\prod_{i\geq 1} A_i) = P(\cap_{i=1}^n \{X_i \in A_i\}) = \prod_{i=1}^n P(X_i \in A_i) = (\prod_{i\geq 1} P^{X_i})(\prod_{i\geq 1} A_i).$$

Así pues, P^X y $\prod_{i\geq 1} P^{X_i}$ son dos probabilidades en el espacio producto $\prod_{i\geq 1} \mathcal{A}_i$ que coinciden sobre los rectángulos medibles. La unicidad en el teorema de la medida producto prueba que también coinciden sobre la σ-álgebra producto $\prod_{i\geq 1} \mathcal{A}_i$.

Recíprocamente, si $P^X = \prod_{i\geq 1} P^{X_i}$ entonces, dados $n \in \mathbb{N}$ y sucesos $A_i \in \mathcal{A}_i$, $1 \leq i \leq n$, se tiene que, haciendo $A_i = \Omega_i$ si $i > n$,

$$P(\cap_{i=1}^n \{X_i \in A_i\}) = P^X(\prod_{i\geq 1} A_i) = (\prod_{i\geq 1} P^{X_i})(\prod_{i\geq 1} A_i) = \prod_{i=1}^n P(X_i \in A_i),$$

que prueba que las X_i son independientes. □

El resultado siguiente afirma que funciones medibles de v.a. independientes son independientes.

Proposición 8.1. *Sean, para cada $i \geq 1$, $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega_i, \mathcal{A}_i)$ y $f_i : (\Omega_i, \mathcal{A}_i) \longrightarrow (\Omega'_i, \mathcal{A}'_i)$, $i = 1, 2, \ldots$, v.a. Si las X_i son independientes, también lo son las $f_i \circ X_i$.*

Demostración. Dados $n \in \mathbb{N}$ y sucesos $A'_i \in \mathcal{A}'_i$, $1 \leq i \leq n$, se tiene que

$$\begin{aligned} P(\cap_{i=1}^n (f_i \circ X_i)^{-1}(A_i)) &= P(\cap_{i=1}^n X_i^{-1}(f_i^{-1}((A_i))) \\ &= \prod_{i=1}^n P(X_i^{-1}(f_i^{-1}((A_i))) \\ &= \prod_{i=1}^n P((f_i \circ X_i)^{-1}(A_i)), \end{aligned}$$

lo que prueba el resultado. □

La noción de independencia es especialmente útil y simple en espacios producto, como pone de manifiesto el siguiente teorema. Recordemos previamente que la σ-álgebra producto es la más pequeña σ-álgebra en el conjunto producto que hace medibles las proyecciones, y que una aplicación a valores en un producto es medible si, y sólo si, al componer con las proyecciones se obtienen funciones medibles.

TEOREMA 8.2. (a) *Sean* $(\Omega_i, \mathcal{A}_i, P_i)_{i\geq 1}$ *una colección numerable de espacios de probabilidad y* $(\Omega, \mathcal{A}, P) = (\prod_{i\geq 1}\Omega_i, \prod_{i\geq 1}\mathcal{A}_i, \prod_{i\geq 1}P_i)$ *el espacio de probabilidad producto. Consideremos las aplicaciones coordenada* $q_i : \omega \in \Omega \to \omega_i \in \Omega_i$, $i \geq 1$, *donde* ω_i *denota la coordenada* i*-ésima de* ω. *Entonces las proyecciones* q_i *son independientes. Además* $P^{q_i} = P_i$.

(b) *Sean* $(\Omega_n, \mathcal{A}_n)_{n\in\mathbb{N}}$ *una sucesión de espacios medibles y* $X_n : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega_n, \mathcal{A}_n)$ *v.a. arbitrarias. Para cada* $n \in \mathbb{N}$ *se define una v.a.* $Y_n : (\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}}, P^{\mathbb{N}}) \longrightarrow (\Omega_n, \mathcal{A}_n)$ *mediante* $Y_n(\omega_1, \omega_2, \dots) = X_n(\omega_n)$. *Entonces* Y_n *tiene respecto a* $P^{\mathbb{N}}$ *la misma distribución que* X_n *respecto a* P *y las* Y_n *son independientes.*

Demostración. (a) Para la primera parte, basta aplicar el teorema anterior a la aplicación identidad $id_\Omega = (q_1, q_2, \dots)$ en Ω. Para la segunda parte, nótese que, dados $i \geq 1$ y $A_i \in \mathcal{A}_i$, se tiene que, haciendo $A_k = \Omega_k$ si $k \neq i$,

$$P^{q_i}(A_i) = P(q_i^{-1}(A_i)) = P(\prod_{k\geq 1} A_k) = (\prod_{k\geq 1} P_k)(\prod_{k\geq 1} A_k) = P_i(A_i).$$

(b) Sabemos que las proyecciones $q_k : (\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}}, P^{\mathbb{N}}) \longrightarrow (\Omega, \mathcal{A})$, $k \in \mathbb{N}$, son independientes respecto a la probabilidad producto $P^{\mathbb{N}}$. Por tanto, las v.a. $Y_k := X_k \circ q_k$, $k \in \mathbb{N}$, son independientes como funciones medibles de v.a. independientes. Además, puesto que $(P^{\mathbb{N}})^{q_k} = P$, se tiene que, para cada suceso $A_k \in \mathcal{A}_k$,

$$\begin{aligned}\left(P^{\mathbb{N}}\right)^{Y_k}(A_k) &= P^{\mathbb{N}}(\{Y_k \in A_k\}) = P^{\mathbb{N}}(\{X_k \circ q_k \in A_k\}) \\ &= \left(P^{\mathbb{N}}\right)(\{q_k \in X_k^{-1}(A_k)\}) = P(\{X_k^{-1}(A_k)\}) = P^{X_k}(A_k),\end{aligned}$$

lo que acaba la prueba. □

***Observación* 8.1.** Consideremos el espacio de probabilidad $([0,1], \mathcal{R}[0,1], P)$, donde P denota la medida de Lebesgue en $[0,1]$. En general, dos funciones medibles arbitrarias $f, g : [0,1] \longrightarrow \mathbb{R}$ no serán independientes; pero, sean quienes sean las funciones medibles f y g, las aplicaciones $F, G : [0,1]^2 \longrightarrow \mathbb{R}$ definidas por $F(x,y) = f(x)$ y $G(x,y) = g(y)$ son P^2-independientes por serlo las proyecciones en $[0,1]^2$ (notar que P^2 coincide con la medida de Lebesgue en $[0,1]^2$).

Presentamos a continuación un resultado que caracteriza la independencia de v.a. en términos de densidades.

TEOREMA 8.3. *Sean $X_i : (\Omega, \mathcal{A}, P) \to (\Omega_i, \mathcal{A}_i, \mu_i)$, $1 \le i \le n$, v.a., donde μ_i es una medida σ-finita en $\mathcal{A}_i$. Sean $X := (X_1, \dots, X_n)$ y $\mu := \prod_i \mu_i$.*

(a) *Si X tiene densidad f respecto a la medida producto μ, entonces X_i tiene respecto a μ_i densidad marginal*

$$f_i(\omega_i) := \int_{\prod_{j\neq i}\Omega_j} f(\omega_1, \dots, \omega_n) d(\textstyle\prod_{j\neq i}\mu_j)(\omega_1, \dots, \omega_{i-1}, \omega_{i+1}, \dots, \omega_n).$$

(b) *Si $f_i \in \frac{dP^{X_i}}{d\mu_i}$, $1 \le i \le n$, las X_i son independientes si, y sólo si,*

$$\prod_{i=1}^n f_i(\omega_i) \in \frac{dP^X}{d\mu}.$$

Demostración. (a) De acuerdo con el teorema de Fubini (Corolario 7.1), f_i así definida es $\mathcal{A}_i$-medible y su integral respecto a μ_i sobre $A_i \in \mathcal{A}_i$ coincide con $P(X_i \in A_i) = P(X \in \prod_j A_j)$, si $A_j = \Omega_j$ para $j \neq i$.

(b) Si las X_i son independientes,

$$P(\cap_i\{X_i \in A_i\}) = \prod_i P(X_i \in A_i) = \prod_i \int_{A_i} f_i d\mu_i = \int_{\prod_i A_i} \prod_{i=1}^n f_i(\omega_i) d\mu(\omega_i, \dots, \omega_n).$$

Eso prueba que $P(X \in A) = \int_A \prod_{i=1}^n f_i(\omega_i) d\mu(\omega_i, \dots, \omega_n)$ si A es un rectángulo medible, y el resultado se sigue de aquí en la forma habitual (ver Problema 3.8). El recíproco es consecuencia inmediata del teorema de Fubini. □

TEOREMA 8.4. (a) *Si $X_1, \dots, X_n$ son v.a. reales o complejas independientes con media finita entonces*

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i).$$

(b) *Dos v.a.r. independientes con media finita son incorreladas (es decir, su covarianza es nula).*

(c) *Si $X_1, \dots, X_n$ son v.a.r. independientes con varianza finita entonces*

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

Demostración. (a) Si $X_1,\dots,X_n$ son v.a.r. no negativas independientes con media finita, entonces, por el teorema de la medida imagen (Teorema 5.2), la suposición de independencia y el teorema de Fubini (Corolario 7.1)

$$E(\prod_i X_i) = \int x_1\cdots x_n dP^{(X_1,\dots,X_n)}(x_1,\dots,x_n) = \int x_1\cdots x_n d\prod_i P^{X_i}(x_i)$$
$$= \prod_i \int x_i dP^{X_i}(x_i) = \prod_i E(X_i).$$

Si $X_1,\dots,X_n$ son v.a.r. no negativas independientes con media finita, el caso no negativo prueba que $E\left(\prod_{i=1}^n |X_i|\right) = \prod_{i=1}^n E(|X_i|) < \infty$ y el teorema de Fubini concluye como antes. El resultado también es cierto si las X_i son v.a. complejas independientes con media finita;([1]) basta probarlo por inducción a partir del caso de dos factores, y este caso se reduce al caso real teniendo en cuenta que si X_1 y X_2 son v.a. complejas independientes, también son independientes $\mathrm{Re}\,X_1$ y $\mathrm{Re}\,X_2$, $\mathrm{Re}\,X_1$ e $\mathrm{Im}\,X_2$, $\mathrm{Im}\,X_1$ y $\mathrm{Re}\,X_2$, e $\mathrm{Im}\,X_1$ y $\mathrm{Im}\,X_2$.

(b) Si X e Y son independientes y con media finita, $Cov(X,Y) = E[(X-EX)(Y-EY)] = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$

(c) $\mathrm{Var}(\sum_i X_i) = E[(\sum_i (X_i - EX_i))^2] = E[\sum_i (X_i - EX_i)^2] + 2E[\sum_{i<j}(X_i - EX_i)(X_j - EX_j)]$, lo que acaba la prueba pues el último sumando es nulo por (b). □

***Observación* 8.2.** No es cierto, sin embargo, que dos v.a.r. incorreladas sean independientes. Basta considerar el espacio de probabilidad $([-\pi,\pi], \mathcal{R}([-\pi,\pi], \frac{1}{2\pi}dx)$ y, sobre él, las v.a.r. $X(\omega) = \cos\omega$ e $Y(\omega) = \operatorname{sen}\omega$; se prueba fácilmente que X e Y son incorreladas, pero no independientes, pues, por ejemplo, $P(\{X \geq 1/2\} \cap \{Y \geq 1/2\}) \neq P(\{X \geq 1/2\})P(\{Y \geq 1/2\})$.

Nos preguntamos ahora por la distribución de la suma de v.a. n-dimensionales independientes. Necesitamos un nuevo concepto.

DEFINICIÓN 8.3. (Convolución) Sean P_1 y P_2 dos probabilidades en $\mathcal{R}^n$. ([2]) Consideremos la aplicación

$$S : (x,y) \in \mathbb{R}^n \times \mathbb{R}^n \longrightarrow S(x,y) := x + y \in \mathbb{R}^n.$$

[1] Recordemos que, a los efectos de la teoría de la medida, $\mathbb{C}$ se identifica con $\mathbb{R}^2$, y que la esperanza de una v.a. compleja X se define como $E(X) = E(\mathrm{Re}\,X) + iE(\mathrm{Im}\,X)$.

[2] No habría ningún problema en considerar que P_1 y P_2 son medidas finitas en $\mathcal{R}^n$.

Se define el producto de convolución $P_1 * P_2$ de P_1 y P_2 mediante

$$P_1 * P_2 := (P_1 \times P_2)^S.$$

***Observación* 8.3.** Si $B \in \mathcal{R}^n$, se verifica que

$$\begin{aligned}(P_1 * P_2)(B) &= (P_1 \times P_2)(S^{-1}(B)) = (P_1 \times P_2)(\{(x,y)\colon x+y \in B\}) \\ &= \int I_B(x+y)d(P_1 \times P_2)(x,y) = \int\int I_{B-x}(y)dP_2(y)dP_1(x) \\ &= \int P_2(B-x)dP_1(x) = \int P_1(B-y)dP_2(y).\end{aligned}$$

TEOREMA 8.5. *Si X_1 y X_2 son v.a. n-dimensionales independientes y $Z = X_1+X_2$, entonces*

$$P^Z = P^{X_1} * P^{X_2}.$$

Demostración. Si $X = (X_1, X_2)$, siendo X_1 y X_2 independientes, se tiene que $P^X = P^{X_1} \times P^{X_2}$. Si $B \in \mathcal{R}^n$,

$$\begin{aligned}P^Z(B) &= P(X_1 + X_2 \in B) = P(S \circ X \in B) \\ &= P(X \in S^{-1}(B)) = (P^{X_1} \times P^{X_2})(S^{-1}(B)) \\ &= (P^{X_1} * P^{X_2})(B).\end{aligned}$$

□

***Observación* 8.4.** Si X_1 y X_2 son v.a. n-dimensionales independientes con densidades f_1 y f_2 (respecto a la medida de Lebesgue λ_n), entonces

$$\frac{d(P^{X_1} \times P^{X_2})}{d(\lambda_n \times \lambda_n)}(x,y) = f_1(x)f_2(y).$$

Si $B \in \mathcal{R}^n$,

$$\begin{aligned}(P^{X_1} * P^{X_2})(B) &= \int\int I_B(x+y)f_1(x)f_2(y)dydx \\ &= \int\left(\int I_B(x+y)f_2(y)dy\right)f_1(x)dx\end{aligned}$$

Puesto que la medida de Lebesgue es invariante por traslaciones

$$\int I_B(x+y)f_2(y)dy = \int I_B(y)f_2(y-x)dy.$$

Luego

$$(P^{X_1} * P^{X_2})(B) = \int \left(\int I_B(y) f_2(y-x) dy \right) f_1(x) dx$$
$$= \int_B \left(\int f_2(y-x) f_1(x) dx \right) dy$$

Por tanto,

$$\frac{d(P^{X_1} * P^{X_2})}{d\lambda_n}(y) = \int f_2(y-x) f_1(x) dx.$$

A la aplicación $(f_1 * f_2)(y) := \int f_2(y-x)f_1(x)dx$, que como acabamos de ver es densidad de $P^{X_1} * P^{X_2}$, le llamaremos producto de convolución de f_1 y f_2. Se verifica que $f_1 * f_2 = f_2 * f_1$, λ_n-c.s.

***Observación* 8.5.** Un argumento análogo prueba que si X_1 y X_2 son v.a. independientes a valores en $\mathbb{Z}$ con funciones de probabilidad respectivas f_1 y f_2, entonces la función de probabilidad de la v.a. discreta $Z = X_1 + X_2$ es la aplicación

$$f : k \in \mathbb{Z} \longrightarrow f(k) := \sum_{m \in \mathbb{Z}} f_2(k-m) f_1(m).$$

PROBLEMAS DEL CAPÍTULO 8

Problema 8.1. En dos lanzamientos de una moneda perfecta, consideremos los sucesos A ="el primer lanzamiento es cara", B ="el segundo lanzamiento es cara"y C ="el primer lanzamiento es igual al segundo lanzamiento". Probar que esos sucesos son dos a dos independientes, pero no globalmente independientes.

Problema 8.2. En dos lanzamientos de un dado perfecto, consideremos los sucesos A ="el segundo lanzamiento es 1,2, o 5", B ="el segundo lanzamiento es 4,5 o 6"y C ="la suma de las caras es 9". Probar que $P(A\cap B\cap C) = P(A)P(B)P(C)$ pero que esos sucesos no son globalmente independientes.

Problema 8.3. Probar que los sucesos A y A^c son independientes si, y sólo si, $P(A) = 0$ o $P(A) = 1$.

Problema 8.4. Sea P la medida de Lebesgue en $[0,1]$. Probar que, en el espacio de probabilidad $([0,1], \mathcal{R}[0,1], P)$ las v.a. $f(x) = x$ y $g(x) = \sin x$ no son independientes. No obstante, en el espacio producto $([0,1]^2, \mathcal{R}[0,1]^2, P^2)$, las v.a. $F(x,y) = f(x)$ y $G(x,y) = \sin y$ sí son independientes.

Problema 8.5. Probar que si A, B y C son sucesos independientes, también lo son $A\cup B$ y C.

Problema 8.6. Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, I un conjunto de índices y, para cada $i \in I$, $\mathcal{C}_i$ un subconjunto de $\mathcal{A}$. Las familias de conjuntos $\mathcal{C}_i$ se dicen independientes si lo son los sucesos de cada familia $(C_i)_{i\in I}$ tales que $C_i \in \mathcal{C}_i$, $\forall i \in I$. Probar que si las familias de conjuntos $\mathcal{C}_i$ son independientes, también lo son las familias obtenidas de ellas añadiendo:

1. Las diferencias propias $A \setminus B$, $A, B \in \mathcal{C}_i$, $A \subset B$.
2. $\emptyset$ y Ω.
3. Complementarios de conjuntos en $\mathcal{C}_i$.
4. Uniones numerables disjuntas de conjuntos en $\mathcal{C}_i$.
5. Límites de sucesiones monótonas en $\mathcal{C}_i$.

Proporcionar un ejemplo que pruebe que las intersecciones finitas no pueden ser añadidas.

Problema 8.7. Si X, Y, Z son v.a. n-dimensionales independientes, probar la independencia de las v.a. $X+Y$ y Z, y la de las v.a. $f(X,Y)$ y Z, donde $f : \mathbb{R}^n \times \mathbb{R}^n \to (\Omega', \mathcal{A}')$ es una función $(\mathcal{R}^n \times \mathcal{R}^n, \mathcal{A}')$-medible.

Problema 8.8. (★) Sean $A_1, \dots, A_n$ sucesos independientes y $p = P(A_1 \cup \cdots \cup A_n)$. Probar que $1 - p \leq \exp\{-\sum_{i=1}^n P(A_i)\}$.

Problema 8.9. (Distribución binomial $b_n(p)$) Consideremos el experimento aleatorio que consiste en n lanzamientos independientes de una moneda en la que la probabilidad de sacar cara es $p \in]0,1[$; describir el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ correspondiente, identificar los lanzamientos con v.a. X_i, $1 \leq i \leq n$, sobre él y probar que la v.a. $S = \sum_{i=1}^n X_i$, que describe el número de caras obtenidas en esos lanzamientos, verifica que

$$P(S = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Denotaremos $b_n(p)$ la distribución de S.

Problema 8.10. (Distribución de Bernouilli) La distribución de Bernouilli de parámetro $p \in]0,1[$ es la distribución binomial $b_1(p)$, es decir, es la distribución de probabilidad de una v.a. que sólo toma los valores 0 y 1 con probabilidades $1 - p$ y p, resp. Probar, por inducción en n, que la distribución de la suma de n v.a. independientes cada una de ellas con distribución de Bernouilli de parámetro p es la distribución binomial $b_n(p)$.

Problema 8.11. (↑) (Distribución geométrica o de Pascal y distribución binomial negativa) Sea (X_n) una sucesión de Bernouilli de parámetro $p \in]0,1[$, es decir, una sucesión de v.a. independientes e idénticamente distribuidas (brevemente, iid) definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ cuya distribución común es la distribución $b_1(p)$.

(a) Se define la distribución geométrica $\bar{b}_1(p)$ de parámetro p como la distribución de la v.a. $Y = \min\{j \geq 1 \colon X_j = 1\} - 1$ (es decir, Y es el tiempo de espera hasta la ocurrencia del primer éxito, entendiendo que éxito=1 y fracaso=0). Probar que $P(Y = k) = p(1-p)^k$, $k = 0, 1, 2, \dots$ y que $E(Y) = p^{-1}$ y $\mathrm{Var}(Y) = (1-p)p^{-2}$.

(b) Dado $m = 1, 2, \dots$, se define la distribución binomial negativa $\bar{b}_m(p)$ como la distribución de la variable $Y = \min\{n \geq 1 \colon \sum_{i=1}^n X_i = m\} - m$ (es decir, Y es el número de fracasos necesarios para la obtención de exactamente m éxitos). Probar que

$$P(Y = k) = \binom{m+k-1}{k} p^m (1-p)^k, \quad k = 0, 1, 2, \dots,$$

y que la media y la varianza de esa distribución son $m(1-p)p^{-1}$ y $m(1-p)p^{-2}$, resp.

(c) Supongamos que X sigue una distribución geométrica $\bar{b}_1(p)$. Probar que cualesquiera que sean n y m enteros no negativos:

$$P(X \geq n+m|X \geq n) = P(X \geq m).$$

Probar que el recíproco también es cierto.

> *Observación:* Esta propiedad es conocida como "pérdida de memoria". La distribución geométrica se utiliza para la distribución del número de fracasos habidos antes de que se produzca el primer éxito, en la realización de pruebas independientes de un experimento, luego la información de que ningún éxito se ha producido en m intentos se olvida en las siguientes realizaciones.

Problema 8.12. (Distribución exponencial y distribución de Poisson) La distribución exponencial de parámetro $\lambda > 0$ en $\mathbb{R}$ es la distribución gamma $G(1, \lambda)$ cuya densidad respecto a la medida de Lebesgue es

$$f_\lambda(x) = \lambda^{-1} \exp\{-\lambda^{-1}x\} I_{[0,+\infty[}(x)$$

Su media es λ y su varianza es λ^2. La distribución de Poisson de parámetro $\lambda > 0$ en $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ queda definida por su densidad respecto a la medida cardinal en $\mathbb{N}_0$:

$$P_\lambda(\{k\}) = e^{-\lambda}\lambda^k/k!, \quad k = 0, 1, 2, \dots$$

(a) Sea $T : (\Omega, \mathcal{A}, \mathcal{P}) \longrightarrow \mathbb{R}$ una v.a.r. con distribución exponencial de parámetro $\lambda > 0$. Probar que, para cada $t \geq 0$, $P(T > t) > 0$, y que, si $t, s \geq 0$, $P(T > t + s|T > t) = P(T > s)$.

(b) Recíprocamente, sea $T : (\Omega, \mathcal{A}, \mathcal{P}) \longrightarrow \mathbb{R}$ una v.a.r. que satisface $P(T > t) > 0$, $\forall t \geq 0$ y $P(T > t + s|T > t) = P(T > s)$, $\forall t, s \geq 0$. Probar que T tiene distribución exponencial de parámetro mayor que cero.

Observación. La distribución exponencial se utiliza en la práctica como distribución del "tiempo de espera" hasta la ocurrencia de un cierto suceso (p. ej., la llegada de un cliente a una ventanilla, de una llamada a un cierto número de teléfono, de un coche a un cierto punto de la calzada). La probabilidad condicional $P(T > t + s|T > t)$ representa la probabilidad de que el tiempo sea superior a $t+s$ supuesto conocido ya que $T > t$; la ecuación $P(T > t+s|T > t) = P(T > s)$ atribuye entonces al tiempo de espera un mecanismo de "pérdida de memoria" en el sentido de que, si después de un cierto tiempo t el suceso aún no ha ocurrido, el tiempo que falta para que ocurra se distribuye condicionalmente de la misma forma que T. Este problema pretende hacernos ver cómo podemos usar ciertos resultados probabilísticos a la hora de justificar el uso de una cierta estructura estadística (en este caso, la distribución desconocida sería exponencial) como modelo para el estudio de un cierto fenómeno aleatorio. En ese mismo sentido habíamos utilizado el teorema del límite central para justificar la hipótesis de normalidad.

(c) Consideremos ahora una sucesión de sucesos. Sea T_1 el tiempo de espera hasta que ocurra el primer suceso, T_2 el tiempo que transcurre entre la ocurrencia del primer suceso y la del segundo,... Hagamos $S_n = T_1 + \cdots + T_n$, para $n \geq 1$. Es conveniente escribir $S_0 = 0$. Supongamos que

$$0 = S_0(\omega) < S_1(\omega) < S_2(\omega) < \dots, \quad \forall \omega,$$

(que viene a decir que dos sucesos no pueden ocurrir simultáneamente). Hagamos también $X_0 = 0$ y, para $t > 0$, denotemos por X_t el número de sucesos que ocurren en el intervalo $[0, t]$, es decir,

$$X_t(\omega) = \text{máx}\{n \in \mathbb{N} \colon S_n(\omega) \leq t\}.$$

Notar que

$$\{\omega \in \Omega \colon X_t(\omega) \geq n\} = \{\omega \in \Omega \colon S_n(\omega) \leq t\}.$$

Supongamos que las v.a.r. T_n tienen la misma distribución y que esa distribución común es la exponencial de parámetro $\lambda > 0$ (el apartado (b) parece querer convencernos de que esa suposición es "razonable"). Supondremos también que las T_n son independientes. Probar que entonces X_t tiene distribución de Poisson de parámetro t/λ.

Indicación: S_n tiene distribución gamma $G(n, \lambda)$ y, para cada $t \geq 0$,

$$P(X_t \geq n) = P(S_n \leq t) = \int_0^t \frac{1}{\lambda^n \Gamma(n)} x^{n-1} e^{-x/\lambda} dx = 1 - e^{-t/\lambda} \sum_{k=0}^{n-1} \frac{(t/k)^k}{k!},$$

donde la última igualdad se prueba derivando ambos miembros de la misma.

Observación. Parece pues justificado usar la distribución de Poisson para describir el número de sucesos que ocurren en un cierto intervalo de tiempo (por ejemplo, número de partículas radiactivas que golpean una placa en un periodo de tiempo, número de llamadas telefónicas recibidas en un día, número de vehículos que pasan por un punto concreto entre las dos y las tres de la tarde). A la familia de v.a.r. $(X_t)_{t \geq 0}$ se le llama proceso de Poisson de parámetro $1/\lambda$.

Problema 8.13. Sea N una variable aleatoria $\mathbb{N}_0$-valorada y no negativa. Supongamos que N bolas son introducidas (independientemente unas de otras), bien en una urna A con probabilidad p, o bien en una urna B con probabilidad $1-p$ ($0<p<1$), resultando N_A bolas en la urna A y $N_B = N - N_A$ en la urna B. Probar que si N sigue una distribución de Poisson, entonces N_A y N_B son variables aleatorias independientes.

Problema 8.14. Probar que el producto de convolución es conmutativo y asociativo.

Problema 8.15. Probar que la medida de Dirac ε_0 en $0 \in \mathbb{R}^n$ es el elemento unidad para el producto de convolución.

Problema 8.16. Probar que, si $0 \leq p \leq 1$ y P_1, P_2 y P son probabilidades, entonces

$$(pP_1 + (1-p)P_2) * P = p(P_1 * P) + (1-p)(P_2 * P).$$

Problema 8.17. Se lanzan al aire un dado con caras numeradas del 1 al 6 y una moneda perfectos y se considera la v.a. S suma de los dos resultados obtenidos. Determinar la distribución de probabilidad de S.

Problema 8.18. Sean X e Y v.a.r. independientes e idénticamente distribuidas con distribución común uniforme en $[0,1]$. Determinar la distribución de $X+Y$.

Problema 8.19. (Entropía) En este problema, y en los próximos, Ω será un conjunto numerable, $\mathscr{P}(\Omega)$ la σ-álgebra de las partes de Ω y P una probabilidad en Ω. Denotamos también $\Omega_0 = \{\omega \in \Omega \colon P(\omega) > 0\}$.

Def.: Se define la entropía $H(P)$ de P mediante

$$H(P) = -\sum_{\omega \in \Omega} P(\omega) \cdot \log P(\omega)$$

(con el convenio $0 \cdot \log 0 = 0$). Probar que $H(P)$ es positiva. ¿Cuándo se anula $H(P)$?

Problema 8.20. (↑) (Información de Kullback) Sea Q otra probabilidad en $\mathscr{P}(\Omega)$ dominada por P. Se define la *Información de Kullback* de P sobre Q por

$$K(P,Q) = -\sum_{\omega \in \Omega} P(\omega) \cdot \log \frac{Q(\omega)}{P(\omega)}$$

(con el convenio $0/0 = 0$, $0 \cdot \log(0/0) = 0$; se admite también la aritmética usual en $\bar{\mathbb{R}}$). Probar que $K(P,Q) \geq 0$ y que, si Ω es finito, entonces $K(P,Q) = 0$ si, y sólo si, $P \equiv Q$.

Problema 8.21. (↑) Sean $\Omega = \{1, \dots, n\}$ y Q la distribución uniforme en Ω. Calcular $H(Q)$ y probar que, si P es otra probabilidad en Ω, entonces $H(P) < H(Q)$.

Indicación: $H(Q) = -n \cdot n^{-1} \cdot \log n^{-1} = \log n$. Si $P(k) = p_k > 0$, $1 \le k \le n$, debemos probar que

$$-\sum_{k=1}^{n} p_k \log p_k \le \log n$$

y para ello utilizaremos el método de los multiplicadores de Lagrange:

> "Sean $m, n \in \mathbb{N}$ tales que $m > n$, U abierto$\subset \mathbb{R}^m$ y $\phi : U \longrightarrow \mathbb{R}$ y $f : U \longrightarrow \mathbb{R}^n$ funciones de clase C^1 en U. Sean $M = \{x \in U : f(x) = 0\}$ y $c \in M$. Supongamos que la restricción $\phi_{|M}$ de ϕ a M tiene un extremo relativo en c y que el rango de la matriz jacobiana $\left(\frac{\partial f_i}{\partial x_j}(c)\right)_{1 \le i \le n, 1 \le k \le m}$ en el punto c es n. Entonces existe un único $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ tal que
>
> $$\frac{\partial \phi}{\partial x_j}(c) + \sum_{i=1}^{n} \lambda_i \cdot \frac{\partial \phi}{\partial x_j}(c) = 0, \quad 1 \le j \le m.$$
>
> Si ϕ y f admiten derivadas parciales continuas en U entonces, para que $\phi_{|M}$ tenga un mínimo (resp., máximo) relativo en el punto $c \in M$ e suficiente que $D^2L(c)(h,h) > 0$ (resp., < 0) cada vez que $h \in \mathbb{R}^m \setminus \{0\}$ satisface que $Df_i(c)(h) = 0$, $1 \le i \le n$, donde $L = \phi + \sum_{i=1}^{n} \lambda_i f_i$."

En nuestro caso hagamos $\phi(x_1, \dots, x_n) = -\sum_{i=1}^{n} x_i \log x_i$, $U = \{x \in \mathbb{R}^n : x_i > 0, \forall i\}$, $f(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i - 1$; se trata de resolver el siguiente sistema, siendo $\lambda, c_1, \dots, c_n$ las incógnitas:

$$\begin{cases} \frac{\partial \phi}{\partial x_j}(c) + \lambda \cdot \frac{\partial f}{\partial x_j}(c) = 0, \quad 1 \le j \le n \\ \sum_{i=1}^{n} c_i = 1 \end{cases}$$

Echando cuentas se obtiene la relación $-1 - \log c_j + \lambda = 0$, $\forall j$ y, entonces, $c_1 = \cdots = c_n$. Por tanto, $c_i = n^{-1}$, $1 \le i \le n$, y $\lambda = -1 + \log n$. Así pues

$$L(x_1, \dots, x_n) = -\sum_{i=1}^{n} x_i \log x_i + (-1 + \log n) \cdot \sum_{i=1}^{n} x_i.$$

Compruébese con el teorema entre comillas que el punto c corresponde a un máximo relativo condicionado y concluir que $-\sum_{k=1}^{n} p_k \log p_k \le \log n$ si los p_k son > 0 y suman 1. Si algún p_k es nulo y $\sum p_k = 1$, entonces $-\sum_{k=1}^{n} p_k \log p_k \le \log(n-1) \le \log n$. Ello prueba que la distribución uniforme en $\{1, \dots, n\}$ da la máxima entropía. Nótese además que, si $P \ne Q$, la desigualdad es estricta pues sólo hay un máximo relativo condicionado que se encuentra en el punto $(1/n, \dots, 1/n)$.

Problema 8.22. ($\uparrow$) Sean $X : \Omega \longrightarrow \Omega'$ una v.a. y $\Omega_0 = \{\omega \colon P(\omega) > 0\}$. Se define $H(X) = H(P^X)$. Probar

(1) Si $X_{|\Omega_0}$ es inyectiva, entonces $H(X) = H(P)$.

(2) Si $X_{|\Omega_0}$ no es inyectiva, entonces $H(X) < H(P)$.

Observación. La entropía de una v.a. se interpreta como una medida de la incertidumbre sobre el valor que toma esa variable. Esa incertidumbre es máxima para la distribución uniforme, según prueba el Problema 8.21.

Problema 8.23. (↑) Sean $X : \Omega \longrightarrow \Omega'$, $Y : \Omega \longrightarrow \Omega$ y $f : \Omega' \longrightarrow \Omega$ tales que $Y = f \circ X$. Probar que $H(Y) \leq H(X)$.

Problema 8.24. (↑) Sean X e Y v.a.r. definidas en Ω y, para $y \in Y(\Omega_0)$, sea $P(X = \cdot|Y = y)$ la distribución condicional de X respecto a $\{Y = y\}$. Probar que

$$H(X,Y) = H(Y) + \sum_{y \in Y(\Omega_0)} P(Y = y) \cdot H(P(X = \cdot|Y = y)).$$

Observación-Definición. Entonces $H(X,Y) - H(Y)$ es una media de las entropías de las distribuciones condicionales de X respecto a los distintos valores de Y; se denotará $H(X|Y) = H(X,Y) - H(Y)$ y se llamará *entropía condicional* de X respecto a Y. $H(X|Y)$ se interpreta como la incertidumbre media que nos queda sobre el valor de X cuando conocemos el valor de Y. Con esas notaciones, la entropía conjunta $H(X,Y)$ coincide con $H(Y) + H(X|Y)$ y con $H(X) + H(Y|X)$. Deducir de ahí que $\sup(H(X), H(Y)) \leq H(X,Y)$.

Problema 8.25. (↑) Con las mismas notaciones que en el Problema 8.24, probar que $H(Y|X)$ se anula si, y sólo si, $Y_{|\Omega_0}$ es función de $X_{|\Omega_0}$.

Problema 8.26. (↑) Si X e Y son independientes, entonces $H(X,Y) = H(X) + H(Y)$.

Problema 8.27. (↑) Un jugador trata de adivinar lo antes posible un número Y que su adversario ha elegido al azar entre 0 y $N-1$. Para ello puede hacer preguntas que su adversario responderá SI o NO. Después de n preguntas, él conoce el valor de una función X a valores en $\{\mathrm{SI}, \mathrm{NO}\}^n$. Probar (comparando la entropía de Y con la entropía máxima de X) que, por muy astuto que sea, el jugador no puede estar seguro de determinar Y antes de haber hecho un número de preguntas $n \geq (\log N)/(\log 2)$.

Problema 8.28. (★) (Una demostración probabilística del teorema de aproximación de Weierstrass) Supondremos conocido que toda función real continua en un compacto es uniformemente continua. Con esto y con la desigualdad de Chebyshev se puede probar que toda función real continua en $[0,1]$ es límite uniforme de una sucesión de funciones polinómicas. Vayamos por partes.

(a) Sean $X_1, \dots, X_n$ v.a. independientes de Bernouilli de parámetro p definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ que fijaremos en lo sucesivo y hagamos $S_n = \sum_{i=1}^n X_i$. Calcular la media y la varianza de S_n/n. Recordar que S_n tiene distribución binomial $b_n(p)$ de parámetros n y p (cosa que, por otra parte, se prueba sin dificultad por inducción en n). Nota: Cabe preguntarse si existen v.a. $X_1, \dots, X_n$ como las requeridas; en el Teorema 8.2, haciendo uso de espacios producto, se ha descrito un método para construir tantas como podamos necesitar.

(b) Sea $u : [0,1] \to \mathbb{R}$ una función continua. Probar que

$$E[u(S_n/n)] = u_n(p),$$

donde $u_n(p) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} u(k/n)$.

(c) Podemos reformular la desigualdad de Chebyshev (ver Problema 5.3) del siguiente modo: si X es una v.a.r. en un espacio de probabilidad y $\eta > 0$, entonces

$$P(|X - E(X)| \geq \eta) \leq \frac{\operatorname{Var}(X)}{\eta^2}.$$

Siendo u uniformemente continua en $[0,1]$, dado $\epsilon > 0$ podemos elegir $\eta > 0$ de forma que

$$|x - y| < \eta, \;\; x, y \in [0,1] \Rightarrow |u(x) - u(y)| < \epsilon.$$

Hagamos $A_\epsilon = \{\omega / |\frac{1}{n} S_n(\omega) - p| \geq \eta\}$. Probar que

$$P(A_\epsilon) \leq \frac{p(1-p)}{n\eta^2} \leq \frac{1}{4n\eta^2}.$$

(d) Sea $M > 0$ tal que $|u(x)| \leq M, \forall x \in [0,1]$. Probar que

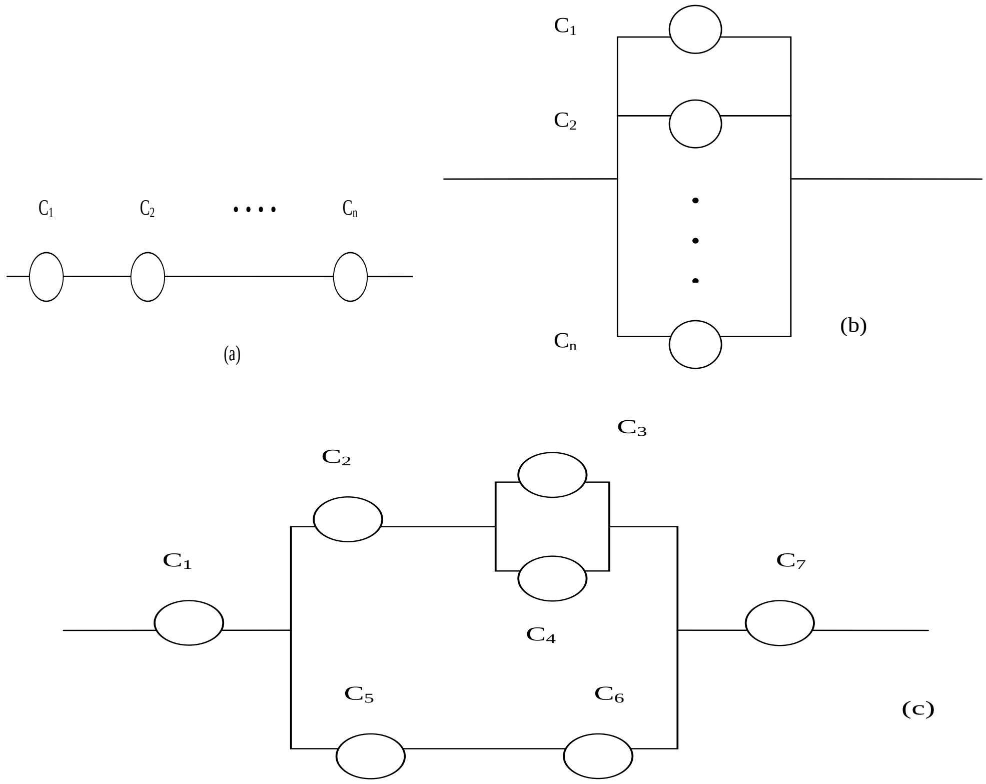
$$u_n(p) - u(p) = E\left[\left(u\left(\frac{1}{n}S_n\right) - u(p)\right) I_{A_\epsilon}\right] + E\left[\left(u\left(\frac{1}{n}S_n\right) - u(p)\right) I_{A_\epsilon^c}\right]$$

y acotar

$$|u_n(p) - u(p)| \leq 2MP(A_\epsilon) + \epsilon P(A_\epsilon^c) \leq \frac{2M}{4n\eta^2} + \epsilon.$$

(e) Notando que $u_n(p)$ es una función polinómica de p concluir que u es límite uniforme de una sucesión de polinomios.

Problema 8.29. (Fiabilidad) El estado de un circuito eléctrico que contiene n componentes $C_1, \dots, C_n$ queda descrito por un punto $\omega = (\omega_1, \dots, \omega_n) \in \{0,1\}^n$, donde $\omega_i = 1$ o 0 según que la componente i-ésima funcione o no, $1 \leq i \leq n$. Dependiendo de cómo se hayan establecido las conexiones, si alguna componente falla, el circuito podrá o no funcionar. Se introduce una función $f : \{0,1\}^n \longrightarrow \{0,1\}$ que vale 1 si el circuito funciona y 0 si no.



1. Probar que, si los componentes se han colocado en serie (circuito (a)), entonces $f(\omega_1,\dots,\omega_n)=\prod_{i=1}^n \omega_i$, y que, si están colocados en paralelo (circuito (b)), entonces $f(\omega_1,\dots,\omega_n)=1-\prod_{i=1}^n(1-\omega_i)$.

2. Probar que, para el circuito (c),
$$f(\omega_1,\dots,\omega_n)=\omega_1\cdot\big[1-(1-\omega_5\cdot\omega_6)\cdot(1-\omega_2\cdot[1-(1-\omega_3)\cdot(1-\omega_4)]\cdot\omega_7)\big]$$

3. Sea $p=(p_1,\dots,p_n)\in]0,1[^n$ y supongamos que el estado de las n componentes del circuito son v.a. independientes $X_1,\dots,X_n$ de Bernouilli de parámetros respectivos $p_1,\dots,p_n$. Determinar la distribución de la v.a. $f(X_1,\dots,X_n)$ y su media para cada uno de los tres circuitos (a), (b) y (c).

Problema 8.30. Sean X e Y v.a.r. independientes y $f:\mathbb{R}^2\to\mathbb{R}$ medible no negativa. Se define $g(x)=E[f(x,Y)]$. Probar que $E[g(X)]=E[f(X,Y)]$.

Problema 8.31. Sean Q la probabilidad en $\{0,1\}$ definida por $Q(0)=Q(1)=1/2$ y P la probabilidad producto $Q^{\mathbb{N}}$. Si T es la aplicación definida en el apartado (f) del Problema 2.13, probar que P^T es la medida de Lebesgue en el intervalo $[0,1]$.([3]) *Observación:* T coincide P-c.s. con la aplicación $T':\omega=(\omega_i)\in\{0,1\}^{\mathbb{N}}\mapsto T'(\omega):=\sum_{i=1}^n \frac{\omega_i}{2^i}\in[0,1]$.

[3] El Problema 10.21 presenta otras descripciones similares de la medida de Lebesgue.

Problema 8.32. (↑) Calcular $\int_0^1 x^2 dx$ utilizando la representación de la medida de Lebesgue en [0,1] del ejercicio anterior, y el teorema de la medida imagen y el de Fubini en el caso de un producto infinito numerable de factores.

Capítulo 9

Función Característica

DEFINICIÓN 9.1. (a) (Función característica de una probabilidad) Si P es una probabilidad en $\mathcal{R}$, se define su función característica φ_P como la aplicación

$$\varphi_P : t \in \mathbb{R} \to \varphi_P(t) := \int e^{itx} dP(x) \in \mathbb{C}.$$

(b) Si X es una v.a.r. en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se define su función característica φ_X como la función característica de su distribución de probabilidad P^X, es decir

$$\varphi_X : t \in \mathbb{R} \to \varphi_X(t) := E(e^{itX}) = \int_\Omega e^{itX(\omega)} dP(\omega) = \int_\mathbb{R} e^{itx} dP^X(x) \in \mathbb{C}.$$

***Observación* 9.1.** e^{itX} es una v.a. compleja acotada (de hecho, tiene módulo 1 en todo punto ω: $e^{itX(\omega)} = \cos tX(\omega) + i \operatorname{sen} tX(\omega))$ y, por tanto, la integral que define $\varphi_X(t)$ existe y es finita, pues, si Z es una v.a. compleja, entonces $|E(Z)| \leq E(|Z|)$, pues, si $\theta \in [0, 2\pi[$ es tal que $|E(Z)| = E(Z)e^{i\theta}$ e $Y = Ze^{i\theta} = Y_1 + iY_2$, entonces $E(Y) = E(Y_1) \in \mathbb{R}$ y, siendo $Y_1 \leq |Y|$, $|E(Z)| = E(Y) = E(Y_1) \leq E(|Y|) = E(|Z|)$.

***Observación* 9.2.** Si X es una v.a.r. discreta que toma los valores $x_1, x_2, \dots$ con probabilidades $p_1, p_2, \dots$, resp., entonces

$$\varphi_X(t) = \sum_{k \geq 1} p_k e^{itx_k}.$$

Si X es una v.a.r. absolutamente continua respecto a la medida de Lebesgue con densidad f, entonces

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx.$$

***Observación* 9.3.** La función característica de una v.a. sólo depende de ésta a través de su distribución, es decir, si X e Y son v.a.r. con la misma distribución de probabilidad, entonces $\varphi_X = \varphi_Y$.

***Observación* 9.4.** En un contexto no probabilístico, la función característica se conoce como transformada de Fourier.

La siguiente proposición resulta útil a menudo a la hora de determinar la distribución de la suma de v.a. independientes.

Proposición 9.1. *Sean $X_1, \ldots, X_n$ v.a.r. independientes y $S = \sum_{j=1}^n X_j$. Entonces*

$$\varphi_S(t) = \prod_{j=1}^n \varphi_{X_j}(t).$$

Demostración. La demostración en consecuencia inmediata de que la esperanza del producto de v.a. complejas independientes es el producto de las esperanzas. □

Proposición 9.2. *Si X es una v.a. compleja con media finita entonces $|E(X)| \leq E(|X|)$.*

Demostración. La aplicación $f : (x, y) \in \mathbb{C} = \mathbb{R}^2 \longrightarrow f(x, y) = \sqrt{x^2 + y^2} \in \mathbb{R}$ es continua y convexa. El resultado se deduce de la desigualdad de Jensen (Teorema 14.1). □

La siguiente proposición recoge algunas propiedades de las funciones características.

Proposición 9.3. *Sea X una v.a.r.*

(a) *$|\varphi_X(t)| \leq 1$, para todo $t \in \mathbb{R}$.*

(b) *$\varphi_X(0) = 1$.*

(c) *$\varphi_X(-t) = \overline{\varphi_X(t)}$, para todo $t \in \mathbb{R}$.*

(d) *La función característica de $aX + b$ es $e^{ibt}\varphi_X(at)$.*

(e) *φ_X es uniformemente continua en $\mathbb{R}$.*

(f) *Si la distribución de X es simétrica respecto al origen (es decir, si $P^X(B) = P^X(-B)$, para cada boreliano B, o, lo que es lo mismo, si $P^X = P^{-X}$), entonces $\varphi_X(t) \in \mathbb{R}$, para todo $t \in \mathbb{R}$.*

(g) *Si X tiene momento absoluto de orden $n \geq 1$ finito, entonces φ_X admite derivada n-ésima continua en $\mathbb{R}$ y*

$$\varphi_X^{(k)}(t) = \int_{\mathbb{R}} (ix)^k e^{itx} dP^X(t), \quad 1 \leq k \leq n.$$

En particular, $\varphi_X^{(k)}(0) = i^k E(X^k)$, $1 \leq k \leq n$.

Demostración. (a), (b), (c) y (d) son triviales.

(e) Si $t, h \in \mathbb{R}$, entonces

$$|\varphi_X(t+h) - \varphi_X(t)| \leq \int |e^{ihx} - 1| dP^X(x),$$

y, por tanto,

$$\sup_{t\in\mathbb{R}} |\varphi_X(t+h) - \varphi_X(t)| \leq \int |e^{ihx} - 1| dP^X(x).$$

El integrando en el segundo miembro está acotado por 2 y tiende a cero cuando $h \to 0$. El teorema de la convergencia dominada (Teorema 4.5) prueba ya que φ_X es uniformemente continua.

(f) Puesto que X y $-X$ tienen la misma distribución, tienen también la misma función característica y, por tanto,

$$\varphi_X(t) = \varphi_{-X}(t) = E(e^{-itX}) = \varphi_X(-t) = \overline{\varphi_X(t)},$$

que prueba que $\varphi_X(t) \in \mathbb{R}$.

(g) Si $E(X^n)$ es finito, también lo es $E(X^k)$, para $1 \leq k \leq n$. Consideremos el caso $k = 1$. Puesto que $\int |x| dP^X(x) < +\infty$, existe y es finita $\int x e^{itx} dP^X(x)$. Entonces, si $h \neq 0$,

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} - \int ixe^{itx} dP^X(x) = \int e^{itx} \left(\frac{e^{ihx} - 1}{h} - ix \right) dP^X(x).$$

Puesto que la función $2 - 2\cos t - t^2$ vale 0 en 0 y tiene un máximo absoluto en 0, se tiene que $(\cos t - 1)^2 + \operatorname{sen}^2 t \leq t^2$, y de aquí se deduce que

$$\left| e^{it} - 1 \right|^2 \leq t^2, \quad t \neq 0,$$

y, reemplazando t por hx,

$$\left|\frac{e^{ihx}-1}{h}\right| \leq |x|, \quad h \neq 0, x \neq 0.$$

Tomando límite cuando $h \to 0$, se deduce del teorema de la convergencia dominada (Teorema 4.5) que

$$\varphi'_X(t) = \int ixe^{itx} dP^X(x),$$

pues

$$\left|\frac{e^{ihx}-1}{h} - ix\right|$$

está acotada por $2|x|$ y converge a cero cuando $h \to 0$, para todo $x \in \mathbb{R}$. Que φ'_X es continua (incluso uniformemente continua) se prueba como en (e). $\square$

A modo de recíproco de la observación 3 anterior, se verifica el siguiente teorema, que no probaremos.

TEOREMA 9.1. *Dos v.a.r. con la misma función tienen característica también la misma distribución de probabilidad.*

***Observación* 9.5.** El teorema anterior es consecuencia de la llamada fórmula de la inversión, que viene a decir lo siguiente: "Sea P una probabilidad en $\mathcal{R}$, F su función de distribución y φ su función característica. Si $a < b$ son puntos de continuidad de F, entonces existe y es finito el límite

$$\lim_{T\to+\infty} \int_{-T}^{T} \mathrm{Re}\left(\varphi(t)\frac{e^{-ita}-e^{-itb}}{it}\right) dt.$$

Además, ese límite $2\pi(F(b) - F(a))$. Por último, si φ es integrable en $\mathbb{R}$ respecto a la medida de Lebesgue, entonces la función

$$f(x) := \frac{1}{2\pi}\int e^{-itx}\varphi(t)dt$$

es una densidad de P respecto a la medida de Lebesgue y $F' = f$." La demostración se puede encontrar en Ash (1972).

***Observación* 9.6.** (Función característica n-dimensional) Si X es una v.a. n-dimensional, su función característica φ_X se define mediante $\varphi_X(t) = E[e^{i<t,X>}] \in \mathbb{C}$, $t \in \mathbb{R}^n$, donde $< \cdot, \cdot >$ denota el producto escalar euclídeo en $\mathbb{R}^n$. Si X es una v.a. n-dimensional e Y una v.a. m-dimensional, se verifica que X e Y son independientes si $\varphi_{(X,Y)}(t,s) = \varphi_X(t)\varphi_Y(s)$, $\forall t \in \mathbb{R}^n$, $\forall s \in \mathbb{R}^m$. También en el caso n-dimensional la función característica de una v.a. determina unívocamente su distribución, y posee propiedades análogas a las probadas en el caso unidimensional: de hecho, vale 1 en 0, es uniformemente continua y si $\int |x_j| dP^X(x) < \infty$, para algún j, entonces φ_X admite derivada parcial j-ésima uniformemente continua y

$$\frac{\partial \varphi_X}{\partial t_j}(t) = i \int x_j e^{i<t,x>} dP^X(x).$$

También en el caso n-dimensional se puede enunciar una fórmula de inversión (ver Billingsley (1986)). Una consecuencia sencilla de lo dicho es la siguiente caracterización de independencia: "Sea $X = (X_1, \dots, X_n)$ una v.a. n-dimensional; para que las v.a. X_i sean independientes es necesario y suficiente que $\varphi_X(t) = \prod_{j=1}^n \varphi_{X_j}(t_j)$, para cada $t \in \mathbb{R}^n$.

***Observación* 9.7.** (Función generatriz de momentos) Si X es una v.a.r., se define su función generatriz de momentos (f.g.m. o transformada de Laplace) como la aplicación

$$M(t) := E(e^{tX}) = \int e^{tx} dP^X(x),$$

definida en el subconjunto de puntos $t \in \mathbb{R}$ que hacen finita la integral, entre los cuales se encuentra el 0; se prueba fácilmente que si la integral es finita para algún punto $t_0 \in \mathbb{R}$, también lo es para cualquier punto t tal que $|t| \le |t_0|$. Puede ocurrir, sin embargo, que M sólo sea finita en 0, y se prueba que si M es finita en un intervalo abierto $]-t_0, t_0[$, entonces M admite derivadas de todos los órdenes en ese intervalo y que $M^{(k)}(0) = E(X^k)$; se verifica incluso en ese intervalo que

$$M(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k).$$

En particular, en ese caso, X tiene momentos finitos de todos los órdenes y éstos se calculan derivando sucesivamente M en 0. Relacionado con lo dicho se encuentra el llamado problema de los momentos que pretende decidir bajo qué condiciones la distribución de probabilidad de una v.a.r. queda unívocamente determinada por sus momentos (supuestos todos ellos finitos); en ese sentido tenemos el siguiente resultado: "Si P es una probabilidad en $\mathcal{R}$ con momento $\alpha_k := \int x^k dP(x)$ finito para todo $k \ge 1$ y la serie de potencias $\sum_k \alpha_k r^k / k!$ tiene radio de convergencia positivo, entonces P es la única probabilidad en $\mathcal{R}$ con momentos $\alpha_1, \alpha_2, \dots$." Este resultado, junto con lo dicho anteriormente, prueba que si la función generatriz de momentos M de una v.a.r. X es finita en un entorno de cero, entonces M determina unívocamente la distribución de X. Como en el caso de la función característica, se prueba ahora que si $X_1, \dots, X_n$ son v.a.r. independientes con f.g.m. finita en un cierto entorno de cero, entonces la f.g.m. de la suma es el producto de las f.g.m. de las X_i. Podemos también definir la f.g.m. de una v.a. n-dimensional y probar que, si es finita en un entorno de cero, determina unívocamente la distribución de la misma. Se obtiene también una caracterización de independencia análoga a la obtenida para funciones características (exigiendo ahora que la f.g.m. de X sea finita en un entorno de cero). Todos estos resultados se pueden consultar en Billingsley (1995).

***Observación* 9.8.** (Función generatriz de probabilidad de una v.a.r. discreta) Si $X : (\Omega, \mathcal{A}, P) \to \mathbb{N}$ es una v.a. y $p_n = P(X = n)$, $n \geq 1$, se define su función generatriz de probabilidad como la aplicación $g(z) := E(z^X) = \sum_n p_n z^n$. Esta serie es convergente en $[-1, 1]$ y define una función de clase C^∞ en $]-1, 1[$. Es obvio que g caracteriza la distribución de X, pues los coeficientes de esa serie de potencias se pueden obtener a partir de las derivadas sucesivas de g en 0: concretamente, $p_n = n! g^{(n)}(0)$. Se prueba incluso que X tiene momento finito de orden $p \geq 1$ si, y sólo si, es finito el límite por la izquierda en 1 de la derivada p-ésima de g. En el caso $p = 1$, si X tiene media finita, $E(X) = \text{lím}_{z \to 1-} g'(z)$.

***Observación* 9.9.** Diremos que una v.a.r. X tiene distribución $N(\mu, \sigma^2)$ (ver Problema 7.5 y siguientes) si su función de densidad es

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

Su función característica es

$$\varphi_X(t) = \exp\left\{it\mu - \frac{1}{2}\sigma^2 t^2\right\}, \quad t \in \mathbb{R}.$$

***Observación* 9.10.** Una v.a.r. X se dice que tiene distribución *gamma* de parámetros α, β ($\alpha > 0$, $\beta > 0$), y se escribirá $X \sim G(\alpha, \beta)$, si su distribución admite como densidad la aplicación

$$[\Gamma(\alpha)\beta^\alpha]^{-1} y^{\alpha-1} e^{-y/\beta} I_{]0,+\infty[}(y).$$

Su función característica es

$$\varphi_X(t) = \frac{1}{(1-\beta i t)^\alpha}, \quad t \in \mathbb{R}.$$

Casos especiales de la distribución gamma son la distribución chi-cuadrado con n grados de libertad $\chi^2(n) := G(\frac{n}{2}, 2)$, y la distribución exponencial de parámetro $\lambda > 0$, $\mathcal{E}(\lambda) := G(1, \lambda)$.

PROBLEMAS DEL CAPÍTULO 9

Problema 9.1. Sea S una v.a. con distribución binomial $b_n(p)$.

1. Determinar su función característica.
2. Probar que si $X_1, \dots, X_n$ son v.a. independientes y $X_i \sim b_{n_i}(p)$ entonces $\sum_i X_i \sim b_{\sum_i n_i}(p)$. Nota: el símbolo $\sim$ en este contexto se leerá "tiene distribución".

Problema 9.2. (Distribución de Poisson $P(\lambda)$) Si $\lambda > 0$, se define la distribución de Poisson de parámetro λ como la distribución de probabilidad en $\mathbb{N}_0 := \{0, 1, 2, \dots\}$ con función de probabilidad

$$P(\lambda)(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, \dots .$$

Determinar la función característica de la distribución de Poisson y probar que si $X_1, \dots, X_n$ son v.a. independientes y $X_i \sim P(\lambda_i)$ entonces $\sum_i X_i \sim P(\sum_i \lambda_i)$.

Problema 9.3. Si $X_1, \dots, X_n$ son v.a.r. independientes y $X_i \sim G(\alpha_i, \beta)$, $1 \leq i \leq n$, entonces $\sum_{i=1}^n X_i$ tiene distribución $G(\sum_{i=1}^n \alpha_i, \beta)$. En particular, si las X_i son independientes y $X_i \sim \chi^2(k_i)$ entonces $\sum_{i=1}^n X_i \sim \chi^2(\sum_{i=1}^n k_i)$. Así, si las X_i son independientes y si $X_i \sim N(\mu_i, \sigma_i^2)$ entonces $\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2$ tiene distribución $\chi^2(n)$. Pruébese también que si $X_1, \dots, X_n$ son independientes y si $X_i \sim N(\mu_i, \sigma_i^2)$ entonces $\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$. Probar, en fin, que si X es $N(\mu, \sigma^2)$ y $a, b \in \mathbb{R}$ entonces $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Problema 9.4. (Distribución beta) Si $\alpha, \beta > 0$ se define la función beta de Euler

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx.$$

Nótese que la integral existe pues el integrando es una función medible no negativa; pruébese que es finita. Probar también que

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = B(\beta, \alpha). \tag{9.1}$$

Def.: Diremos que una v.a.r. X tiene distribución *beta* con parámetros $\alpha, \beta > 0$ (y escribiremos $X \sim B(\alpha, \beta)$) si P^X admite como densidad respecto a la medida de Lebesgue la aplicación

$$f(x) = B(\alpha, \beta)^{-1}x^{\alpha-1}(1-x)^{\beta-1}I_{]0,1[}(x).$$

Pruébese en fin que si $X \sim B(\alpha, \beta)$ entonces $1 - X \sim B(\beta, \alpha)$.

Indicación: Para probar (9.1) nótese que

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta) &= \left(\int_0^\infty x^{\alpha-1}e^{-x}dx\right)\left(\int_0^\infty y^{\beta-1}e^{-y}dy\right)\\ &= 4\left(\int_0^\infty u^{2\alpha-1}e^{-u^2}du\right)\left(\int_0^\infty v^{2\beta-1}e^{-v^2}dv\right)\\ &= 4\int_0^\infty\int_0^\infty u^{2\alpha-1}v^{2\beta-1}e^{-(u^2+v^2)}dudv;\end{aligned}$$

efectuar ahora el cambio de variables

$$T : (r,\theta) \in]0,\infty[\times]0,\pi/2[\longrightarrow (r\cos\theta, r\sin\theta) \in]0,\infty[^2$$

para obtener

$$\Gamma(\alpha)\Gamma(\beta) = 4\left(\int_0^\infty r^{2(\alpha+\beta)-1}e^{-r^2}\right)\left(\int_0^{\pi/2}(\cos\theta)^{2\alpha-1}(\sin\theta)^{2\beta-1}d\theta\right);$$

hacer ahora los cambios $r^2 = x$ y $\sin^2\theta = y$.

Problema 9.5. Probar que, si X e Y son v.a.r. independientes e idénticamente distribuidas con distribución $N(0,1)$, entonces X/Y tiene distribución de Cauchy.

Problema 9.6. (La distribución normal multivariante) En este problema y en los siguientes estudiamos la distribución normal en $\mathbb{R}^n$. En lo que sigue utilizaremos notación matricial como se indica a continuación: un punto $u \in \mathbb{R}^n$ y una v.a. n-dimensional X se escribirán en forma de matrices columna, y usaremos los símbolos u^t y X^t para las matrices fila correspondientes. Def.: (V.a. normal n-dimensional) Una v.a. n-dimensional X se dice normal si su función característica es de la forma

$$h(u) = \exp\{iu^tb - \frac{1}{2}u^tCu\}$$

donde $b \in \mathbb{R}^n$ y $C = (C_{jk})$ es una matriz real cuadrada de orden n simétrica (es decir, $C = C^t$) y semidefinida positiva (es decir, $u^tCu \geq 0$, $\forall u \in \mathbb{R}^n$). Se dice, en concreto, que X tiene distribución normal de media b y matriz de covarianzas C ([1]) y se escribe $X \sim N(b,C)$ o $X \sim N_n(b,C)$.

Necesitaremos el siguiente resultado de algebra lineal: "Si C es una matriz cuadrada simétrica, entonces existen una matriz ortogonal A (es decir, A es inversible y $A^{-1} = A^t$) y una matriz diagonal D tales que $A^tCA = D$; los elementos de la diagonal en D son los autovalores de C". Nótese que habremos de asegurarnos de que

[1] Si X es una v.a. n-dimensional con componentes $X_1,\dots,X_n$, se define la matriz de covarianzas $\mathrm{Cov}(X)$ de X como la matriz $(\mathrm{Cov}(X_i,X_j))_{i,j}$; en notación matricial, $\mathrm{Cov}(X) = E[(X-EX)(X-EX)^t]$.

hay v.a. n-dimensionales con función característica h; no cualquier función compleja definida en $\mathbb{R}^n$ es la función característica de alguna v.a. n-dimensional.

Sean A una matriz de orden $m \times n$, $b \in \mathbb{R}^m$ y X' una v.a. n-dimensional cuyas componentes son v.a.r. independientes y normalmente distribuidas con media cero. Probar que la v.a. m-dimensional $X = AX' + b$ tiene distribución $N_m(b, C)$ donde $C = ADA^t$, siendo D la matriz diagonal en la que los elementos de la diagonal son las varianzas λ_k de las X'_k, $1 \leq k \leq n$.

Problema 9.7. (↑) Probar que, si X es una v.a. normal n-dimensional $N_n(b, C)$, entonces existen $b \in \mathbb{R}^n$, una matriz cuadrada A de orden n que podemos elegir ortogonal y una v.a. n-dimensional X' cuyas componentes son v.a.r. independientes normalmente distribuidas con media cero tales que $X = AX' + b$.

Problema 9.8. (↑) Deducir del Problema 9.7 que, si $X \sim N_n(b, C)$, entonces $E(X) = b$ y $C_{jk} = \mathrm{Cov}(X_j, X_k)$.

Problema 9.9. (↑) Probar que, si b y C son como en la definición de la distribución normal multivariante, entonces existe una v.a. n-dimensional con función característica $h(u) = \exp\{iu^t b - \frac{1}{2}u^t Cu\}$.

Problema 9.10. (↑) Sean X, X', A ortogonal, C y b como en los problemas precedentes tales que $X = AX' + b$. Probar que, si C es inversible, entonces X es absolutamente continua (respecto a la medida de Lebesgue en $\mathbb{R}^n$) y su densidad es

$$f(x) = (2\pi)^{-\frac{n}{2}} \cdot (\det C)^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x-b)^t C^{-1}(x-b)\right\}, \quad x \in \mathbb{R}^n.$$

Problema 9.11. (↑) Sea X una v.a. n-dimensional. Probar que X es normal si, y sólo si, $u^t X$ es normal unidimensional (posiblemente degenerada) para cada $u \in \mathbb{R}^n$.

Observaciones. 1) Una distribución normal unidimensional degenerada es una distribución normal con varianza cero, que viene a coincidir con la distribución de Dirac concentrada en la media.

2) Este resultado sugiere una forma natural de definir la distribución normal en el caso infinito dimensional.

Problema 9.12. (↑) Sea $X := (X_1, X_2)$ una v.a. normal bidimensional con media (μ_1, μ_2) y matriz de covarianzas

$$C = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},$$

donde $\sigma_i^2 = \mathrm{Var}(X_i)$ y $\sigma_{12} = \mathrm{Cov}(X_1, X_2)$. Determinar la densidad de X cuando $\det C \neq 0$. Hacer

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

(:=coeficiente de correlación entre X_1 y X_2).

Problema 9.13. (↑) Sea $X_1 \sim N(0,1)$. Sea X_3 una v.a.r. independiente de X_1 tal que

$$P(X_3 = 0) = P(X_3 = 1) = \frac{1}{2}.$$

Sea $X_2(\omega) = X_1(\omega)$ si $X_3(\omega) = 0$ y $X_2(\omega) = -X_1(\omega)$ si $X_3(\omega) = 1$. Probar que X_2 tiene distribución $N(0,1)$. Probar que $P(X_1 + X_2 = 0) = \frac{1}{2}$ y concluir que $X_1 + X_2$ no es una v.a.r. normal. Concluir que (X_1, X_2) no tiene por qué ser normal aunque X_1 y X_2 lo sean.

Observación. (X_1, X_2) es normal si X_1 y X_2 lo son y son, además, independientes.

Problema 9.14. (↑) (Invarianza por rotaciones de las v.a. normales) Sean X_1, $\dots, X_n$ v.a.r. independientes tales que $X_j \sim N(\mu_j, \sigma^2)$, y A una matriz ortogonal de orden n. Sea $Y = AX$ donde $X^t = (X_1, \dots, X_n)$. Probar que las componentes de Y son v.a.r. independientes y normalmente distribuidas con varianza común σ^2.

Capítulo 10

Muestras

Definición 10.1. Llamaremos muestra a una colección numerable de v.a. independientes e idénticamente distribuidas (brevemente, iid); en el caso finito, llamamos tamaño de la muestra al número de v.a. consideradas.

***Observación* 10.1.** El último teorema de la lección anterior garantiza la existencia de sucesiones de v.a. independientes teniendo cada una de ellas una distribución fijada de antemano.

***Observación* 10.2.** Se deduce de lo dicho en el tema anterior que, si $X_1, X_2, \dots$ es una colección infinita numerable de v.a. independientes en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ a valores en un espacio medible $(\Omega', \mathcal{A}')$, entonces el espacio de probabilidad imagen de la v.a. $X = (X_1, X_2, \dots)$ es el espacio de probabilidad producto $(\Omega'^{\mathbb{N}}, \mathcal{A}'^{\mathbb{N}}, \prod_{i\in\mathbb{N}} P^{X_i})$. En particular, si las X_i son iid (es decir, si $(X_n)_{n\in\mathbb{N}}$ es una muestra infinita) y si Q denota la distribución de probabilidad común de las X_i, entonces el espacio de probabilidad imagen de esa muestra es $(\Omega'^{\mathbb{N}}, \mathcal{A}'^{\mathbb{N}}, Q^{\mathbb{N}})$.

***Observación* 10.3.** Una primera aproximación al concepto de muestra en Estadística Aplicada nos la presenta como una parte de la población. El objetivo de la Inferencia Estadística consiste, precisamente, en obtener información sobre toda la población a partir de los datos obtenidos para una muestra de la misma, y se pretende, entonces, que la muestra sea "representativa" de la población. Al proceso de extracción de una muestra de una población se le suele llamar muestreo. La definición anterior se puede considerar como la aproximación probabilística al concepto de muestra. Supongamos que estamos interesados en el estudio de una cierta característica de los individuos de una población Ω y que esa característica viene definida por la v.a. $X : \Omega \longrightarrow \Omega'$. Siendo a veces muy costoso (o, incluso, imposible) estudiar todos los individuos de la población, cabe pensar en seleccionar adecuadamente un cierto número n de ellos, estudiar para ellos la característica que nos interesa y extrapolar las conclusiones extraídas a toda la población. Por ejemplo, si pretendemos obtener información sobre la estatura media de los españoles de 18 años, cabe pensar en seleccionar una muestra del conjunto de españoles que tienen esa edad y aproximar la estatura media buscada por la estatura media de los individuos de la muestra (es decir, la media muestral). El adverbio adecuadamente hace alusión al hecho de que, para que la extrapolación de conclusiones sea válida, en el proceso de selección de individuos no debe aparecer ningún tipo de preferencias por parte del seleccionador; en lugar de hablar de seleccionar adecuadamente hablaremos de seleccionar al azar, en un sentido que precisaremos posteriormente. Para el caso de poblaciones finitas, el muestreo aleatorio simple queda caracterizado por el hecho de que cada muestra posible de n individuos tiene la misma probabilidad de ser extraída de la población. Podemos distinguir dos tipos de muestreo aleatorio simple: el muestreo aleatorio simple con reemplazamiento (en el que cada individuo seleccionado es devuelto a la población tras haber sido estudiado y puede volver a ser seleccionado) y el muestreo aleatorio simple sin reemplazamiento (en el que un individuo que haya sido seleccionado en la muestra es retirado de la población y no podrá ser seleccionado en lo sucesivo). Consideraremos ahora solamente el muestreo con reemplazamiento. Elegir con reemplazamiento n individuos de la población finita Ω equivale a elegir un elemento $\omega = (\omega_1, \dots, \omega_n)$ de Ω^n; si a la hora de elegir el individuo i-ésimo de la población no se tiene en cuenta cuáles fueron las elecciones precedentes (es decir, si los n individuos se extraen independientemente los unos de los otros) y si la probabilidad de elegir un individuo cualquiera de la población no se modifica de una extracción a otra, la probabilidad P en Ω induce en el conjunto producto Ω^n una probabilidad P^n (que llamaremos probabilidad producto) definida por

$$P^n(\omega_1, \dots, \omega_n) = P(\omega_1) \cdots P(\omega_n).$$

Suponiendo además que todos los individuos de la población finita Ω tienen la misma probabilidad de ser elegidos (esa probabilidad será entonces igual a $1/N$ si N es el número de individuos de la población Ω), entonces

$$P^n(\omega_1, \dots, \omega_n) = P(\omega_1) \cdots P(\omega_n) = \frac{1}{N^n},$$

y todas las n-uplas $(\omega_1, \dots, \omega_n) \in \Omega^n$ tienen la misma probabilidad $(1/N^n)$ de ser seleccionadas; por tanto, el espacio de probabilidad $(\Omega^n, \mathcal{A}^n, P^n)$ corresponde al muestreo aleatorio simple con reemplazamiento de n individuos de la población Ω. Estudiar las características $X(\omega_i)$ de los individuos ω_i, $1 \leq i \leq n$, equivale a considerar las v.a. X_i, $1 \leq i \leq n$, definidas en Ω^n por $X_i(\omega_1, \dots, \omega_n) = X(\omega_i)$. Se prueba que las v.a. X_i, $1 \leq i \leq n$, son P^n-iid, y que cada X_i tiene la misma distribución que X (se dice por ello que las X_i son copias independientes de la v.a. X). Todo esto justifica la definición precedente de muestra, que no se circunscribe al caso finito de sucesos elementales equiprobables. Se habla, a veces, de muestreo probabilístico cuando se seleccionan al azar y con reemplazamiento individuos de una población Ω que no tiene por qué ser finita y que, aunque lo fuese, no considere a todos sus individuos equiprobables; sólo es imprescindible la hipótesis de independencia en la selección de los individuos –lo que queda garantizado si a la hora de elegir al individuo i-ésimo no tenemos en cuenta lo ocurrido en elecciones precedentes– y la hipótesis de que la probabilidad de elegir a un individuo se mantenga inalterable de una elección a otra.

Para el caso de muestras de una distribución real podemos dar la siguiente definición:

DEFINICIÓN 10.2. (Función de distribución empírica) Sea $X_1, \dots, X_n$ una muestra de tamaño n de una distribución real. Llamaremos función de distribución empírica (o muestral) a la aplicación

$$F(x, \omega) := \frac{1}{n} \sum_{i=1}^{n} I_{]-\infty,x]}(X_i(\omega)).$$

***Observación* 10.4.** $n \cdot F(x, \omega)$ es el número de índices $i \in \{1, \dots, n\}$ tales que $X_i(\omega) \leq x$.

***Observación* 10.5.** Dado $\omega \in \Omega$, $F(\cdot, \omega)$ es la función de distribución de la v.a. discreta

$$Z_\omega : i \in (\{1, \dots, n\}, \mathscr{P}(\{1, \dots, n\}), Q) \to Z_\omega(i) := X_i(\omega) \in \mathbb{R},$$

donde $Q(i) = 1/n$, $1 \leq i \leq n$. En efecto, si $x \in \mathbb{R}$,

$$Q(Z_\omega \leq x) = \frac{1}{n} \operatorname{card} \{j \in \{1, \dots, n\} \colon X_j(\omega) \leq x\} = F(x, \omega).$$

***Observación* 10.6.** Si $n \in \mathbb{N}$ y $x \in \mathbb{R}$, $F_n(x, \cdot)$ es, evidentemente, una v.a.r. Denotando $I_{(i,x)} = I_{]-\infty,x]} \circ X_i$, $i \in \mathbb{N}$, $x \in \mathbb{R}$, entonces, a x fijo, las $I_{(i,x)}$, $i \in \mathbb{N}$, son independientes (por serlo las X_i) e idénticamente distribuidas; notar que

$$I_{(i,x)}(\omega) = \begin{cases} = 1 & \text{si } X_i(\omega) \leq x \\ = 0 & \text{si } X_i(\omega) > x \end{cases},$$

y que $E(I_{(i,x)}) = P(X_i \leq x) = F(x)$, si F denota la función de distribución común de las X_i. Entonces, $I_{(i,x)}$ es una v.a. de Bernouilli que toma los valores 1 y 0 con probabilidades $F(x)$ y $1-F(x)$, resp. La ley fuerte de los grandes números ([1]) prueba entonces que $F_n(x,\cdot)$ converge P-c.s. a $F(x)$ cuando $n\to\infty$, es decir, para cada $x\in\mathbb{R}$, existe $A_x\in\mathcal{A}$ tal que $P(A_x)=0$ y $F_n(x,\omega)\longrightarrow_{n\to\infty} F(x)$, si $\omega\notin A_x$. El teorema siguiente mejora esta última conclusión.

TEOREMA 10.1. (Glivenko-Cantelli) *Sea $(X_n)_{n\geq 1}$ una sucesión de v.a.r. iid definidas en $(\Omega,\mathcal{A},P)$ con función de distribución común F. Si F_n denota la función de distribución empírica obtenida de $X_1,\dots,X_n$, entonces*

$$\sup_{x\in\mathbb{R}}|F_n(x,\omega)-F(x)|\underset{n\to\infty}{\longrightarrow}0,\qquad P\text{-c.s.}$$

Demostración. Hemos probado en la observación 3) anterior que, para cada $x\in\mathbb{R}$, existe $A_x\in\mathcal{A}$ tal que $P(A_x)=0$ y $\lim_n F_n(x,\omega)=F(x)$ si $\omega\notin A_x$. Se sigue análogamente de la ley fuerte de los grandes números (tomando ahora $I_{]-\infty,x[}$ en lugar de $I_{]-\infty,x]}$) que, para cada $x\in\mathbb{R}$, existe $B_x\in\mathcal{A}$ tal que $P(B_x)=0$ y $\lim_n F_n(x-,\omega)=F(x-)$ si $\omega\notin B_x$, donde $F(x-)$ (resp., $F_n(x-,\omega)$) denota el límite por la izquierda de la función de distribución F (resp., $F_n(\cdot,\omega)$) en el punto x.

Consideremos la función auxiliar (una especie de inversa de F, llamada también función cuantil)

$$\varphi: u\in]0,1[\longrightarrow\varphi(u)=\inf\{x\in\mathbb{R}\colon u\leq F(x)\}.$$

Nótese que $\varphi(u)$ está bien definida habida cuenta que, para cada $u\in]0,1[$, $\{x\in\mathbb{R}\colon u\leq F(x)\}\neq\emptyset$ (pues $F(x)\to 1$ cuando $x\to+\infty$); además $\varphi(u)\in\mathbb{R}$ (si el ínfimo fuese $-\infty$ existiría una sucesión x_n en $\mathbb{R}$ convergente a $-\infty$ tal que $u\leq F(x_n)$, para cada $n\in\mathbb{N}$, lo cual no es posible pues $u>0$ y $F(x)\to 0$ cuando $x\to-\infty$). Además, φ es creciente en $]0,1[$, pues, si $0<u\leq v<1$, entonces $\{x\in\mathbb{R}\colon v\leq F(x)\}\subset\{x\in\mathbb{R}\colon u\leq F(x)\}$ y $\varphi(u)\leq\varphi(v)$. Nótese también que $\{x\in\mathbb{R}\colon u\leq F(x)\}$ es un intervalo

[1] Teorema: (Ley fuerte de los grandes números: caso iid) Si $(X_n)_n$ es una sucesión de v.a.r. iid con media común μ, entonces

$$\frac{1}{n}\sum_{i=1}^n X_i\longrightarrow_{n\to\infty}\mu,\qquad P-\text{c.s.}$$

con extremo superior $+\infty$, por ser F no decreciente, y cerrado por la izquierda, por ser F continua por la derecha. Así pues, si $0 < u < 1$,

$$\{x \in \mathbb{R} \colon u \leq F(x)\} = [\varphi(u), +\infty[,$$

es decir, si $0 < u < 1$ y $x \in \mathbb{R}$, entonces

$$\varphi(u) \leq x \Longleftrightarrow u \leq F(x). \tag{10.1}$$

Siendo φ creciente, podemos definir $\varphi(1) = \lim_{u \to 1} \varphi(u)$; nótese que $\varphi(1)$ puede ser $+\infty$. Se verifica, para cada $u \in]0, 1]$, que

$$F(\varphi(u)-) \leq u \leq F(\varphi(u)); \tag{10.2}$$

la segunda desigualdad es cierta por (10.1) para $u \in]0, 1[$, y trivialmente cierta para $u = 1$; para la primera desigualdad, nótese que, si fuese $u < F(\varphi(u)-)$, existiría (por definición de límite por la izquierda) $x < \varphi(u)$ tal que $u < F(x)$ y, por (10.1), debería ser $\varphi(u) \leq x$, llegando así a un absurdo.

Para $m \in \mathbb{N}$ y $1 \leq k \leq m$, se define

$$x_{m,k} = \varphi(k/m).$$

Se siguen inmediatamente de (10.2) las relaciones siguientes:

$$F(x_{m,k}-) - F(x_{m,k-1}) \leq 1/m, \quad 2 \leq k \leq m, \tag{10.3}$$

$$F(x_{m,1}-) \leq \frac{1}{m}, \tag{10.4}$$

$$F(x_{m,m}) \geq 1 - \frac{1}{m}. \tag{10.5}$$

Consideremos la v.a.r.

$$D_{m,n}(\omega) = \max_{1 \leq k \leq m} \left(|F_n(x_{m,k}, \omega) - F(x_{m,k})|, |F_n(x_{m,k}-, \omega) - F(x_{m,k}-)|\right).$$

Probemos que

$$\sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)| \leq \frac{1}{m} + D_{m,n}(\omega), \tag{10.6}$$

para cada $m, n \in \mathbb{N}$ y cada $\omega \in \Omega$.

Si $2 \leq k \leq m$ y $x_{m,k-1} \leq x < x_{m,k}$, se sigue de (10.3) y de la definición de $D_{m,n}(\omega)$ que

$$F_n(x, \omega) \leq F_n(x_{m,k}-, \omega) \leq F(x_{m,k}-) + D_{m,n}(\omega) \leq F(x) + \frac{1}{m} + D_{m,n}(\omega)$$

y

$$F_n(x, \omega) \geq F_n(x_{m,k-1}, \omega) \geq F(x_{m,k-1}) - D_{m,n}(\omega) \geq F(x) - \frac{1}{m} - D_{m,n}(\omega);$$

entonces

$$|F_n(x, \omega) - F(x)| \leq \frac{1}{m} + D_{m,n}(\omega) \tag{10.7}$$

si $x \in \cup_{k=2}^{m} [x_{m,k-1}, x_{m,k}[$, $m, n \in \mathbb{N}$. Análogamente, si $x < x_{m,1}$, se sigue de (10.4) y de la definición de $D_{m,n}(\omega)$ que

$$0 \leq F_n(x, \omega) \leq F_n(x_{m,1}-, \omega) \leq F(x_{m,1}-) + D_{m,n}(\omega) \leq \frac{1}{m} + D_{m,n}(\omega)$$

y

$$0 \leq F(x) \leq F(x_{m,1}-) \leq \frac{1}{m} \leq \frac{1}{m} + D_{m,n}(\omega),$$

de donde se sigue que (10.7) también es cierta para $x < x_{m,1}$. En fin, si $x \geq x_{m,m}$ $(= \varphi(1) \in \mathbb{R} \cup \{+\infty\})$, entonces

$$1 \geq F_n(x, \omega) \geq F(x_{m,m}) - D_{m,n}(\omega) \geq 1 - \frac{1}{m} - D_{m,n}(\omega)$$

y

$$1 \geq F(x) \geq F(x_{m,m}) \geq 1 - \frac{1}{m} - D_{m,n}(\omega),$$

de donde se sigue que (10.7) también es cierta para $x \geq x_{m,m}$. Queda así probado (10.6).

Se define ahora

$$A = \cup_{m \in \mathbb{N}} \cup_{k=1}^{m} (A_{x_{m,k}} \cup B_{x_{m,k}}).$$

Entonces $A \in \mathcal{A}$ y $P(A) = 0$. Además, si $\omega \in \Omega \setminus A$, entonces, para cada $m \in \mathbb{N}$, $\lim_{n \to \infty} D_{m,n}(\omega) = 0$. Así, si $\omega \notin A$, dado $\epsilon > 0$ podemos elegir $m = m(\epsilon) \in \mathbb{N}$ tal que $1/m \le \epsilon/2$ y, para este m, existe $n_m \in \mathbb{N}$ tal que $D_{m,n}(\omega) \le \epsilon/2$ si $n \ge n_m$. Entonces, si $\omega \notin A$ y $n \ge n_{m(\epsilon)}$,

$$\sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)| \le \frac{1}{m} + D_{m,n}(\omega) \le \epsilon,$$

lo que prueba que $\sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)|$ converge a cero puntualmente P-c.s. □

***Observación* 10.7.** La función de distribución empírica $F_n(x,\omega)$ es fácilmente calculable a partir de las observaciones $X_1(\omega),\dots,X_n(\omega)$ (sólo hay que contar cuántos de estos números son $\le x$ y dividir por n) y, según acabamos de ver, para casi todas las observaciones ω, $F_n(\cdot,\omega)$ converge uniformemente en $\mathbb{R}$ a F. El teorema de Glivenko-Cantelli se llama a veces teorema fundamental de la Estadística Matemática debido a que, desde el punto de vista estadístico, la función de distribución F que siguen las observaciones X_i es desconocida y ese teorema permite aproximar F uniformemente en $\mathbb{R}$ mediante funciones fácilmente calculables a partir de las observaciones (téngase en cuenta además que el objetivo de la inferencia estadística consiste en extraer información sobre F a partir de las observaciones).

Los momentos de la v.a. discreta Z_ω, considerados como función de ω, son v.a. que llamaremos momentos muestrales de la muestra considerada. Tenemos, pues la siguiente definición:

DEFINICIÓN 10.3. Si $X_1,\dots,X_n$ es una muestra de tamaño n de una distribución real, llamaremos media muestral a la v.a.

$$\bar{X}(\omega) := \frac{1}{n}\sum_{i=1}^n X_i(\omega).$$

Llamaremos momento muestral (resp., momento muestral absoluto) de orden k a la v.a.

$$a_k(\omega) := \frac{1}{n}\sum_{i=1}^n X_i^k(\omega) \quad (\text{resp., } A_k(\omega) := \tfrac{1}{n}\textstyle\sum_{i=1}^n |X_i(\omega)|^k).$$

Llamaremos momento central muestral (resp., momento central muestral absoluto) de orden k a la v.a.

$$b_k(\omega) := \frac{1}{n}\sum_{i=1}^n (X_i(\omega) - \bar{X}(\omega))^k \quad (\text{resp., } B_k(\omega) := \tfrac{1}{n}\textstyle\sum_{i=1}^n |X_i(\omega) - \bar{X}(\omega)|^k)$$

Llamaremos varianza muestral a la v.a.

$$S^2(\omega) := \frac{1}{n-1}\sum_{i=1}^n (X_i(\omega) - \bar{X}(\omega))^2.$$

Observación **10.8.** Haber llamado varianza muestral a S^2, en vez de b_2, es debido a que la media de S^2 es igual a la varianza común de las X_i (si ésta es finita), según prueba la siguiente proposición. A veces se llama cuasi-varianza muestral a S^2.

Proposición 10.1. *Sea $X_1, \dots, X_n$ una muestra de tamaño n de una distribución real.*

(a) *Si X_1 tiene media finita, $E(\bar{X}) = E(X_1)$.*

(b) *Si X_1 tiene varianza finita, $\mathrm{Var}(\bar{X}) = \frac{1}{n}\mathrm{Var}(X_1)$.*

(c) *Si X_1 tiene momento finito de orden k, $E(a_k) = E(X_1^k)$.*

(d) *Si X_1 tiene momento finito de orden $2k$, $\mathrm{Var}(a_k) = \frac{1}{n}[E(X_1^{2k}) - E(X_1^k)^2]$.*

(e) *Si X_1 tiene momento finito de orden 2, $E(S^2) = \mathrm{Var}(X_1)$.*

Demostración. (a), (b), (c) y (d) son triviales. Probemos (e):

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} E\left[\sum_i X_i^2 - n\bar{X}^2\right] \\ &= \frac{1}{n-1}\left[\sum_i E(X_i^2) - nE(\bar{X}^2)\right] \\ &= \frac{1}{n-1}\left[n(\mathrm{Var}(X_1) + E(X_1)^2) - n\left(\frac{1}{n}\mathrm{Var}(X_1) + E(X_1)^2\right)\right] \\ &= \frac{n}{n-1}\mathrm{Var}(X_1) - \frac{1}{n-1}\mathrm{Var}(X_1) = \mathrm{Var}(X_1). \end{aligned}$$

□

Para muestras de una distribución normal podemos precisar algunas de las afirmaciones de la proposición anterior. Antes de ello definiremos dos distribuciones de probabilidad que aparecen de forma natural en el muestreo de poblaciones normales y probaremos un lema.

DEFINICIÓN 10.4. Sea $X_1, \dots, X_n$ una muestra de tamaño n de una distribución $N(\mu, \sigma^2)$.

(a) (Distribución χ^2) La distribución $\chi^2(n)$ –llamada chi-cuadrado con n grados de libertad– es la definición gamma $G(n/2, 2)$.

(b) (Distribución $t(n)$ de Student) Se define la distribución $t(n)$ de Student con n grados de libertad como la distribución del cociente

$$\sqrt{n}\frac{X}{\sqrt{Y}},$$

donde X e Y son v.a.r. independientes tales que $X \sim N(0,1)$ e $Y \sim \chi^2(n)$.

(c) (Distribución F de Fisher) Se define la distribución $F(m,n)$ de Fisher con (m,n) grados de libertad como la distribución del cociente

$$\frac{X/m}{Y/n},$$

donde X e Y son v.a.r. independientes tales que $X \sim \chi^2(m)$ e $Y \sim \chi^2(n)$.

Lema 10.1. (a) *Si $X \sim N(\mu,\sigma^2)$ y $a,b \in \mathbb{R}$ entonces $aX+b \sim N(a\mu+b, a^2\sigma^2)$.*

(b) *Si $X_1,\dots,X_n$ son independientes y $X_i \sim N(\mu_i,\sigma_i^2)$, $1 \le i \le n$, entonces $\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.*

(c) *Si $X_1,\dots,X_n$ son v.a.r. independientes y $X_i \sim G(\alpha_i,\beta)$, $1 \le i \le n$, entonces $\sum_{i=1}^n X_i$ tiene distribución $G(\sum_{i=1}^n \alpha_i, \beta)$.*

(d) *En particular, si las X_i son independientes y $X_i \sim \chi^2(k_i)$ entonces $\sum_{i=1}^n X_i \sim \chi^2(\sum_{i=1}^n k_i)$.*

(e) *Si X es una v.a.r. con distribución $N(0,1)$, entonces X^2 tiene distribución $\chi^2(1)$. Así, si las X_i son independientes y si $X_i \sim N(\mu_i,\sigma_i^2)$ entonces $\sum_{i=1}^n \left(\frac{X_i-\mu_i}{\sigma_i}\right)^2$ tiene distribución $\chi^2(n)$.*

Demostración. (a), (b), (c) y (d) son consecuencia inmediata de que la función característica de la suma de v.a.r. independientes es el producto de las funciones características de los sumandos. Para probar (e), notar que, si $V = X^2$, entonces $P(V \le v) = 0$ si $v \le 0$, y, si $v > 0$

$$P(V \le v) = P(-\sqrt{v} \le X \le \sqrt{v}) = 2P(0 \le X \le \sqrt{v}) = 2\int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\} dx.$$

Basta hacer el cambio de variables $t = \sqrt{x}$ (Teorema 6.2). □

Proposición 10.2. *Sea $X_1,\dots,X_n$ una muestra de tamaño n de una distribución $N(\mu,\sigma^2)$.*

(a) *$\bar X$ tiene distribución $N(\mu,\sigma^2/n)$.*

(b) *Las v.a. $\bar X$ y $(X_1-\bar X,\dots,X_n-\bar X)$ son independientes. En particular, son independientes $\bar X$ y S^2.*

(c) *$(n-1)\sigma^{-2}S^2$ tiene distribución $\chi^2(n-1)$.*

(d) *La v.a. $\sqrt{n}(\bar X-\mu)/S$ tiene distribución $t(n-1)$.*

Demostración. (a) Es consecuencia inmediata del lema anterior.

(b) Será suficiente probar que

$$\begin{aligned}&\varphi_{(\bar X,X_1-\bar X,\dots,X_n-\bar X)}(t,t_1,\dots,t_n)\\ &=\varphi_{(\bar X)}(t)\cdot\varphi_{(X_1-\bar X,\dots,X_n-\bar X)}(t_1,\dots,t_n).\end{aligned}$$

Podemos escribir

$$\begin{aligned}&\varphi_{(\bar X,X_1-\bar X,\dots,X_n-\bar X)}(t,t_1,\dots,t_n)\\ &=E\left(\prod_{j=1}^n \exp\{in^{-1}(nt_j-n\bar t+t)X_j\}\right)=(*),\end{aligned}$$

donde $\bar t=n^{-1}\sum_{j=1}^n t_j$; utilizando que las X_j son independientes, notar que

$$\begin{aligned}(*)&=\prod_{j=1}^n E\left(\exp\{in^{-1}(nt_j-n\bar t+t)X_j\}\right)\\ &=\prod_{j=1}^n \exp\left\{in^{-1}[n(t_j-\bar t\,)+t]\mu-\frac12\sigma^2[t+n(t_j-\bar t\,)]^2\right\}\\ &=\exp\left\{i\mu t-\frac12\sigma^2t^2/n\right\}\cdot\exp\left\{-\frac12\sigma^2\sum_{j=1}^n(t_j-\bar t\,)^2\right\}\\ &=\varphi_{(\bar X)}(t)\cdot\varphi_{(X_1-\bar X,\dots,X_n-\bar X)}(t_1,\dots,t_n).\end{aligned}$$

(c) Nótese que $\sum_{j=1}^n\left(\frac{X_j-\mu}{\sigma}\right)^2$ tiene distribución $\chi^2(n)$ y que $n\left(\frac{\bar X-\mu}{\sigma}\right)^2$ tiene distribución $\chi^2(1)$. Por otra parte, se verifica que

$$\sigma^{-2}\sum_{j=1}^n(X_j-\mu)^2=n\left(\frac{\bar X-\mu}{\sigma}\right)^2+(n-1)\sigma^{-2}S^2. \tag{10.8}$$

Siendo $\bar{X}$ y S^2 independientes se sigue que la función característica del primer miembro de (10.8) es el producto de las funciones características de los sumandos del segundo miembro. Puesto que $(1-2it)^{-\frac{n}{2}}$ es la función característica de la distribución $\chi^2(n)$, se tiene

$$(1-2it)^{-\frac{n}{2}} = (1-2it)^{-\frac{1}{2}}\varphi_Y(t),$$

donde $Y = (n-1)\sigma^{-2}S^2$. Así pues, la función característica de Y es $(1-2it)^{-(n-1)/2}$, de donde se sigue el resultado.

(d) Es consecuencia inmediata de los apartados anteriores. □

Proposición 10.3. *Si $X_1,\dots,X_m,Y_1,\dots,Y_n$ son v.a.r. independientes tales que $X_i \sim N(\mu_X,\sigma_X^2)$, $1\le i\le m$, e $Y_j \sim N(\mu_Y,\sigma_Y^2)$, $1\le j\le n$, es decir, dos muestras independientes de sendas distribuciones normales. Sean $S_X^2 := \frac{1}{m-1}\sum_{i=1}^m (X_i-\mu_X)^2$ y $S_Y^2 := \frac{1}{n-1}\sum_{j=1}^n (Y_j-\mu_Y)^2$ las varianzas muestrales de ambas muestras. La v.a.*

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

tiene distribución $F(m-1,n-1)$. En particular, si $\sigma_X=\sigma_Y$ se verifica que el cociente S_X^2/S_Y^2 tiene distribución $F(m-1,n-1)$.

Si $\sigma_X^2=\sigma_Y^2$, entonces

$$\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{S\sqrt{\frac{1}{m}+\frac{1}{n}}}$$

tiene distribución $t(m+n-2)$, donde

$$S^2 = \frac{1}{m+n-2}\left[\sum_{i=1}^m (X_i-\bar{X})^2 + \sum_{j=1}^n (Y_j-\bar{Y})^2\right]$$

es la varianza muestral combinada.

Demostración. Es consecuencia inmediata de la definición de la distribución F y de la proposición anterior. □

PROBLEMAS DEL CAPÍTULO 10

Problema 10.1. Decidir si son muestreos aleatorios simples los procedimientos siguientes para extraer una muestra de tamaño 10 de una población finita formada por N persona numeradas del 1 al N.

1. Supongamos $N = 100$. Se extrae una bola de una urna con 10 bolas idénticas numeradas del 1 al 10: si se ha extraído la bola k, se propone seleccionar a los individuos k, $k + 10$, ..., $k + 90$.

2. Supongamos $N = 100$. Se extrae una bola de una urna con 100 bolas idénticas numeradas del 1 al 100: si se ha extraído la bola k, se propone seleccionar a los individuos k, $k + 1$, ..., $k + 9$ (entendiendo que si pasamos de 100 volvemos al principio).

3. Supongamos $N = 60$. Se extraen 10 bolas de una urna con 100 bolas idénticas numeradas del 1 al 100: si en una extracción sale la bola $k \leq 60$ seleccionamos al individuo número k; si sale la bola $k > 60$ seleccionamos al individuo número $101 - k$.

Problema 10.2. Probar que, si $X_1, \dots, X_M, Y_1, \dots, Y_n$ son v.a.r. independientes normalmente distribuidas de forma que $X_i \sim N(\mu_X, \sigma_X^2)$, $1 \leq i \leq m$, e $Y_j \sim N(\mu_Y, \sigma_Y^2)$, entonces, la v.a.

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{(m-1)\sigma_X^{-2} S_X^2 + (n-1)\sigma_Y^{-2} S_Y^2}} \sqrt{\frac{m+n-2}{m^{-1}\sigma_X^2 + n^{-1}\sigma_Y^2}}$$

tiene distribución $t(m + n - 2)$.

Indicación: Basta notar que

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, m^{-1}\sigma_X^2 + n^{-1}\sigma_Y^2\right)$$

y

$$(m-1)\sigma_X^{-2} S_X^2 + (n-1)\sigma_Y^{-2} S_Y^2 \sim \chi^2(m+n-2),$$

y que ambas variables son independientes.

Problema 10.3. (Distribución t de Student) Si X e Y son v.a.r. independientes tales que $X \sim N(0, 1)$ e $Y \sim \chi^2(n)$, llamaremos distribución t de Student con n grados de libertad, y la denotaremos $t(n)$, la distribución de la v.a.r.

$$T = \frac{X}{\sqrt{Y/n}}.$$

Si Y está definida en un espacio probabilístico $(\Omega, \mathcal{A}, P)$ nótese que $P(Y \leq 0) = 0$, con lo cual T está bien definida P-c.s. Probar que la distribución $t(n)$ queda definida por la densidad

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma(n/2)(1+t^2/n)^{-\frac{n+1}{2}}}.$$

Indicación: $Z = \sqrt{Y/n}$ es una v.a.r. no negativa. Para determinar su densidad f_Z notar que si $t \leq 0$ entonces $P(Z \leq t) = 0$ y que si $t > 0$ entonces $P(Z \leq t) = P(0 \leq Y \leq nt^2) = \int_0^{nt^2} f_Y(y)dy$; hacer aquí el cambio de variables $y = nx^2$. Notar que X y Z son independientes y, entonces, la densidad conjunta de X y Z es $f_{(X,Z)}(x,z) = f_X(x)f_Z(z)$ (para probar esto nótese que $P^{(X,Z)} = P^X \times P^Z$ y aplicar el teorema de Fubini (Teorema 7.2)). Notar ahora que $T = X/Z$ y probar que la aplicación

$$g : (x,z) \in \mathbb{R} \times \mathbb{R}^+ \longrightarrow (x/z, z) \in \mathbb{R} \times \mathbb{R}^+$$

es biyectiva y g y g^{-1} son de clase C^1. Nótese que $g \circ (X,Z) = (T,S)$ donde $S(\omega) = Z(\omega)$, $\forall \omega$; así, si B es un boreliano de $\mathbb{R} \times \mathbb{R}^+$ se verifica que

$$\begin{aligned} P((T,S) \in B) &= P(g \circ (X,Z) \in B) = P((X,Z) \in g^{-1}(B)) \\ &= \iint_{g^{-1}(B)} f_{(X,Z)}(x,z)dxdz; \end{aligned}$$

hacer aquí el cambio de variables $(x,z) = g^{-1}(t,s)$ y determinar la densidad $f_{(T,S)}$ conjunta de T y S. Notando ahora que $P^T(A) = P(T \in A) = P((T,S) \in A \times \mathbb{R})$ determínese la densidad f_T de T. Se obtendrá concretamente

$$\begin{aligned} f_T(t) &= \int_0^\infty f_{(T,S)}(t,s)ds \\ &= \frac{2(n/2)^{n/2}}{\sqrt{2\pi}\Gamma(n/2)} \int_0^\infty e^{-\frac{1}{2}(t^2+n)s^2} s^n ds; \end{aligned}$$

haciendo ahora el cambio de variables $u = (t^2+n)s^2/2$ se obtiene la expresión deseada para f_T.

Problema 10.4. (Distribución F de Fisher) Sean X e Y v.a.r. independientes tales que $X \sim \chi^2(m)$ e $Y \sim \chi^2(n)$. Llamaremos distribución F de Fisher con (m,n) grados de libertad, y la denotaremos $F(m,n)$, la distribución de la v.a.r.

$$S = \frac{X/m}{Y/n}.$$

Nótese que S está bien definida salvo quizás en un conjunto de probabilidad nula. Probar que la densidad de la distribución $F(m,n)$ es

$$f_S(s) = \frac{m^{m/2}n^{n/2}\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{s^{\frac{m}{2}-1}}{(ms+n)^{(m+n)/2}} I_{]0,\infty[}(s).$$

Indicación: Para ello determínese la densidad conjunta $f_{(X,Y)}$ de X e Y y considérese la aplicación biyectiva $g : (x,y) \in \mathbb{R}^+ \times \mathbb{R}^+ \longrightarrow g(x,y) = (\frac{x/m}{y/n}, y) \in \mathbb{R}^+ \times \mathbb{R}^+$; probar que g y g^{-1} son de clase C^1, hacer $(S,V) = g \circ (X,Y)$ y determinar la densidad $f_{(S,V)}$ conjunta de S y V; notar que

$$f_S(s) = \int_0^\infty f_{(S,V)}(s,v)dv = C_s \int_0^\infty e^{-\frac{1}{2}v(msn^{-1}+1)} v^{\frac{1}{2}(m+n)-1} dv$$

donde C_s es una constante dependiente de s; hacer ahí el cambio de variables $y = (2n)^{-1}v(ms+n)$ y obtener la expresión deseada para f_S.

Problema 10.5. ($\uparrow$) Probar que

1. $X \sim F(m,n) \Longrightarrow X^{-1} \sim F(n,m)$.
2. $X \sim F(1,n) \Longrightarrow \sqrt{X} \sim t(n)$.
3. $X \sim F(m,n) \Longrightarrow Y = (1 + \frac{m}{n}X)^{-1} \sim B(n/2, m/2)$, $1 - Y \sim B(m/2, n/2)$.

Indicación: Para probar 3) notar que $Y = g(X)$ donde $g(x) = (1 + \frac{m}{n}X)^{-1}$ si $x > 0$; notar que $1 = P(X > 0) = P(Y \in]0,1[)$; si B es un boreliano de $]0,1[$,

$$P^Y(B) = P(X \in g^{-1}(B)) = \int_{g^{-1}(B)} f_{(m,n)}(x)dx$$

donde $f_{(m,n)}$ es la densidad de X; hacer el cambio $y = g(x)$ en esa integral.

Problema 10.6. ($\bigstar$) (Estadísticos de orden)([2]) Sean $X_1, \dots, X_n$ v.a.r. en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Para cada $\omega \in \Omega$ sea

$$X_{(1)}(\omega) \leq \cdots \leq X_{(n)}(\omega) \tag{10.9}$$

una reordenación creciente de $X_1(\omega), \dots, X_n(\omega)$. Probar que las $X_{(i)}$ son también v.a. en $(\Omega, \mathcal{A}, P)$. Concluir que la aplicación n-dimensional $X_{(\cdot)} := (X_{(1)}, \dots, X_{(n)})$ es una v.a. Def.: Las aplicaciones $X_{(i)} : \Omega \longrightarrow \mathbb{R}, 1 \leq i \leq n$, así definidas se llaman estadísticos de orden asociados a las v.a. X_i.

Problema 10.7. ($\uparrow$) (Rangos) Se define el estadístico R_i de rango i mediante

$$R_i(\omega) = \sum_{j=1}^n I_{[0,+\infty[}(X_i(\omega) - X_j(\omega)).$$

Así, $R_i(\omega)$ cuenta el número de índices j tales que $X_j(\omega) \leq X_i(\omega)$ o, dicho de otro modo, $R_i(\omega)$ proporciona el lugar que ocupa $X_i(\omega)$ en la sucesión reordenada (10.9)

[2] Entiéndase el término estadístico como sinónimo de variable aleatoria.

del problema anterior. Debe notarse que el valor $X_i(\omega)$ puede aparecer repetido en dicha sucesión; en ese caso, $R_i(\omega)$ da la posición del último de esos valores. Probar que R_i es una v.a. a valores en $\{1,\ldots,n\}$ y que, para cada $\omega \in \Omega$,

$$X_i(\omega) = X_{(R_i(\omega))}(\omega).$$

Entonces, el vector rango $R = (R_1,\ldots,R_n)$ es una v.a. y los vectores $X_{(\cdot)}$ y R determinan el vector $X = (X_1,\ldots,X_n)$.

Problema 10.8. ($\uparrow$) Sea $N^k = \cup_{1\le i<j\le k}\{x = (x_1,\ldots,x_k) \in \mathbb{R}^k : x_i = x_j\}$. Probar que N^k tiene medida de Lebesgue nula en $\mathbb{R}^k$. Denotemos $\mathbb{R}^{(k)} = \{x \in \mathbb{R}^k : x_1 < \ldots < x_k\}$. Concluir que, si X es una v.a. k-dimensional en $(\Omega,\mathcal{A},P)$ que admite una densidad respecto a la medida de Lebesgue en $\mathbb{R}^k$, entonces

$$P(X \in \mathbb{R}^k \setminus N^k) = 1,$$
$$P(X_{(\cdot)} \in \mathbb{R}^{(k)}) = 1,$$
$$P(R \in S_k) = 1,$$

donde S_k es el conjunto de las permutaciones sobre $\{1,\ldots,k\}$.

Observación. Así pues, cuando X es una v.a. k-dimensional absolutamente continua, podemos suponer que el vector $X_{(\cdot)}$ de los estadísticos de orden construidos a partir de X toma siempre valores en $\mathbb{R}^{(k)}$ y que el vector R de los rangos toma siempre valores en S_k.

Problema 10.9. ($\uparrow$) Probar que, si $B \subset \mathbb{R}^{(n)}$, $\omega \in \Omega$ y $r = (r(1),\ldots,r(n)) \in S_n$, entonces

$$[X_{(\cdot)}(\omega) \in B\,,\ R(\omega) = r] \Longleftrightarrow [X(r^{-1})(\omega) \in B]$$

donde r^{-1} denota la permutación inversa de r y $X(r) = (X_{r(1)},\ldots,X_{r(n)})$.

Problema 10.10. ($\uparrow$) En las hipótesis precedentes, probar que R tiene distribución uniforme sobre S_n, es decir, para cada permutación $r \in S_n$, $P(R = r) = 1/n!$ Pruébese además que $X_{(\cdot)}$ y R son independientes.

Problema 10.11. ($\uparrow$) La densidad de $X_{(\cdot)}$ respecto a la medida de Lebesgue en $\mathbb{R}^n$ es

$$\bar{g}(x) = n! \cdot g(x) \cdot I_{\mathbb{R}^{(n)}}(x), \quad x \in \mathbb{R}^n.$$

Problema 10.12. ($\uparrow$) (Distribución de los estadísticos de orden) Probar que $X_{(i)}$ es absolutamente continua y que su densidad es

$$f_{(i)}(x) = n \cdot \binom{n-1}{i-1} \cdot [F(x)]^{i-1} \cdot [1-F(x)]^{n-i} \cdot f(x), \quad x \in \mathbb{R}.$$

Problema 10.13. (↑) Las mismas ideas utilizadas en el Problema 10.12 prueban que, si $1 \leq n_1 < \ldots < n_k \leq n$, la densidad de la v.a. k-dimensional $Z = (X_{(n_1)}, \ldots, X_{(n_k)})$ es

$$f_Z(z_{n_1}, \ldots, z_{n_k}) = n! \cdot \left(\prod_{i=1}^{k} f(z_{n_i}) \right)$$
$$\cdot \frac{\prod_{i=0}^{k} [F(z_{n_{i+1}}) - F(z_{n_i})]^{n_{i+1}-n_i-1}}{\prod_{i=0}^{k} (n_{i+1} - n_i - 1)!} \cdot I_{\mathbb{R}^{(k)}}(z_{n_1}, \ldots, z_{n_k})$$

donde $n_0 = 0$, $n_{k+1} = n + 1$, $z_{n_0} = -\infty$, $z_{n+1} = +\infty$.

Problema 10.14. (↑) Si $f(x) = I_{[0,1]}$ (densidad de la distribución uniforme en $[0, 1]$), entonces $X_{(i)}$ tiene distribución beta de parámetros $(i, n - i + 1)$. Véase la definición de la distribución beta en el Problema 9.4.

Problema 10.15. (↑) (Distribución de los rangos) Determinar la distribución de R_i. Calcular también $E(R_i)$ y $\mathrm{Var}(R_i)$.

Problema 10.16. (↑) Sean $p \in \{1, \ldots, n\}$ e $i_1, \ldots, i_p \in \mathbb{N}$ tales que $1 \leq i_1 < \ldots < i_p \leq n$. Describir la imagen del estadístico $(R_{i_1}, \ldots, R_{i_p})$ y su distribución.

Problema 10.17. (↑) (Estadístico de Wilcoxon) Si $1 \leq p \leq n$, denotemos $W_p = R_1 + \cdots + R_p$. Determinar la esperanza y la varianza de W_p.

Problema 10.18. (★) (Números Aleatorios) Sean $(\Omega, \mathcal{A}, \mathcal{P})$ un espacio probabilístico y U una v.a.r. con distribución uniforme en $[0, 1]$. Si F es una función de distribución en $\mathbb{R}$ entonces la v.a.r. $F^{-1} \circ U$ tiene función de distribución F, donde $F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$, $t \in [0, 1]$.

> *Indicación:* Recordar de la demostración del teorema de Glivenko-Cantelli que F^{-1} es creciente (entonces, medible) y que $F^{-1}(t) \leq x \iff t \leq F(x)$, con lo cual
> $$P(F^{-1} \circ U \leq x) = P(U \leq F(x)) = F(x)$$
> pues U tiene distribución uniforme en $[0, 1]$.

Problema 10.19. (↑) Describir explícitamente un espacio probabilístico y una sucesión (X_n) de v.a. sobre él independientes e idénticamente distribuidas con distribución común uniforme en $\{0, 1, \ldots, 9\}$.

Problema 10.20. (↑) Con las notaciones del problema 12, ¿cuál es la distribución de la v.a.r.

$$0'\underline{X_1}\,\underline{X_2} \ldots \underline{X_n} \left(:= \sum_{i=1}^{n} 10^{-i} X_i \right)?$$

Indicación: $\sum_{i=1}^{n} 10^{-i} X_i$ toma valores en $A_n = \{i \cdot 10^{-n} : 0 \le i < 10^n\}$ y cada valor de A_n lo toma con la misma probabilidad, pues si $x_i, y_i \in \{0, 1, \dots, 9\}$ y $\sum_{i=1}^{n} 10^{-i} x_i = \sum_{i=1}^{n} 10^{-i} y_i$ entonces $x_i = y_i$, $1 \le i \le n$.

Problema 10.21. ($\uparrow$) ¿Cuál es la distribución de la v.a.r.

$$0'\underline{X_1}\,\underline{X_2}\dots (:= \textstyle\sum_{i=1}^{\infty} 10^{-i} X_i)?$$

Indicación: Si $x \in [0, 1] - \cup_{n=1}^{\infty} A_n$ entonces x admite una única representación en la forma $x = \sum_{i=1}^{\infty} 10^{-i} x_i$, $x_i \in \{0, 1, \dots, 9\}$, $\forall i$. Utilizando que $\{\omega : \sum_{i=1}^{\infty} 10^{-i} X_i(\omega) < x\}$ coincide con $\cup_{n=1}^{\infty} \{\omega : X_1(\omega) = x_1, \dots, X_{n-1}(\omega) = x_{n-1}, X_n(\omega) < x_n\}$ (pruébese esa igualdad) probar que la probabilidad de ese suceso es exactamente x. Puesto que la probabilidad del suceso $\{\sum_{i=1}^{\infty} 10^{-i} X_i = x\}$ es 0 para cada x, se verifica que la función de distribución de $\sum_{i=1}^{\infty} 10^{-i} X_i$ coincide con la identidad en $[0, 1] - \cup_{n=1}^{\infty} A_n$ y, entonces, por continuidad a la derecha, con la identidad en $[0, 1]$; es decir, $\sum_{i=1}^{\infty} 10^{-i} X_i$ tiene distribución uniforme en $[0, 1]$.([3])

Problema 10.22. ($\uparrow$) A partir de una tabla de números aleatorios sobre $\{0, 1, \dots, 9\}$, ¿cómo puede fabricarse una tabla aproximada de números aleatorios sobre $[0, 1]$?

Indicación: $\sum_{i=1}^{n} 10^{-i} X_i$ converge puntualmente a $\sum_{i=1}^{\infty} 10^{-i} X_i$ y, entonces, hay convergencia en distribución; ello permite aproximar la función de distribución de $\sum_{i=1}^{\infty} 10^{-i} X_i$ por la distribución uniforme en A_n si n es grande. En fin, siendo $(X_k)_{k \ge 1}$ una muestra de extensión infinita de la distribución uniforme en A_n, la sucesión

$$\sum_{i=1}^{n} 10^{-i} X_i, \quad \sum_{i=n+1}^{2n} 10^{n-i} X_i, \quad \sum_{i=2n+1}^{3n} 10^{2n-i}, \dots$$

es una muestra de extensión infinita de la distribución uniforme en A_n (y, por tanto, una tabla aproximada de números aleatorios en $[0, 1]$).

Problema 10.23. ($\uparrow$) ¿Cómo se puede fabricar, a partir de una muestra de extensión n de una distribución uniforme en $[0, 1]$, una muestra de extensión n de una distribución F real dada?

Problema 10.24. (Distribución hipergeométrica: muestreo sin reemplazamiento) Sean $N, N_1, n \in \mathbb{N}$ tales que $N_1, n \le N$. La distribución hipergeométrica $H(N, N_1, n)$

[3] He aquí una nueva descripción de la medida de Lebesgue similar -ahora en base 10- a la que habíamos dado en el Problema 8.31 en base 2. El resultado se puede generalizar a otras bases: Sean $k \in \mathbb{N}$, $\Omega = \{0, 1, \dots, k-1\}$, Q la distribución uniforme discreta en Ω, $P = Q^{\mathbb{N}}$, $X_i : \Omega^{\mathbb{N}} \to \Omega$ la proyección i-ésima y $T := \sum_{i=1}^{\infty} X_i / k^i$; entonces la distribución de T respecto a P es la medida de Lebesgue en $[0, 1]$.

es la distribución de una v.a. discreta que toma los valores $k = 0, 1, \dots, n$ con probabilidades

$$p_k = \frac{\binom{N_1}{n}\binom{N-N_1}{n-k}}{\binom{N}{n}},$$

donde se ha denotado $\binom{t}{r} = \frac{t(t-1)\cdots(t-r+1)}{r!}$ si $t \in \mathbb{R}$ y $r \in \mathbb{N}$. Probar que ésta es la distribución correspondiente al muestreo sin reemplazamiento, es decir, la distribución del número de bolas blancas obtenidas al extraer sin reemplazamiento n bolas de una urna que contiene N_1 bolas blancas y $N - N_1$ bolas negras. Probar que la media y la varianza de esta distribución son, respectivamente, $\frac{nN_1}{N}$ y $n \cdot \frac{N-n}{n-1} \cdot \frac{N_1}{N} \cdot \left(1 - \frac{N_1}{N}\right)$.

Problema 10.25. (Distribución multinomial) Supongamos que los individuos de una población pueden ser clasificados en una y sólo una de r clases disjuntas, y que la probabilidad de que un individuo esté en la clase j es p_j; entonces $p_1, \dots, p_r \geq 0$ y $\sum_{k=1}^r p_k = 1$. Seleccionamos al azar y con reemplazamiento n individuos de la población y denotamos por X_i la clase a la que pertenece el individuo i-ésimo. Para $1 \leq j \leq r$, denotaremos por Z_j el número de individuos de la muestra que caen en la clase j, es decir, $Z_j = \sum_{i=1}^n I_{\{X_i=j\}}$. Llamaremos distribución multinomial de orden n y parámetro $p = (p_1, \dots, p_r)$ a la distribución de la v.a. discreta r-dimensional $(Z_1, \dots, Z_r)$; la denotaremos por $M(n; p_1, \dots, p_r)$.

(a) Probar que es una distribución concentrada en $E := \{(i_1, \dots, i_r) \in \mathbb{N}_0^r : i_1 + \cdots + i_r = n\}$ y que

$$P(Z_1 = i_1, \dots, Z_r = i_r) = \frac{n!}{i_1! \cdots i_r!} p_1^{i_1} \cdots p_r^{i_r}.$$

Problema 10.26. (★) (Un conjunto Lebesgue-medible que no es boreliano) En este problema y los siguientes pretendemos utilizar el conjunto de Cantor para construir un conjunto Lebesgue-medible que no es Borel-medible. Consideremos la aplicación $T' : \omega = (\omega_i) \in \{0,1,2\}^{\mathbb{N}} \mapsto T'(\omega) := \sum_{i=1}^n \frac{\omega_i}{3^i} \in [0,1]$. Se define $K := T'(\{0,2\}^{\mathbb{N}}) \subset [0,1]$. K es el llamado conjunto de Cantor; siendo T' continua y $\{0,2\}^{\mathbb{N}}$ compacto, K es compacto. Por tanto, K está formado por los puntos del intervalo $[0,1]$ que admiten en base 3 una representación decimal en la que solo se utilizan las cifras 0 y 2. Geométricamente: sea $K_0 = [0,1]$; K_1 se obtiene dividiendo K_0 en 3 intervalos iguales y retirando el intervalo central (es decir, retirando aquellos puntos que en base 3 se escriben en la forma 0'1...); K_2 se obtiene dividiendo cada uno de los intervalos que conforman K_1 en 3 partes iguales y retirando el intervalo central de cada uno de ellos (es decir, retirando aquellos puntos que en base 3 se escriben en la forma 0'01... o 0'21...); procediendo de ese modo obtenemos una sucesión (K_n) de compactos y se define $K = \cap_n K_n$ (aritméticamente, lo que hacemos es eliminar sucesivamente el dígito 1 de la primera, segunda,... posición decimal). Sean Q la probabilidad en $\{0,2\}$ definida por $Q(\{0\}) = Q(\{2\}) = 1/2$, $P = Q^{\mathbb{N}}$ y T la restricción de T' a $\{0,2\}^{\mathbb{N}}$. P^T es una probabilidad en $[0,1]$ concentrada en el conjunto de Cantor que llamaremos medida de Cantor.([4]) La función F de distribución de T respecto a P es la llamada función de Cantor; puesto que la medida de Cantor es trivialmente continua, la función de Cantor es continua; es, también, como cualquier función de distribución, creciente. De hecho, F vale $1/2$ en el intervalo central retirado de K_0 para definir K_1; vale $1/4$ y $3/4$, resp., en los dos intervalos centrales retirados de K_1 para obtener K_2; y así sucesivamente. Sea φ la correspondiente función cuantil,([5]) es decir, $\varphi(y) = \text{ínf}\{x \in \mathbb{R}\colon F(x) \geq y\} = \text{ínf}\{x \in \mathbb{R}\colon F(x) = y\}$. Siendo F continua, $F(\varphi(y)) = y$, para cada $y \in [0,1]$; por tanto, φ es inyectiva. Se prueba que φ toma valores en el conjunto de Cantor; además, siendo creciente, φ es Borel-medible. Aprovechando lo anterior y la existencia de un subconjunto de $[0,1]$ que no es Lebesgue-medible -ver ejemplo de Vitali en la pág. 36-, construir un conjunto Lebesgue-medible de $[0,1]$ que no sea Borel-medible.

> *Indicación:* Sean $A \subset [0,1]$ no Lebesgue-medible y $B = \varphi(A) \subset K$; K tiene medida de Lebesgue nula([6]) y, por tanto, B es Lebesgue-medible -ver Teorema 3.1- y tiene medida de Lebesgue nula. Pero B no puede ser Borel-medible, pues, en ese caso, también lo sería $\varphi^{-1}(B)$, que coincide con A por ser φ inyectiva.

[4] El teorema de la medida imagen y el de Fubini en el caso de un producto infinito numerable de factores nos permiten calcular integrales respecto a la medida de Cantor -otra cuestión es que los cálculos sean más o menos simples-.

[5] Ver propiedades de función cuantil en la demostración del Teorema 10.1 de Glivenko-Cantelli.

[6] Para probar esta afirmación, se considera la distribución Q' uniforme discreta en $\{0,1,2\}$, y la probabilidad producto $P' = Q'^{\mathbb{N}}$ en $\{0,1,2\}^{\mathbb{N}}$, y se prueba, con argumentos análogos a los del Problema 8.31, que $P'^{T'}$ es la medida de Lebesgue λ en $[0,1]$; basta ya notar que $\lambda(K) = P'^{T'}(\{0,2\}^{\mathbb{N}})$=0

Capítulo 11

Teorema de Radon-Nikodym

El teorema de Radon-Nikodym (Teorema 11.2), especialmente útil en el capítulo dedicado a la esperanza condicional, es una especie de recíproco de un resultado probado anteriormente en el que se introducía el concepto de derivada de Radon-Nikodym. Para probarlo necesitaremos el concepto de medida real y algunos resultados previos.

DEFINICIÓN 11.1. (Medida real) Una medida real en un espacio medible $(\Omega, \mathcal{A})$ es una función de conjuntos $\mu : \mathcal{A} \to \bar{\mathbb{R}}$ nula en el vacío y numerablemente aditiva.

***Observación* 11.1.** Forma parte de la definición que, p.ej., para cada $A \in \mathcal{A}$, $\mu(A)+\mu(A^c)$ debe estar bien definido (es decir, no puede ser de la forma $(+\infty) - (+\infty)$ o $(+\infty) + (-\infty)$) y ser igual a $\mu(\Omega)$; por tanto, si existe $A \in \mathcal{A}$ tal que $\mu(A) = +\infty$ debe ser $\mu(\Omega) = +\infty$ y si existe $A \in \mathcal{A}$ tal que $\mu(A) = -\infty$ debe ser $\mu(\Omega) = -\infty$. Por tanto, los valores $+\infty$ y $-\infty$ no pueden estar simultáneamente en el rango de una medida real. Un argumento análogo prueba que si $B \in \mathcal{A}$ y $\mu(B)$ es finito entonces $\mu(A)$ es finito para cada suceso $A \subset B$.

DEFINICIÓN 11.2. (Conjuntos positivos y negativos para una medida real) Si μ es una medida real en un espacio medible $(\Omega, \mathcal{A})$, un subconjunto A de Ω se dice un conjunto positivo para μ si $A \in \mathcal{A}$ y cada subconjunto medible de A tiene medida positiva. Análogamente, un subconjunto A de Ω se dice un conjunto negativo para μ si $A \in \mathcal{A}$ y cada subconjunto medible de A tiene medida negativa.

Lema 11.1. *Sea μ una medida real en $(\Omega, \mathcal{A})$ y A un conjunto medible tal que $-\infty < \mu(A) < 0$. Entonces A contiene un conjunto negativo B tal que $\mu(B) \leq \mu(A)$.*

Demostración. Sea $\delta_1 = \sup\{\mu(E) \colon E \in \mathcal{A},\ E \subset A\}$, y tomemos $A_1 \in \mathcal{A}$ tal que $A_1 \subset A$ y $\mu(A_1) \geq \text{mín}(\delta_1/2, 1)$. Nótese que si no existiera A_1 en esas condiciones, se tendría que $\mu(E) < \text{mín}(\delta_1/2, 1)$ para cada subconjunto medible E de A y, por definición de supremo, no podría ocurrir que $\delta_1 > 0$; se tendría entonces que $\delta_1 = 0$ (el vacío es un subconjunto medible de A y $\delta_1 \geq \mu(\emptyset) = 0$), y el propio A sería un subconjunto negativo de A tal que $\mu(A) \leq \mu(A)$. Procedemos por inducción construyendo sucesiones $(\delta_n)_n$ y $(A_n)_n$ haciendo, para $n \geq 2$, $\delta_n = \sup\{\mu(E) \colon E \in \mathcal{A},\ E \subset A \setminus \cup_{i=1}^{n-1} A_i\}$ y eligiendo $A_n \in \mathcal{A}$, $A_n \subset A \setminus \cup_{i=1}^{n-1} A_i$, tal que $\mu(A_n) \geq \text{mín}(\delta_n/2, 1)$. Como antes, si no pudiésemos encontrar un tal A_n, se tendría que $\delta_n = 0$ y $A \setminus \cup_{i=1}^{n-1} A_i$ sería un subconjunto negativo de A tal que $\mu(A) = \sum_{i=1}^{n-1} \mu(A_i) + \mu(A \setminus \cup_{i=1}^{n-1} A_i) \geq \mu(A \setminus \cup_{i=1}^{n-1} A_i)$ pues $\mu(A_i) \geq 0$ para $1 \leq i \leq n-1$.

Así, en el peor de los casos, podemos definir $A_\infty = \cup_{n=1}^{\infty} A_n$ y $B = A \setminus A_\infty$. Veamos que B conviene. Puesto que los A_n son dos a dos disjuntos y $\mu(A_n) \geq 0$, se tiene que $\mu(A_\infty) \geq 0$ y, entonces, $\mu(B) \leq \mu(B) + \mu(A_\infty) = \mu(A)$. Veamos que B es un conjunto negativo. Siendo $\mu(A)$ finito, también lo es $\mu(A_\infty)$. Puesto que $\mu(A_\infty) = \sum_n \mu(A_n)$, se tiene que $\text{lím}_n \mu(A_n) = 0$. Entonces $\text{lím}_n \delta_n = 0$. Puesto que un subconjunto medible E de B arbitrario verifica que $\mu(E) \leq \delta_n$, para cada n, debe ser $\mu(E) \leq 0$. $\square$

TEOREMA 11.1. (Teorema de descomposición de Hahn) *Sea μ una medida real en $(\Omega, \mathcal{A})$. Existen entonces sucesos disjuntos P y N de Ω tales que P es un conjunto positivo para μ, N es un conjunto negativo para μ y $\Omega = P \cup N$.*

Demostración. Puesto que el rango de μ no puede contener simultáneamente a $+\infty$ y $-\infty$, podemos suponer que, p.ej., no contiene a $-\infty$. Hagamos

$$L = \text{ínf}\{\mu(A) \colon A \text{ es un conjunto negativo para } \mu\}.$$

Estamos considerando el ínfimo de un conjunto que es no vacío, pues contiene al menos a $\mu(\emptyset) = 0$. Sea $(A_n)_n$ una sucesión de conjuntos negativos tal que $L = \text{lím}_n \mu(A_n)$

y sea $N = \cup_n A_n$. Nótese que cada subconjunto medible de N se puede escribir como unión disjunta de conjuntos medibles cada uno de ellos contenido en algún A_n, lo que prueba que N es un conjunto negativo para μ. Entonces $L \leq \mu(N) \leq \mu(A_n) + \mu(N \setminus A_n) \leq \mu(A_n)$ para cada n, y, por tanto, $L = \mu(N)$. Además, puesto que μ no toma el valor $-\infty$, $\mu(N)$ debe ser finito. Sea $P = N^c$ y veamos que P es positivo para μ. Si no, existiría un suceso $A \subset P$ tal que $-\infty < \mu(A) < 0$ y, por el lema anterior, A contendría un conjunto negativo B tal que $\mu(B) < 0$, y $N \cup B$ sería un conjunto negativo tal que $\mu(N \cup B) = \mu(N) + \mu(B) < \mu(N) = L$ (L es finito) en contra de la definición de L. □

DEFINICIÓN 11.3. Una descomposición de Hahn de una medida real μ es una partición medible (P, N) de Ω tal que P es positivo para μ y N es negativo para μ.

Corolario 11.1. (Teorema de descomposición de Jordan) *Cada medida real es diferencia de dos medidas positivas, una de las cuales, al menos, es finita.*

Demostración. Si (P, N) es una descomposición de Hahn para μ, se definen funciones de conjunto μ^+ y μ^- en $\mathcal{A}$ por

$$\mu^+(A) := \mu(A \cap P) \quad \text{y} \quad \mu^-(A) := -\mu(A \cap N).$$

Es claro que μ^+ y μ^- son medidas positivas en $\mathcal{A}$ y que $\mu = \mu^+ - \mu^-$. Puesto que $+\infty$ y $-\infty$ no están simultáneamente en el rango de μ, o $\mu(P)$ es finito o lo es $\mu(N)$, o lo son ambos, lo que prueba que al menos una de las medidas μ^+ o μ^- es finita. □

TEOREMA 11.2. (Radon-Nikodym) *Sean λ y μ medidas σ-finitas en $\mathcal{A}$. Si λ está dominada por μ, entonces existe una función Borel-medible $g : \Omega \longrightarrow [0, +\infty[$ (que llamaremos densidad o derivada de Radon-Nikodym de λ respecto a μ) no negativa tal que $\lambda(A) = \int_A g d\mu$, para cada $A \in \mathcal{A}$. La función g es, además, esencialmente única, en el sentido de que cualquier otra densidad h de λ respecto a μ coincide con g μ-c.s. El teorema sigue siendo cierto si λ es una medida real arbitraria y μ es una medida σ-finita que domina a λ, pero en ese caso g es $\bar{\mathbb{R}}$-valorada.*

Demostración. Primer caso: Consideremos en primer lugar el caso en que λ y μ son finitas. Sea

$$\mathcal{F} := \left\{ f : \Omega \to [0, +\infty] \colon \int_A f d\mu \leq \lambda(A),\ \forall A \in \mathcal{A} \right\}.$$

Veamos que $\mathcal{F}$ contiene una función g tal que $\int g d\mu = \sup\{\int f d\mu \colon f \in \mathcal{F}\}$ y después que esa función g conviene.

- $\mathcal{F}$ es no vacío pues $0 \in \mathcal{F}$.

- Si $f_1, f_2 \in \mathcal{F}$ entonces $\sup(f_1, f_2) \in \mathcal{F}$: dado $A \in \mathcal{A}$, hagamos $A_1 = \{f_1 > f_2\} \cap A$ y $A_2 = \{f_1 \leq f_2\} \cap A$; entonces

$$\int_A \sup(f_1, f_2) d\mu = \int_{A_1} f_1 d\mu + \int_{A_2} f_2 d\mu \leq \lambda(A_1) + \lambda(A_2) = \lambda(A).$$

- Sea $(f_n)_n$ una sucesión en $\mathcal{F}$ tal que

$$\lim_n \int f_n d\mu = \sup\left\{ \int f d\mu \colon f \in \mathcal{F} \right\}.$$

Podemos suponer que $(f_n)_n$ es creciente reemplazando f_n por $\sup(f_1, \dots, f_n)$. Sea $g = \lim_n f_n$. El teorema de la convergencia monótona prueba que

$$\int_A g d\mu = \lim_n \int_A f_n d\mu \leq \lambda(A),$$

para cada $A \in \mathcal{A}$. Entonces $g \in \mathcal{F}$. Además $\int g d\mu = \sup\{\int f d\mu \colon f \in \mathcal{F}\}$.

Veamos ahora que $\lambda(A) = \int_A g d\mu$ para cada $A \in \mathcal{A}$. Puesto que $g \in \mathcal{F}$, la expresión $\lambda_0(A) := \lambda(A) - \int_A g d\mu$ define una medida positiva en $\mathcal{A}$. Basta probar que $\lambda_0 = 0$. Si no, siendo μ finita, existe $\epsilon > 0$ tal que

$$\lambda_0(\Omega) > \epsilon\mu(\Omega).$$

Sea (P, N) una descomposición de Hahn para la medida real $\lambda_0 - \epsilon\mu$. Nótese que para cada $A \in \mathcal{A}$ se tiene que $\lambda_0(A \cap P) \geq \epsilon\mu(A \cap P)$, y entonces

$$\begin{aligned} \lambda(A) &= \int_A g d\mu + \lambda_0(A) \geq \int_A g d\mu + \lambda_0(A \cap P) \\ &\geq \int_A g d\mu + \epsilon\mu(A \cap P) = \int_A (g + \epsilon I_P) d\mu; \end{aligned}$$

nótese además que $\mu(P) > 0$, pues si $\mu(P) = 0$ entonces $\lambda_0(P) = 0$ y $\lambda_0(\Omega) - \epsilon\mu(\Omega) = \lambda_0(N) - \epsilon\mu(N) \leq 0$ en contra de la elección de ϵ. Por tanto, $\int g d\mu \leq \lambda(\Omega) < +\infty$, $g + \epsilon I_P \in \mathcal{F}$ y $\int (g + \epsilon I_P) d\mu > \int g d\mu$. Ello contradice la maximalidad de g en $\mathcal{F}$ y, por tanto, $\lambda_0 = 0$. Así pues, $\lambda(A) = \int_A g d\mu$ para cada $A \in \mathcal{A}$. Siendo λ finita, g es μ-integrable y, por tanto, finita μ-c.s., y puede modificarse de forma que sólo tome valores reales. Esto acaba la prueba en el caso de que λ y μ sean finitas.

Segundo caso: Supongamos ahora que λ y μ son σ-finitas. Existe entonces una sucesión disjunta $(B_n)_n$ en $\mathcal{A}$ que recubre Ω formada por sucesos con medidas λ y μ finitas. Para cada $n \in \mathbb{N}$, la primera parte asegura la existencia de una función $g_n : B_n \to [0, +\infty[$ tal que $\lambda(A) = \int_A g_n d\mu$ para cada $A \in \mathcal{A}$ tal que $A \subset B_n$. Hagamos $g(\omega) = g_n(\omega)$ si $\omega \in B_n$ (es decir, $g = \sum_n g_n$). Es claro que g conviene.

Tercer caso: Supongamos ahora que μ es finita y que λ es una medida positiva arbitraria. Denotemos por $\mathcal{C}$ la clase de los conjuntos $C \in \mathcal{A}$ sobre los que λ es σ-finita (es decir, tal que C puede ser recubierto por una sucesión de conjuntos medibles de λ-medida finita). Puesto que $\emptyset \in \mathcal{C}$, $\mathcal{C}$ es no vacío. Sea $s = \sup\{\mu(A) \colon A \in \mathcal{C}\}$ y tomemos una sucesión $(C_n)_n$ en $\mathcal{C}$ tal que $\mu(C_n)$ converge a s. Si $C = \cup_n C_n$, es claro que $C \in \mathcal{C}$ y que $\mu(C_n) \leq \mu(C) \leq s$ para cada n; por tanto, $\mu(C) = s$. Por el segundo caso, existe una función medible no negativa $g' : C \to [0, +\infty[$ tal que

$$\lambda(A \cap C) = \int_{A \cap C} g' d\mu$$

para cada $A \in \mathcal{A}$. Consideremos ahora un suceso arbitrario $A \in \mathcal{A}$. Podemos distinguir dos subcasos:

(a) $\mu(A \cap C^c) > 0$. Entonces $\lambda(A \cap C^c) = \infty$ pues si $\lambda(A \cap C^c) < \infty$, entonces $C \cup (A \cap C^c) \in \mathcal{C}$ y tendríamos

$$s \geq \mu(C \cup (A \cap C^c)) = \mu(C) + \mu(A \cap C^c) > \mu(C) = s,$$

lo que es una contradicción.

(b) $\mu(A \cap C^c) = 0$. Entonces debe ser $\lambda(A \cap C^c) = 0$ por continuidad absoluta.

Así pues, en cualquier caso se tiene que $\lambda(A \cap C^c) = \int_{A\cap C^c} \infty d\mu$ y

$$\lambda(A) = \lambda(A \cap C) + \lambda(A \cap C^c) = \int_A g d\mu$$

donde $g = g'$ sobre C y $g = \infty$ sobre C^c.

Cuarto caso: Supondremos ahora que μ es σ-finita y que λ es una medida arbitraria. Descomponemos Ω como unión numerable disjunta de sucesos A_n de μ-medida finita y aplicamos el caso anterior para obtener funciones g_n medibles $\bar{\mathbb{R}}$-valoradas tales que, para cada $A \in \mathcal{A}$ y cada n, $\lambda(A \cap A_n) = \int_{A\cap A_n} g_n d\mu$. Es claro que $g \sum_n g_n$ conviene.

Quinto caso: En el último caso suponemos que μ es σ-finita y que λ es una medida real arbitraria. Escribamos $\lambda = \lambda^+ - \lambda^-$ donde, supongamos, λ^- es finita. Existen entonces funciones medibles no negativas g_1 y g_2 tales que $\lambda^+(A) = \int_A g_1 d\mu$ y $\lambda^-(A) = \int_A g_2 d\mu$ para cada $A \in \mathcal{A}$. Puesto que λ^- es finita, g_2 es integrable y finita μ-c.s. Entonces $g = g_1 - g_2$ conviene. □

***Observación* 11.2.** (Densidades en el caso discreto: función de probabilidad) Una v.a. a valores en un espacio discreto (es decir, en un conjunto numerable provisto de la σ-álgebra de todos sus subconjuntos) o en un subespacio discreto de otro espacio medible, se llamará una v.a. discreta. Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, Ω' un conjunto numerable y $X : \Omega \longrightarrow \Omega'$ una v.a. La distribución P^X de X respecto a P queda determinada por los valores $P^X(\{\omega'\})$, $\omega' \in \Omega'$, y está obviamente dominada por la medida cardinal μ en Ω'. En este caso, la densidad de P^X respecto a μ es la llamada función de probabilidad de la v.a. X y se verifica que

$$\frac{dP^X}{d\mu}(\omega') = P^X(\{\omega'\}) = P(X = \omega'), \quad \omega' \in \Omega'.$$

***Observación* 11.3.** (El caso absolutamente continuo) Cuando se habla de la función de densidad de una v.a. X a valores en $\mathbb{R}^n$ (o lo que es lo mismo, de la densidad de su distribución P^X) sin hacer alusión a medida dominante alguna, se entenderá que ésta es la medida de Lebesgue en $\mathbb{R}^n$.

***Observación* 11.4.** (Relación con resultados de análisis real) Una aplicación $f : [a, b] \longrightarrow \mathbb{R}$ se dice absolutamente continua si para cada $\epsilon > 0$ existe $\delta > 0$ tal que para cada colección finita $]a_i, b_i[$, $1 \leq i \leq n$, de subintervalos abiertos disjuntos de $[a, b]$ cuya suma de longitudes es $\leq \delta$ se tiene que $\sum_{i=1}^n |f(b_i) - f(a_i)| \leq \epsilon$. Se prueba que si F y G son funciones de distribución en $[a,b]$ con medidas de Lebesgue-Stieljes (finitas) μ y ν y si $H = F - G$ y $\lambda = \mu - \nu$ (luego, $\lambda]x,y] = H(y) - H(x)$) entonces λ es absolutamente continua respecto a la medida de Lebesgue si y sólo si H es absolutamente continua. Se prueba también que una aplicación $f : [a,b] \longrightarrow \mathbb{R}$ es absolutamente continua si y sólo si es una integral indefinida, es decir, $f(x) - f(a) = \int_a^x g(t)dt$ para cada $x \in [a,b]$, donde $g : [a,b] \longrightarrow \mathbb{R}$ es una función Borel-medible e integrable respecto a la medida de Lebesgue; además, en ese caso, f es derivable en casi todo punto de $[a,b]$ y $f' = g$ c.s. (respecto a la medida de Lebesgue). Completamos este comentario con el llamado teorema de diferenciación de Lebesgue: "Cada medida de Lebesgue-Stieljes μ en $\mathbb{R}$ se puede descomponer en la forma $\mu = \mu_a + \mu_s$, siendo μ_a una medida absolutamente continua respecto a la medida de Lebesgue y μ_s una medida singular respecto a la medida de Lebesgue (es decir, concentrada en un suceso de medida de Lebesgue nula); además, si F denota la función de distribución asociada a μ y $f \in d\mu_a/d\lambda$, entonces $F'(x) = f(x)$, λ-c.s."

PROBLEMAS DEL CAPÍTULO 11

Problema 11.1. a) Probar que la medida de Dirac ε_x en un punto $x \in \mathbb{R}$ no tiene densidad respecto a la medida de Lebesgue en $\mathbb{R}$. b) Calcular la densidad de una distribución $N(\mu, 1)$ respecto a una distribución $N(0, 1)$. c) Probar que la distribución uniforme discreta en $\{1, \ldots, n\}$ tiene densidad respecto a la distribución binomial $b_n(1/2)$ y calcularla.

Problema 11.2. Sean λ, μ y ν medidas σ-finitas en $(\Omega, \mathcal{A})$ tales que $\lambda \ll \mu \ll \nu$. Probar que

$$\frac{d\lambda}{d\nu} = \frac{d\lambda}{d\mu} \cdot \frac{d\mu}{d\nu}, \quad \nu - \text{c.s.}$$

En particular, si $\lambda \ll \mu$ y $\mu \ll \lambda$, entonces $d\lambda/d\mu = (d\mu/d\lambda)^{-1}$, μ-c.s.

Problema 11.3. Sean $(\Omega, \mathcal{A}, \mu)$ un espacio de medida σ-finito y $T : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$. Probar que $\mu^T \ll \mu$ si, y sólo si, $T^{-1}(N)$ es un suceso μ-nulo para cada suceso μ-nulo N.

Problema 11.4. Poner un ejemplo de dos v.a.r. X e Y sobre un mismo espacio de probabilidad de forma que cada una de ellas tenga densidad (respecto a la medida de Lebesgue en $\mathbb{R}$), pero que (X, Y) no tenga densidad respecto a la medida de Lebesgue en $\mathbb{R}^2$.

Problema 11.5. Sea X una v.a.r. con distribución uniforme en $[-1, 1]$. Determinar las distribuciones de $|X|$ y de X^2.

Capítulo 12

Definición de Esperanza Condicional

Este capítulo pretende hacer un estudio sistemático y riguroso de la noción probabilística de dependencia siguiendo las pautas marcadas por Kolmogorov en los años treinta.

En lo que sigue $(\Omega, \mathcal{A}, P)$ será un espacio de probabilidad. Si $A, B \in \mathcal{A}$ y $P(A) > 0$, se define la probabilidad condicional $P(B|A)$ como $P(A \cap B)/P(A)$. Se obtiene así una probabilidad P_A en $(\Omega, \mathcal{A})$ definida por $P_A(B) = P(B|A)$, $B \in \mathcal{A}$, totalmente concentrada en A (es decir, $P_A(A) = 1$) y que no es más que la probabilidad P restringida al subespacio medible (normalizada por el factor $P(A)^{-1}$ para que P_A sea efectivamente una probabilidad). Si Y es una v.a.r. no negativa o P-integrable en $(\Omega, \mathcal{A})$, se puede definir la esperanza condicional de Y respecto al suceso A (ver el Problema 4.6), que denotaremos $E(Y|A)$, mediante $E(Y|A) = \int Y dP_A = P(A)^{-1} \int_A Y dP$, de tal suerte que $P(B|A) = E(I_B|A)$; si podemos realizar un experimento aleatorio del que se obtienen números reales de acuerdo con la distribución de Y, la ley fuerte de los grandes números nos permite aproximar $E(Y|A)$ por el valor medio de Y en una sucesión de pruebas independientes del experimento si sólo tenemos en cuenta el valor de Y después de conocer que ha ocurrido el suceso A. Estas definiciones de probabilidad condicional y de esperanza condicional respecto a un suceso de probabilidad no nula nos permiten detectar la posible dependencia entre dos sucesos o entre un suceso y una v.a.

Nos proponemos en esta sección estudiar la dependencia entre otro tipo de objetos probabilísticos, como pueden ser dos v.a. o una v.a. y una σ-álgebra; para ello necesitamos extender las definiciones de probabilidad y esperanza condicionales a un contexto más amplio.

El teorema siguiente, que es consecuencia inmediata del teorema de Radon-Nikodym (Teorema 11.2) , nos permitirá obtener la definición deseada de esperanza condicional.

En el resto de esta sección, mientras no se diga lo contrario, $(\Omega', \mathcal{A}')$ será un espacio medible, $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ una v.a. e Y será una v.a.r. definida en $(\Omega, \mathcal{A})$ no negativa o con esperanza finita.

TEOREMA 12.1. *En las condiciones precedentes, si Y es no negativa (resp., si Y tiene esperanza finita), existe una v.a. $g : (\Omega', \mathcal{A}') \longrightarrow \bar{\mathbb{R}}$ no negativa (resp., con esperanza finita respecto a P^X) que, para cada $A' \in \mathcal{A}'$, verifica que*

$$\int_{X^{-1}(A')} Y(\omega)dP(\omega) = \int_{A'} g(x)dP^X(x). \tag{12.1}$$

Además, g es esencialmente única, en el sentido de que cualquier otra v.a. que satisfaga lo mismo que g coincide con g, P^X-c.s. Por último, si Y tiene esperanza finita, entonces g es finita P^X-c.s.

Demostración. Supongamos en primer lugar que Y es no negativa. La función de conjuntos

$$\lambda : A' \in \mathcal{A}' \longrightarrow \lambda(A') = \int_{X^{-1}(A')} Y(\omega)dP(\omega) \in [0, +\infty]$$

está bien definida (por ser Y medible no negativa), y define una medida en la σ-álgebra $\mathcal{A}'$; en efecto, si $(A'_n)_n$ es una sucesión disjunta en $\mathcal{A}'$, entonces $X^{-1}(\cup_n A'_n) = \cup_n X^{-1}(A'_n)$ y $(X^{-1}(A'_n))_n$ es una sucesión disjunta en $\mathcal{A}$; se sigue entonces del teorema de Beppo-Levy que

$$\lambda(\cup_n A'_n) = \int_{X^{-1}(\cup_n A'_n)} Y(\omega)dP(\omega) = \int_{\cup_n X^{-1}(A'_n)} Y(\omega)dP(\omega)$$
$$= \int_\Omega Y(\omega)I_{\cup_n X^{-1}(A'_n)}(\omega)dP(\omega) = \int_\Omega Y(\omega)(\sum_n I_{X^{-1}(A'_n)}(\omega))dP(\omega)$$
$$= \sum_n \int_\Omega Y(\omega)I_{X^{-1}(A'_n)}(\omega)dP(\omega) = \sum_n \lambda(A'_n).$$

Además, λ está dominada por P^X pues, si $P^X(A') = 0$ (es decir, si $P(X^{-1}(A')) = 0$), entonces $\int_{X^{-1}(A')} Y(\omega)dP(\omega) = 0$ pues $YI_{X^{-1}(A')}$ es nula P-c.s. El teorema de Radon-Nikodym (Teorema 11.2) garantiza la existencia de una v.a. $g : (\Omega', \mathcal{A}') \longrightarrow [0, +\infty]$ tal que $\lambda(A') = \int_{A'} g(x)dP^X(x)$.

Si Y es P-integrable, podemos escribir $Y = Y^+ - Y^-$ (tómese como Y^+ e Y^- las partes positiva y negativa de Y, resp.) donde Y^+ e Y^- son no negativas y P-integrables; del caso no negativo se sigue la existencia de v.a.r. $g_+, g_- : (\Omega', \mathcal{A}') \longrightarrow [0, +\infty]$ tales que

$$\int_{X^{-1}(A')} Y^\pm(\omega)dP(\omega) = \int_{A'} g_\pm(x)dP^X(x),$$

para cada $A' \in \mathcal{A}'$. Siendo Y^+ e Y^- P-integrables, g_+ y g_- son P^X-integrables y, por tanto, finitas P^X-c.s., y la aplicación $g := g_+ - g_-$ está bien definida P^X-c.s. (hágase, por ejemplo, $g = 0$ en el conjunto excepcional donde la indeterminación $(+\infty) - (+\infty)$ pueda tener lugar), es P^X-integrable y verifica (12.1).

La unicidad esencial se sigue sin dificultad de la siguiente propiedad conocida de la integral: "Sean h, g funciones reales medibles sobre un espacio de medida σ-finito $(\Omega, \mathcal{A}, \mu)$; si existen las integrales de h y g respecto a μ y $\int_A hd\mu \leq \int_A gd\mu$, para cada $A \in \mathcal{A}$, entonces $h \leq g$, μ-c.s." □

DEFINICIÓN 12.1. (Esperanza y probabilidad condicionales respecto a una v.a.) Sean $Y : (\Omega, \mathcal{A}) \longrightarrow \mathbb{R}$ una v.a.r. no negativa (resp., P-integrable) y $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$. Llamaremos (una versión de la) esperanza condicional de Y respecto a X, y se denotará por $E(Y|X)$, una v.a. $\bar{\mathbb{R}}$-valorada no negativa (resp., $\mathbb{R}$-valorada con esperanza finita respecto a P^X) g en $(\Omega', \mathcal{A}')$ que satisface (12.1). Para cada suceso $A \in \mathcal{A}$ se define la probabilidad condicional de A respecto a X mediante $P(A|X) = E(I_A|X)$.

***Observación* 12.1.** Nótese que la función g sólo es única esencialmente; cualquier función medible h que coincida P^X-c.s. con g también verifica (12.1) y, según afirma el teorema, si h verifica (12.1), entonces coincide con g, P^X-c.s. ¿A cuál de las funciones g que verifican (12.1) hemos llamado esperanza condicional de Y respecto a X? Debemos considerar $E(Y|X)$, más que como una v.a., como una clase de equivalencia de v.a., siendo equivalentes dos v.a. si, y sólo si, coinciden P^X-c.s. Por ello, a veces, diremos que g es un representante o una versión de la esperanza condicional de Y respecto a X. En otras ocasiones, en las que el resultado no dependa del representante elegido en la clase, denotaremos por $E(Y|X)$ cualquiera de los representantes de esta clase de equivalencia. Observaciones análogas se pueden hacer para la probabilidad condicional $P(A|X)$ que, correctamente hablando, es una clase de v.a. g que podemos suponer reales (incluso $[0,1]$-valoradas si conviene, pues I_A toma valores en $[0,1]$) que verifican $P(A \cap X^{-1}(A')) = \int_{A'} g(x) dP^X(x)$, para cada $A' \in \mathcal{A}'$.

***Observación* 12.2.** Seguiremos la costumbre habitual de denotar por $E(Y|X = x)$ el valor $g(x)$, $x \in \Omega'$, si g es un representante de la clase $E(Y|X)$ (sin olvidar, por supuesto, que esta notación depende del representante elegido). Asumiendo este abuso de notación, se verifica, para cada $A' \in \mathcal{A}'$, que $\int_{X^{-1}(A')} Y dP = \int_{A'} E(Y|X = x) dP^X(x)$. A veces escribiremos también $g \in E(Y|X)$ para enfatizar que g es una versión de la esperanza condicional de Y respecto a X. La notación $E(Y|X = x)$ ya había sido utilizada anteriormente en el caso de que $P(X = x) > 0$, pero es fácil ver que, en ese caso, ambas definiciones coinciden.

***Observación* 12.3.** Denotemos por $\mathcal{L}^1(\Omega, \mathcal{A}, P; \mathbb{R})$ el espacio vectorial de las funciones reales medibles y P-integrables en $(\Omega, \mathcal{A})$, y por $L^1(\Omega, \mathcal{A}, P; \mathbb{R})$ el cociente de $\mathcal{L}^1(\Omega, \mathcal{A}, P; \mathbb{R})$ por la relación de equivalencia que identifica dos funciones que coinciden P-c.s. Puesto que dos representantes de una misma clase tienen la misma esperanza condicional respecto a la v.a. $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ (dos funciones que coinciden P-c.s. tienen la misma integral sobre cualquier suceso de $\mathcal{A}$) se puede hablar sin problemas de la esperanza condicional de Y respecto a X si $Y \in L^1(\Omega, \mathcal{A}, P; \mathbb{R})$ (Y es ahora una clase de equivalencia); formalmente, debe entenderse que $E(Y|X)$ (sea Y una clase o cualquiera de sus representantes) es un elemento de $L^1(\Omega', \mathcal{A}', P^X; \mathbb{R})$, aunque, como se ha dicho, se denotará también de ese modo cualquier representante de esa clase cuando no haya lugar a confusión.

***Observación* 12.4.** Si Y es una v.a. n-dimensional, se define

$$E(Y|X) = (E(Y_1|X), \ldots, E(Y_n|X))$$

si existen las esperanzas condicionales de las componentes de Y.

La definición siguiente recoge la noción de dependencia de una v.a. respecto a los sucesos de una cierta sub-σ-álgebra de $\mathcal{A}$.

DEFINICIÓN 12.2. (Esperanza y probabilidad condicionales respecto a una σ-álgebra) Sean $\mathcal{B}$ una sub-σ-álgebra de $\mathcal{A}$ e Y una v.a.r. en Ω no negativa (resp., con esperanza finita). Llamaremos (una versión de la) esperanza condicional de Y respecto a $\mathcal{B}$ a cualquier función $g : \Omega \longrightarrow \bar{\mathbb{R}}$ $\mathcal{B}$-medible no negativa (resp., con esperanza finita) tal que

$$\int_B Y(\omega)dP(\omega) = \int_B g(\omega)dP(\omega),$$

para cada $B \in \mathcal{B}$. La esperanza condicional de Y respecto a $\mathcal{B}$ se denotará por $E(Y|\mathcal{B})$. Si $A \in \mathcal{A}$, se define la probabilidad condicional de A respecto a $\mathcal{B}$, y se denota $P(A|\mathcal{B})$, mediante

$$P(A|\mathcal{B}) = E(I_A|\mathcal{B}).$$

***Observación* 12.5.** La existencia de $E(Y|\mathcal{B})$ queda asegurada por el teorema anterior si, con las notaciones allí utilizadas, hacemos $\Omega' = \Omega$, $\mathcal{A}' = \mathcal{B}$ y tomamos como X la v.a. identidad en Ω. Podemos hacer también aquí observaciones análogas a las que siguen a la definición de esperanza condicional respecto a una v.a.

***Observación* 12.6.** Notemos explícitamente que, si $A \in \mathcal{A}$, $P(A|\mathcal{B})$ es una clase de funciones reales g (que podemos suponer incluso $[0,1]$-valoradas) $\mathcal{B}$-medibles tales que $P(A \cap B) = \int_B g dP$, para cada $B \in \mathcal{B}$.

***Ejemplo* 12.1.** Consideremos la probabilidad condicional $P(B|I_A)$, $A, B \in \mathcal{A}$. Puesto que I_A es una v.a. en $(\Omega, \mathcal{A})$ que toma valores en $\{0,1\}$, $P(B|I_A)$ es una v.a. a valores en $\Omega' = \{0,1\}$ que verifica

$$P(B \cap I_A^{-1}(i)) = \int_{\{i\}} P(B|I_A)dP^{I_A}, \quad i = 0,1. \tag{12.2}$$

Es claro que $P^{I_A}(\{1\}) = P(A)$ y que $P^{I_A}(\{0\}) = P(A^c)$. Así pues, (12.2) no dice otra cosa que

$$P(B \cap A) = P(B|A)P(A) \qquad \text{y} \qquad P(B \cap A^c) = P(B|A^c)P(A^c).$$

Se recupera así la definición clásica de probabilidad condicional si $0 < P(A) < 1$.

***Ejemplo* 12.2.** Es claro que, si Y es una v.a.r. $\mathcal{B}$-medible, entonces $E(Y|\mathcal{B}) = Y$. Por otra parte, si $f : (\Omega', \mathcal{A}') \longrightarrow \mathbb{R}$ es una v.a.r. no negativa o P^X-integrable, el teorema de la medida imagen (Teorema 5.2) prueba que $E(f \circ X|X) = f$.

***Ejemplo* 12.3.** $E(Y|\{\emptyset, \Omega\}) = E(Y)$, pues una función es $\{\emptyset, \Omega\}$-medible si, y sólo si, es constante, y esa constante debe ser $E(Y)$ según se prueba fácilmente.

El siguiente teorema será útil en diferentes ocasiones. De momento, nos ayudará a profundizar en la relación que existe entre las nociones de esperanza condicional respecto a una v.a. y respecto a la σ-álgebra inducida por dicha v.a.

TEOREMA 12.2. *Sean $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ una v.a., $\mathcal{B}$ la sub-σ-álgebra $X^{-1}(\mathcal{A}')$ inducida por X y h una v.a. $\bar{\mathbb{R}}$-valorada en Ω. Entonces h es $\mathcal{B}$-medible si, y sólo si, existe una v.a. $f : (\Omega', \mathcal{A}') \longrightarrow \bar{\mathbb{R}}$ tal que $h = f \circ X$.*

Demostración. Puesto que la composición de v.a. es una v.a., sólo es necesario probar que, si h es $\mathcal{B}$-medible, entonces existe una v.a. $f : (\Omega', \mathcal{A}') \longrightarrow \bar{\mathbb{R}}$ tal que $h = f \circ X$. Si $h = I_C$, $C \in \mathcal{B}$, existe $A' \in \mathcal{A}'$ tal que $C = X^{-1}(A')$ (por def. de $\mathcal{B}$). Haciendo $f = I_{A'}$ es claro que $h = f \circ X$. Si h es una función simple $\mathcal{B}$-medible, entonces admite una representación de la forma $h = \sum_{k=1}^n y_k I_{C_k}$, siendo los C_k $\mathcal{B}$-medibles; si escribimos $I_{C_k} = f_k \circ X$ como antes y hacemos $f = \sum_{k=1}^n y_k f_k$, se verifica que $h = f \circ X$. En el caso general, existe una sucesión $(h_n)_n$ de funciones simples $\mathcal{B}$-medibles que converge puntualmente (en todo punto $\omega \in \Omega$) a h. Por el caso anterior, podemos escribir $h_n = f_n \circ X$, para cada $n \in \mathbb{N}$. Sea $A' = \{x \in \Omega' : \exists \lim_n f_n(x)\}$ y hagamos $f(x) = \lim_n f_n(x)$ si $x \in A'$ y $f(x) = 0$ si $x \notin A'$. Puesto que las f_n son $\mathcal{A}'$-medibles, se verifica que $A' \in \mathcal{A}'$. Podemos escribir $f(x) = I_{A'}(x) \limsup_n f_n(x)$, lo que prueba que f es $\mathcal{A}'$-medible. Además, para cada $\omega \in \Omega$, $h(\omega) = \lim_n h_n(\omega) = \lim_n f_n(X(\omega)) = f(X(\omega))$. $\square$

Corolario 12.1. *Sean X una v.a. como en el teorema anterior e Y una v.a.r. no negativa o con esperanza finita en Ω. Si $\mathcal{B} = X^{-1}(\mathcal{A}')$, entonces $g \longrightarrow g \circ X$ es una aplicación suprayectiva de la clase $E(Y|X)$ sobre la clase $E(Y|\mathcal{B})$.*

Demostración. Si $g \in E(Y|X)$, entonces $g : (\Omega', \mathcal{A}') \longrightarrow \bar{\mathbb{R}}$ es medible y

$$\int_{X^{-1}(A')} Y dP = \int_{A'} g dP^X, \quad \forall A' \in \mathcal{A}'.$$

Es claro que $g \circ X$ es $\mathcal{B}$-medible. Además, si $B \in \mathcal{B}$, existe $A' \in \mathcal{A}'$ tal que $B = X^{-1}(A')$. El teorema de la medida imagen (Teorema 5.2) prueba que

$$\int_B (g \circ X) dP = \int_{A'} g dP^X = \int_B Y dP,$$

es decir, que $g \circ X \in E(Y|\mathcal{B})$.

Por otra parte, si $h \in E(Y|\mathcal{B})$, entonces es $\mathcal{B}$-medible y $\int_B Y dP = \int_B h dP$, para cada $B \in \mathcal{B}$. Por el teorema anterior existe una v.a. $g : (\Omega', \mathcal{A}') \longrightarrow \bar{\mathbb{R}}$ tal que $h = g \circ X$. El teorema de la medida imagen (Teorema 5.2) prueba, como antes, que $g \in E(Y|X)$. □

***Observación* 12.7.** Este corolario, junto con las observaciones precedentes, nos permiten escribir $E(Y|\mathcal{B}) = E(Y|X) \circ X$ si $\mathcal{B} = X^{-1}(A')$.

Presentamos a continuación algunas de las principales propiedades de la esperanza condicional. Muchas de ellas recuerdan propiedades ya conocidas para la esperanza, aunque la esperanza condicional es una (clase de) v.a., y no un número como es la esperanza.

Recordemos en primer lugar que $E(Y|X)$ tiene esperanza finita respecto a P^X si Y tiene esperanza finita respecto a P; se verifica incluso la siguiente proposición, que se sigue de (12.1) tomando $A' = \Omega'$.

Proposición 12.1. *Sea Y una v.a.r. no negativa o integrable. Entonces*

$$E_P(Y) = E_{P^X}(E(Y|X)).$$

TEOREMA 12.3. *Sean Y_1, Y_2 v.a. en Ω a valores en $\mathbb{R}$ (resp., $\bar{\mathbb{R}}$) con esperanza finita (resp., no negativas) y $a_1, a_2 \in \mathbb{R}$ (resp., $a_1, a_2 \geq 0$).*

(a) *Si $Y_1 \leq Y_2$, entonces*

$$E(Y_1|X) \leq E(Y_2|X), \quad P^X - c.s.$$

En particular, si Y_1 es P-integrable,

$$|E(Y_1|X)| \leq E(|Y_1||X), \quad P^X - c.s.$$

(b) $E(a_1Y_1 + a_2Y_2|X) = a_1E(Y_1|X) + a_2E(Y_2|X), \quad P^X - c.s.$

Demostración. La linealidad de la integral prueba (b) de forma inmediata. En cuanto a (a), nótese que, para cada $A' \in \mathcal{A}'$, se verifica que

$$\int_{A'} E(Y_1|X)dP^X = \int_{X^{-1}(A')} Y_1 dP \leq \int_{X^{-1}(A')} Y_2 dP = \int_{A'} E(Y_2|X)dP^X,$$

lo que prueba que $E(Y_1|X) \leq E(Y_2|X)$, P^X-c.s. Puesto que $Y_1, -Y_1 \leq |Y_1|$ y que $E(-Y_1|X) = -E(Y_1|X)$, se tiene que $|E(Y_1|X)| \leq E(|Y_1||X)$ si Y_1 es P-integrable. $\square$

***Observación* 12.8.** Podemos escribir, en lugar de (a) en el teorema anterior, que

$$E(Y_1|X = x) \leq E(Y_2|X = x), \quad P^X - \text{c.s.},$$

haciendo uso del convenio de notación que hemos introducido, expresión que resulta probablemente más intuitiva que la dada en el enunciado. Quede claro en cualquier caso que es conveniente leer una expresión de ese tipo como sigue: si f_1, f_2 son versiones cualesquiera de $E(Y_1|X), E(Y_2|X)$, resp., entonces $f_1 \leq f_2$, P^X-c.s. Observaciones análogas se pueden hacer para los próximos teoremas.

Existen, para la esperanza condicional, teoremas límite análogos a los de convergencia de Lebesgue.

TEOREMA 12.4. (Teorema de la convergencia monótona para la esperanza condicional) *Sea (Y_n) una sucesión creciente P-c.s. de v.a.r. en Ω no negativas tal que $Y = \text{lím}_n Y_n$, P-c.s. Entonces $(E(Y_n|X))$ es una sucesión de v.a. en Ω, $\bar{\mathbb{R}}$-valoradas, creciente P^X-c.s. Además*

$$E(Y|X) = \lim_n E(Y_n|X), \quad P^X - c.s.$$

Demostración. Si $A' \in \mathcal{A}'$, se sigue del teorema de la convergencia monótona de Lebesgue y de la definición de esperanza condicional que

$$\int_{A'} E(Y|X)dP^X = \int_{X^{-1}(A')} YdP = \lim_n \int_{X^{-1}(A')} Y_n dP = \lim_n \int_{A'} E(Y_n|X)dP^X.$$

Por el teorema anterior, $(E(Y_n|X))_n$ es una sucesión creciente P^X-c.s. de v.a. no negativas. Puesto que toda sucesión creciente en $\bar{\mathbb{R}}$ es convergente (posiblemente a $+\infty$) y que el límite puntual de funciones medibles es medible, existe una función medible $h : (\Omega', \mathcal{A}') \longrightarrow \bar{\mathbb{R}}$ tal que $h(x) = \lim_n E(Y_n|X = x)$, P^X-c.s. Una nueva aplicación del teorema de la convergencia monótona (Teorema 4.1) prueba que

$$\int_{A'} hdP^X = \lim_n \int_{A'} E(Y_n|X)dP^X.$$

Por tanto, $h = E(Y|X)$, P^X-c.s. □

Corolario 12.2. (a) *Si (Y_n) es una sucesión de v.a.r. no negativas en Ω, se verifica que*

$$E\big(\sum_n Y_n|X\big) = \sum_n E(Y_n|X), \quad P^X - c.s.$$

(b) *Si (A_n) es una sucesión de sucesos dos a dos disjuntos, entonces*

$$P(\cup_n A_n|X) = \sum_n P(A_n|X), \quad P^X - c.s.$$

TEOREMA 12.5. (Teorema de la convergencia dominada para la esperanza condicional) *Sea (Y_n) una sucesión de v.a.r. en Ω. Supongamos que existe una v.a.r. Z en Ω con esperanza finita tal que $|Y_n| \leq Z$, P-c.s., para cada n. Supongamos también que Y_n converge puntualmente P-c.s. a una v.a.r. Y. Entonces $E(Y_n|X)$ converge puntualmente P^X-c.s. a $E(Y|X)$.*

Demostración. Sea $Z_n = \sup_{k\geq n} |Y_k - Y|$. (Z_n) es pues una sucesión decreciente de v.a.r. no negativas que converge puntualmente a cero P-c.s. Siendo Z P-integrable, $E(Z|X)$ es P^X-integrable y, por tanto, finita P^X-c.s., con lo cual $E(Y_n|X)$ es finita P^X-c.s., para cada $n \in \mathbb{N}$. Análogamente, $E(Y|X)$ es finita P^X-c.s. Además,

$$|E(Y_n|X) - E(Y|X)| \leq E(|Y_n - Y||X) \leq E(Z_n|X).$$

Basta pues probar que $E(Z_n|X)$ es una sucesión puntualmente convergente a cero P^X-c.s. Puesto que Z_n es decreciente P-c.s., $E(Z_n|X)$ es decreciente P^X-c.s. Sea h una función medible en $(\Omega', \mathcal{A}')$ tal que $h = \lim_n E(Z_n|X)$, P^X-c.s. Siendo $0 \leq Z_n \leq 2Z$ y $2Z$ P- integrable, se sigue del teorema de la convergencia dominada (Teorema 4.5) que

$$0 \leq \int h dP^X \leq \int E(Z_n|X) dP^X = \int Z_n dP \to_{n\to\infty} 0$$

y, entonces, $h = 0$, P^X-c.s. □

Observación **12.9.** Habida cuenta que la esperanza condicional respecto a una sub-σ-álgebra se puede expresar como la esperanza condicional respecto a una cierta v.a., según se ha hecho notar anteriormente, resultados análogos se pueden enunciar para la esperanza condicional respecto a una sub-σ-álgebra.

TEOREMA 12.6. (a) *Sean $\mathcal{B}$ una sub-σ-álgebra de $\mathcal{A}$ e $Y : (\Omega, \mathcal{A}) \longrightarrow \mathbb{R}$ y $Z : (\Omega, \mathcal{B}) \longrightarrow \mathbb{R}$ v.a. Supongamos que Y e YZ tienen esperanza finita. Entonces*

$$E(YZ|\mathcal{B}) = Z \cdot E(Y|\mathcal{B}), \quad P\text{-c.s.}$$

(b) *Si $f : (\Omega', \mathcal{A}') \longrightarrow \mathbb{R}$ es una v.a. e Y e $Y(f \circ X)$ son v.a.r. con esperanza finita, entonces*

$$E(Y(f \circ X)|X) = f \cdot E(Y|X), \quad P^X\text{-c.s.}$$

Demostración. (a) Si $Z = I_B$, $B \in \mathcal{B}$, entonces, para cada $C \in \mathcal{B}$,

$$\int_C YZ\,dP = \int_{C\cap B} Y\,dP = \int_{C\cap B} E(Y|\mathcal{B})\,dP = \int_C I_B E(Y|\mathcal{B})\,dP = \int_C ZE(Y|\mathcal{B})\,dP;$$

siendo $ZE(Y|\mathcal{B})$ $\mathcal{B}$-medible, queda probado que el resultado es cierto para indicadores.

Si Z es simple $\mathcal{B}$-medible, el resultado es cierto por la linealidad de la esperanza condicional.

En el caso general, existe una sucesión (Z_n) de funciones simples $\mathcal{B}$-medibles que converge puntualmente a Z y verifican que $|Z_n| \leq |Z|$. Entonces $|YZ_n| \leq |YZ|$, para cada n, y, por el teorema de la convergencia dominada para la esperanza condicional (Teorema 12.5), $E(YZ|\mathcal{B}) = \lim_n E(YZ_n|\mathcal{B})$, P-c.s.

Por otra parte, siendo Y P-integrable, $E(Y|\mathcal{B})$ es P-integrable y, entonces, finita P-c.s. Se sigue pues que

$$ZE(Y|\mathcal{B}) = \lim_n Z_n E(Y|\mathcal{B}), \quad P\text{-c.s.}$$

Puesto que el resultado es cierto para funciones simples, se tiene que $E(YZ_n|\mathcal{B}) = Z_n E(Y|\mathcal{B})$ P-c.s., para cada n. De ello y de lo anterior se sigue la tesis.

La demostración de (b) es totalmente análoga. □

***Observación* 12.10.** El teorema anterior es una generalización de los resultados ya conocidos $E(Z|\mathcal{B}) = Z$ y $E(f \circ X|X) = f$.

PROBLEMAS DEL CAPÍTULO 12

Problema 12.1. Probar que $E(Y|\mathcal{A}) = Y$, P-c.s. y que $E(Y|\{\emptyset,\Omega\}) = E(Y)$, P-c.s.

Problema 12.2. Sea $A \in \mathcal{A}$ tal que $P(A) > 0$. Determinar la esperanza condicional $E(Y|I_A)$.

Problema 12.3. Sean $A_1, A_2, \ldots \in \mathcal{A}$ dos a dos disjuntos tales que $\Omega = \cup_n A_n$ y $P(A_n) > 0$, $\forall n$. Sea $\mathcal{B}$ la σ-álgebra engendrada por los A_n. Probar que cada elemento de $\mathcal{B}$ es unión de ciertos A_n y que $E(Y|\mathcal{B})$ coincide P-c.s. con la aplicación $\omega \in \Omega \longrightarrow [P(A_n)]^{-1}\int_{A_n} Y dP$ si $\omega \in A_n$, $n = 1, 2, \ldots$

Problema 12.4. Consideremos el espacio de probabilidad $([0,1], \mathcal{B}, P)$ donde $\mathcal{B}$ es la σ-álgebra de Borel en $[0,1]$ y P es la medida de Lebesgue en $[0,1]$. Sean $C = [\frac{1}{3}, \frac{2}{3}]$ y $\mathcal{C}$ la σ-álgebra engendrada por C. Consideremos también la v.a. $Y : x \in [0,1] \longrightarrow Y(x) = x^2 \in \mathbb{R}$. Calcular $E(Y|\mathcal{C})$ y $P(C|\mathcal{C})$.

Problema 12.5. Probar que, si $\mathcal{B}$ y $\mathcal{C}$ son sub-σ-álgebras de $\mathcal{A}$ tales que $\mathcal{C} \subset \mathcal{B}$, entonces

$$E[E(Y|\mathcal{B})|\mathcal{C}] = E(Y|\mathcal{C}), \quad P\text{-c.s.}$$

y

$$E[E(Y|\mathcal{C})|\mathcal{B}] = E(Y|\mathcal{C}), \quad P\text{-c.s.}$$

Problema 12.6. (a) (★) (Imagen de una esperanza condicional) Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, $(\Omega', \mathcal{A}')$ un espacio medible, $T : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ una v.a. y $\mathcal{B}'$ una sub-σ-álgebra de $\mathcal{A}'$. Denotando por P^T la distribución de T respecto a P (como es habitual), probar que, si $X : (\Omega', \mathcal{A}') \longrightarrow \mathbb{R}$ es una v.a.r. no negativa o P^T-integrable, entonces

$$E_{P^T}(X|\mathcal{B}') \circ T = E_P(X \circ T|T^{-1}(\mathcal{B}')), \quad P\text{-c.s.}$$

(b) Sean $T : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ una v.a., $\mathcal{B}'$ una sub-σ-álgebra de $\mathcal{A}'$ y $A' \in \mathcal{A}'$. ¿Qué relación existe entre las probabilidades condicionales

$$P^T(A'|\mathcal{B}') \qquad \text{y} \qquad P(T^{-1}(A')|T^{-1}(\mathcal{B}'))\ ?$$

Problema 12.7. (★) (Varianza condicional) Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad e $Y : (\Omega, \mathcal{A}, P) \longrightarrow \mathbb{R}$ una v.a.r. de cuadrado integrable. Si $\mathcal{B}$ es una sub-σ-álgebra de $\mathcal{A}$, se define la varianza condicional de Y respecto a $\mathcal{B}$ mediante

$$\mathrm{Var}(Y|\mathcal{B}) = E(Y^2|\mathcal{B}) - E(Y|\mathcal{B})^2.$$

Probar que

$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y|\mathcal{B})] + \mathrm{Var}[E(Y|\mathcal{B})].$$

Probar también que $\mathrm{Var}(X|\mathcal{B}) = E[(X - E(X|\mathcal{B}))^2|\mathcal{B}]$.

Problema 12.8. Sean $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1, 2$, dos espacios de probabilidad y

$$f : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2) \longrightarrow \mathbb{R}$$

una v.a.r. no negativa o integrable respecto a la probabilidad producto $P_1 \times P_2$. Denotemos por X la proyección de $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$ sobre $(\Omega_1, \mathcal{A}_1)$. Calcular $E(f|X)$.

Problema 12.9. (★) (Martingalas) Def.: Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, $(\mathcal{A}_n)_n$ una sucesión creciente de sub-σ-álgebras de $\mathcal{A}$ y (X_n) una sucesión de v.a.r. P-integrables. Diremos que $(X_n, \mathcal{A}_n)$ es una martingala si X_n es $\mathcal{A}_n$-medible y $E(X_{n+1}|\mathcal{A}_n) = X_n$, para cada $n \in \mathbb{N}$. Si $\mathcal{A}_n$ es la más pequeña σ-álgebra que hace medibles a $X_1, \dots, X_n$, se dice simplemente que (X_n) es una martingala.

a) Si $(Y_n)_n$ es una sucesión de v.a.r. independientes con media nula y $X_n = \sum_{i=1}^n Y_i$, entonces $(X_n)_n$ es una martingala.

b) Sean $(\mathcal{A}_n)_n$ una sucesión creciente de sub-σ-álgebras de $\mathcal{A}$ e Y una v.a.r. P-integrable en $(\Omega, \mathcal{A})$. Entonces $(E(Y|\mathcal{A}_n), \mathcal{A}_n)_n$ es una martingala.

Problema 12.10. (Esperanza condicional e invarianza) A lo largo de este problema, G será un grupo de biyecciones bimedibles de $(\Omega, \mathcal{A})$ sobre sí mismo.

(a) Def.: Un suceso $A \in \mathcal{A}$ se dice G-invariante si $g(A) = A$, $\forall g \in G$. Probar que la familia $\mathcal{A}_G$ de los sucesos G-invariantes es una sub-σ-álgebra de $\mathcal{A}$.

(b) Def.: Una v.a. $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ se dice G-invariante si $f(g(\omega)) = f(\omega)$, $\forall g \in G$, $\forall \omega \in \Omega$. Supongamos que $\{\omega'\} \in \mathcal{A}'$, para cada $\omega' \in \Omega'$. Probar que una v.a. $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ es G-invariante si, y sólo si, f es $(\mathcal{A}_G, \mathcal{A}')$-medible.

(c) Def.: Una probabilidad P en $(\Omega, \mathcal{A})$ se dice G-invariante si $P^g = P$, $\forall g \in G$. Supongamos que G es un grupo finito de cardinal $|G|$ y que P es una probabilidad invariante en $(\Omega, \mathcal{A})$. Si f es una v.a.r. acotada, probar que

$$E_P(f|\mathcal{A}_G) = |G|^{-1} \sum_{g \in G} f \circ g.$$

(d) Hagamos ahora $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{R})$ y $G = \{I, S\}$ donde I es la identidad en $\mathbb{R}$ y $S(x) = -x$ si $x \in \mathbb{R}$. Los elementos de $\mathcal{A}_G$ se llaman en este caso borelianos simétricos. Probar que $\mathcal{A}_G = Y^{-1}(\mathcal{R})$ donde $Y(x) = x^2$, $x \in \mathbb{R}$. Si P es una probabilidad Borel G-invariante (P se dice, entonces, simétrica) y si f es una v.a.r. acotada definida en $(\mathbb{R}, \mathcal{R})$, calcular $E_P(f|Y)$.

Problema 12.11. (★) (Átomos) Sea $\mathcal{B}$ una sub-σ-álgebra de $\mathcal{A}$. Def.: Un átomo de $\mathcal{B}$ relativo a P es un suceso $A \in \mathcal{B}$ que verifica: i) $P(A) > 0$; ii) si $B \in \mathcal{B}$ y $B \subset A$, entonces $P(B) = 0$ o $P(A \setminus B) = 0$ (ver Problema 2.10).

Probar que, si $X : (\Omega, \mathcal{A}) \longrightarrow \mathbb{R}$ es una v.a.r. y si $A \in \mathcal{A}$ es un átomo de $\mathcal{A}$ relativo a P, entonces X es constante P-c.s. sobre A.

Problema 12.12. ($\uparrow$) Sean $\mathcal{B}$ una sub-σ-álgebra de $\mathcal{A}$ y A un átomo de $\mathcal{B}$ relativo a P. Probar que $E(Y|\mathcal{B})$ es constante P-c.s. sobre A; concretamente,

$$E(Y|\mathcal{B})(\omega) = [P(A)]^{-1} \int_A Y dP$$

para casi todo punto ω de A.

Problema 12.13. ($\uparrow$) Sean $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ una v.a., $x_0 \in \Omega'$ y $B = \{\omega \in \Omega \colon X(\omega) = x_0\}$. Supongamos que $P(B) > 0$. Probar que B es un átomo de $\mathcal{B} := X^{-1}(\mathcal{A}')$ relativo a P. Recordar que $E(Y|X) \circ X = E(Y|\mathcal{B})$ y concluir que

$$E(Y|X = x_0) = [P(B)]^{-1} \int_B Y dP.$$

Problema 12.14. (★) (Independencia condicional) Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $\mathcal{B}, \mathcal{C}, \mathcal{D}$ sub-σ-álgebras de $\mathcal{A}$. Probar que son equivalentes las proposiciones siguientes:

$$\text{(i)} \quad \forall B \in \mathcal{B}, \forall C \in \mathcal{C}, \quad P(B \cap C|\mathcal{D}) = P(B|\mathcal{D})P(C|\mathcal{D}), \quad P - \text{c.s.}$$

$$\text{(ii)} \quad \forall B \in \mathcal{B}, P(B|\mathcal{C} \vee \mathcal{D}) = P(B|\mathcal{D}), \quad P - \text{c.s.},$$

donde $\mathcal{C} \vee \mathcal{D}$ es la más pequeña σ-álgebra que contiene a $\mathcal{C} \cup \mathcal{D}$. Se dice, en ese caso, que $\mathcal{B}$ y $\mathcal{C}$ son condicionalmente independientes respecto a $\mathcal{D}$.

Problema 12.15. Si X es una v.a.r. no negativa, entonces

$$E(X|\mathcal{B}) = \int_0^\infty P(X > t|\mathcal{B}) dt, \quad P\text{-c.s.}$$

Problema 12.16. (Desigualdad de Markov generalizada) Si $k > 0$ entonces

$$P(|X| \geq \alpha|\mathcal{B}) \leq \alpha^{-k} E[|X|^k|\mathcal{B}], \quad P\text{-c.s.}$$

Capítulo 13

Distribución Condicional y Distribución Conjunta de V.A.

Recordemos una vez más que las probabilidades condicionales $P(A|X)$ y $P(A|\mathcal{B})$ son clases de equivalencia de v.a., y que es habitual (aquí lo hemos hecho) denotar así cualquiera de los representantes de esa clase. Hay pues que tener cuidado a la hora de escribir, p.ej., $P(A|\mathcal{B})(\omega)$ o $P(A|X = x)$, pues expresiones como éstas dependen del representante elegido: dado un representante de $P(A|X)$, hemos interpretado $P(A|X = x)$ como la probabilidad condicional de A respecto al suceso $\{\omega \in \Omega \colon X(\omega) = x\}$, tenga este suceso o no probabilidad nula (argumentos ya conocidos prueban que, si ese suceso tiene probabilidad > 0, $P(A|X = x)$ coincide con la definición clásica). Ocurre en el caso clásico que, si $P(B) > 0$, entonces la función de conjuntos $A \in \mathcal{A} \longrightarrow P(A|B)$ es una probabilidad, pero aún no hemos podido probar que, para cada x (o, al menos, para casi todo x), $P(\cdot|X = x)$ sea una probabilidad. En este sentido, el corolario del teorema de la convergencia monótona para la esperanza condicional (Teorema 12.4) prueba que, si (A_n) es una sucesión de sucesos dos a dos disjuntos en $\mathcal{A}$, entonces $P(\cup_n A_n|X) = \sum_n P(A_n|X)$, P^X-c.s., es decir, dada la sucesión (A_n), existe un suceso $N \in \mathcal{A}'$ tal que $P^X(N) = 0$ y $P(\cup_n A_n|X = x) = \sum_n P(A_n|X = x)$ si $x \notin N$; pero el conjunto excepcional N

depende de la sucesión (A_n).

En esta sección comenzamos planteándonos el problema de decidir bajo qué condiciones es posible suponer que $P(A|X = x)$ sea, a x fijo, una probabilidad en $(\Omega, \mathcal{A})$, en cuyo caso obtendríamos la extensión realmente deseada de la definición clásica de probabilidad condicional. Veremos que ello es posible en condiciones bastante generales.

En lo que sigue, utilizaremos las siguientes notaciones: $(\Omega, \mathcal{A}, P)$ será un espacio de probabilidad, $\mathcal{B}$ una sub-σ-álgebra de $\mathcal{A}$, $(\Omega', \mathcal{A}')$ un espacio medible y $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ una v.a.

DEFINICIÓN 13.1. (Función de distribución condicional regular) Sea Y un v.a.r. en $(\Omega, \mathcal{A}, P)$. Llamaremos función de distribución condicional regular de Y respecto a X a una aplicación $F : \Omega' \times \mathbb{R} \longrightarrow \mathbb{R}$ que verifique las dos proposiciones siguientes

(i) Para cada $x \in \Omega'$, $F(x, \cdot)$ es una función de distribución en $\mathbb{R}$ que se anula en $-\infty$ y vale 1 en $+\infty$.

(ii) Para cada $y \in \mathbb{R}$, $F(\cdot, y)$ es una versión de la probabilidad condicional $P(Y \leq y|X)$.

El teorema siguiente prueba que esta definición no está desprovista de contenido.

TEOREMA 13.1. *Si Y es una v.a.r. en $(\Omega, \mathcal{A})$, entonces existe una función de distribución condicional regular de Y respecto a X.*

Demostración. Para cada racional $r \in \mathbb{Q}$, denotemos por F_r una versión de $P(Y \leq r|X)$ tal que $0 \leq F_r(x) \leq 1$, para cada $x \in \Omega'$. Sea $r_1, r_2, \ldots$ una enumeración de $\mathbb{Q}$.

Hagamos $A_{ij} = \{x \in \Omega' \colon F_{r_j}(x) < F_{r_i}(x)\}$, $i, j \in \mathbb{N}$, y $A = \cup_{r_i < r_j} A_{ij}$. Es claro que $A \in \mathcal{A}'$ y que $P^X(A) = 0$, pues si $r_i < r_j$ entonces $P(Y \leq r_i|X) \leq P(Y \leq r_j|X)$, P^X-c.s.

Se definen también $B_i = \{x \in \Omega' \colon \text{lím}_{n\to\infty} F_{r_i + \frac{1}{n}}(x)$ o no existe o es $\neq F_{r_i}(x)\}$, $i \in \mathbb{N}$, y $B = \cup_i B_i$. Entonces $B_i, B \in \mathcal{A}'$. Puesto que $I_{\{Y \leq r_i + \frac{1}{n}\}} \rightarrow_{n\to\infty} I_{\{Y \leq r_i\}}$, el teorema de la convergencia dominada para la esperanza condicional (Teorema 12.5) prueba que $F_{r_i+\frac{1}{n}}(x) \to_{n\to\infty} F_{r_i}(x)$, P^X-c.s. Entonces $P^X(B) = 0$.

Sea $C = \{x \in \Omega' : \text{lím}_{n\to+\infty} F_n(x) \text{ o no existe o es } \neq 1\}$. Entonces $C \in \mathcal{A}'$ y $P^X(C) = 0$, pues $\{Y \leq n\}$ es una sucesión creciente de sucesos que recubren Ω y, por tanto, $P(Y \leq n|X)$ converge a 1 puntualmente P^X-c.s.

Análogamente, si $D = \{x \in \Omega' : \text{lím}_{n\to-\infty} F_n(x) \text{ o no existe o es } \neq 0\}$ entonces $D \in \mathcal{A}'$ y $P^X(D) = 0$.

Sea G una función de distribución en $\mathbb{R}$ que vale 0 en $-\infty$ y 1 en $+\infty$. Se define, para $x \in \Omega'$ e $y \in \mathbb{R}$,

$$F(x,y) = \begin{cases} = \text{lím}_{r\to y+, r\in\mathbb{Q}} F_r(x) & \text{si } x \notin A \cup B \cup C \cup D \\ = G(y) & \text{si } x \in A \cup B \cup C \cup D. \end{cases}$$

F está bien definida pues, si $x \notin A$, $F_r(x)$ es creciente en r, y, por tanto,

$$\lim_{r\to y+, r\in\mathbb{Q}} F_r(x) = \inf_{r>y, r\in\mathbb{Q}} F_r(x). \tag{13.1}$$

Además, si $x \notin A \cup B$ e $y \in \mathbb{Q}$, entonces $\text{lím}_{r\to y+} F_r(x) = F_y(x)$. Así pues, $F(x,y) = F_y(x)$ si $y \in \mathbb{Q}$ y $x \notin A \cup B \cup C \cup D$. Análogamente, si $x \notin A \cup C \cup D$, entonces $\text{lím}_{r\to+\infty} F_r(x) = 1$ y $\text{lím}_{r\to-\infty} F_r(x) = 0$.

Veamos que F es una función de distribución condicional regular de Y respecto a X.

Si $x \notin A \cup B \cup C \cup D$, entonces $F(x,\cdot)$ es creciente pues, si $y, y' \in \mathbb{R}$, $y < y'$, entonces $\{r \in \mathbb{Q} : r > y'\} \subset \{r \in \mathbb{Q} : r > y\}$ y basta aplicar (13.1). Por otra parte, si $y < y' \leq r \in \mathbb{Q}$, entonces

$$F(x,y) \leq F(x,y') \leq F(x,r) = F_r(x) \to_{r\to y+} F(x,y),$$

lo que prueba que $F(x,\cdot)$ es continua por la derecha en cada punto $y \in \mathbb{R}$. Si $r \leq y$, $r \in \mathbb{Q}$, entonces $F(x,y) \geq F(x,r) = F_r(x) \to_{r\to+\infty} 1$ y, por tanto, $F(x,y) \to_{y\to+\infty} 1$. Análogamente, $F(x,y) \to_{y\to-\infty} 0$. Queda probado que $F(x,\cdot)$ es una función de distribución en $\mathbb{R}$ si $x \notin A \cup B \cup C \cup D$. Eso es también evidentemente cierto si $x \in A \cup B \cup C \cup D$ pues, en ese caso, $F(x, \cdot) = G$.

Además, el teorema de la convergencia dominada para la esperanza condicional (Teorema 12.5) prueba que F_r, como versión de $P(Y \leq r|X)$, converge puntualmente P^X-c.s. a $P(Y \leq y|X)$ si $r \to y+$; por continuidad a la derecha, se verifica también que $F_r(x) = F(x, r)$ converge a $F(x, y)$ si $r \to y+$. Así pues, para cada $y \in \mathbb{R}$, $F(\cdot, y)$ es una versión de $P(Y \leq y|X)$. $\square$

***Observación* 13.1.** Análogamente, si $\mathcal{B}$ es una sub-σ-álgebra de $\mathcal{A}$ y si Y es una v.a.r. en $(\Omega, \mathcal{A}, P)$, se llama función de distribución condicional regular de Y respecto a $\mathcal{B}$ a una aplicación $F : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ tal que $F(\omega, \cdot)$ es una función de distribución en $\mathbb{R}$, para cada $\omega \in \Omega$, y $F(\cdot, y)$ es una versión de $P(Y \leq y|\mathcal{B})$, para cada $y \in \mathbb{R}$. La existencia de una función de distribución condicional regular de Y respecto a $\mathcal{B}$ se deduce del teorema anterior haciendo $X = id_\Omega\colon (\Omega, \mathcal{A}) \longrightarrow (\Omega, \mathcal{B})$.

DEFINICIÓN 13.2. (Probabilidad condicional regular respecto a una v.a.) Sean $Y : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega_1, \mathcal{A}_1)$ y $X : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega', \mathcal{A}')$ v.a. Llamaremos probabilidad condicional regular de Y respecto a X a una aplicación $Q : \Omega' \times \mathcal{A}_1 \longrightarrow [0, 1]$ que satisfaga

(i) $\forall x \in \Omega'$, $Q(x, \cdot)$ es una probabilidad en $(\Omega_1, \mathcal{A}_1)$.

(ii) $\forall A_1 \in \mathcal{A}_1$, $Q(\cdot, A_1)$ es una versión de $P(Y^{-1}(A_1)|X)$.

***Observación* 13.2.** Una probabilidad condicional regular Q de Y respecto a X es, en particular, una probabilidad de transición en $\Omega' \times \mathcal{A}_1$.

TEOREMA 13.2. *Si Y es una v.a.r. en $(\Omega, \mathcal{A})$, entonces existe una probabilidad condicional regular de Y respecto a X.*

Demostración. Sea F una función de distribución condicional regular de Y respecto a X y, para cada $x \in \Omega'$, denotemos por $Q(x, \cdot)$ la medida (probabilidad) de Lebesgue-Stieljes asociada a $F(x, \cdot)$. Denotemos $\mathcal{C} = \{B \in \mathcal{R}\colon Q(\cdot, B) \in P(Y \in B|X)\}$. $\mathcal{C}$ contiene al π-sistema formado por los intervalos de la forma $]-\infty, y]$, $y \in \mathbb{R}$. Si $A, B \in \mathcal{C}$ y $A \subset B$, entonces $B \setminus A \in \mathcal{C}$, pues $Q(x, B \setminus A) = Q(x, B) - Q(x, A) =$

$P(Y \in B|X)(x) - P(Y \in A|X)(x) = P(Y \in B \setminus A|X)(x)$, P^X-c.s. Además $\mathcal{C}$ es estable frente a uniones numerables crecientes, por el teorema de la convergencia monótona para la esperanza condicional (Teorema 12.4). Puesto que $\mathbb{R} \in \mathcal{C}$, queda probado que $\mathcal{C}$ es un d-sistema. Por el teorema de Dynkin (Teorema 1.1), $\mathcal{C}$ contiene a la σ-álgebra engendrada por los intervalos de la forma $]-\infty, y]$, $y \in \mathbb{R}$, es decir, $\mathcal{C} = \mathcal{R}$. □

***Observación* 13.3.** Si Y es una v.a. a valores en $(\Omega_1, \mathcal{A}_1)$, una probabilidad condicional regular de Y respecto a una sub-σ-álgebra $\mathcal{B}$ es una aplicación $Q : \Omega \times \mathcal{A}_1 \longrightarrow [0,1]$ tal que $Q(\omega, \cdot)$ es una probabilidad en $\mathcal{A}_1$, para cada $\omega \in \Omega$, y $Q(\cdot, A_1)$ es una versión de $P(Y \in A_1|\mathcal{B})$, para cada $A_1 \in \mathcal{A}_1$. Como en el teorema anterior se prueba que, si Y es una v.a.r. en $(\Omega, \mathcal{A}, P)$, entonces existe una probabilidad condicional regular de Y respecto a $\mathcal{B}$.

El teorema anterior puede extenderse a otros tipos de v.a. Y (por ejemplo, si Y es discreta o n-dimensional). Necesitaremos la siguiente definición:

DEFINICIÓN 13.3. (Espacios de Borel) Un espacio medible $(\Omega_1, \mathcal{A}_1)$ se llamará un espacio de Borel si existen un boreliano E de $\mathbb{R}$ y una biyección $h : \Omega_1 \longrightarrow E$ bimedible cuando E se supone provisto de su σ-álgebra de Borel.

TEOREMA 13.3. *Sean $Y : (\Omega, \mathcal{A}) \longrightarrow (\Omega_1, \mathcal{A}_1)$ y $X : (\Omega, \mathcal{A}, P) \longrightarrow (\Omega', \mathcal{A}')$ v.a. Si $(\Omega_1, \mathcal{A}_1)$ es un espacio de Borel, entonces existe una probabilidad condicional regular de Y respecto a X.*

Demostración. Sean E un boreliano de $\mathbb{R}$, $h : (\Omega_1, \mathcal{A}_1) \longrightarrow E$ una biyección bimedible como en la definición anterior y $Q_0 : \Omega' \times \mathcal{R} \longrightarrow [0,1]$ una probabilidad condicional regular de $h \circ Y$ respecto a X. Si denotamos por $\mathcal{R}(E)$ la σ-álgebra de Borel de E, se define una probabilidad condicional regular Q_0' en $\Omega' \times \mathcal{R}(E)$ por $Q_0'(x, B) = Q_0(x, B)/Q_0(x, E)$ si $Q_0(x, E) > 0$, $= Q'(B)$ si $Q_0(x, E) = 0$, donde Q' es una probabilidad fija en $\mathcal{R}(E)$. Es fácil ver que, para cada $x \in \Omega'$, $Q_0'(x, \cdot)$ es una probabilidad en $\mathcal{R}(E)$. Además, puesto que $P(X \in A') = P(\{X \in A'\} \cap \{h \circ Y \in E\}) = \int_{A'} Q_0(x, E) P^X(dx)$, se sigue que $Q_0(x, E) = 1$, P^X-c.s., y entonces, para cada

$B \in \mathcal{R}(E)$, $Q_0'(x,B) = Q_0(x,B)$, P^X-c.s., lo que prueba que $Q_0'(\cdot, B)$ es una versión de $P(h \circ Y \in B|X)$. Si $A_1 \in \mathcal{A}_1$ y $x \in \Omega'$, se define

$$Q(x, A_1) = Q_0'(x, h(A_1))$$

($h(A_1) \in \mathcal{R}(E)$ pues h^{-1} es medible). Es claro que $Q(x,\cdot)$ es una probabilidad en $\mathcal{A}_1$ (es, concretamente, la distribución de h^{-1} respecto a $Q_0'(x,\cdot)$). Además, si $A_1 \in \mathcal{A}_1$, entonces $Q(x, A_1) = P(h \circ Y \in h(A_1)|X = x) = P(Y \in A_1|X = x)$, P^X-c.s. □

Observación **13.4.** Si $(\Omega_1, \mathcal{A}_1)$ es un espacio discreto (es decir, si Ω_1 es numerable y $\mathcal{A}_1$ es el conjunto de las partes de Ω_1), entonces es trivialmente un espacio de Borel. También $\mathbb{R}^n$ es un espacio de Borel, para cada $n \in \mathbb{N}$.

La noción de probabilidad condicional regular que hemos introducido es clave en el estudio de la dependencia entre dos v.a., en general, y de lo que conoceremos como distribuciones condicionales de una v.a. respecto a otra, en particular.

Las notaciones serán las habituales, es decir, $(\Omega, \mathcal{A}, P)$ será un espacio de probabilidad, $(\Omega', \mathcal{A}')$ y $(\Omega_1, \mathcal{A}_1)$ espacios medibles y $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ e $Y : (\Omega, \mathcal{A}) \longrightarrow (\Omega_1, \mathcal{A}_1)$ v.a.

DEFINICIÓN 13.4. (Distribución condicional) Con las notaciones precedentes, y supuesto que existe una probabilidad condicional regular Q de Y respecto a X, llamaremos distribución condicional de Y respecto a $X = x$ la probabilidad $P^{Y|X=x}$ en $\mathcal{A}_1$ definida por $P^{Y|X=x}(A_1) := Q(x, A_1)$. En las condiciones de esta definición, diremos también que existe la distribución condicional de Y respecto a $X = x$ para todo x.

Observación **13.5.** Las distribuciones condicionales $P^{Y|X=x}$ de Y respecto a X existen para una gran clase de v.a. Y (entre ellas las que normalmente aparecen en estadística); basta para ello, según sabemos, que $(\Omega_1, \mathcal{A}_1)$ sea un espacio de Borel. En particular, si $(\Omega, \mathcal{A})$ es un espacio de Borel, entonces existe una probabilidad condicional regular de la identidad $id_\Omega : (\Omega, \mathcal{A}) \longrightarrow (\Omega, \mathcal{A})$ respecto a X. También, si $(\Omega, \mathcal{A})$ es un espacio de Borel e $Y : (\Omega, \mathcal{A}) \longrightarrow (\Omega_1, \mathcal{A}_1)$ es una v.a., entonces existe una probabilidad condicional regular de Y respecto a X; basta considerar, para cada $x \in \Omega'$, la distribución de Y respecto a $Q(x, \cdot)$, donde Q es una probabilidad condicional regular de id_Ω respecto a X.

Observación **13.6.** Si existen las distribuciones condicionales de Y respecto a $X = x$, $x \in \Omega'$, entonces, a x fijo, $P^{Y|X=x}(\cdot)$ es una probabilidad en $\mathcal{A}_1$ y, a A_1 fijo, $P^{Y|X=\cdot}(A_1)$ es una v.a. $[0,1]$-valorada en $(\Omega', \mathcal{A}')$. Así pues, $P^{Y|X=x}(A_1)$ es, en particular, una probabilidad de transición en $\Omega' \times \mathcal{A}_1$. Además $P^{Y|X=\cdot}(A_1)$ es una versión de la probabilidad condicional del suceso $Y^{-1}(A_1)$ respecto a X; por tanto, para cada $A' \in \mathcal{A}'$, se verifica que

$$P(\{X \in A'\} \cap \{Y \in A_1\}) = \int_{A'} P^{Y|X=x}(A_1) dP^X(x).$$

El teorema generalizado de la medida producto (Teorema 7.1) nos ayuda a precisar más la observación 2 anterior.

TEOREMA 13.4. *Supongamos que existen las distribuciones condicionales $P^{Y|X=x}$, $x \in \Omega'$. Entonces, para cada $C \in \mathcal{A}' \times \mathcal{A}_1$, la aplicación*

$$x \in (\Omega', \mathcal{A}') \longrightarrow \int I_C(x,y)\, dP^{Y|X=x}(y)$$

es una v.a.r. no negativa. Además, si $P^{(X,Y)}$ denota la distribución conjunta de X e Y, entonces

$$P^{(X,Y)}(C) = \int \left(\int I_C(x,y)\, dP^{Y|X=x}(y) \right) dP^X(x).$$

En general, si $f : \Omega' \times \Omega_1 \longrightarrow \mathbb{R}$ tiene esperanza finita respecto a $P^{(X,Y)}$, entonces la aplicación

$$x \in (\Omega', \mathcal{A}') \longrightarrow \int f(x,y)\, dP^{Y|X=x}(y)$$

está bien definida P^X-c.s., y (si se define como cero en el conjunto excepcional de puntos x donde la integral no exista) es una v.a.r. con esperanza finita respecto a P^X y se verifica que

$$\int f(x,y)\, dP^{(X,Y)}(x,y) = \int \left(\int f(x,y)\, dP^{Y|X=x}(y) \right) dP^X(x).$$

Demostración. Si C es un rectángulo medible en $\mathcal{A}' \times \mathcal{A}_1$, digamos $C = A' \times A_1$, entonces

$$P^{(X,Y)}(C) = P(X \in A', Y \in A_1) = \int_{A'} P^{Y|X=x}(A_1)\, P^X(dx).$$

Por otra parte, P^X es una probabilidad en $(\Omega', \mathcal{A}')$ y $P^{Y|X=x}(A_1)$ es una probabilidad de transición en $\Omega' \times \mathcal{A}_1$. El teorema generalizado de la medida producto (Teorema 7.1) prueba la existencia de una única medida μ en $\mathcal{A}' \times \mathcal{A}_1$ tal que $\mu(A' \times A_1) = \int_{A'} P^{Y|X=x}(A_1)\, P^X(dx)$, para cada $A' \in \mathcal{A}'$ y cada $A_1 \in \mathcal{A}_1$. Entonces $\mu = P^{(X,Y)}$. El teorema generalizado de Fubini (Teorema 7.2) prueba la segunda mitad del enunciado.

□

Observación **13.7.** El teorema anterior determina la distribución conjunta de las v.a. X e Y en términos de la distribución de X y las distribuciones condicionales de Y respecto a X. La última fórmula expresa la esperanza de f respecto a $P^{(X,Y)}$ en función de la esperanza de $f(x,\cdot)$ respecto a $P^{Y|X=x}$ y de la esperanza de la integral interior respecto a P^X. Sea $A' \in \mathcal{A}'$ y hagamos $C = A' \times \Omega_1$; cambiemos f por $f \cdot I_C$ en el teorema anterior. Puesto que $I_C(x,y) = I_{A'}(x)$, se sigue que

$$\begin{aligned}\int_{\{X\in A'\}} f(X(\omega), Y(\omega))\, dP(\omega) &= \int f \cdot I_C\, dP^{(X,Y)} \\ &= \int_{A'} \left(\int f(x,y)\, dP^{Y|X=x}(y) \right) dP^X(x),\end{aligned}$$

lo que prueba que la aplicación $x \in \Omega' \longrightarrow \int f(x,y) dP^{Y|X=x}(y)$ es una versión de la esperanza condicional $E(Z|X)$ donde $Z = f(X,Y)$. Se verifica entonces que

$$E(Z|X=x) = \int f(x,y) dP^{Y|X=x}(y), \quad P^X\text{-c.s.}$$

En particular, si ϕ es una v.a.r. en $(\Omega_1, \mathcal{A}_1)$ y $f(x,y) = \phi(y)$, se verifica que, si $\phi \circ Y$ es P-integrable,

$$E(\phi \circ Y|X=x) = \int \phi(y) dP^{Y|X=x}(y), \quad P^X\text{-c.s.} \tag{13.2}$$

Para $\phi(y) = y$ se obtiene que $E(Y|X=x) = \int y dP^{Y|X=x}(y)$, P^X-c.s., si Y es P-integrable. La existencia P^X-c.s. de esas integrales quedan garantizadas por el teorema anterior. Necesitaremos también un resultado análogo a éste para el caso de que Y sea una v.a. n-dimensional a la hora de probar la desigualdad de Jensen para la esperanza condicional (Teorema 14.2); se trata, en concreto, de probar que, si Y es una v.a. n-dimensional P-integrable (es decir, sus componentes Y_i son P-integrables), entonces $E(Y_i|X=x) = \int y_i dP^{Y|X=x}(y)$, P^X-c.s., $1 \le i \le n$; podríamos escribir entonces $E(Y|X=x) = \int y dP^{Y|X=x}(y)$, P^X-c.s. Que eso es así, se sigue de (13.2) sin más que tomar, para $1 \le i \le n$, $\phi(y) = y_i$.

Corolario 13.1. (Caracterización de independencia) *Sean* $X : \Omega \longrightarrow \Omega'$ *e* $Y : \Omega \longrightarrow \Omega_1$ *v.a. Entonces* X *e* Y *son independientes si, y sólo si, para cada* $A_1 \in \mathcal{A}_1$, $P(Y \in A_1|X = x)$ *es una constante dependiente únicamente de* A_1 *(no de* x*). Se verifica en ese caso que*

$$P(Y \in A_1|X = x) = P(Y \in A_1), \quad P^X\text{-c.s.}$$

Además, en ese caso, si Y *es real y con esperanza finita, entonces* $E(Y|X = x) = E(Y)$, P^X*-c.s.*

Demostración. Si X e Y son independientes, entonces, para cada $A' \in \mathcal{A}'$ y cada $A_1 \in \mathcal{A}_1$,

$$P(\{X \in A'\} \cap \{Y \in A_1\}) = P(X \in A') \cdot P(Y \in A_1).$$

Pero $P(\{X \in A'\} \cap \{Y \in A_1\}) = \int_{X^{-1}(A')} I_{\{Y \in A_1\}} dP$. Así pues, si X e Y son independientes,

$$\int_{X^{-1}(A')} I_{\{Y \in A_1\}} dP = \int_{A'} P(Y \in A_1) dP^X,$$

lo que prueba que $P(Y \in A_1|X) = P(Y \in A_1)$, P^X-c.s.

Recíprocamente, si $P(Y \in A_1|X) = k(A_1)$, P^X-c.s., entonces

$$\int_{X^{-1}(A')} I_{\{Y \in A_1\}} dP = \int_{A'} k(A_1) dP^X = k(A_1) \cdot P(X \in A'),$$

para cada $A' \in \mathcal{A}'$. En particular, si $A' = \Omega'$, $k(A_1) = P(Y \in A_1)$.

Veamos en fin que, si X e Y son independientes e Y es una v.a.r. P-integrable, entonces $E(Y|X) = E(Y)$, P^X-c.s. Puesto que $P(Y \in A_1|X) = P(Y \in A_1)$, P^X-c.s., para cada $A_1 \in \mathcal{A}_1$, se sigue que $Q(x, A_1) := P^Y(A_1)$ es una probabilidad condicional regular de Y respecto a X (cuya existencia queda garantizada al margen de que $(\Omega_1, \mathcal{A}_1)$ sea o no un espacio de Borel). Entonces $P^{Y|X=x} = P^Y$ y, por la observación anterior, $\int_{\mathbb{R}} y dP^{Y|X=x}(y)$ es una versión de $E(Y|X)$. De ambas cosas se sigue que la función constante $E(Y)$ es una versión de $E(Y|X)$. $\square$

Acabamos de ver que, si X e Y son independientes, entonces existe una versión regular de la probabilidad condicional de Y respecto a X, al margen de que $(\Omega_1, \mathcal{A}_1)$

sea o no un espacio de Borel (en ese caso, las distribuciones condicionales de Y respecto a $X = x$ no dependen de x y coinciden con la distribución de Y). En el teorema siguiente, que nos permite calcular las densidades condicionales a partir de la densidad conjunta, tampoco es necesario suponer que $(\Omega_1, \mathcal{A}_1)$ sea un espacio de Borel.

TEOREMA 13.5. *Sean $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ e $Y : (\Omega, \mathcal{A}) \longrightarrow (\Omega_1, \mathcal{A}_1)$ v.a. y μ' y μ_1 medidas σ-finitas en $(\Omega', \mathcal{A}')$ y $(\Omega_1, \mathcal{A}_1)$, resp. Supongamos que la v.a. (X, Y) admite una densidad $f(x, y)$ respecto a la medida producto $\mu' \times \mu_1$. Entonces existen las distribuciones condicionales de Y respecto a $X = x$, para cada $x \in \Omega'$; además, la distribución condicional $P^{Y|X=x}$ admite una densidad $f^{Y|X=x}$ respecto a μ_1 definida por*

$$f^{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}, \quad P^X\text{-c.s.},$$

donde $f_X(x) = \int_{\Omega_1} f(x, y) d\mu_1(y)$ es la densidad marginal de X respecto a μ'.

Demostración. Si $A' \in \mathcal{A}'$, entonces

$$P^X(A') = P(X \in A', Y \in \Omega_1) = P^{(X,Y)}(A' \times \Omega_1) =$$
$$\int_{A' \times \Omega_1} f(x, y) d(\mu' \times \mu_1)(x, y) = \int_{A'} \left(\int_{\Omega_1} f(x, y) d\mu_1(y) \right) d\mu'(x),$$

lo que prueba que la aplicación $f_X : x \in \Omega' \longrightarrow f_X(x) = \int_{\Omega_1} f(x, y) d\mu_1(y)$ (no negativa y $\mathcal{A}'$-medible de acuerdo con el teorema de Fubini (Teorema 7.2)) es una densidad de P^X respecto a μ'. Denotemos $B' = \{x \in \Omega' \colon f_X(x) \neq 0\}$ y sea $h : (\Omega_1, \mathcal{A}_1) \longrightarrow [0, +\infty[$ una v.a con integral 1 respecto a μ_1. Para $x \in \Omega'$ e $y \in \Omega_1$, se define

$$g_x(y) = \begin{cases} = f(x, y)/f_X(x) & \text{si } x \in B' \\ = h(y) & \text{si } x \notin B'. \end{cases}$$

Para cada $x \in \Omega'$, g_x es una v.a.r. no negativa en $(\Omega_1, \mathcal{A}_1)$ con integral 1 respecto a μ_1. Si $A_1 \in \mathcal{A}_1$ y $x \in \Omega'$, se define $Q(x, A_1) = \int_{A_1} g_x(y) d\mu_1(y)$. Veamos que Q es una probabilidad condicional regular de Y respecto a X. A x fijo, $Q(x, \cdot)$ es evidentemente una probabilidad en $(\Omega_1, \mathcal{A}_1)$. Fijo $A_1 \in \mathcal{A}_1$, $Q(\cdot, A_1)$ es una función real medible en $(\Omega', \mathcal{A}')$, pues es constante en $B'^c \in \mathcal{A}'$ y cociente de dos funciones medibles (Fubini, de nuevo) con denominador no nulo en B'. Basta ver que $Q(\cdot, A_1)$ es una versión de $P(Y \in A_1|X)$ para cada $A_1 \in \mathcal{A}_1$; si $A' \in \mathcal{A}'$, entonces

$$P(X \in A', Y \in A_1) = P^{(X,Y)}(A' \times A_1) = \int_{A' \times A_1} f(x,y)d(\mu' \times \mu_1)(x,y)$$

y

$$\int_{A'} Q(x, A_1)dP^X(x) = \int_{A'} \left(\int_{A_1} g_x(y)d\mu_1(y) \right) f_X(x)d\mu'(x) =$$
$$\int_{A' \cap B'} \left(\int_{A_1} \frac{f(x,y)}{f_X(x)} d\mu_1(y) \right) f_X(x)d\mu'(x) = \int_{A' \cap B'} \left(\int_{A_1} f(x,y)d\mu_1(y) \right) d\mu'(x)$$
$$= \int_{A'} \left(\int_{A_1} f(x,y)d\mu_1(y) \right) d\mu'(x).$$

Podemos, entonces, escribir $P^{Y|X=x} = Q(x, \cdot)$ y $f^{Y|X=x} = g_x$, con lo cual

$$f^{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)}, \quad P^X\text{-c.s.},$$

pues $P^X(B') = 1$. □

***Observación* 13.8.** Se deduce de la demostración del teorema que, en el caso de que la distribución conjunta de X e Y admita una densidad respecto al producto de dos medidas σ-finitas, existen las distribuciones condicionales $P^{Y|X=x}$ de Y respecto a X sin necesidad de hipótesis de regularidad alguna sobre los espacios medibles implicados.

***Observación* 13.9.** De acuerdo con el teorema, y en las hipótesis del mismo, la densidad condicional de Y respecto a $X = x$ se obtiene dividiendo la densidad conjunta de (X, Y) por la densidad marginal de X. Si ambas variables son discretas, Ω' y Ω_1 son a lo sumo numerables y μ' y μ_1 son las respectivas medidas cardinales; en ese caso, $\mu' \times \mu_1$ es la medida cardinal en el producto y el teorema anterior se aplica. También queda dentro del teorema anterior el caso en que X e Y son v.a. a valores en $\mathbb{R}^m$ y $\mathbb{R}^n$, resp., y la distribución conjunta de ambas admite una densidad respecto a la medida λ_{m+n} de Lebesgue en $\mathbb{R}^{m+n}$; téngase en cuenta que $\lambda_{m+n} = \lambda_m \times \lambda_n$.

***Observación* 13.10.** Consecuencia de los resultados precedentes es la llamada fórmula de Bayes generalizada que, con las notaciones precedentes, podemos escribir en la forma

$$f^{X|Y=y}(x) = \frac{f_X(x) f^{Y|X=x}(y)}{\int_{\Omega'} f_X(t) f^{Y|X=t}(y)d\mu'(t)}, \quad P^{(X,Y)} - c.s.$$

PROBLEMAS DEL CAPÍTULO 13

Problema 13.1. Sean X e Y v.a.r. en $(\Omega, \mathcal{A}, P)$ y supongamos que la densidad conjunta de X e Y (respecto a la medida de Lebesgue en $\mathbb{R}^2$) es

$$f(x,y) = \begin{cases} = 2 \text{ si } 0 < x < y < 1 \\ = 0 \text{ en otro caso.} \end{cases}$$

Determinar las densidades marginales f_X y f_Y de X e Y, resp. Determinar la distribución condicional de Y respecto a $X = x$ y la de X respecto a $Y = y$. Determinar $E(Y|X)$ y $E(X|Y)$.

Problema 13.2. Sea $X : (\Omega, \mathcal{A}, P) \longrightarrow \mathbb{R}$ una v.a.r. con distribución uniforme en $]0,1[$ (es decir, $dP^X(x) = I_{]0,1[}(x)dx$). Sea Y una v.a.r. tal que la distribución condicional de Y respecto a $X = x$ ($x \in]0,1[$) es binomial de parámetros n y x, es decir, $P^{Y|X=x}$ es una distribución discreta en $\{0, 1, \dots, n\}$ que admite una densidad respecto a la medida cardinal dada por

$$f^{Y|X=x}(i) = \binom{n}{i} x^i (1-x)^{n-i}, \quad i = 0, 1, \dots, n.$$

Determinar $E(Y)$ y la distribución P^Y de Y.

Problema 13.3. Sean X e Y dos variables aleatorias reales en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ con distribución conjunta uniforme en el rombo de extremos $(1,0)$, $(0,1)$, $(-1,0)$ y $(0,-1)$. Determinar la función de densidad conjunta de X e Y y las densidades marginales de cada una de ellas. ¿Son independientes las variables X e Y?.

Problema 13.4. Consideremos el espacio de probabilidad $(\mathbb{R}^2, \mathcal{R}^2, P)$, donde $P = N(0,1) \times N(2,1)$, y denotemos por X e Y las aplicaciones coordenada en $\mathbb{R}^2$. Calcular $E(X^2 + Y|X)$.

Problema 13.5. Consideremos tres lanzamientos independientes de una moneda y se definen las siguientes variables aleatorias: X = número de caras en los tres lanzamientos, Y = número de cruces antes de la primera cara. Se pide:

(a) Determinar el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ correspondiente a este experimento y la distribución $P^{(X,Y)}$, P^X y P^Y.

(b) Calcular la $E[Y|X]$. Sea $A = \{\text{Salen al menos dos caras}\}$, determinar $P(A|Y)$.

Problema 13.6. Se dispone de una moneda con canto gordo de la que sabemos que se obtiene cruz (=0) con probabilidad $0,3$, cara (=1) con probabilidad $0,5$ y cae de canto (=2) con probabilidad $0,2$.

(a) Describir el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ correspondiente a tres lanzamientos independientes de esa moneda.

(b) Sea X la v.a. "suma de los resultados obtenidos en los tres lanzamientos". Determinar su distribución P^X, su media $E(X)$ y su varianza $Var(X)$.

(c) ¿Son independientes los sucesos $A =$"la suma de los dos primeros resultados es mayor o igual que 1" y $B =$"el resultado del tercer lanzamiento es igual a 2"?

(d) Sea Y la v.a. "suma de los dos primeros lanzamientos". Calcular las distribuciones condicionales $P^{Y|X=x}$.

(e) Calcular también $E(Y|X)$.

(f) ¿Cuál es la distribución del números de caras obtenidas en 20 lanzamientos independientes de la moneda?

Problema 13.7. (★) (Caracterización de rectángulos medibles: imagen geométrica de la independencia) (a) Sea (X, Y) una v.a. bidimensional con distribución uniforme sobre un boreliano de medida finita A de $\mathbb{R}^2$ tal que $\lambda_2(A) > 0$, donde λ_2 denota la medida de Lebesgue en $\mathbb{R}^2$. Probar que la distribución condicional $P^{Y|X=x}$ coincide P^X-c.s. con la distribución uniforme sobre $A_x = \{y \in \mathbb{R} \colon (x, y) \in A\}$. (b) En las mismas hipótesis que en el apartado anterior, y con las notaciones allí utilizadas, decidir bajo qué hipótesis sobre A se verifica que X e Y son independientes. Probar de hecho que X e Y son independientes si, y sólo si, A es, salvo conjuntos de medida nula, un producto de dos borelianos de $\mathbb{R}$.

Problema 13.8. Sean X e Y v.a. independientes sobre un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ a valores en $(\Omega', \mathcal{A}')$ y $(\Omega_1, \mathcal{A}_1)$ y $f : (\Omega' \times \Omega_1, \mathcal{A}' \times \mathcal{A}_1) \longrightarrow \mathbb{R}$ una función medible tal que $f(X, Y)$ es de cuadrado integrable. Probar que

$$E[f(X, Y)|X = x] = E[f(x, Y)], \quad P^X - c.s.$$

Indicación: Considerar, en primer lugar, el caso en que f es el indicador de un rectángulo medible.

Capítulo 14

Desigualdad de Jensen. Problema General de Regresión

En esta sección nos proponemos probar la desigualdad de Jensen en el caso n-dimensional para la esperanza ordinaria y para la esperanza condicional. La aplicaremos en esta sección para resolver el problema general de regresión. La demostración se apoya en una serie de lemas previos.

Recordemos que un subconjunto C de un espacio vectorial se dice convexo si, para cada $\lambda \in]0,1[$ y cada par de puntos x, y de C, se verifica que $\lambda x + (1-\lambda)y \in C$, y que una función real $\varphi : C \longrightarrow \mathbb{R}$ se dice convexa si $\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y)$, cualesquiera que sean $\lambda \in]0,1[$ y $x, y \in C$.

Lema 14.1. *Sean C un convexo de $\mathbb{R}^n$ y $\varphi : C \longrightarrow \mathbb{R}$ una función convexa. Sean $p_1, \ldots, p_k$ números reales no negativos tales que $\sum_{i=1}^k p_i = 1$ y $x^1, \ldots, x^k$ puntos de C. Entonces*

$$\varphi(\textstyle\sum_{i=1}^k p_i x^i) \leq \sum_{i=1}^k p_i \varphi(x^i).$$

Demostración. Se procederá por inducción en k. Para $k = 2$ es cierto por definición. Supuesto cierto para $k - 1 \geq 2$, consideremos $p_1, \ldots, p_k \geq 0$ tales que $\sum_i p_i = 1$ y puntos $x^1, \ldots, x^k \in C$. Si algún p_i es nulo, la hipótesis de inducción concluye.

Supongamos entonces que $p_i > 0$, para cada i. En ese caso, basta escribir

$$\sum_{i=1}^{k} p_i x^i = (1-p_n)\left(\sum_{i=1}^{k-1} \frac{p_i}{1-p_n} x^i\right) + p_n x^n,$$

y aplicar la definición de convexidad y la hipótesis de inducción a la suma entre paréntesis. □

Lema 14.2. *Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, C un convexo de $\mathbb{R}^k$ y $Z : (\Omega, \mathcal{A}) \longrightarrow \mathbb{R}^k$ una v.a. P-integrable tal que $Z(\omega) \in C$, P-c.s. Entonces*

$$E(Z) \in C.$$

Demostración. Procederemos de nuevo por inducción en k. En el caso $k = 1$, recordemos que los convexos de $\mathbb{R}$ son los intervalos, con lo cual basta considerar los casos $C =]-\infty, a]$ y $C =]-\infty, a[$, $a \in \mathbb{R}$. El primer caso es consecuencia inmediata de la monotonía de la integral; supuesto que $Z(\omega) < a$, P-c.s., debe ocurrir que $E(Z) < a$ pues, en otro caso, $E(Z - a) \geq 0$ y, siendo $Z - a < 0$, debería ser $E(Z - a) = 0$ y, por tanto, $Z = a$, P-c.s., contra lo supuesto. Supongamos ahora el lema cierto para $k - 1$ y probemos que también lo es para k. Reemplazando, si fuera preciso, Z por $Z - E(Z)$ y C por $C - E(Z) := \{x - E(Z) \colon x \in C\}$, puede suponerse sin pérdida de generalidad que $E(Z) = 0$. Basta probar, entonces, que $0 \in C$. Si $0 \notin C$, siendo C convexo, podemos separar 0 de C por un hiperplano (es el teorema de Hahn-Banach en el caso finito dimensional), es decir, existe $p = (p_i)_{1 \leq i \leq k} \in \mathbb{R}^k \setminus \{0\}$ tal que $\sum_{i=1}^k p_i y_i \geq 0$, para cada $y \in C$. Hagamos $U = \sum_{i=1}^k p_i Z_i$. Entonces $E(U) = 0$ y $P(U \geq 0) = 1$ pues Z tiene media cero y toma valores en C con probabilidad 1; se sigue de ello que $P(U = 0) = 1$. Denotemos $N = \{\omega \in \Omega \colon U(\omega) \neq 0\}$; entonces $P(N) = 0$. Siendo p no nulo, alguno de los p_i es no nulo. Supongamos $p_k \neq 0$ y hagamos $C' = \{(x_1, \dots, x_{k-1}) \in \mathbb{R}^{k-1} \colon (x_1, \dots, x_{k-1}, \sum_{i=1}^{k-1} q_i x_i) \in C\}$, donde $q_i = -p_i/p_k$, $1 \leq i \leq k-1$ (entonces $Z_k(\omega) = \sum_{i=1}^{k-1} q_i Z_i(\omega)$ si $\omega \notin N$). C' es trivialmente un convexo de $\mathbb{R}^{k-1}$. Si $Y = (Z_1, \dots, Z_{k-1})$, entonces Y es P-integrable y toma valores en C' con probabilidad 1, pues, si $\omega \notin N$, entonces $(Z_1(\omega), \dots, Z_{k-1}(\omega)) \in C'$. La hipótesis de inducción prueba que $E(Y) \in C'$, lo que significa que $(E(Z_1), \dots, E(Z_{k-1}), \sum_{i=1}^{k-1} q_i E(Z_i)) \in C$, es decir, $0 \in C$. De esta contradicción se sigue la tesis. $\square$

Lema 14.3. *Sean K un convexo compacto en $\mathbb{R}^k$ y $\varphi : K \longrightarrow \mathbb{R}$ una función convexa y continua. Sea μ una probabilidad en la σ-álgebra $\mathcal{R}^k(K)$ de los borelianos de K. Entonces*

$$\int_K x d\mu(x) \in K$$

y

$$\varphi\left(\int_K x d\mu(x)\right) \le \int_K \varphi(x) d\mu(x),$$

donde $\int_K x d\mu(x) = (\int_K x_1 d\mu(x_1, \dots, x_k), \dots, \int_K x_k d\mu(x_1, \dots, x_k))$.

Demostración. La identidad en K es medible y μ-integrable (K es acotado) y el lema anterior prueba que $\int_K x d\mu(x) \in K$. Siendo φ continua y K compacto, φ es uniformemente continua en K y, dado $\epsilon > 0$, existe $\delta > 0$ tal que, si $x, y \in K$ y $\|x - y\|_\infty \le \delta$, entonces $|\varphi(x) - \varphi(y)| < \epsilon$. Sean $E_1, \dots, E_p$ borelianos disjuntos en K con diámetro $\le \delta$ tales que $K = \cup E_i$ (si $B(x, \delta)$ denota la bola abierta de centro x y radio δ, entonces $(B(x, \delta/2) \cap K)_{x \in K}$ es un recubrimiento abierto de K del que, siendo K compacto, se puede extraer un subrecubrimiento finito $(B(x^i, \delta/2) \cap K)_{1 \le i \le p}$; basta tomar $E_i = [B(x^i, \delta/2) \cap K] \setminus [\cup_{j=1}^{i-1} B(x^j, \delta/2) \cap K]$). Tomemos puntos $y^i \in E_i$, $1 \le i \le p$. Entonces

$$\left| \int_K \varphi(x) d\mu(x) - \sum_{i=1}^p \varphi(y^i) \mu(E_i) \right| \le \sum_{i=1}^p \int_{E_i} |\varphi(x) - \varphi(y^i)| d\mu(x) \le \epsilon \sum_i \mu(E_i) = \epsilon.$$

También

$$\left\| \int_K x d\mu(x) - \sum_{i=1}^p \mu(E_i) y^i \right\|_\infty \le \sum_{i=1}^p \int_{E_i} \|x - y^i\|_\infty d\mu(x) \le \delta \sum_i \mu(E_i) = \delta,$$

y, por tanto, $|\varphi(\int_K x d\mu(x)) - \varphi(\sum_{i=1}^p \mu(E_i) y^i)| < \epsilon$. Entonces

$$\varphi\left(\int_K x d\mu(x)\right) \le \epsilon + \varphi\left(\sum_{i=1}^p \mu(E_i) y^i\right) \le \epsilon + \sum_{i=1}^p \varphi(y^i) \mu(E_i) \le 2\epsilon + \int_K \varphi(x) d\mu(x),$$

y eso acaba la prueba. $\square$

Lema 14.4. *Sea C un convexo de $\mathbb{R}^k$ que se puede escribir como unión numerable creciente de convexos compactos. Sea μ una probabilidad en la σ-álgebra $\mathcal{R}^k(C)$ de los borelianos de C tal que $\int_C \text{máx}_i(|x_i|)d\mu(x) < \infty$. Si $\varphi : C \longrightarrow \mathbb{R}$ es convexa y continua y tiene esperanza finita respecto a μ, entonces $\int_C xd\mu(x) \in C$ y*

$$\varphi\left(\int_C xd\mu(x)\right) \leq \int_C \varphi(x)d\mu(x).$$

Demostración. Ya sabemos que $\int_C xd\mu(x) \in C$. Escribamos ahora $C = \cup_{n\in\mathbb{N}} K_n$, siendo los K_n convexos compactos tales que $K_n \subset K_{n+1}$, $\forall n \in \mathbb{N}$. Entonces $\text{lím}_n \mu(K_n) = \mu(C) = 1$. Para $n \in \mathbb{N}$ se define $\mu_n(A) = \mu(A)/\mu(K_n)$ si A es un boreliano de K_n. μ_n es una probabilidad en K_n. Se sigue del lema anterior que $\varphi(\int_{K_n} xd\mu_n(x)) \leq \int_{K_n} \varphi(x)d\mu_n(x)$. Las hipótesis de integrabilidad prueban además que $\text{lím}_n \int_{K_n} xd\mu_n(x) = \text{lím}_n \frac{1}{\mu(K_n)} \int_{K_n} xd\mu(x) = \int_C xd\mu(x)$ y que $\text{lím}_n \int_{K_n} \varphi(x)d\mu_n(x) = \int_C \varphi(x)d\mu(x)$. La continuidad de φ implica que

$$\varphi\left(\int_C xd\mu(x)\right) = \text{lím}_n \varphi\left(\int_{K_n} xd\mu_n(x)\right) \leq \text{lím}_n \int_{K_n} \varphi(x)d\mu_n(x) = \int_C \varphi(x)d\mu(x).$$

□

TEOREMA 14.1. (Desigualdad de Jensen) *Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, C un convexo unión numerable creciente de convexos compactos de $\mathbb{R}^k$, $\varphi : C \longrightarrow \mathbb{R}$ una función convexa y continua y Z una v.a. k-dimensional en $(\Omega, \mathcal{A})$ tal que $Z(\omega) \in C$, P-c.s. Si $E(\text{máx}_i |Z_i|) < \infty$ y $E(|\varphi \circ Z|) < \infty$, entonces $E(Z) \in C$ y*

$$\varphi(E(Z)) \leq E(\varphi \circ Z).$$

Demostración. Hagamos $\mu = P^Z$. μ es una probabilidad en C. Además, $\int_C xd\mu = E(Z)$ y, por tanto, $\int_C xd\mu$ es finita. También $\int_C \varphi(x)d\mu(x) = \int_\Omega \varphi \circ ZdP$ es finita. Del lema anterior se sigue que $E(Z) \in C$ y que $\varphi(E(Z)) = \varphi(\int_C xd\mu) \leq \int_C \varphi(x)d\mu = E(\varphi \circ Z)$. □

TEOREMA 14.2. (Desigualdad de Jensen para la esperanza condicional) *Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ una v.a., C un convexo* *unión numerable creciente de convexos compactos de* $\mathbb{R}^k$, $\varphi : C \longrightarrow \mathbb{R}$ *una función continua y convexa y* Z *una v.a.* k*-dimensional que toma valores en* C, P*-c.s. Si las v.a.* $\varphi \circ Z$ *y* Z *tienen esperanza finita, entonces la v.a.* $E(Z|X)$ *toma valores en* C, P^X*-c.s., y*

$$\varphi(E(Z|X)) \leq E(\varphi \circ Z|X).$$

Demostración. Por definición, $E(Z|X) = (E(Z_1|X), \dots, E(Z_k|X))$. Siendo C un espacio de Borel existen las distribuciones condicionales $P^{Z|X=x}$ de Z respecto a X. Siendo $\varphi \circ Z$ y las componentes de Z P-integrables, se deduce de la observación que sigue al teorema 13.4 que

$$E(Z|X)(x) = \int_C z P^{Z|X=x}(dz), \quad P^X\text{-c.s.}$$

$$E(\varphi \circ Z|X)(x) = \int_C \varphi(z) P^{Z|X=x}(dz), \quad P^X\text{-c.s.}$$

De esta representación integral y del teorema anterior se sigue que

$$\varphi(E(Z|X)(x)) = \varphi\left(\int_C z P^{Z|X=x}(dz)\right)$$
$$\leq \int_C \varphi(z) P^{Z|X=x}(dz) = E(\varphi \circ Z|X)(x), \quad P^X\text{-c.s.}$$

□

***Observación* 14.1.** Se prueba que todo convexo cerrado o abierto C en $\mathbb{R}^k$ es unión de una sucesión (K_n) creciente de convexos compactos. En el caso cerrado, basta tomar $K_n = C \cap B'(0, n)$, donde $B'(0, n)$ denota la bola cerrada de centro 0 y radio n; en el caso abierto basta tomar $K_n = B'(0, n) \cap \{x \in \mathbb{R}^k : d(x, C^c) \geq 1/n\}$.

Una primera consecuencia de la desigualdad de Jensen es el siguiente resultado que nos permite resolver el que llamaremos problema general de regresión.

TEOREMA 14.3. *Sean* $1 \leq p < \infty$, $Y \in \mathcal{L}^p(\Omega, \mathcal{A}, P; \mathbb{R})$ *y* $\mathcal{B}$ *una sub-*σ*-álgebra de* $\mathcal{A}$*. Entonces* $E(Y|\mathcal{B}) \in \mathcal{L}^p(\Omega, \mathcal{B}, P; \mathbb{R})$ *y* $\|E(Y|\mathcal{B})\|_p \leq \|Y\|_p$*. Además, si* $p = 2$,

$$\|Y - E(Y|\mathcal{B})\|_2 \leq \|Y - Z\|_2, \quad \forall Z \in \mathcal{L}^2(\Omega, \mathcal{B}, P; \mathbb{R}).$$

Demostración. De la desigualdad de Jensen (Teorema 14.1) para la función continua y convexa $t \in \mathbb{R} \longrightarrow \varphi(t) = |t|^p \in \mathbb{R}$, se sigue que

$$E(|Y|^p) = E(E(|Y|^p|\mathcal{B})) \geq E(|E(Y|\mathcal{B})|^p),$$

lo que prueba que $E(Y|\mathcal{B}) \in \mathcal{L}^p(\Omega, \mathcal{B}, P; \mathbb{R})$ y que $\|E(Y|\mathcal{B})\|_p \leq \|Y\|_p$.

Si $Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P; \mathbb{R})$ y $Z \in \mathcal{L}^2(\Omega, \mathcal{B}, P; \mathbb{R})$, entonces, por la desigualdad de Cauchy-Schwarz,

$$ZY,\ ZE(Y|\mathcal{B}) \in \mathcal{L}^1(\Omega, \mathcal{A}, P; \mathbb{R});$$

además,

$$\int_\Omega ZE(Y|\mathcal{B})\, dP = E(ZE(Y|\mathcal{B})) = E(E(ZY|\mathcal{B})) = E(ZY) = \int_\Omega ZY\, dP$$

(siendo Z $\mathcal{B}$-medible, se verifica que $ZE(Y|\mathcal{B}) = E(ZY|\mathcal{B})$). Se sigue de ello que

$$\int_\Omega Z(Y - E(Y|\mathcal{B}))\, dP = 0, \quad \forall Z \in \mathcal{L}^2(\Omega, \mathcal{B}, P; \mathbb{R}). \tag{*}$$

Así, si $Z \in \mathcal{L}^2(\Omega, \mathcal{B}, P; \mathbb{R})$ y $Z' = E(Y|\mathcal{B}) - Z$, entonces

$$\begin{gathered}\|Y - Z\|_2^2 = \int |Y - Z|^2 dP = \\ \int |Y - E(Y|\mathcal{B}) + Z'|^2 dP = \int (Y - E(Y|\mathcal{B}) + Z')(Y - E(Y|\mathcal{B}) + Z') dP = \\ \int |Y - E(Y|\mathcal{B})|^2 dP + 2\int (Y - E(Y|\mathcal{B}))Z' dP + \int |Z'|^2 dP \\ \geq \int |Y - E(Y|\mathcal{B})|^2 dP = \|Y - E(Y|\mathcal{B})\|_2^2,\end{gathered}$$

pues $\int (Y - E(Y|\mathcal{B}))Z' dP = 0$ en virtud de (*). □

***Observación* 14.2.** (La esperanza condicional como aplicación lineal continua entre espacios L^p) Si $1 \leq p \leq \infty$, la aplicación $E(\cdot|\mathcal{B}) : L^p(\Omega, \mathcal{A}, P; \mathbb{R}) \longrightarrow L^p(\Omega, \mathcal{B}, P; \mathbb{R})$ es lineal y continua de norma ≤ 1 de acuerdo con el teorema anterior (incluso $E(\cdot|\mathcal{B})$ es de norma 1 según se prueba fácilmente). En el caso $p = 2$, $L^2(\Omega, \mathcal{A}, P; \mathbb{R})$ es un espacio de Hilbert (el producto escalar es $\langle X, Y \rangle = \int XY dP$) y $E(\cdot|\mathcal{B})$ es una proyección ortogonal sobre el subespacio([1]) $L^2(\Omega, \mathcal{B}, P; \mathbb{R})$, es decir, para cada $Y \in$

[1] Estrictamente hablando $L^2(\Omega, \mathcal{B}, P; \mathbb{R})$ no es un subespacio de $L^2(\Omega, \mathcal{A}, P; \mathbb{R})$ a menos que $\mathcal{B}$ contenga los sucesos P-nulos de $\mathcal{A}$; de hecho, si f es una función real $\mathcal{B}$-medible, las clases de f en ambos espacios pueden ser diferentes, pues la clase de f en $L^2(\Omega, \mathcal{A}, P; \mathbb{R})$ puede contener funciones que no sean $\mathcal{B}$-medibles que coincidan P-c.s. con f. No obstante, se prueba que si $\hat{\mathcal{B}}$ es la más pequeña σ-álgebra que contiene a $\mathcal{B}$ y a los sucesos P-nulos de $\mathcal{A}$, entonces los espacios de Hilbert $L^2(\Omega, \mathcal{B}, P; \mathbb{R})$ y $L^2(\Omega, \hat{\mathcal{B}}, P; \mathbb{R})$ son isométricos, y éste último sí es un subespacio de $L^2(\Omega, \mathcal{A}, P; \mathbb{R})$.

$L^2(\Omega, \mathcal{A}, P; \mathbb{R})$, $Y - E(Y|\mathcal{B})$ es ortogonal al subespacio $L^2(\Omega, \mathcal{B}, P; \mathbb{R})$ ($E(Y|\mathcal{B})$ es el punto de $L^2(\Omega, \mathcal{B}, P; \mathbb{R})$ más próximo a Y). Si $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ es una v.a. e $Y \in L^2(\Omega, \mathcal{A}, P; \mathbb{R})$, entonces $E(Y|X) \circ X = E(Y|\mathcal{B})$ si $\mathcal{B} = X^{-1}(\mathcal{A}')$. $E(Y|\mathcal{B})$ es, entonces, la función de X de cuadrado integrable que mejor aproxima a Y.

Consideremos una v.a. X (en regresión, ésta sería la llamada variable independiente que, a menudo, se supone controlada por el investigador) y nos planteamos el problema de "aproximar" una v.a.r. Y (en regresión Y se suele conocer como variable dependiente y podemos estar interesados en predecir el valor que tomará Y una vez conocido el que toma X) por una cierta función de la v.a. X. Para fijar la notación, sean Y una v.a.r. en $(\Omega, \mathcal{A}, P)$ y $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ una v.a., y supongamos que estamos interesados en encontrar una relación funcional entre Y y X, es decir, en encontrar una aplicación $h : \Omega' \longrightarrow \mathbb{R}$ tal que $Y(\omega) = h(X(\omega))$, para cada $\omega \in \Omega$. El problema, así planteado, no tendrá siempre solución (debería ocurrir, en particular, que, si $X(\omega) = X(\omega')$, entonces $Y(\omega) = Y(\omega')$, cosa que, en principio, no tiene por qué ser cierta). Cabe entonces replantear el problema en los siguiente términos: dadas dos v.a. Y y X como antes, encontrar una aplicación $h : \Omega' \longrightarrow \mathbb{R}$ tal que $h \circ X$ sea la función de X que mejor aproxima a Y en el sentido de los mínimos cuadrados, es decir, encontrar la función h que haga mínima la expresión $\|Y - h \circ X\|_2$. Debemos pues exigir que Y sea de cuadrado integrable y que h sea $\mathcal{A}'$-medible. El teorema anterior nos permite resolver el problema así replanteado. Hagamos, en efecto, $\mathcal{B} = X^{-1}(\mathcal{A}')$. $E(Y|\mathcal{B})$ es una función de X (concretamente, $E(Y|\mathcal{B}) = E(Y|X) \circ X$) y, si $Z = h \circ X$ es otra v.a. función de X con momento de orden 2 finito, entonces se verifica

$$\|Y - E(Y|X) \circ X\|_2 \leq \|Y - Z\|_2,$$

para cualquier función real $\mathcal{B}$-medible de cuadrado integrable o, equivalentemente,

$$\|Y - E(Y|X) \circ X\|_2 \leq \|Y - h \circ X\|_2,$$

para cualquier v.a.r. $h : \Omega' \longrightarrow \mathbb{R}$ de cuadrado P^X-integrable (esa desigualdad es también trivialmente cierta si h es una v.a.r. arbitraria, no necesariamente de cuadrado integrable). La solución del problema general de regresión es pues $h = E(Y|X)$. $E(Y|X)$ suele llamarse, por ello, regresión de Y sobre X. Si X es además una v.a.r., $E(Y|X)$ es una v.a.r. y la aplicación $x \in \mathbb{R} \longrightarrow E(Y|X = x)$ se llama también curva de regresión de Y sobre X. Si X es una v.a. n-dimensional, $E(Y|X)$ es una v.a.r. y la aplicación $(x_1, \ldots, x_n) \in \mathbb{R}^n \longrightarrow E[Y|X = (x_1, \ldots, x_n)]$ se llama superficie de regresión de Y sobre X.

Lo dicho resuelve teóricamente el problema general de regresión, que consiste en encontrar la función de la variable independiente X que mejor aproxima a la variable dependiente Y en el sentido de los mínimos cuadrados. Pero en una situación concreta, éste es, en general, un problema difícil, en la medida en que no es sencillo calcular esperanzas condicionales. En el caso de que X sea una v.a.r., el problema se simplifica si nos conformamos con buscar la función lineal de X que mejor aproxima a Y; éste es el llamado problema de regresión lineal, que pretende determinar constantes a y b tales que $a + bX$ aproxime lo mejor posible a Y en el sentido de los mínimos cuadrados (supuesto que tanto X como Y son de cuadrado integrable), es decir, se trata de minimizar la función $f(a, b) = E[(Y - a - bX)^2]$; es sencillo probar que la solución a ese problema es:

$$b = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} \qquad a = E(Y) - b \cdot E(X);$$

estos son, resp., la pendiente de regresión de Y sobre X y la ordenada en el origen de esa recta de regresión. El caso lineal tiene la ventaja de la sencillez; el caso general proporciona mejor aproximación de Y que el caso lineal.

En el caso de que X e Y tengan distribución conjunta normal, se verifica que la mejor función medible coincide con la mejor recta, con lo cual tenemos ambas ventajas a un tiempo. Puesto que la hipótesis de normalidad bidimensional queda justificada por el teorema del límite central en $\mathbb{R}^2$ como la unidimensional por el teorema límite central en $\mathbb{R}$, el resultado obtenido en este problema debe considerarse como un importante tanto a favor del uso de la técnica de regresión lineal en el estudio de la relación entre variables, tan utilizada en la práctica.

Supongamos que X e Y son v.a.r. en $(\Omega, \mathcal{A}, P)$ con distribución conjunta normal bidimensional (ver Apéndice) que viene dada por su densidad

$$\begin{aligned} f(x,y) =& \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ & \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \end{aligned}$$

donde $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$, $\mu_X = E(X)$, $\mu_Y = E(Y)$, $\rho = \mathrm{Cov}(X,Y)/(\sigma_X\sigma_Y)$ (coeficiente de correlación entre X e Y). Es sencillo comprobar que la distribución marginal de X es

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}\right\},$$

y que la densidad condicional de Y respecto a $X = x$ es

$$f^{Y|X=x}(y) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)\sigma_Y^2}\left[y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right]^2\right\},$$

que es la densidad de una distribución normal unidimensional de media $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$ y desviación típica $\sigma_Y\sqrt{1-\rho^2}$. Por tanto, la curva de regresión de Y sobre X es una recta; concretamente,

$$E(Y|X=x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X).$$

PROBLEMAS DEL CAPÍTULO 14

Problema 14.1. Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, $\mathcal{B}_1$ y $\mathcal{B}_2$ sub-σ-álgebras de $\mathcal{A}$ tales que $\mathcal{B}_1 \subset \mathcal{B}_2$ y $X : (\Omega, \mathcal{A}) \longrightarrow \mathbb{R}$ una v.a.r. tal que $E(X^2) < \infty$. Probar que

$$E[(X - E(X|\mathcal{B}_2))^2] \leq E[(X - E(X|\mathcal{B}_1))^2],$$

es decir, que la dispersión de X alrededor de su media condicional decrece a medida que la σ-álgebra crece. En particular, $E[(X - E(X|\mathcal{B}_2))^2] \leq \mathrm{Var}(X)$.

Problema 14.2. Consideremos el espacio de probabilidad $(\mathbb{R}^3, \mathcal{R}^3, P)$, donde $P = N(0,1) \times N(1,1) \times N(0,1)$, y denotemos por X, Y y Z las aplicaciones coordenada en $\mathbb{R}^3$.

(a) Calcular la esperanza condicional $E[(XZ + Y)Z^2|Z]$.

(b) Probar que $E(X|X + Y + Z)^2 \leq E(X^2|X + Y + Z)$, P-c.s.

Problema 14.3. Dadas dos v.a.r. X e Y, determinar la parábola de regresión de Y sobre X, es decir, determinar las constantes a, b, c tales que $\|Y - a - bX - cX^2\|_2$ sea lo más pequeño posible.

Problema 14.4. Sea (X, Y) una variable aleatoria bidimensional con función de densidad

$$f(x, y) = \begin{cases} \frac{1}{3} x e^{-x} & \text{si } x < 2y, \ y < 2x \\ 0 & \text{en el resto} \end{cases}$$

(a) Calcular la curva de regresión de Y sobre X. ¿Cuál es la recta de regresión de Y sobre X?

(b) Decidir si las variables aleatorias son incorreladas e independientes.

(c) Calcular la distribución conjunta de (Z, T), siendo $Z = X^2$ y $T = X + Y$.

APÉNDICES

DISTRIBUCIONES DISCRETAS UNIVARIANTES EN TEORÍA DE LA PROBABILIDAD Y EN ESTADÍSTICA

Distribución	Notación	Parámetros	$P(X=k)$	Media	Varianza	F. Característica	Observaciones
Dirac	ε_x	x	$P(X=x)=1$	x	0	e^{iux}	Distribución de una variable que sólo toma el valor x
Bernouilli	$b_1(p)$	$p \in [0,1]$	$p^k(1-p)^{1-k}$ $k=0,1$	p	$p(1-p)$	$1-p+pe^{it}$	Distribución de una variable que sólo toma los valores 0 y 1
Binomial	$b_n(p)$	$p \in [0,1]$	$\binom{n}{k}p^k(1-p)^{1-k}$ $0 \le k \le n$	np	$np(1-p)$	$[1-p+pe^{it}]^n$	Dist. del número de éxitos en n realizaciones independientes del experimento si la probabilidad de éxito es p
Uniforme discreta	$UD(n)$	$n \in \mathbb{N}$	$\frac{1}{n}$	$\frac{n(n+1)}{2}$	$\frac{(n+1)(2n+1)}{6}$	$e^{it}\frac{e^{int}-1}{e^{it}-1}$	Distribución de una v.a. que toma los valores $1,\dots,n$ con la misma probabilidad
Poisson	$P(\lambda)$	λ	$e^{-\lambda}\frac{\lambda^k}{k!}$ $k \ge 0$	λ	λ	$\exp\{\lambda(e^{it}-1)\}$	Distribución del número de ocurrencias de un suceso dado en un intervalo de tiempo
Binomial negativa	$BN(r,p)$	$r \in \mathbb{N}$ $p \in [0,1]$	$\binom{k+r-1}{k}p^r(1-p)^k$ $k \ge 0$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	$\frac{p^r}{(1-(1-p)e^{it})^r}$	Dist. del número de fracasos que preceden a r éxitos en realizaciones sucesivas independientes del experimento si la probabilidad de éxito es p
Geométrica	$BN(1,p)$	$p \in [0,1]$	$p(1-p)^k$ $k \ge 0$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)e^{it}}$	Dist. del número de fracasos que preceden al primer éxito en realizaciones sucesivas independientes del experimento si la probabilidad de éxito es p
Hipergeo-métrica	$H(N,N_1,n)$	$N,N_1,n \in \mathbb{N}$	$\frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}}$ $\text{máx}(0,n-N+N_1)$ $\le k \le \text{mín}(n,N_1)$	$\frac{nN_1}{N}$	$n\frac{N_1}{N}\frac{N-N_1}{N}\frac{N-n}{N-1}$		Si en una urna hay N_1 bola blancas y $N-N_1$ bola negras, la distribución del número de bolas blancas entre n elegidas al azar sin reemplazamiento de la urna es $H(N,N_1,n)$

DISTRIBUCIONES CONTINUAS UNIVARIANTES EN TEORÍA DE LA PROBABILIDAD Y EN ESTADÍSTICA (I)

Distribución	Notación	Parámetros	Densidad	Media	Varianza	F. Característica	Observaciones
Uniforme	$U[a,b]$	$a<b$	$\frac{1}{b-a}I_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$	Es la medida de Lebesgue en el intervalo normalizada para que sea una probabilidad
Normal	$N(\mu,\sigma^2)$	$\mu\in\mathbb{R},\sigma^2>0$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$	μ	σ^2	$\exp\{it\mu-\frac{1}{2}t^2\sigma^2\}$	Útil en teoría de errores y como modelo para otros muchos fenómenos aleatorios, recibe su importancia especialmente del Teorema del Límite Central, que permite considerar como normal la distribución de la suma de un gran número de v.a.r. independientes e idénticamente distribuidas
Exponencial	$G(1,\beta)$	$\beta>0$	$\frac{1}{\beta}e^{-x/\beta}I_{]0,\infty[}(x)$	β	β^2	$(1-i\beta t)^{-1}$	Es la distribución del tiempo de espera hasta la ocurrencia de un cierto suceso
Gamma	$G(\alpha,\beta)$	$\alpha,\beta>0$	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}I_{]0,\infty[}(x)$	$\alpha\beta$	$\alpha\beta^2$	$(1-i\beta t)^{-\alpha}$	Es una generalización de la distribución exponencial, que se utiliza cuando $\alpha\in\mathbb{N}$, en lugar de aquella, como distribución del tiempo de espera que transcurre hasta la ocurrencia de α sucesos. En ese caso se suele conocer como ley de Erlang.
Log-normal	$LN(\mu,\sigma^2)$	$\mu\in\mathbb{R},\sigma^2>0$	$\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\left(\frac{\log x-\mu}{\sigma\sqrt{2}}\right)^2\right\}$ $x>0$	$\exp\{\mu+\frac{1}{2}\sigma^2\}$	$\sigma^2(e^{\sigma^2}-1)$		Una v.a.r. X tiene distribución $LN(\mu,\sigma^2)$ si la v.a.r. $\log X$ tiene distribución normal $N(\mu,\sigma^2)$. Se usa en teoría de fiabilidad.
Cauchy	$C(\mu,\sigma)$	$\sigma>0$	$\frac{1}{\pi\sigma}\left(1+\left(\frac{x-\mu}{\sigma}\right)\right)^{-1}$	No tiene	No tiene	$\exp\{it\mu-\sigma\|t\|\}$	Si X es una v.a.r. con distribución uniforme en $[-\pi/2,\pi/2]$, la v.a. $\sigma\cdot\mathrm{tg}X$ tiene distribución $C(o,\sigma)$.

DISTRIBUCIONES CONTINUAS UNIVARIANTES EN TEORÍA DE LA PROBABILIDAD Y EN ESTADÍSTICA (II)

Distribución	Notación	Parámetros	Densidad	Media	Varianza	Observaciones
Laplace	$\mathcal{L}(c)$	$c>0$	$\frac{1}{2c}\exp\{-\lvert x\rvert/c\}$	0	$2c^2$	
Logística	$L(\mu,\sigma)$	$\mu\in\mathbb{R},\sigma>0$	$\frac{1}{\sigma}\frac{e^{-(x-\mu)/\sigma}}{[1+e^{-(x-\mu)/\sigma}]^2}$	μ	$\sigma^2\pi^2/3$	
Pareto	$Pa(a,\theta)$	$a,\theta>0$	$\theta a^\theta x^{-(1+\theta)}I_{]a,\infty[}(x)$	$\frac{a\theta}{\theta-1}$ $(\theta>1)$	$\frac{a^2\theta}{(\theta-1)^2(\theta-2)}$ $(\theta>2)$	Distribución usada a veces para variables que son superiores a un valor dado a; p.ej., distribución de rentas superiores a un nivel de ingresos dado.
Doble exponencial	$DE(\mu,\sigma)$	$\mu\in\mathbb{R},\sigma>0$	$\frac{1}{2\sigma}\exp\left\{-\frac{\lvert x-\mu\rvert}{\sigma}\right\}$	μ	$2\sigma^2$	Generaliza a la de Laplace.
Weibull	$W(\alpha,\beta)$	$\alpha,\beta>0$	$\frac{\alpha}{\beta}x^{\alpha-1}e^{-x^\alpha/\beta}I_{]0,\infty[}(x)$	$\beta^{1/\alpha}\Gamma\left(1+\frac{1}{\alpha}\right)$	$\beta^{2/\alpha}\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2$	Generaliza a la exponencial
Beta	$B(\alpha,\beta)$	$\alpha,\beta>0$	Ver problemas	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha}{(\alpha+\beta+1)(\alpha+\beta)^2}$	
Maxwell Rayleigh	$MR(c)$	$c>0$	$\frac{x}{c^2}\exp\left\{-\frac{x^2}{2c^2}\right\}I_{[0,\infty[}(x)$	$c\sqrt{\pi/2}$	$2c^2(1-\pi^2/4)$	Es, como la exponencial, un caso particular de la de Weibull, y se utiliza en teoría de la fiabilidad y en sistemas de comunicaciones.
Gumbel	$Gum(\alpha,\beta)$	$\alpha,\beta>0$	$\beta\alpha\exp\{\alpha x-\beta(e^{\alpha t-1})\}$			Se usa en fiabilidad.
Chi cuadrado	$\chi^2(n)$	$n\in\mathbb{N}$	Ver problemas	n	$2n$	Es la distribución gamma $G(n/2,2)$.
t de Student	$t(n)$	$n\in\mathbb{N}$	Ver problemas	0	$\frac{n}{n-2}$ $(n>2)$	
F	$F(m,n)$	$m,n\in\mathbb{N}$	Ver problemas	$\frac{n}{n-2}$ $(n>2)$	$\frac{2n^2(m+n-2)}{(m(n-2)^2(n-4))}$ $(n>4)$	

DISTRIBUCIONES MULTIVARIANTES EN TEORÍA

DE LA PROBABILIDAD Y EN ESTADÍSTICA

Distribución	Notación	Parámetros	$P(X=(k_1,\dots,k_n))$	Observaciones
Multinomial	$M(N,(N_i)_{1\le i\le r},n)$	$N,(N_i)_{1\le i\le r},n$ $N_1+\cdots+N_r=N$	$\frac{n!}{k_1!\cdots k_r!}q_1^{k_1}\cdots q_r^{k_r}$ $k_1+\cdots+k_r=n$ $q_i=\frac{N_i}{N}$	En una urna hay $N=N_1+\cdots+N_r$ bolas de las que N_i son del color c_i, $1\le i\le r$; se extraen al azar y con reemplazamiento n bolas de la urna; la probabilidad p_k obtenida es la de sacar k_1 bolas del color c_1, ..., y k_r bolas del color c_r entre las n bolas sacadas
Hipergeométri-ca múltiple	$H(N,(N_i)_{1\le i\le r},n)$	$N,(N_i)_{1\le i\le r},n$ $N_1+\cdots+N_r=N$	$\binom{N}{n}^{-1}\prod_{i=1}^r\binom{N_i}{k_i}$ $k_1+\cdots+k_r=n$	En una urna hay $N=N_1+\cdots+N_r$ bolas de las que N_i son del color c_i, $1\le i\le r$; se extraen al azar y sin reemplazamiento n bolas de la urna; la probabilidad p_k obtenida es la de sacar k_1 bolas del color c_1, ..., y k_r bolas del color c_r entre las n bolas sacadas

Distribución	Notación	Parámetros	Densidad	Media	Matriz de Cova-rianzas
Normal Multivariante	$N_n(\mu,C)$	$\mu\in\mathbb{R}^n$ C matriz simétrica definida positiva	$(2\pi)^{-\frac{n}{2}}\cdot(\det C)^{-\frac{1}{2}}\cdot\exp\left\{-\frac{1}{2}(x-\mu)^tC^{-1}(x-\mu)\right\}$	μ	C
		Función característica	$\exp\{iu^t\mu-\frac{1}{2}u^tCu\}$		

Índice alfabético

D