

Notas de Probabilidad y Estadística  
(en construcción- versión 0.7)

©2006-16 - Pablo L. De Nápoli

21 de junio de 2016

### **Prólogo a la versión 2016:**

La versión 2016 de este apunte ha sido completada con muchos temas que estaban sólo parcialmente desarrollados en las versiones anteriores (funciones características y convergencia en distribución, el teorema del límite central, etc.). Además se incluyeron muchos ejemplos, y gráficos ilustrativos realizados utilizando los paquetes *Matplotlib* y *Scipy* para el lenguaje *Python*, y el programa *Sagemath*.

### **Prólogo a la versión 2012:**

Actualmente estoy actualizando las notas, preparándolas para la versión 2012. Estoy corrigiéndoles algunos errores que me reportaron varios estudiantes y agregándole paulatinamente los temas faltantes.

Por otra parte, como se decidió que Análisis Real sea correlativa de Probabilidad y Estadística (para matemáticos), el enfoque en esta cursada será (aún) más elemental que en las anteriores. Por dicha razón, algunos temas como las leyes fuertes de Kolmogorov (que usan conceptos de análisis real) pasaron a los apéndices (cuyo contenido NO formará parte del programa del curso).

### **Prólogo a la versión 2010:**

Estas son las notas del curso de Probabilidades y Estadística (para matemáticos) que di en 2006. Mi intención es ir las completando a lo largo de la presente cursada del primer cuatrimestre de 2010.

Seguramente también, introduciré algunas modificaciones al desarrollo de los temas, teniendo en cuenta la experiencia de aquella cursada. Sin embargo, espero que las presentes notas sigan siendo de utilidad.

### **Prólogo a la versión 2006:**

El objetivo de estas notas es ser una ayuda para facilitar el seguimiento y la comprensión de las clases teóricas de Probabilidad y Estadística (para matemáticos).

Sin embargo se advierte que no contienen todos los temas vistos en clase (sino solamente algunos, en la medida que he podido ir escribiéndolas; algunas demostraciones están incompletas), ni los ejemplos vistos en las prácticas; y que su lectura no sustituye la consulta de la bibliografía de la materia.

En particular, los siguientes temas no están desarrollados en estas notas:

- Teorema del Límite. Central.

- Distribución Normal Multivariada.
- Estadística: estimadores de máxima verosimilitud e intervalos de confianza.

y algunas partes (probabilidad y esperanza condicionales, convergencia de variables aleatorias, ley fuerte de los grandes números) están incompletas.

Así mismo se advierte que estas notas aún están en contrucción, por lo que seguramente contienen algunos errores de tipeo o de redacción (que procuraré ir corrigiendo).

Una aclaración: Las demostraciones de los teoremas de análisis real que se incluyen en los apéndices no serán tomadas en el final (ya que forman parte del contenido de dicha materia). Sin embargo, los teoremas de convergencia monótona y mayorada están en el programa de Probabilidad y Estadística, y son necesarios para poder realizar algunas demostraciones de los teoremas sobre convergencia de variables aleatorias (ley fuerte de los grandes números, etc.), por lo que sí es necesario conocer al menos los enunciados, para poder comprender dichas demostraciones.

Agradecimientos: aún a riesgo de olvidarme de alguien, no quiero dejar de agradecer a todos los que de alguna manera me ayudaron a dar la materia, y a redactar este apunte. A N. Fava y V. Yohai (con quienes en su momento cursé esta materia, dado que el curso estuvo inspirado en gran parte en lo que aprendí de ellos); a G. Boente Boente (quien generosamente me prestó el material de sus clases, y me reemplazó durante algunas clases que estuve de licencia); a M. Svarc, S. Laplagne y J. Molina (que fueron mis ayudantes durante el curso, y me brindaron en todo momento una inestimable colaboración), también a M. A. García Alvarez (por regalarme su excelente libro); y finalmente, a todos mis alumnos, quienes en muchas veces han aportado correcciones u observaciones que han contribuido a mejorar este apunte.

Pablo L. De Nápoli

# Índice general

<b>1. El Espacio Muestral</b>	<b>7</b>
1.1. Experimentos Aleatorios . . . . .	7
1.2. La definición clásica de Laplace . . . . .	8
1.3. Definición axiomática de la probabilidad (provisional) . . . . .	10
1.4. Probabilidad Condicional . . . . .	12
1.5. Independencia . . . . .	15
1.6. El marco de Kolmogorov . . . . .	15
<b>2. Variables Aleatorias Discretas</b>	<b>19</b>
2.1. La Esperanza . . . . .	19
2.1.1. Esperanzas infinitas . . . . .	23
2.1.2. Propiedades de la Esperanza . . . . .	23
2.1.3. Independencia . . . . .	26
2.1.4. Desigualdad de Jensen . . . . .	27
2.2. Momentos - Varianza . . . . .	29
2.2.1. Desigualdades de Tchesbychev y de Markov . . . . .	31
2.2.2. Covarianza . . . . .	32
2.3. Ensayos de Bernoulli - La Distribución Binomial . . . . .	33
2.4. El método de las funciones generatrices . . . . .	36
2.4.1. El teorema de Bernoulli . . . . .	39
2.5. Ley débil de los grandes números: caso general . . . . .	41
2.6. Polinomios de Bernstein: Una prueba del teorema de Weierstrass	45
2.7. La aproximación de Poisson a la distribución binomial . . . . .	47
2.8. Otras distribuciones relacionadas con los ensayos de Bernoulli	50
<b>3. Distribuciones Continuas</b>	<b>55</b>
3.1. Variables aleatorias continuas . . . . .	55
3.1.1. Propiedades de las funciones de distribución . . . . .	59

3.2.	La integral de Riemann-Stieltjes y la definición de esperanza .	62
3.3.	La definición de Esperanza . . . . .	64
3.4.	Vectores Aleatorios . . . . .	73
3.4.1.	Densidades y distribuciones marginales . . . . .	76
3.4.2.	Esperanza de funciones de vectores aleatorios. Cova- rianza . . . . .	77
3.4.3.	Independencia . . . . .	80
3.4.4.	Vectores aleatorios $n$ -dimensionales . . . . .	83
3.5.	Cambio de variable . . . . .	84
3.5.1.	Cambios de variables unidimensionales . . . . .	84
3.5.2.	Cambios de variables $n$ -dimensionales . . . . .	86
3.6.	Suma de variables aleatorias independientes . . . . .	87
3.6.1.	Suma de variables normales independientes . . . . .	88
3.7.	Las Distribuciones Gama . . . . .	90
3.8.	La Distribución Exponencial y la propiedad de Falta de Memoria	93
3.8.1.	Tiempos de espera y procesos de Poisson . . . . .	95
3.9.	Algunas densidades útiles en estadística . . . . .	97
3.9.1.	Las densidades $\chi^2$ . . . . .	97
3.9.2.	Densidad del cociente de dos variables aleatorias inde- pendientes . . . . .	98
3.9.3.	La densidad $t$ de Student . . . . .	99
3.10.	Distribución Normal Multivariada . . . . .	100
<b>4.</b>	<b>Convergencia de Variables Aleatorias, y Ley Fuerte de los Grandes Números</b>	<b>101</b>
4.1.	Los diferentes tipos de convergencia . . . . .	101
4.2.	Relación entre los modos de convergencia . . . . .	106
4.3.	El lema de Borel-Cantelli . . . . .	107
4.4.	La ley fuerte de los grandes números . . . . .	109
4.4.1.	Un ejemplo: La ley fuerte de Borel para ensayos de Bernoulli . . . . .	111
4.4.2.	Números Normales . . . . .	113
4.5.	El Teorema de Helly . . . . .	114
4.6.	El Recíproco del teorema de Helly . . . . .	117
4.7.	El Principio de Selección de Helly . . . . .	119
4.8.	Funciones Características . . . . .	121
4.8.1.	La Función Característica de la Distribución Normal .	124

4.8.2.	La Fórmula de Inversión: unicidad de la función característica . . . . .	126
4.9.	El Teorema de Continuidad de Paul Levy . . . . .	127
<b>5.</b>	<b>El Teorema del Límite Central</b>	<b>130</b>
5.1.	El Teorema Local de De Moivre-Laplace . . . . .	130
5.2.	El Teorema de De Moivre-Laplace . . . . .	134
5.3.	Una Aplicación a la Estadística . . . . .	139
5.4.	El Teorema del Límite Central . . . . .	142
5.4.1.	Aplicación a las distribuciones $\chi_n^2$ . . . . .	144
5.5.	Generalizaciones y comentarios adicionales . . . . .	145
5.6.	Una Aplicación a la Teoría de Números . . . . .	147
<b>6.</b>	<b>Esperanza Condicional</b>	<b>150</b>
6.1.	Esperanza condicional respecto de un evento . . . . .	150
6.2.	Esperanzas condicionales en el caso discreto . . . . .	151
6.3.	Esperanzas condicionales en el caso continuo . . . . .	153
<b>A.</b>	<b>La Fórmula de Stirling</b>	<b>155</b>
A.1.	La fórmula de Wallis para $\pi$ . . . . .	155
A.1.1.	Otra fórmula de la fórmula de Wallis . . . . .	157
A.2.	Prueba de la fórmula de Stirling . . . . .	158
<b>B.</b>	<b>Construcción de la Integral de Lebesgue, y equivalencia de las distintas definiciones de esperanza</b>	<b>161</b>
B.1.	Funciones Medibles . . . . .	162
B.1.1.	Funciones Simples . . . . .	166
B.2.	Integral de Funciones Simples . . . . .	167
B.3.	Integral de funciones no negativas . . . . .	168
B.4.	Funciones Integrables . . . . .	172
B.5.	Equivalencia de las distintas definiciones de Esperanza . . . . .	175
B.5.1.	Vectores Aleatorios . . . . .	179
<b>C.</b>	<b>Independencia</b>	<b>180</b>
C.1.	El teorema $\pi - \lambda$ de Dynkin . . . . .	180
C.2.	Variables independientes . . . . .	182
C.3.	Esperanza del producto de variables independientes . . . . .	184
<b>D.</b>	<b>Existencia de las Integrales de Riemann-Stieltjes</b>	<b>186</b>

<b>E. Las Leyes Fuertes de Kolmogorov</b>	<b>190</b>
E.1. La Desigualdad de Kolmogorov . . . . .	190
E.2. La Ley Fuerte de los Grandes Números . . . . .	192
E.2.1. La Primera Ley Fuerte de Kolmogorov . . . . .	192
E.2.2. Algunos Lemas Preparatorios . . . . .	195
E.2.3. La Segunda Ley Fuerte de Kolmogorov . . . . .	198

# Capítulo 1

## El Espacio Muestral

### 1.1. Experimentos Aleatorios

La teoría de probabilidades trata con experimentos aleatorios, es decir con experimentos cuyo resultado no resulta posible prever de antemano. Denominamos **espacio muestral** al conjunto de los posibles resultados de un experimento aleatorio, y lo simbolizamos con la letra  $\Omega$ .

Históricamente, la teoría de probabilidades se desarrolló para estudiar los juegos de azar, pero posteriormente encontró otras innumerables aplicaciones. En estos casos el espacio muestral es usualmente finito:

**Ejemplos de experimentos aleatorios:**

- Se arroja una moneda. Hay dos resultados posibles:

$$\Omega = \{cara, ceca\}$$

- Se arroja un dado. Hay seis resultados posibles:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Sin embargo, en otras aplicaciones del cálculo de probabilidades, aparecen espacios muestrales de naturaleza más compleja. Veamos algunos ejemplos:

- Se elije un individuo al azar de una población humana y se mide su altura. El resultado es un número real positivo (dentro de un cierto



rango). Podemos por lo tanto pensar que el espacio muestral  $\Omega$  es un intervalo de la recta real.

- Se observa la trayectoria de una partícula que se mueve sobre la superficie de un líquido siguiendo una trayectoria de apariencia caótica durante un cierto intervalo de tiempo  $[0, T]$  (movimiento Browniano). En este caso, cada posible resultado del experimento es una curva continua. Por ello el espacio muestral podría tomarse como el espacio de funciones continuas  $C([0, T], \mathbb{R}^2)$ .

Un **evento** o **suceso** es algo que puede ocurrir o no ocurrir en cada realización del experimento aleatorio. Los eventos corresponden a subconjuntos del espacio muestral. Por ejemplo: si el experimento consiste en arrojar un dado, el evento “sale un número par” está representado por el subconjunto  $A = \{2, 4, 6\}$  del espacio muestral.

## 1.2. La definición clásica de Laplace

La idea básica del cálculo de probabilidades será asignar a cada evento  $A \subset \Omega$ , un número real entre 0 y 1 que llamaremos su probabilidad y simbolizaremos por  $P(A)$ . Este número medirá qué tan probable es que ocurra el evento  $A$ .

El matemático francés Pierre-Simon Laplace (1749–1827) propuso la siguiente definición del concepto de probabilidad: consideremos un experimento aleatorio que tiene un número finito de resultados posibles

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$$

y supongamos que dichos resultados son *equiprobables* (es decir que consideramos que cada uno de ellos tiene las mismas chances de ocurrir o no que los demás), entonces la probabilidad de un evento  $A \subset \Omega$  se define por

$$P(A) = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{\#(A)}{\#(\Omega)}$$

Por ejemplo, supongamos que nos preguntamos ¿cuál es la probabilidad de obtener un número par al arrojar un dado?. En este caso hay 6 casos posibles, que corresponden a los elementos del espacio muestral

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

y 3 casos posibles, que corresponden a los elementos del evento

$$A = \{2, 4, 6\}$$

Si suponemos que el dado no está cargado (de modo que asumimos que los seis resultados posibles del experimento son equiprobables), entonces

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

¿Cuál es el significado intuitivo de esta probabilidad?. Intuitivamente, esperamos que si repetimos el experimento muchas veces, observemos que aproximadamente la mitad de las veces sale un número par (y la otra mitad de las veces sale un número impar).

Notemos algunas propiedades de la noción de probabilidad, introducida por la definición de Laplace:

1. La probabilidad de un evento es un número real entre 0 y 1.

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

2. La probabilidad de un evento imposible es 0:

$$P(\emptyset) = 0$$

mientras que la probabilidad de un evento que ocurre siempre es 1:

$$P(\Omega) = 1$$

Por ejemplo; al tirar un dado, la probabilidad de sacar un 7 es cero mientras que la probabilidad de sacar un número menor que 10 es uno (Los eventos imposibles corresponden como conjuntos al conjunto vacío, y los que ocurren siempre corresponden a todo el espacio muestral  $\Omega$ ).

Notemos que para el concepto de probabilidad introducido por la definición clásica de Laplace, es cierta la recíproca de esta afirmación: si  $P(A) = 0$ , el suceso  $A$  es imposible, mientras que si  $P(A) = 1$  el suceso ocurre siempre. Sin embargo, esto no será cierto para otras extensiones del concepto de probabilidad que introduciremos más adelante.

3. Si  $A$  y  $B$  son dos eventos que no pueden ocurrir simultáneamente, entonces la probabilidad de que ocurra  $A$  u ocurra  $B$  (lo que corresponde como conjunto a  $A \cup B$ ), es cero

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

### 1.3. Definición axiomática de la probabilidad (provisional)

La definición clásica de Laplace, aunque tiene un claro significado intuitivo presenta algunas limitaciones. En primer lugar, su aplicación está limitada a problemas donde el espacio muestral es finito. Sin embargo como hemos mencionado al comienzo, en muchas aplicaciones importantes del cálculo de probabilidades, nos encontramos con espacios muestrales que no lo son.

Por otra parte, la definición clásica de Laplace hace la suposición de que los posibles resultados del experimento aleatorio (los puntos del espacio muestral) son equiprobables, pero es fácil imaginar experimentos en los que esta suposición no se verifica, por ejemplo si arrojamus un dado que no está equilibrado (“está cargado”).

Por los motivos expresados, será conveniente generalizar la noción de probabilidad. Por ello, introduciremos la siguiente definición axiomática (provisional).

**Definición 1.3.1** Sea  $\Omega$  un espacio muestral, por una probabilidad definida en  $\Omega$  entenderemos una función  $P$  que a cada parte de  $\Omega$  (evento) le asigna un número real de modo que se cumplen las propiedades enunciadas en la sección anterior:

1. La probabilidad de un evento  $A$  es un número real entre 0 y 1:

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

2. La probabilidad del evento imposible es 0:

$$P(\emptyset) = 0$$

mientras que la probabilidad de un evento que ocurre siempre es 1:

$$P(\Omega) = 1$$

3. *La probabilidad es finitamente aditiva:*

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

Más adelante, nos veremos obligados a modificar esta definición, ya que en muchos ejemplos no es posible asignar probabilidades a todas las posibles partes de  $\Omega$  (por lo que deberemos restringir la noción de evento).

Veamos algunos ejemplos:

Supongamos que tenemos un espacio muestral finito

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$$

pero que no queremos asumir que los posibles resultados de nuestro experimento aleatorio son equiprobables. Entonces supondremos que cada uno de ellos tiene una probabilidad  $p_i \in [0, 1]$ :

$$P(\{\omega_i\}) = p_i$$

Entonces dado un evento  $A \subset \Omega$ , le asignamos la probabilidad

$$P(A) = \sum_{\omega_i \in A} p_i$$

Si suponemos que

$$\sum_{i=1}^n p_i = 1$$

entonces la probabilidad así definida, verifica los axiomas de nuestra definición axiomática de probabilidad.

Notemos que en particular, si los resultados  $\omega_i$  ( $1 \leq i \leq n$ ) son equiprobables:

$$p_1 = p_2 = \dots = p_n$$

entonces  $p_i = \frac{1}{n}$  para todo  $i$ , y recuperamos la definición clásica de Laplace:

$$P(A) = \frac{\#(A)}{n}$$

El ejemplo anterior, fácilmente puede generalizarse al caso de un espacio muestral numerable

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$$

Nuevamente supongamos que a cada resultado  $r_i$  (con  $i \in \mathbb{N}$ ) le hemos asignado una probabilidad  $p_i \in [0, 1]$ , de modo que

$$\sum_{i=1}^{\infty} p_i = 1$$

entonces si definimos

$$P(A) = \sum_{r_i \in A} p_i$$

obtenemos una probabilidad definida en  $\Omega$ .

Es importante notar, que para esta nueva noción de probabilidad que hemos definido ya no se verifica en general que  $P(A) = 0$  implique que  $A$  sea un evento imposible, o que si  $P(A) = 1$  entonces  $A$  es un evento que ocurre siempre.

Veamos algunas consecuencias de estas definiciones:

**Proposición 1.3.1** *Si  $A$  es un evento y  $A^c = \Omega - A$  su complemento, entonces*

$$P(A^c) = 1 - P(A)$$

En efecto:  $\Omega = A \cup A^c$  (unión disjunta), en consecuencia  $1 = P(A) + P(A^c)$ , luego  $P(A^c) = 1 - P(A)$ .

**Proposición 1.3.2** *Si  $A$  y  $B$  son dos eventos, entonces*

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

## 1.4. Probabilidad Condicional

En muchas situaciones tendremos que estimar la probabilidad de un evento pero disponemos de alguna información adicional sobre su resultado.

Por ejemplo supongamos que arrojamus un dado (equilibrado) y nos preguntamos ¿Qué probabilidad le asignaríamos a sacar un dos, si supiéramos de antemano que el resultado será un número par?. Para formalizar esta pregunta consideramos en el espacio muestral

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

los eventos

$$A = \text{sale un 2} = \{2\}$$

$$B = \text{sale un número par} = \{2, 4, 6\}$$

Entonces vamos a definir la probabilidad condicional de que ocurra el evento  $A$  sabiendo que ocurre el evento  $B$  que notaremos  $P(A/B)$ .

Si estamos en una situación como la anterior donde la definición clásica de Laplace se aplica podemos pensarlo del siguiente modo: los resultados posibles de nuestro experimento son ahora sólo los elementos de  $B$  (es decir: hemos restringido nuestro espacio muestral a  $B$ ), mientras que los casos favorables son ahora los elementos de  $A \cap B$  luego

$$P(A/B) = \frac{\#(A \cap B)}{\#(B)}$$

Si dividimos numerador y denominador por  $\#(\Omega)$ , tenemos:

$$P(A/B) = \frac{\frac{\#(A \cap B)}{\#(\Omega)}}{\frac{\#(B)}{\#(\Omega)}} = \frac{P(A \cap B)}{P(B)}$$

Aunque hemos deducido esta fórmula de la definición clásica de Laplace, la misma tiene sentido en general siempre que  $P(B) > 0$ . Adoptamos pues la siguiente definición:

**Definición 1.4.1** *La probabilidad condicional  $P(A/B)$  de un evento  $A$  suponiendo que ocurre el evento  $B$  se define por:*

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \tag{1.1}$$

*siempre que  $P(B) > 0$ .*

Otra manera de comprender esta definición es la siguiente: para definir la probabilidad condicional  $P(A/B)$  queremos reasignar probabilidades a los eventos  $A \subset \Omega$  de modo que se cumplan tres condiciones:

1. La función  $A \mapsto P(A/B)$  debe ser una probabilidad (o sea satisfacer los requisitos de nuestra definición axiomática).
2.  $P(A \cap B/B) = P(A/B)$  (Esta fórmula dice que la probabilidad condicional de que ocurran los eventos  $A$  y  $B$  simultáneamente sabiendo que ocurre  $B$  debe ser igual a la probabilidad condicional de  $A$  sabiendo que ocurre  $B$ ).
3. Si  $A \subset B$  la probabilidad condicional  $P(A/B)$  debe ser proporcional a la probabilidad de  $A$  de modo que

$$P(A/B) = kP(A) \text{ si } A \subset B$$

siendo  $k$  una constante de proporcionalidad fija.

Entonces a partir de estas dos condiciones tenemos:

$$P(A/B) = P(A \cap B/B) = kP(A \cap B)$$

y como queremos que  $P(A/B)$  sea una probabilidad debe ser  $P(\Omega/A) = 1$ , luego

$$1 = kP(\Omega \cap B) = kP(B)$$

con lo que:

$$k = \frac{1}{P(B)}$$

y vemos que la definición (1.1) es la única que satisface estas condiciones.

Si ahora consideramos una partición del espacio muestral  $\Omega$  en eventos disjuntos  $B_1, B_2, \dots, B_n$  con  $P(B_k) > 0$  para todo  $k$  tenemos que:

$$P(A) = \sum_{k=1}^n P(A \cap B_k)$$

por la aditividad de la probabilidad, y como

$$P(A \cap B_k) = P(B_k)P(A/B_k)$$

en virtud de la definición de probabilidad condicional, deducimos la siguiente fórmula:

$$P(A) = \sum_{k=1}^n P(B_k)P(A/B_k)$$

(fórmula de la probabilidad total)

## 1.5. Independencia

**Definición 1.5.1** Decimos que el evento  $A$  es independiente del evento  $B$  con  $P(B) > 0$  si

$$P(A/B) = P(A)$$

Intuitivamente este concepto significa que saber si el evento  $B$  ocurre o no, no nos dará una mejor estimación de la probabilidad de que ocurre el evento  $B$  que si no lo supiéramos.

Teniendo en cuenta la definición de la probabilidad condicional, vemos que la condición para que el evento  $A$  sea independiente de  $B$  es que:

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

Esta manera de escribir la definición tiene dos ventajas: se ve que tiene sentido aún si  $P(B) = 0$ , y muestra que los roles de los eventos  $A$  y  $B$  son simétricos. Reescribimos pues la definición en la siguiente forma:

**Definición 1.5.2** Decimos que los eventos  $A$  y  $B$  son (estocásticamente) independientes si

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

Esta definición admite la siguiente generalización:

**Definición 1.5.3** Decimos que una familia cualquiera de eventos  $(A_i)_{i \in I}$  es independiente si

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_n})$$

para cualquier elección de una cantidad finita  $A_{i_1}, \dots, A_{i_n}$  de eventos distintos de la familia.

## 1.6. El marco de Kolmogorov

Como hemos dicho, en muchas situaciones importantes, no es posible asignar probabilidades a todos los subconjuntos del espacio muestral.

El ejemplo más sencillo de esta situación es el siguiente: supongamos que realizamos el experimento de elegir un número real del intervalo  $[0, 1]$  con



“distribución uniforme”. Con esto queremos decir que si  $I \subset [0, 1]$  es un intervalo, queremos que:

$$P(I) = |I| \quad (1.2)$$

donde  $I$  designa la longitud del intervalo  $I$ .

Un experimento equivalente es el siguiente (ruleta continua): imaginemos que tenemos una rueda y la hacemos girar. Nos interesa medir cual es la posición de la rueda. Dado que esta está determinada por un ángulo  $\theta \in [0, 2\pi)$  respecto de la posición inicial, podemos pensar este experimento como elegir un número al azar en el intervalo  $[0, 2\pi)$ . La distribución uniforme, corresponde a postular que todas las posiciones finales de la rueda son igualmente probables.

Se demuestra en análisis real que no es posible definir una medida (probabilidad)  $\sigma$ -aditiva, que esté definida para todos los posibles subconjuntos del intervalo  $[0, 1]$  de modo que se verifique la relación (1.2) para cada subintervalo  $I \subset [0, 1]$ .

Lebesgue propuso la siguiente solución a este problema: restringir la clase de los conjuntos a los que asignaremos medida (probabilidad) a lo que se llama una  $\sigma$ -álgebra.

**Definición 1.6.1** Sea  $\Omega$  un conjunto (espacio muestral). Una  $\sigma$ -álgebra de partes de  $\Omega$ , es una colección de partes de  $\Omega$  con las siguientes propiedades:

1.  $\emptyset \in \mathcal{E}$ .
2. Si  $A$  está en  $\mathcal{E}$ , entonces su complemento  $A^c = \Omega - A \in \mathcal{E}$ .
3. Si  $(A_n)_{n \in \mathbb{N}}$  es una familia numerable de conjuntos de  $\Omega$  entonces  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$ .

Obviamente, el conjunto de todas las partes de  $\Omega$ ,  $\mathcal{P}(\Omega)$  es una  $\sigma$ -álgebra, pero existen  $\sigma$ -álgebras más pequeñas.

**Algunas observaciones importantes:**

Si  $\mathcal{E}$  es una  $\sigma$ -álgebra de partes de  $\Omega$ , entonces

1.  $\Omega \in \mathcal{E}$ .

2. Si  $(A_n)_{n \in \mathbb{N}}$  es una familia numerable de subconjuntos de  $\Omega$  entonces  $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{E}$  Prueba: por la fórmula de De Morgan

$$\bigcap_{n \in \mathbb{N}} A_n = \left( \bigcup_{n \in \mathbb{N}} A_n^c \right)^c$$

3. Si  $A, B \in \mathcal{E}$  entonces  $A - B \in \mathcal{E}$ .

**Definición 1.6.2** *Observemos que la intersección de una familia cualquiera de  $\sigma$ -álgebras de partes de  $\Omega$ , también es una  $\sigma$ -álgebra. Deducimos que para cualquier  $\mathcal{A} \subset \mathcal{P}(\Omega)$ , existe una menor  $\sigma$ -álgebra que la contiene. Dicha  $\sigma$ -álgebra se denomina la  $\sigma$ -álgebra generada por  $\mathcal{A}$ .*

*Definimos la  $\sigma$ -álgebra de Borel de  $\mathbb{R}$ , como la  $\sigma$ -álgebra generada por los intervalos abiertos de  $\mathbb{R}$ . **Notación:**  $\mathcal{B}(\mathbb{R})$*

**Definición 1.6.3** *Sean  $\Omega$  un conjunto y  $\mathcal{E} \subset \mathcal{P}(\Omega)$ . Una medida sobre  $\mathcal{E}$  es una función  $\mu : \mathcal{E} \rightarrow [0, +\infty]$ . con las siguientes propiedades:*

1.

$$\mu(\emptyset) = 0$$

2. Si  $(A_n)_{n \in \mathbb{N}}$  es una familia disjunta numerable de conjuntos de  $\mathcal{E}$ , entonces:

$$\mu \left( \bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n)$$

*Si además se verifica que  $\mu(\Omega) = 1$ ,  $\mu$  se denomina una medida de probabilidad sobre  $\Omega$ .*

**Definición 1.6.4** *Un espacio de probabilidad es una terna  $(\Omega, \mathcal{E}, P)$  donde  $\Omega$  es un conjunto (espacio muestral),  $\mathcal{E}$  es una  $\sigma$ -álgebra de partes de  $\Omega$  (la  $\sigma$ -álgebra de los eventos) y  $P$  es una medida de probabilidad sobre  $\Omega$ .*

El siguiente es un resultado fundamental de análisis real:

**Teorema 1.6.1** *(Existencia de la medida de Lebesgue) Existen una única  $\sigma$ -álgebra  $\mathcal{M}$  de partes de  $\mathbb{R}$  y una única medida  $m : \mathcal{M} \rightarrow [0, +\infty)$  con las siguientes propiedades:*

1.  $\mathcal{M}$  contiene a los intervalos abiertos (por lo tanto  $\mathcal{M}$  contiene a la  $\sigma$ -álgebra de Borel).
2.  $m(I) = |I|$  para cualquier intervalo de la recta.
3. Para cualquier conjunto  $A \in \mathcal{M}$ , la medida de  $A$  es el supremo de las medidas de los compactos contenidos en  $A$ :

$$m(A) = \sup\{m(K) : K \text{ compacto}, K \subset A\}$$

y es el ínfimo de las medidas de los abiertos que contienen a  $A$ :

$$m(A) = \inf\{m(U) : U \text{ abierto}, U \supset A\}$$

(Se dice que la medida  $m$  es regular).

4. La medida  $m$  es invariante por traslaciones:

$$m(A + x) = m(A) \quad \forall A \in \mathcal{M}$$

5. Si  $A \in \mathcal{M}$ ,  $m(A) = 0$  y  $B \subset A$ ; entonces  $B \in \mathcal{M}$  y  $m(B) = 0$ . (se dice que la  $\sigma$ -álgebra de Lebesgue es completa).

$\mathcal{M}$  se denomina la  $\sigma$ -álgebra de Lebesgue y  $m$  se denomina la medida de Lebesgue. Los conjuntos de la  $\sigma$ -álgebra  $\mathcal{M}$  se denominan conjuntos medibles Lebesgue.

**Corolario 1.6.1** Si consideramos la restricción de la medida de Lebesgue y de la  $\sigma$ -álgebra de Lebesgue al intervalo  $[0, 1]$ , entonces obtenemos un espacio de probabilidad.

# Capítulo 2

## Variables Aleatorias Discretas

### 2.1. La Esperanza

En muchas situaciones, nos interesa un número asociado al resultado de un experimento aleatorio: por ejemplo, el resultado de una medición.

Para evitar por el momento, algunas dificultades técnicas, comenzaremos con el caso de variables aleatorias discretas, que resulta más sencillo de entender.

**Definición 2.1.1** Sea  $(\Omega, \mathcal{E}, P)$  un espacio de probabilidad. Una variable aleatoria discreta es una función  $X : \Omega \rightarrow \mathbb{R}$  tal que la imagen de  $X$  es un conjunto finito o numerable de  $\mathbb{R}$ :

$$Im(X) = \{x_1, x_2, \dots, x_i, \dots\}$$

(donde la sucesión  $(x_i)$  puede ser finita o infinita), y tal que  $X^{-1}(\{x_i\}) \in \mathcal{E}$  sea un evento para cada  $x_i \in Im(X)$ .

Como  $X^{-1}(\{x_i\}) = \{\omega \in \Omega : X(\omega) = x_i\}$  es un evento para cada  $i$ , esto significa que están definidas las probabilidades:

$$p_i = P(\{X = x_i\})$$

Dichas probabilidades se conocen como la *distribución de probabilidades* de la variable  $X$ .

Un concepto de fundamental importancia asociado a las variables aleatorias, es el de esperanza (o valor esperado). Para variables aleatorias discretas, este concepto resulta sencillo de definir:

**Definición 2.1.2** Sea  $X : \Omega \rightarrow \mathbb{R}$  una variable aleatoria discreta. Diremos que  $X$  es integrable (o que tiene esperanza finita) si la serie

$$\sum_i p_i x_i$$

es absolutamente convergente, es decir si:

$$\sum_i p_i |x_i| < +\infty$$

En este caso definimos, la esperanza de  $X$  como el valor de dicha suma.

$$E[X] = \sum_i p_i x_i$$

Hagamos algunas observaciones sobre esta definición:

- Una variable aleatoria cuya imagen es finita siempre es integrable.
- Una variable aleatoria discreta no negativa con imagen finita (o sea: que tome sólo un número finito de valores) siempre es integrable.

**Ejemplo:** Supongamos que arrojamos un dado ¿cuál es la esperanza del valor obtenido  $X$  ?

$$E[X] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3,5$$

**Ejemplo:** Supongamos que jugamos un peso a la ruleta y apostamos a un color (por ej. negro). Sea  $X$  nuestra ganancia (o pérdida) ¿cuánto debemos esperar ganar (o perder) ?

Aquí

$$X = \begin{cases} 1 & \text{si sale negro} & (\text{con probabilidad } \frac{18}{37}) \\ -1 & \text{si sale rojo o cero} & (\text{con probabilidad } \frac{19}{37}) \end{cases}$$

En consecuencia:

$$E[X] = \frac{18}{37} - \frac{19}{37} = \frac{-1}{37} = -0,027 \dots$$

Así pues, al jugar a la ruleta, debemos esperar perder un 27 por mil.

**Ejemplo:** Sea  $A$  un evento, consideramos la función  $I_A : \Omega \rightarrow \mathbb{R}$  definida por

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

Intuitivamente  $I_A$  vale 1 cuando el evento  $A$  ocurre, y 0 sino. Se denomina el indicador del evento  $A$ . (En la teoría de la medida, esta función se llama la función característica del conjunto  $A$  y se suele denotar por  $\chi_A$ , pero en la teoría de probabilidades la expresión “función característica” tiene un significado diferente).

$I_A$  es una variable aleatoria discreta pues su imagen consta de dos valores (0 y 1) y sus pre-ímagenes son  $X^{-1}(0) = \Omega - A$  y  $X^{-1}(1) = A$ , que son eventos.

La esperanza de  $I_A$  es:

$$E[I_A] = 0 \cdot P(\Omega - A) + 1 \cdot P(A) = P(A)$$

Es decir, la esperanza del indicador de un evento, coincide con su probabilidad.

**Ejemplo:**(un ejemplo de una variable aleatoria que toma infinitos valores). Consideremos el experimento consistente en arrojar infinitas veces una moneda (en forma independiente).

Como vimos anteriormente, podemos modelizar este experimento utilizando el espacio muestral  $\Omega = \{0, 1\}^{\mathbb{N}}$  de las sucesiones de ceros y unos, y representando cada realización del experimento por la sucesión  $\omega = (X_i)_{i \in \mathbb{N}}$  donde

$$X_i = \begin{cases} 1 & \text{si en la } i\text{-ésima realización del experimento sale cara} \\ 0 & \text{si en la } i\text{-ésima realización del experimento sale ceca} \end{cases}$$

Notemos que las  $X_i$  son variables aleatorias. Estamos interesados ahora en la siguiente variable aleatoria,  $T$  = cuántas tiradas tengo que esperar hasta que salga una cara por primera vez. Formalmente

$$T(\omega) = \min_{x_i=1} i$$

Hay un caso especial, que es cuando siempre sale ceca, esto es: ¿qué valor de  $T$  le asignaremos a la sucesión  $\omega = (0, 0, 0, \dots, 0, \dots)$  ? Lo razonable es poner:

$$T((0, 0, 0, \dots, 0, \dots)) = +\infty$$

Esto muestra que a veces resulta conveniente admitir variables aleatorias que pueden tomar el valor  $+\infty$  (o también  $-\infty$ ).

Ahora debemos calcular cuál es la distribución de probabilidades de  $T$ , es decir cuál es la probabilidad de que  $T$  tome cada valor.

$$P\{T = k\} = P\{X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1\}$$

y dado que los ensayos son independientes a este evento le asignamos la probabilidad dada por el producto de las probabilidades:

$$P\{T = k\} = P\{X_1 = 0\} \cdot P\{X_2 = 0\} \cdot \dots \cdot P\{X_{k-1} = 0\} \cdot P\{X_k = 1\} = \frac{1}{2^k}$$

Mientras que al evento “siempre sale ceca” le asignamos probabilidad 0,

$$P\{T = +\infty\} = P\{T((0, 0, 0, \dots, 0, \dots))\} = 0$$

Entonces la esperanza de  $T$  se calcularía por:

$$E[T] = \sum_{k=1}^{\infty} kP\{T = k\} + (+\infty) \cdot P\{T = +\infty\} = \sum_{k=1}^{\infty} \frac{k}{2^k} + (+\infty) \cdot 0$$

Hacemos la convención de que:

$$0 \cdot (+\infty) = 0$$

Entonces la esperanza de  $T$  es:

$$E[T] = \sum_{k=1}^{\infty} \frac{k}{2^k}$$

Utilizando la fórmula,

$$\sum_{k=1}^{\infty} kx^k = \frac{x}{(1-x)^2} \text{ si } |x| < 1$$

que se deduce de derivar la serie geométrica, con  $x = \frac{1}{2}$ , deducimos que  $E[T] = 2$ .

Así pues, en promedio, habrá que esperar dos tiradas, para que salga cara.

### 2.1.1. Esperanzas infinitas

A veces resulta conveniente admitir esperanzas infinitas. Si  $X \geq 0$  diremos que  $E[X] = +\infty$  si

$$\sum_i x_i P\{X = x_i\}$$

diverge.

Si  $X$  es una variable aleatoria discreta cualquiera, escribimos

$$X = X^+ - X^-$$

donde

$$X^+ = \begin{cases} X & \text{si } X \geq 0 \\ 0 & \text{si } X < 0 \end{cases}$$

y

$$X^- = \begin{cases} -X & \text{si } X < 0 \\ 0 & \text{si } X \geq 0 \end{cases}$$

Notamos que  $X^+$  y  $X^-$  son variables aleatorias no negativas.

Decimos que  $E[X] = +\infty$  si  $E[X^+] = +\infty$  y  $E[X^-] < \infty$ . Similarmente diremos que  $E[X] = -\infty$  si  $E[X^-] = +\infty$  y  $E[X^+] < \infty$ . Si  $E[X^+]$  y  $E[X^-]$  son ambas infinitas,  $E[X]$  no está definida.

### 2.1.2. Propiedades de la Esperanza

**Proposición 2.1.1 (linealidad de la esperanza)** 1. Si  $X, Y : \Omega \rightarrow \mathbb{R}$  son variables aleatorias discretas con esperanza finita, entonces

$$E[X + Y] = E[X] + E[Y]$$

2. Si  $X : \Omega \rightarrow \mathbb{R}$  es una variable aleatoria discreta con esperanza finita, entonces:

$$E[\lambda X] = \lambda E[X]$$

**Prueba:** Sean  $(x_i)$  los valores que toma  $X$ , e  $(y_j)$  los valores que toma  $Y$ : entonces

$$E[X] = \sum_i x_i P\{X = x_i\} = \sum_{i,j} x_i P\{X = x_i, Y = y_j\}$$



ya que

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\} \text{ (unión disjunta)}$$

y el reordenamiento de la serie está justificado por la convergencia absoluta, de la serie:

$$\sum_{i,j} x_i P\{X = x_i, Y = y_j\}$$

Similarmente,

$$E[Y] = \sum_j y_j P\{X = x_i\} = \sum_{i,j} y_j P\{X = x_i, Y = y_j\}$$

En consecuencia,

$$E[X] + E[Y] = \sum_{i,j} (x_i + y_j) P\{X = x_i, Y = y_j\}$$

Sea  $Z = X + Y$  y sean  $z_1, z_2, \dots, z_k, \dots$  los valores de  $Z$ . Entonces los  $z_k$  son exactamente los valores  $x_i + y_j$  (pero estos últimos pueden repetirse). Entonces,

$$E[Z] = \sum_k z_k P\{Z = z_k\} = \sum_k \sum_{i,j: x_i + y_j = z_k} z_k P\{X = x_i, Y = y_j\}$$

pues

$$\{Z = z_k\} = \bigcup_{i,j: x_i + y_j = z_k} \{X = x_i, Y = y_j\} \text{ (unión disjunta)}$$

Deducimos que

$$E[Z] = \sum_k (x_i + y_j) P\{X = x_i, Y = y_j\} = E[X] + E[Y]$$

Esto completa la prueba de la primera afirmación. En cuanto a la segunda afirmación,  $\lambda X$  es una variable aleatoria discreta que toma los valores  $\lambda x_i$ , por lo tanto:

$$E[\lambda X] = \sum_i \lambda x_i P\{\lambda X = \lambda x_i\} = \lambda \sum_i x_i P\{X = x_i\} = \lambda E[X]$$

□

**Proposición 2.1.2 (Monotonía de la esperanza)** 1. Si  $X$  es una variable aleatoria con esperanza finita y  $X \geq 0$  con probabilidad 1, entonces  $E[X] \geq 0$ .

2. Sean  $X$  e  $Y$  variables aleatorias con esperanza finita. Entonces, si  $X \leq Y$  con probabilidad 1, tenemos que  $E[X] \leq E[Y]$

3. Si  $X$  es una variable aleatoria acotada, entonces:

$$\inf_{\Omega} X \leq E[X] \leq \sup_{\Omega} X$$

4. Si  $X$  es una variable aleatoria discreta con esperanza finita, entonces:

$$|E[X]| \leq E[|X|]$$

**Proposición 2.1.3** Sean  $X$  una variable aleatoria discreta y  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ . Entonces

$$E[\varphi(X)] = \sum_i g(x_i)P\{X = x_i\}$$

siempre que esta serie sea absolutamente convergente.

**Prueba:** Sea  $Y = \varphi(X)$ , y sean  $(y_j)$  los valores de  $Y$ , entonces:

$$E[Y] = \sum_j y_j P\{Y = y_j\} = \sum_j y_j \sum_{i: \varphi(x_i)=y_j} P\{X = x_i\} = \sum_i \varphi(x_i) P\{X = x_i\}$$

(El reordenamiento se justifica usando la convergencia absoluta de la serie.)

□

Esta propiedad se puede generalizar a funciones de **vectores aleatorios**. Este concepto es una generalización natural del de variable aleatoria discreta:

**Definición 2.1.3** Un vector aleatorio discreto  $n$ -dimensional es una función  $X : \Omega \rightarrow \mathbb{R}^n$  tal que  $\text{Im}(X)$  sea finita o infinita numerable, y  $P\{X = x\}$  sea un evento  $x \in \mathbb{R}^n$ . Dar un vector aleatorio discreto  $X = (X_1, X_2, \dots, X_n)$  es equivalente a dar  $n$  variables aleatorias discretas  $x_1, x_2, \dots, x_n$

Con esta terminología tenemos [con la misma demostración de antes]:

**Proposición 2.1.4** Sean  $X$  un vector aleatorio  $n$ -dimensional y  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ , entonces

$$E[\varphi(X)] = \sum_i g(x_i)P\{X = x_i\}$$

donde  $x_i$  recorre la imagen de  $X$ , siempre que esta serie sea absolutamente convergente.

### 2.1.3. Independencia

**Definición 2.1.4** Sean  $X$  e  $Y$  dos variables aleatorias discretas definidas en un mismo espacio muestral. Diremos que son **independientes**, si para cada  $x_i, y_j$  los eventos  $\{X = x_i\}$  e  $\{Y = y_j\}$  son independientes, es decir de acuerdo a la definición de eventos independientes si,

$$P\{X = x_i, Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j\}$$

**Observación:** Remarcamos que esta definición solamente se aplica a variables discretas, cuando generalicemos esta noción a variables aleatorias no discretas, nos veremos en la necesidad de adoptar una definición diferente.

**Proposición 2.1.5** Si  $X$  e  $Y$  son variables aleatorias discretas independientes, y  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  son funciones, entonces  $Z = f(X)$  y  $W = g(Y)$  también son variables aleatorias discretas independientes.

**Prueba:** Calculemos la distribución conjunta de  $Z$  y  $W$ :

$$\begin{aligned} P\{Z = z, W = w\} &= \sum_{x,y: f(x)=z, g(y)=w} P\{X = x, Y = y\} \\ &= \sum_{x,y: f(x)=z, g(y)=w} P\{X = x\}P\{Y = y\} \\ &= \left( \sum_{x: f(x)=z} P\{X = x\} \right) \left( \sum_{y: g(y)=w} P\{Y = y\} \right) = P\{Z = z\}P\{W = w\} \end{aligned}$$

□

**Proposición 2.1.6** Si  $X$  e  $Y$  son variables aleatorias discretas independientes con esperanza finita, entonces:

$$E(XY) = E(X)E(Y)$$

**Prueba:**

$$E[XY] = \sum_{i,j} x_i y_j P\{X = x_i, Y = y_j\} = \sum_{i,j} x_i y_j P\{X = x_i\} P\{Y = y_j\}$$

$$\left( \sum_i x_i P\{X = x_i\} \right) \left( \sum_j y_j P\{Y = y_j\} \right) = E[X] E[Y]$$

**Observación:** En el caso en que  $X$  e  $Y$  toman infinitos valores, la aplicación de la propiedad distributiva, está justificada por el hecho de que las series que intervienen son absolutamente convergentes, por hipótesis.  $\square$

#### 2.1.4. Desigualdad de Jensen

**Definición 2.1.5** Sea  $f : \mathbb{R} \rightarrow \mathbb{R}$  una función. Diremos que  $f$  es convexa, si dados  $x, y \in \mathbb{R}$  y  $\alpha \in [0, 1]$ , se verifica que:

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

**Observación:** Si  $f$  es de clase  $C^2$ , entonces  $f$  es convexa, si y sólo si  $f''(x) \geq 0$ .

**Observación:** Una función convexa en  $\mathbb{R}$  es necesariamente continua. Además es posible probar que su derivada  $f'(x)$  existe salvo quizás para un conjunto a lo sumo numerable de valores de  $x$ , y que  $f'$  es creciente (ver [22], teorema 7.40).

**Ejercicio:** Una **combinación convexa** de los  $x_i$  es una combinación lineal

$$\sum_{i=1}^n \alpha_i x_i$$

en la que  $0 \leq \alpha_i$  y  $\sum_{i=1}^n \alpha_i = 1$ . Probar que si  $f : \mathbb{R} \rightarrow \mathbb{R}$  es una función convexa y  $\sum_{i=1}^n \alpha_i x_i$  es una combinación convexa, entonces:

$$f\left(\sum_{i=1}^n \alpha_i x_i\right) \leq \sum_{i=1}^n \alpha_i f(x_i)$$

**Proposición 2.1.7 (Desigualdad de Jensen)** Si  $g : \mathbb{R} \rightarrow \mathbb{R}$  es una función convexa, entonces:

$$g(E[X]) \leq E[g(X)]$$

en los siguientes casos: si  $X$  es no negativa y  $g(x) \geq 0$  para  $x \geq 0$ , o si  $X$  y  $g$  son arbitrarias y  $E(|g(X)|) < \infty$ .

**Prueba:** Hagamos la demostración primero, en el caso que  $X$  toma sólo finitos valores. Sea  $p_i = P\{X = x_i\}$ . Entonces

$$E[X] = \sum_{i=1}^n p_i x_i$$

es una combinación convexa de los valores de  $X$ . Como  $X$  es una función convexa,

$$g(E[X]) = g\left(\sum_{i=1}^n p_i x_i\right) \leq \sum_{i=1}^n p_i g(x_i) = E[g(X)]$$

Si  $X$  toma un número numerable de valores,  $x_i$  con probabilidades  $p_i$ , entonces hacemos lo siguiente: para cada  $n \in \mathbb{N}$  definamos,

$$s_n = \sum_{i=1}^n p_i$$

y notamos que

$$\sum_{i=1}^n \frac{p_i}{s_n} x_i$$

es una combinación convexa. Entonces, como  $g$  es convexa:

$$g\left(\sum_{i=1}^n \frac{p_i}{s_n} x_i\right) \leq \sum_{i=1}^n \frac{p_i}{s_n} g(x_i)$$

Cuando  $n \rightarrow +\infty$ , tenemos que  $s_n \rightarrow 1$ . Entonces, utilizando la continuidad de  $g$ , obtenemos que:

$$g(E[X]) = g\left(\sum_{i=1}^{\infty} p_i x_i\right) \leq \sum_{i=1}^{\infty} p_i g(x_i) = E[g(X)]$$

□

**Ejemplo:**  $f(x) = |x|^p$  es una función convexa si  $p \geq 1$ . En consecuencia, en este caso:

$$|E[X]|^p \leq E[|X|^p]$$

## 2.2. Momentos - Varianza

**Definición 2.2.1** Sea  $X$  una variable aleatoria (discreta). Definimos el  $k$ -ésimo momento de  $X$  entorno de  $b$  como  $E[(X - b)^k]$ . El  $k$ -ésimo momento absoluto entorno de  $b$  se define como  $E[|X - b|^k]$ .

### Algunas observaciones:

1. Si  $E[|X|^t] < \infty$  y  $0 \leq s \leq t$ , entonces  $E[|X|^s] < +\infty$ . En efecto según la desigualdad de Jensen,

$$(E[|X|^s])^p \leq E[|X|^t]$$

donde  $p = \frac{t}{s} \geq 1$ . Es más, vemos que:

2.  $E[|X|^p]^{1/p}$  es una función creciente de  $p$ .
3. Si  $E[|X|^p] < +\infty$  y  $E[|Y|^p] < +\infty$  entonces  $E[|X + Y|^p]^{1/p} < +\infty$

### Prueba:

$$\begin{aligned} |X + Y|^p &\leq (|X| + |Y|)^p = (2 \max(|X|, |Y|))^p \\ &\leq 2^p \max(|X|^p, |Y|^p) \leq 2^p(|X|^p + |Y|^p) \end{aligned}$$

Por lo tanto,

$$E[|X + Y|^p] \leq 2^p(E[|X|^p] + E[|Y|^p]) < +\infty$$

□

4. En consecuencia, el conjunto

$$L_d^p(\Omega, \mathcal{E}, P) = \{X : \Omega \rightarrow \overline{\mathbb{R}} \text{ variable aleatoria discreta} : E[|X|^p] < +\infty\}$$

(siendo  $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ ) es un espacio vectorial.

5. Si  $p \geq 1$ , es posible probar que

$$\|X\|_p = E[|X|^p]^{\frac{1}{p}}$$

es una norma en dicho espacio.

En lo sucesivo, nos van a interesar especialmente dos clases  $L^p$ :

$$L_d^1(\Omega) = \{X : \Omega \rightarrow \overline{\mathbb{R}} : \text{variable aleatoria (discreta) con esperanza finita}\}$$

$$L_d^2(\Omega) = \{X : \Omega \rightarrow \overline{\mathbb{R}} : \text{variable aleatoria (discreta) con segundo momento finito}\}$$

**Ejemplo:** Notemos que  $L_d^2 \subset L_d^1$  por lo anterior. Veamos un ejemplo de una variable aleatoria que está en  $L_d^1$  pero no en  $L_d^2$ : Consideramos un espacio muestral numerable

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$$

en el que

$$P\{\omega_n\} = \frac{1}{n(n+1)}$$

Verifiquemos que esta asignación efectivamente define una distribución de probabilidades en  $\Omega$ :

$$\sum_{n=1}^{\infty} P\{\omega_n\} = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \sum_{n=1}^{\infty} \left[ \frac{1}{n} - \frac{1}{n+1} \right] = 1$$

(serie telescópica). Definamos la variable aleatoria  $X : \Omega \rightarrow \mathbb{R}$ , dada por  $X(\omega_n) = \sqrt{n}$ . Entonces,

$$E(X) = \sum_{n=1}^{\infty} X(\omega_n) P\{\omega_n\} = \sum_{n=1}^{\infty} \frac{\sqrt{n}}{n(n+1)} \leq \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} < +\infty$$

pero

$$E(X^2) = \sum_{n=1}^{\infty} X(\omega_n)^2 P\{\omega_n\} = \sum_{n=1}^{\infty} \frac{n}{n(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n+1} = +\infty$$

**Definición 2.2.2** *El segundo momento de  $X$  entorno de su media se llama la **varianza** (o **variancia**<sup>1</sup>) de  $X$ , es decir:*

$$\text{Var}(X) = E[(X - E(X))^2]$$

*Por lo anterior  $\text{Var}(X) < +\infty$  si y sólo si el segundo momento de  $X$  es finito, es decir si  $X \in L_d^2$ .*

---

<sup>1</sup>Según el diccionario de la RAE, ambas gráficas son aceptables.

**Ejemplo:** Sea  $A$  un evento con probabilidad  $p$ , e  $I_A$  su indicador. Calculemos su varianza. Ya vimos que:

$$E[I_A] = P(A) = p$$

En consecuencia:

$$\text{Var}(I_A) = E[(I_A - p)^2]$$

La distribución de probabilidades de  $(I_A - p)^2$  es:

$$(I_A - p)^2 = \begin{cases} (1-p)^2 & \text{si ocurre } A \quad (\text{con probabilidad } p) \\ p^2 & \text{si no ocurre } A \quad (\text{con probabilidad } q = 1-p) \end{cases}$$

En consecuencia,

$$\text{Var}(I_A) = (1-p)^2 p + p^2(1-p) = p - p^2 = pq$$

**Proposición 2.2.1** 1. Si  $X = c$  es constante, entonces  $\text{Var}(X) = 0$ .

2.  $\text{Var}(aX + b) = a^2 \text{Var}(X)$ .

### 2.2.1. Desigualdades de Tchesbychev y de Markov

**Proposición 2.2.2 (Desigualdad básica)** Sea  $X$  una variable aleatoria no negativa, entonces

$$P(X \geq \lambda) \leq \frac{1}{\lambda} E(X) \quad (2.1)$$

**Prueba:** Sea  $A = \{\omega \in \Omega : X(\omega) \geq \lambda\}$ . Entonces  $X \geq \lambda I_A$ , en consecuencia:  $E[X] \geq \lambda E[I_A] = \lambda P(A)$   $\square$

**Proposición 2.2.3 (Desigualdad de Markov)** Si  $X$  es una variable aleatoria (discreta) entonces

$$P\{|X| \geq \lambda\} \leq \frac{1}{\lambda^p} E(|X|^p)$$

**Prueba:** Si cambiamos  $X$  por  $|X|^p$  en la desigualdad anterior tenemos que:

$$P\{|X| \geq \lambda\} = P\{|X|^p \geq \lambda^p\} \leq \frac{1}{\lambda^p} E(|X|^p)$$

$\square$



**Proposición 2.2.4 (desigualdad de Tchebyshev clásica)** Sea  $X$  una variable (discreta) entonces

$$P\{|X - E(X)| > \lambda\} \leq \frac{\text{Var}(X)}{\lambda^2}$$

**Prueba:** Usamos la desigualdad anterior con  $p = 2$  y cambiamos  $X$  por  $X - E(X)$ .  $\square$

Intuitivamente, la desigualdad de Tchebyshev dice que la varianza de la variable  $X$  nos da una estimación de la probabilidad de que  $X$  tome valores alejados de su esperanza. Si  $\text{Var}(X)$  es pequeña, entonces es poco probable que  $X$  tome un valor alejado de  $E(X)$ .

## 2.2.2. Covarianza

**Definición 2.2.3** Sean  $X$  e  $Y$  dos variables aleatorias. Definimos la **covarianza** o **covariancia** de  $X$  e  $Y$  por

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

**Observación:** Si  $X$  e  $Y$  son variables aleatorias independientes entonces  $\text{Cov}(X, Y) = 0$ . La recíproca no es cierta, como muestra el siguiente ejemplo:

**Ejemplo** (Barry James, pag. 130) Sean  $X$  e  $Y$  dos variables aleatorias con valores  $-1, 0, 1$  con la siguiente función de probabilidad conjunta:

	-1	0	1
-1	$\frac{1}{5}$	0	$\frac{1}{5}$
0	0	$\frac{1}{5}$	0
1	$\frac{1}{5}$	0	$\frac{1}{5}$

entonces  $E[XY] = E[X] = E[Y] = 0$ , pero  $X$  e  $Y$  no son independientes pues

$$P\{X = 0, Y = 0\} = \frac{1}{5} \neq \frac{1}{25} = \frac{1}{5} \frac{1}{5} = P\{X = 0\}P\{Y = 0\}$$

**Definición 2.2.4** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias discretas. Diremos que no están correlacionadas si  $\text{Cov}(X_i, X_j) = 0$  para  $i \neq j$ .

**Proposición 2.2.5** Si  $X$  e  $Y$  son variables aleatorias (discretas) con segundo momento finito:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

**Prueba:**

$$\begin{aligned} \text{Var}(X+Y) &= E[(X+Y-E[X]-E[Y])^2] = E[((X-E(X))+(Y-E(Y)))^2] = \\ &= E[(X-E(X))^2] + E[(Y-E(Y))^2] + 2E[(X-E(X))(Y-E(Y))] = \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \end{aligned}$$

□

**Corolario 2.2.1** Si  $X_1, X_2, \dots, X_n$  son variables aleatorias (discretas) con segundo momento finito, que no están correlacionadas, entonces

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i)$$

**Dem:** Sale de la fórmula anterior por inducción.

## 2.3. Ensayos de Bernoulli - La Distribución Binomial

En esta sección presentaremos un esquema conceptual, que fue introducido por Bernoulli, y que es útil para modelizar muchas situaciones.

El esquema de ensayos de Bernoulli consiste en lo siguiente: Consideramos un experimento aleatorio con dos resultados, que convencionalmente llamamos “éxito” y “fracaso”. Supongamos que la probabilidad de obtener un éxito en una realización del experimento es  $p \in [0, 1]$ , y naturalmente la de obtener un fracaso será  $q = 1 - p$

Imaginemos que repetimos el experimento una cantidad  $n$  de veces, de manera independiente. Para modelizar este experimento consideramos el espacio muestral  $\Omega = \{0, 1\}^n$  compuesto por las  $n$ -uplas de números 0 y 1 con la siguiente interpretación: codificaremos una realización del experimento por una  $n$ -upla  $\omega = (x_1, x_2, \dots, x_n) \in \Omega$  de modo que:

$$x_i = \begin{cases} 1 & \text{si la } i\text{-ésima realización del experimento fue un “éxito”} \\ 0 & \text{si la } i\text{-ésima realización del experimento fue un “fracaso”} \end{cases}$$

Es un espacio muestral finito, con cardinal  $2^n$ . Notemos que las funciones  $X_i : \Omega \rightarrow \mathbb{R}$  (proyecciones) dadas por  $X_i(\omega) = x_i$  son variables aleatorias.

¿De qué modo asignaremos las probabilidades en este espacio?. Puesto que consideramos que los ensayos son independientes, a una determinada  $n$ -upla  $\omega = (x_1, x_2, \dots, x_n)$  le asignamos la probabilidad

$$P\omega = P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = \prod_{i=1}^n P\{X_i = x_i\}$$

Ahora la probabilidad de que  $X_i = x_i$  es  $p$  si  $x_i = 1$  (es un éxito) y  $q$  si  $x_i = 0$  (es un fracaso). De modo que

$$P\{\omega\} = p^k q^{n-k}$$

donde  $k = \sum_{i=1}^n x_i$  es el número de éxitos que ocurren en esa realización del experimento. Notemos que esta forma de asignar las probabilidades dice precisamente que las  $X_i$  son variables aleatorias independientes.

Por otra parte, notemos que si definimos  $S_n : \Omega \rightarrow \mathbb{R}$  como el número de éxitos en los  $n$  ensayos de Bernoulli, es una variable aleatoria (en la notación anterior  $S_n(\omega) = k$ ). Tenemos que:

$$S_n = X_1 + X_2 + \dots + X_n \quad (2.2)$$

Nos interesa cuál es la distribución de probabilidades de  $S_n$ , es decir queremos determinar para cada  $k$  (con  $0 \leq k \leq n$ ) cuál es la probabilidad de que  $S_n$  tome el valor  $k$ .

Observamos que el evento  $\{S_n = k\} = \{\omega \in \Omega : S_n(\omega) = k\}$  se compone de las  $n$ -uplas que tienen exactamente  $k$  éxitos y  $n - k$  fracasos, y que hay exactamente

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

de tales  $n$ -uplas, y cada una de ellas tiene probabilidad  $p^k q^{n-k}$ . En consecuencia la probabilidad del evento  $S_n = k$  será

$$P\{S_n = k\} = \binom{n}{k} p^k q^{n-k}$$

Esta distribución de probabilidades se conoce como la **distribución binomial**, dado que viene dada por los términos del desarrollo del binomio de Newton:

$$(p + q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}$$

**Definición 2.3.1** Sea  $X : \Omega \rightarrow \mathbb{N}_0$  una variable aleatoria con valores enteros. Diremos que  $X$  tiene **distribución binomial** si:

$$P\{X = k\} = b(k, n, p) = \binom{n}{k} p^k q^{n-k}$$

y  $P\{X = k\} = 0$  si  $k \notin \{0, 1, \dots, n\}$ . **Notación:**  $X \sim Bi(n, p)$

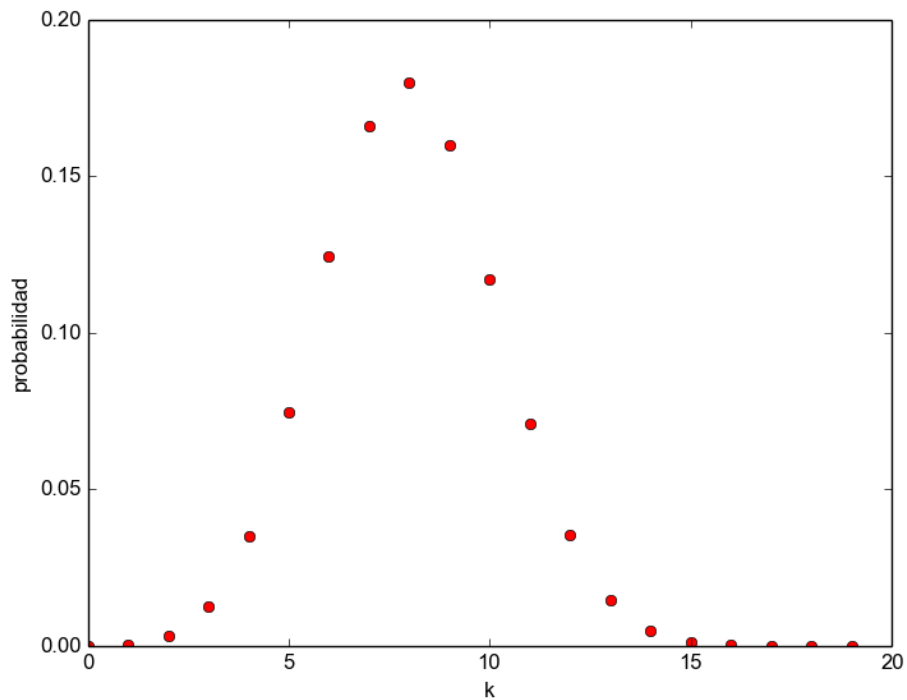


Figura 2.1: La distribución binomial con  $n = 20$  y  $p = 0,4$

Necesitamos calcular la esperanza y la varianza de  $S_n$ . Para ello utilizamos la representación (2.2) de  $S_n$  como suma de las variables  $X_i$ . Notamos que

cada  $X_i$  es de hecho el indicador del evento “ocurre un éxito en la  $i$ -ésima realización del experimento”. En consecuencia:

$$E[X_i] = p, \quad \text{Var}(X_i) = pq$$

Por la linealidad de la esperanza,

$$E[S_n] = np$$

y por otro lado, como las  $X_i$  son variables aleatorias independientes, también se verifica que

$$\text{Var}(S_n) = npq$$

## 2.4. El método de las funciones generatrices

En algunas situaciones, el método que expondremos a continuación resulta de utilidad para operar con distribuciones de probabilidad discretas. Lo usaremos para obtener de otro modo la distribución binomial, y calcular su esperanza y su varianza.

**Definición 2.4.1** Sea  $X : \Omega \rightarrow \mathbb{N}_0$  una variable aleatoria que toma valores enteros. Llamamos función generatriz de la distribución de probabilidades de  $X$  a

$$g_X(z) = \sum_{k=0}^{\infty} P\{X = k\} z^k \quad (z \in \mathbb{C})$$

suponiendo que esta serie tenga un radio de convergencia  $r_X > 0$  (entonces convergerá absolutamente en  $|z| < r_X$ ). **Observación:** La notación  $g_X$  que usaremos en estas notas, no es una notación estándar.<sup>2</sup>

Notemos que si  $0 < |z| < r_X$ ,

$$g_X(z) = E[z^X]$$

(Cuando  $z = 0$  esta fórmula es problemática si  $X$  toma el valor 0, pues  $0^0$  no está definido. Se tiene que  $g_X(0) = P\{X = 0\}$ )

---

<sup>2</sup>En clase y en versiones anteriores de estas notas utilicé la notación  $f_X$ , pero decidí cambiarla por  $g_X$ , ya que en la teoría de probabilidades la notación  $f_X$  suele utilizarse para la densidad de probabilidad para variables aleatorias absolutamente continuas.

**Observación:** En virtud de la unicidad del desarrollo en serie de potencias, la distribución de probabilidades de una variable aleatoria entera está unívocamente determinada por su función generatriz.

**Proposición 2.4.1** Si  $X$  e  $Y$  son variables aleatorias independientes, entonces:

$$g_{X+Y}(z) = g_X(z) \cdot g_Y(z)$$

para  $|z| < \min(r_X, r_Y)$ .

**Prueba:** Como  $X$  e  $Y$  son independientes,  $z^X$  y  $z^Y$  son independientes. En consecuencia, si  $0 < |z| < r_X$ :

$$g_{X+Y}(z) = E[z^{X+Y}] = E[z^X \cdot z^Y] = E[z^X] \cdot E[z^Y] = g_X(z) \cdot g_Y(z)$$

Cuando  $z = 0$ ,

$$\begin{aligned} g_{X+Y}(0) &= P\{X + Y = 0\} = P\{X = 0, Y = 0\} \\ &= P\{X = 0\} \cdot P\{Y = 0\} = g_X(0) \cdot g_Y(0) \end{aligned}$$

□

Esta proposición puede generalizarse sin dificultad a varias variables independientes: si  $X_1, X_2, \dots, X_n$  son independientes, entonces

$$g_{X_1+X_2+\dots+X_n}(z) = g_{X_1}(z) \cdot g_{X_2}(z) \cdots g_{X_n}(z)$$

**Aplicación:** Otra prueba de que el número de éxitos  $S_n$  en  $n$  ensayos de Bernoulli tiene distribución binomial.

Utilicemos la representación (2.2) de  $S_n$  como suma de  $n$  variables independientes que valen 1 con probabilidad  $p$  y 0 con probabilidad  $q = 1 - p$ . La función generatriz de cada  $X_i$  es:

$$g_{X_i}(z) = pz + q$$

y como  $S_n$  es la suma de las  $X_i$  y son independientes:

$$g_{S_n}(z) = (pz + q)^n = \sum_{k=0}^n \binom{n}{k} p^k z^k q^{n-k} \quad (2.3)$$

Notemos que la probabilidad de que  $S_n$  tome el valor  $k$  viene dado por el coeficiente de  $z^k$  en  $g_{S_n}$ . En consecuencia:

$$P\{S_n = k\} = \binom{n}{k} p^k q^{n-k} \quad (0 \leq k \leq n)$$

Las funciones generatrices pueden usarse para calcular esperanzas y varianzas (y más generalmente momentos) de variables aleatorias enteras:

**Proposición 2.4.2** *Si la serie que define la función generatriz  $g_X$  tiene radio de convergencia  $r_X > 1$ , entonces*

$$E(X) = g'_X(1)$$

$$Var(X) = g''_X(1) + g'_X(1) - g'_X(1)^2$$

**Prueba:** Como las series de potencia pueden derivarse término a término en el interior de su disco de convergencia, tenemos que:

$$g'_X(z) = \sum_{k=1}^{\infty} k P\{X = k\} z^{k-1}$$

con convergencia absoluta si  $|z| < r_X$ . En particular si  $z = 1$ ,

$$g'_X(1) = \sum_{k=1}^{\infty} k P\{X = k\} = E[X]$$

Volviendo a derivar tenemos que

$$g''_X(z) = \sum_{k=2}^{\infty} k(k-1) P\{X = k\} z^{k-2}$$

con convergencia absoluta si  $|z| < r_X$ , y haciendo  $z = 1$ ,

$$g''_X(1) = \sum_{k=2}^{\infty} k(k-1) P\{X = k\} = E[X(X-1)] = E[X^2] - E[X]$$

Luego

$$Var(X) = E[X^2] - E[X]^2 = g''_X(1) + g'_X(1) - g'_X(1)^2$$

□

**Aplicación:** Cálculo de la esperanza y la varianza de la distribución binomial (de otra manera).

Sea como antes  $S_n$  el número de éxitos en  $n$  ensayos de Bernoulli. Como vimos antes  $g_{S_n}(z) = (pz + q)^n$ . En consecuencia, como

$$g'_{S_n}(z) = n(pz + q)^{n-1}p$$

$$g''_{S_n}(z) = n(n-1)(pz + q)^{n-2}p^2$$

deducimos que

$$E[S_n] = np$$

y que:

$$\text{Var}(S_n) = n(n-1)p^2 + np - n^2p^2 = -np^2 + np = np(1-p) = npq$$

**Ejercicio:** Si  $X \sim \text{Bi}(n, p)$  e  $Y \sim \text{Bi}(m, p)$  y son independientes, entonces  $X + Y \sim \text{Bi}(n + m, p)$ .

### 2.4.1. El teorema de Bernoulli

Imaginemos que realizamos una sucesión ilimitada de ensayos de Bernoulli. Sea  $f_n = \frac{S_n}{n}$  la frecuencia de éxitos que obtenemos en los  $n$  primeros ensayos. Es intuitivamente razonable que conforme  $n \rightarrow +\infty$ ,  $f_n$  tienda a la probabilidad  $p$  de obtener un éxito.

Nos gustaría transformar esta idea intuitiva en un teorema matemático. El siguiente teorema debido a Jacques Bernoulli, y publicado en 1713 en su libro *Ars Conjectandi*, constituye una formalización de esta idea:

**Teorema 2.4.1 (Teorema de J. Bernoulli)** *Sea  $f_n$  la frecuencia de éxitos en los  $n$  primeros ensayos de una sucesión ilimitada de ensayos de Bernoulli. Entonces dado cualquier  $\delta > 0$ ,*

$$P\{|f_n - p| > \delta\} \rightarrow 0 \text{ conforme } n \rightarrow \infty$$

**Prueba:** Notemos que  $E[f_n] = p$ . Luego, por la desigualdad de Tchebyshev,

$$P\{|f_n - p| > \delta\} \leq \frac{\text{Var}(f_n)}{\delta^2}$$



pero

$$\text{Var}(f_n) = \text{Var}\left(\frac{S_n}{n}\right) = \frac{pq}{n}$$

En consecuencia:

$$P\{|f_n - p| > \delta\} \leq \frac{pq}{n\delta^2} \rightarrow 0 \text{ cuando } n \rightarrow +\infty \quad (2.4)$$

□

Una generalización del teorema de Bernoulli (que se prueba con el mismo argumento) es la siguiente, conocida (al igual que a veces el teorema de Bernoulli) como la ley débil de los grandes números:

**Teorema 2.4.2 (Ley débil de los grandes números - caso de variancia finita)**

*Sean  $X_1, X_2, \dots, X_n, \dots$  una secuencia infinita de variables aleatorias independientes e idénticamente distribuidas, con*

$$E[X_i] = \mu$$

$$\text{Var}(X_i) = \sigma^2 < +\infty$$

*Entonces si llamamos*

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

*y tomamos cualquier  $\delta > 0$ , tenemos que*

$$P\{|\bar{X}_n - \mu| > \delta\} \rightarrow 0 \text{ cuando } n \rightarrow +\infty$$

**Prueba:** Por linealidad de la esperanza,  $E[\bar{X}_n] = \mu$ , y por otro lado

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

ya que las  $X_i$  son independientes. La desigualdad de Tchebyshev, dice entonces que:

$$P\{|\bar{X}_n - \mu| > \delta\} \leq \frac{\sigma^2}{n\delta^2} \rightarrow 0 \text{ cuando } n \rightarrow +\infty$$

□

**Algunas observaciones sobre el teorema de Bernoulli:**

- Si bien la prueba del teorema de Bernoulli, resulta muy sencilla hoy en día, J. Bernoulli dice en su libro que estuvo pensando en este teorema durante más de 20 años, lo cuál muestra que el resultado no es para nada trivial.
- Como todo teorema matemático, el teorema de Bernoulli no afirma nada sobre la realidad, es solamente una afirmación sobre el modelo matemático

(La cuestión de la validez práctica de un modelo matemático sólo se puede decidir sobre bases empíricas, es decir contrastándolo con la experiencia). Sin embargo, podemos interpretarlo como una muestra de la consistencia interna de nuestro modelo matemático.

- La ley débil de los grandes números recibe este nombre, porque, como veremos más adelante, existe otro teorema conocido como la ley fuerte de los grandes números, que afirma que en realidad  $S_n \rightarrow p$  (o  $\bar{X}_n \rightarrow \mu$ ) con probabilidad 1.

(Pero notemos que para darle sentido a la afirmación de que  $S_n \rightarrow p$  con probabilidad 1, debemos asignar probabilidades a secuencias de infinitos ensayos de Bernoulli, como en el experimento que consideramos anteriormente de arrojar infinitas veces una moneda. Esto introduce ciertas dificultades relacionadas con la teoría de la medida, como por ejemplo que ya no podremos asignarle probabilidad a cualquier parte del espacio muestral  $\Omega$ , y que por lo tanto debemos restringir el dominio de la función probabilidad a una  $\sigma$ -álgebra de eventos.)

## 2.5. Ley débil de los grandes números: caso general

La hipótesis de que las variables aleatorias  $X_i$  tengan varianza finita no es realmente necesaria para la validez de la ley débil de los grandes números, pudiéndose probar para variables que tengan solamente esperanza finita, por medio de un método de truncamiento. Sin embargo, para fijar ideas, hemos optado por enunciarla y demostrarla primero en este caso en el que la demostración resulta más sencilla. Veamos ahora el caso general:

**Teorema 2.5.1 (Ley débil de los grandes números - caso general)** Sean  $X_1, X_2, \dots, X_n, \dots$  una secuencia infinita de variables aleatorias independientes e idénticamente distribuidas, con

$$E[X_i] = \mu < +\infty$$

Entonces si llamamos

$$S_n = X_1 + X_2 + \dots + X_n$$

y tomamos cualquier  $\delta > 0$ , tenemos que

$$P \left\{ \left| \frac{S_n}{n} - \mu \right| > \delta \right\} \rightarrow 0 \text{ cuando } n \rightarrow +\infty$$

**Prueba:** Para simplificar la notación, notemos que podemos asumir sin pérdida de generalidad, que

$$E(X_i) = 0 \quad \forall i$$

(cambiando si no  $X_i$  por  $X_i - \mu$ ).

La demostración en el caso de variancia infinita, se basa en el **método de truncamiento**, que consiste en descomponer  $X_i$  como suma de dos variables aleatorias. Para cada  $k = 1, 2, \dots, n$ , escribimos:

$$X_k = U_{n,k} + V_{n,k} \quad (k = 1, 2, \dots, n) \quad (2.5)$$

donde

$$U_{n,k} = \begin{cases} X_k & \text{si } |X_k| \leq \lambda n \\ 0 & \text{si } |X_k| > \lambda n \end{cases}$$

y

$$V_{n,k} = \begin{cases} 0 & \text{si } |X_k| \leq \lambda n \\ X_k & \text{si } |X_k| > \lambda n \end{cases}$$

donde  $\delta > 0$  es una constante que especificaremos después. Y pongamos:

$$U_n = U_{n,1} + U_{n,2} + \dots + U_{n,n}$$

$$V_n = V_{n,1} + V_{n,2} + \dots + V_{n,n}$$

De la desigualdad triangular  $|S_n| \leq |U_n| + |V_n|$ , y de la subaditividad de la probabilidad, deducimos que:

$$\begin{aligned} & P\{|S_n| > \delta n\} \\ & \leq P\{|U_n| > \delta n/2\} + P\{|V_n| > \delta n/2\} \end{aligned} \quad (2.6)$$

Entonces hemos de probar que cada una de las probabilidades del segundo miembro tiende a cero cuando  $n \rightarrow +\infty$ .

Comencemos acotando:

$$P\{|U_n| > \delta n/2\}$$

Observemos que las variables  $U_{n,k}$  están acotadas ( $|U_{n,k}| \leq \lambda n$ ) y en consecuencia tienen segundo momento finito. Más explícitamente, si llamemos  $a = E(|X_i|)$ , tenemos que

$$E(U_{n,k}^2) \leq n\lambda a$$

En consecuencia las  $U_{k,n}$  tienen variancia finita:

$$\text{Var}(U_{n,k}) \leq E(U_{n,k}^2) \leq n\lambda a$$

Por otra parte las  $U_{n,k}$  son variables independientes e idénticamente distribuidas (pues  $U_{n,k}$  es función de  $X_k$ , y las  $X_k$  eran independientes e idénticamente distribuidas). En consecuencia:

$$\text{Var}(U_n) = \text{Var}(U_{n,1} + U_{n,2} + \dots + U_{n,n}) = \sum_{k=1}^n \text{Var}(U_{n,k}) \leq n^2 \lambda a$$

Además de la definición de las  $U_{n,k}$  deducimos que

$$E(U_{n,k}) = E(U_{n,1}) = \sum_{i:|x_i| > \lambda n} x_i P\{X_1 = x_i\} \rightarrow E(X_1) = 0$$

conforme  $n \rightarrow +\infty$ . En consecuencia para  $n \geq n_0(\varepsilon)$  será:

$$E(U_n^2) = \text{Var}(U_n) + E(U_n)^2 < 2\lambda n^2 a$$

y entonces por la desigualdad de Tchebyshev, tenemos que:

$$P\{|U_n| > \delta n/2\} < \frac{8a\lambda}{\delta^2} < \frac{\varepsilon}{2}$$

si elegimos  $\lambda$  suficientemente pequeño.

En cuanto al segundo término: obviamente

$$P\{|V_n| > \delta n/2\} \leq P\{V_{n,1} + V_{n,2} + \dots + V_{n,n} \neq 0\}$$

y como

$$\{V_{n,1} + V_{n,2} + \dots + V_{n,n} \neq 0\} \subset \bigcup_{k=1}^n \{V_{n,k} \neq 0\}$$

tenemos que:

$$P\{|V_n| > \delta n/2\} \leq \sum_{k=1}^n P\{V_{n,k} \neq 0\} = nP\{V_1 \neq 0\}$$

ya que las  $V_k$  tienen todas la misma distribución de probabilidades. Pero por definición de  $V_1$ , esto dice que

$$P\{|V_n| > \delta n/2\} \leq nP\{|X_1| > \lambda n\} = n \sum_{i: |x_i| > \lambda n} P\{X_1 = x_i\}$$

donde  $\text{Im}(X_1) = \{x_1, x_2, \dots, x_n \dots\}$ . Deducimos que:

$$P\{|V_n| > \delta n/2\} \leq \frac{1}{\lambda} \sum_{|x_i| > \lambda n} |x_i| P\{X_1 = x_i\}$$

Dado entonces cualquier  $\varepsilon > 0$ , como la esperanza de  $X_1$  es finita por hipótesis, deducimos que si elegimos  $n$  suficientemente grande, digamos si  $n \geq n_0(\varepsilon)$ , tendremos que:

$$P\{|V_n| > \delta n/2\} < \frac{\varepsilon}{2}$$

(ya que las colas de una serie convergente tienden a cero).

Por (2.6), deducimos que:

$$P\{|S_n| > \delta n\} \leq \varepsilon$$

si  $n \geq n_0(\varepsilon)$ .

□

## 2.6. Polinomios de Bernstein: Una prueba del teorema de Weierstrass

En esta sección expondremos una prueba del teorema de Weierstrass sobre aproximación a funciones continuas por polinomios, debida a S.N. Bernstein:

**Teorema 2.6.1 (Weierstrass)** *Sea  $f \in C[0, 1]$  una función continua  $f : [0, 1] \rightarrow \mathbb{R}$ , entonces existe una sucesión de polinomios  $P_n(t)$  tal que  $P_n(t) \rightarrow f(t)$  uniformemente para  $t \in [0, 1]$ .*

En un lenguaje más moderno, el teorema de Weierstrass dice que los polinomios son densos en el espacio  $C[0, 1]$  de las funciones continuas (con la norma del supremo).

La prueba de S.N. Bernstein (1912) de este teorema, consiste en utilizar la distribución binomial, para construir explícitamente una sucesión de polinomios que converge uniformemente a  $f$ .

Veamos primero la idea intuitiva de la demostración: sea  $p \in [0, 1]$  y sea como antes  $S_n$  el número de éxitos en  $n$  ensayos de Bernoulli con probabilidad  $p$ . La ley de los grandes números afirma que:

$$\frac{S_n}{n} \rightarrow p \text{ (en probabilidad)}$$

y como  $f$  es continua es razonable esperar que:

$$f\left(\frac{S_n}{n}\right) \rightarrow f(p)$$

(De vuelta, esto no es estrictamente cierto para toda sucesión de ensayos de Bernoulli, pero sí vale en probabilidad.) Por lo que esperamos que:

$$E\left[f\left(\frac{S_n}{n}\right)\right] \rightarrow E[f(p)] = f(p)$$

Notemos que:

$$\begin{aligned} B_n(p) &= E\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^n f\left(\frac{k}{n}\right) b(k, n, p) \\ &= \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) p^k (1-p)^{n-k} \end{aligned}$$

es un polinomio en la variable  $p$ . Se lo denomina el  $n$ -ésimo **polinomio de Bernstein**.

La demostración de S.N. Bernstein, consiste en probar que  $B_n(p) \rightarrow f(p)$  uniformemente para  $p \in [0, 1]$  (Los argumentos anteriores no constituyen una prueba rigurosa, pero explican intuitivamente por qué esta afirmación es cierta).

De hecho, la demostración de esta afirmación se basa en argumentos muy similares a los que nos llevaron a la prueba del teorema de Bernoulli

Para la prueba del teorema de Weierstrass utilizaremos, dos propiedades claves de las funciones continuas en un intervalo cerrado de la recta, a saber:

1. Una función continua en un intervalo cerrado de la recta, es acotada: existe una constante  $M > 0$  tal que:

$$|f(p)| \leq M \quad \forall p \in [0, 1]$$

2. Una función continua en un intervalo cerrado de la recta, es uniformemente continua: dado  $\varepsilon > 0$  existe  $\delta > 0$  tal que si  $x, y \in [0, 1]$  y si  $|x - y| \leq \delta$ , entonces  $|f(x) - f(y)| < \varepsilon$ .

Necesitaremos una acotación de las colas de la distribución binomial: de acuerdo a la desigualdad (2.4):

$$P \left\{ \left| \frac{S_n}{n} - p \right| > \delta \right\} \leq \frac{pq}{n\delta^2} \leq \frac{1}{4n\delta^2}$$

ya que:

$$pq = p(1 - p) \leq \frac{1}{4} \quad \forall p \in [0, 1]$$

Más explícitamente podemos escribir esto como:

$$\sum_{|k/n - p| > \delta} b(k, n, p) = \sum_{|k/n - p| > \delta} P\{S_n = k\} \leq \frac{1}{4n\delta^2}$$

Queremos acotar la diferencia:

$$B_n(p) - f(p) = \sum_{k=0}^n \left[ f\left(\frac{k}{n}\right) b(k, n, p) \right] - f(p) = \sum_{k=0}^n \left[ f\left(\frac{k}{n}\right) - f(p) \right] b(k, n, p)$$

pues

$$\sum_{k=0}^n b(k, n, p) = 1$$

(¡Es una distribución de probabilidades!). En consecuencia,

$$|B_n(p) - f(p)| \leq \sum_{k=0}^n \left| f\left(\frac{k}{n}\right) - f(p) \right| b(k, n, p)$$

En esta suma separamos dos partes, la suma sobre los  $k$  donde  $|k/n - p| \leq \delta$  (con el  $\delta$  dado por la continuidad uniforme), y la parte donde  $|k/n - p| > \delta$ .

La primer parte la acotamos, fácilmente:

$$\sum_{k: |k/n - p| \leq \delta} \left| f\left(\frac{k}{n}\right) - f(p) \right| b(k, n, p) \leq \sum_{k: |k/n - p| \leq \delta} \varepsilon b(k, n, p) \leq \varepsilon$$

pues los  $b(k, n, p)$  suman 1.

La otra parte de la suma la acotamos usando nuestra estimación de las colas de la distribución binomial:<sup>3</sup>

$$\sum_{k: |k/n - p| > \delta} \left| f\left(\frac{k}{n}\right) - f(p) \right| b(k, n, p) \leq 2M \sum_{|k/n - p| > \delta} b(k, n, p) < \frac{2M}{4n\delta^2} < \varepsilon$$

si  $n \geq n_0(\varepsilon)$ . En consecuencia,  $|B_n(p) - f(p)| < 2\varepsilon$  si  $n \geq n_0(\varepsilon)$ , para todo  $p \in [0, 1]$ . Esto concluye la prueba del teorema de Weierstrass.

## 2.7. La aproximación de Poisson a la distribución binomial

La aproximación de Poisson es una aproximación de la distribución binomial para el caso en que  $k$  es pequeño comparado con  $n$  y  $p$  es también pequeño pero  $\lambda = np$  es moderado.

---

<sup>3</sup>Si en lugar de utilizar la desigualdad de Tchebyshev, utilizamos otra herramienta de probabilidades conocida como la “teoría de grandes desviaciones”, es posible obtener una acotación más precisa del error de aproximar  $f$  por  $B_n$ . Ver el artículo [14] citado en la bibliografía



Empecemos desarrollando el combinatorio que aparece en la distribución binomial:

$$b(k, n, p) = \binom{n}{k} p^k q^{n-k} = \frac{n(n-1)(n-2) \dots (n-k+1)}{k!} p^k (1-p)^{n-k} =$$

Notamos que en el desarrollo del combinatorio, hay  $k$  factores en el numerador. Multiplicando y dividiendo por  $n^k$  queda:

$$b(k, n, p) = \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \cdot \frac{(np)^k}{k!} (1-p)^{n-k}$$

Pongamos  $\lambda = np$ , entonces

$$b(k, n, p) = \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Como

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

deducimos que si  $k$  es pequeño en comparación con  $n$ , entonces

$$b(k, n, p) \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

Como formalización de esta idea, obtenemos el siguiente teorema:

**Teorema 2.7.1 (Teorema de Poisson)** *Si  $k$  está fijo, y  $n \rightarrow +\infty$  de modo que  $\lambda = np$  permanece fijo, entonces:*

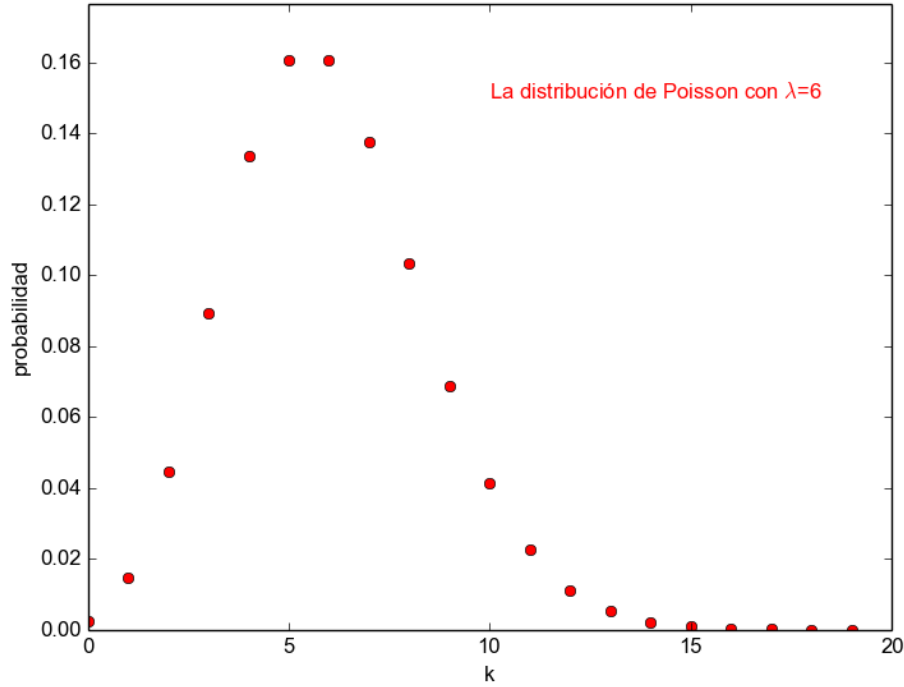
$$\lim_{n \rightarrow +\infty} b(k, n, p) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Lo que obtuvimos en el límite, es otra distribución de probabilidades que se utiliza con frecuencia y se conoce como **distribución de Poisson**:

**Definición 2.7.1** *Sea  $X : \Omega \rightarrow \mathbb{N}_0$  una variable aleatoria entera. Diremos que  $X$  tiene distribución de Poisson de parámetro  $\lambda > 0$ , si*

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$$

**Notación:**  $X \sim \mathcal{P}(\lambda)$ .

Figura 2.2: La distribución de Poisson con  $\lambda = 6$ .

Hay que verificar que efectivamente tenemos una distribución de probabilidades, es decir que:

$$\sum_{k=0}^{\infty} P\{X = k\} = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1$$

pero esto es inmediato, considerando el desarrollo en serie de  $e^{\lambda}$ .

Vamos a calcular ahora la esperanza y la varianza de la distribución de Poisson: para ello utilizaremos el método de las funciones generatrices, que desarrollamos anteriormente: Si  $X$  tiene distribución de Poisson de parámetro  $\lambda$ , la función generatriz de su distribución de probabilidades es:

$$g_X(z) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k z^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)} \quad (2.7)$$

Tenemos que

$$g'_X(z) = \lambda e^{\lambda(z-1)}$$

$$g''_X(z) = \lambda^2 e^{\lambda(z-1)}$$

En consecuencia por la proposición 2.4.2, deducimos que:

$$E(X) = g'_X(1) = \lambda$$

$$\text{Var}(X) = g''(1) + g'(1) - g'(1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Otra consecuencia es la siguiente:

**Proposición 2.7.1** Si  $X \sim \mathcal{P}(\lambda_1)$ ,  $Y \sim \mathcal{P}(\lambda_2)$  y son independientes, entonces  $X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2)$ .

**Prueba:** Por la proposición 2.4.1,

$$g_{X+Y}(z) = g_X(z) \cdot g_Y(z) = e^{\lambda_1(z-1)} e^{\lambda_2(z-1)} = e^{(\lambda_1+\lambda_2)(z-1)}$$

En consecuencia,  $X+Y \sim \mathcal{P}(\lambda_1+\lambda_2)$ , ya que la distribución de probabilidades de  $X + Y$  está determinada por su función generatriz.  $\square$

## 2.8. Otras distribuciones relacionadas con los ensayos de Bernoulli

### Distribución Geométrica

Supongamos que realizamos una secuencia infinita de ensayos de Bernoulli, con probabilidad de éxito  $p$ . Sea  $T_1$  la cantidad de ensayos que tenemos que realizar hasta obtener el primer éxito (esto generaliza el ejemplo de la página 21 que corresponde al caso  $p = 1/2$ ).

Entonces, si  $T_1 = k$  significa que los primeros  $k-1$  ensayos fueron fracasos y el  $k$ -ésimo fue un éxito, y como los ensayos son independientes obtenemos como antes que:

$$P\{T_1 = k\} = q^{k-1}p = (1-p)^{k-1}p$$

(y  $T_1 = +\infty$  con probabilidad cero). Esta distribución se conoce con el nombre de **distribución geométrica** de parámetro  $p$ .

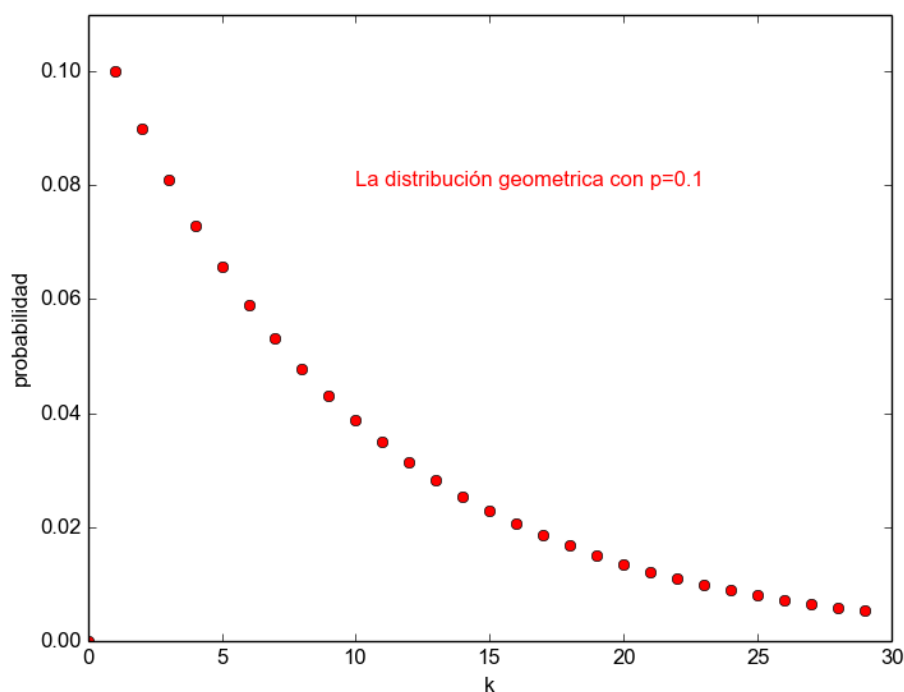


Figura 2.3: La distribución geométrica con  $p = 0,1$ .

**Notación:**  $X \sim \mathcal{G}(p)$  significa que  $X$  se distribuye con la distribución geométrica de parámetro  $p$ .

Con una cuenta análoga a la que hicimos antes para el caso  $p = 1/2$  podemos probar que  $E[X] = \frac{1}{p}$ .

La función generatriz de la distribución de probabilidades de  $X$  se obtiene justamente sumando una serie geométrica:

$$g_X(z) = \sum_{k=1}^{\infty} q^{k-1} p z^k = \frac{pz}{1 - qz} \text{ si } |z| < \frac{1}{q} \quad (2.8)$$

## Distribución binomial negativa

Más generalmente podemos considerar la variable  $T_r$  definida como el número de ensayos que tenemos que realizar hasta obtener  $r$  éxitos. Queremos

calcular la distribución de  $T_r$ :

Para ello notamos que,

$$T_r = E_1 + E_2 + \dots + E_r$$

donde  $E_1 = T_1$  y  $E_j$  = número de ensayos que debemos realizar después del éxito  $j - 1$  para obtener el siguiente éxito. Notamos que las variables  $E_j$  son independientes (ya que el tiempo que tenemos que esperar para obtener el siguiente éxito después de obtener  $j - 1$  éxitos no depende de cuánto tardamos en obtener  $j$  éxitos) y que por la discusión anterior, cada  $E_j$  tiene distribución geométrica de parámetro  $p$ .

Podemos entonces calcular la distribución de  $T_r$  utilizando el método de las funciones generatrices, ya que por la independencia de las  $E_j$ , la función generatriz de la distribución de probabilidades de  $T_r$  es:

$$g_{T_r}(z) = g_{E_1}(z)g_{E_2}(z) \cdots g_{E_r}(z) = \left( \frac{pz}{1 - qz} \right)^r$$

Por lo tanto, utilizando el desarrollo del binomio  $(1 - qz)^{-r}$  y haciendo el cambio de índice  $k = j + r$ ,

$$g_{T_r}(z) = (pz)^r \sum_{j=0}^{\infty} \binom{-r}{j} (-qz)^j = \sum_{k=r}^{\infty} \binom{-r}{k-r} p^r (-q)^{k-r} z^k$$

En consecuencia,

$$P\{T_r = k\} = \binom{-r}{k-r} p^r (-q)^{k-r} \quad (k = r, r+1, \dots)$$

**Notación:**  $X \sim BN(r, p)$

**Falta:** distribucion hipergeometrica

## Distribución Multinomial

Es una generalización de la distribución binomial donde consideramos experimentos con muchos varios posibles, en lugar de un experimento con sólo dos resultados.

Consideramos un experimento con  $N$  resultados posibles, y supongamos que la probabilidad de que ocurra el  $i$ -ésimo resultado en una realización del experimento es  $p_i$ , de modo que:

$$\sum_{i=1}^N p_i = 1$$

Supongamos que repetimos el experimento  $n$  veces en condiciones independientes, y llamemos  $X_i$  a la cantidad de veces que ocurre el  $i$ -ésimo resultado, de modo que:

$$X_1 + X_2 + \dots + X_N = n$$

Entonces, la distribución de probabilidades conjunta de las  $X_i$  viene dada por:

$$P\{X_1 = k_1, X_2 = k_2, \dots, X_N = k_N\} = \frac{n!}{k_1! k_2 \dots k_N!} p_1^{k_1} p_2^{k_2} \dots p_N^{k_N} \quad (2.9)$$

si  $k_1 + k_2 + \dots + k_N = N$  (y cero en caso contrario). Notamos que  $X = (X_1, X_2, \dots, X_N)$  es un vector aleatorio  $N$ -dimensional.

**Notación:**  $X \sim \mathcal{M}(n, p_1, p_2, \dots, p_N)$

Esta distribución recibe este nombre, debido a su relación con el desarrollo multinomial:

$$(x_1 + x_2 + \dots + x_N)^n = \sum_{\substack{k_N: k_1 + k_2 + \dots + k_N = n \\ 0 \leq k_i \leq n}} \frac{n!}{k_1! k_2 \dots k_N!} x_1^{k_1} x_2^{k_2} \dots x_N^{k_N}$$

(Tomando  $x_i = p_i$  se ve que las probabilidades en (2.9) suman 1, por lo que se trata efectivamente de una distribución de probabilidades).

Una propiedad interesante de la distribución multinomial es que las distribuciones de cada una de las  $X_i$  por separado (distribuciones marginales) son binomiales:

**Proposición 2.8.1** Si  $X \sim \mathcal{M}(n, p_1, p_2, \dots, p_N)$ , entonces

$$X_i \sim Bi(n, p_i) \quad 0 \leq i \leq N$$

**Prueba:** Por simetría, basta verlo para la distribución de  $X_1$ . Si  $0 \leq k_1 \leq n$ ,

$$\begin{aligned}
 P\{X_1 = k_1\} &= \sum_{\substack{k_N:k_2+\dots+k_N=n-k_1 \\ 0 \leq k_i \leq n}} P\{X_1 = k_1, X_2 = k_2, \dots, X_N = k_N\} \\
 &= \sum_{\substack{k_N:k_2+\dots+k_N=n-k_1 \\ 0 \leq k_i \leq n}} \frac{n!}{k_1!k_2 \dots k_N!} p_1^{k_1} p_2^{k_2} \dots p_N^{k_N} \\
 &= \frac{n!}{k_1!(n-k_1)!} p_1^{k_1} \sum_{\substack{k_N:k_2+\dots+k_N=n-k_1 \\ 0 \leq k_i \leq n}} \frac{(n-k_1)!}{k_2! \dots k_N!} p_2^{k_2} \dots p_N^{k_N} \\
 &= \frac{n!}{k_1!(n-k_1)!} p_1^{k_1} (p_2 + p_3 + \dots + p_N)^{n-k_1} \\
 &= \frac{n!}{k_1!(n-k_1)!} p_1^{k_1} (1-p_1)^{n-k_1}
 \end{aligned}$$

luego

$$X_1 \sim \text{Bi}(n, p_1)$$

□

# Capítulo 3

## Distribuciones Continuas

### 3.1. Variables aleatorias continuas

En este capítulo estudiaremos variables aleatorias no discretas, en particular variables continuas. La idea básica es la misma que antes: una variable aleatoria es un número asociado al resultado de un experimento aleatorio, por lo que será una función  $X$  definida sobre el espacio muestral  $\Omega$ . Nuevamente, hay un requerimiento técnico, derivado del hecho de que en general no resulta posible asignar probabilidades a todas las partes de  $\Omega$ ; a saber que podamos calcular las probabilidades asociadas a dicha función. En el caso de variables discretas, pedíamos que estuvieran definidas las probabilidades de que  $X$  tome un determinado valor. En el caso de variables no discretas, esto no será suficiente: requeriremos que podamos calcular la probabilidad de que el valor de  $X$  caiga en un intervalo dado de la recta.

**Definición 3.1.1** Sea  $(\Omega, \mathcal{E}, P)$  un espacio de probabilidad. Una variable aleatoria será una función  $X : \Omega \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ , con la siguiente propiedad: para cualquier intervalo de la recta  $(a, b]$  ( $a, b \in \overline{\mathbb{R}}$ ) la preimagen  $X^{-1}(a, b] = \{\omega \in \Omega : a < X(\omega) \leq b\}$  pertenece a  $\mathcal{E}$ , es decir está definida la probabilidad  $P(X^{-1}(a, b]) = P\{a < X \leq b\}$  de que  $X$  tome un valor entre  $a$  y  $b$ .

**Observación:** En análisis real, el concepto análogo es el de función medible (ver apéndice B).

**Definición 3.1.2** Diremos que la variable  $X$  es (absolutamente) continua si



existe una función integrable<sup>1</sup> no negativa  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  tal que

$$P\{a < X \leq b\} = \int_a^b f(x) dx$$

La función  $f$  debe verificar que:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Se dice que  $f$  se distribuye según la densidad de probabilidades  $f(x)$  (o que  $f$  es la densidad de probabilidad de  $X$ ). A veces se nota,  $X \sim f(x)$ .

**Definición 3.1.3** Si  $X : \Omega \rightarrow \overline{\mathbb{R}}$  es una variable aleatoria, su función de distribución<sup>2</sup> será la función  $F : \mathbb{R} \rightarrow \mathbb{R}$  dada por:

$$F_X(x) = P\{X \leq x\}$$

Si  $X$  es absolutamente continua, y se distribuye según la densidad  $f(x)$  tendremos:

$$F_X(x) = \int_{-\infty}^x f(t) dt$$

**Ejemplo 3.1.1 Variables aleatorias discretas:** Sea  $X$  una variable aleatoria discreta que toma una sucesión a lo sumo numerable de valores  $(x_i)$ . Entonces,  $X$  es una variable aleatoria de acuerdo a nuestra nueva definición (es decir, realmente estamos extendiendo el concepto) ya que:

$$\{\omega \in \Omega : a < X(\omega) \leq b\} = \bigcup_{a < x_i \leq b} \{\omega \in \Omega : X(\omega) = x_i\}$$

Por definición de variable aleatoria discreta,  $\{\omega \in \Omega : X(\omega) = x_i\} \in \mathcal{E}$ , y como siendo la clase  $\mathcal{E}$  una  $\sigma$ -álgebra, es cerrada por uniones numerables, deducimos que  $\{\omega \in \Omega : a < X(\omega) \leq b\} \in \mathcal{E}$ .

---

<sup>1</sup>Quiere decir que en algún sentido sea posible calcular la integral de  $f$  sobre un intervalo de la recta. Los que no conozcan la teoría de la integral de Lebesgue pueden pensar integrable Riemann, los que cursaron análisis real pueden pensar que es integrable Lebesgue

<sup>2</sup>También llamada a veces función de distribución acumulada en la literatura

La función de distribución de  $X$  viene dada por la función “en escalera”

$$F_X(x) = \sum_{x_i < x} P\{X = x_i\}$$

que tiene un salto de magnitud  $p_i = P\{X = x_i\}$  en el punto  $x_i$  (y que es constante en cada intervalo entre dos  $x_i$ ).

**Ejemplo 3.1.2** Volvamos a considerar el experimento de elegir un número real en el intervalo  $[0, 1]$  con distribución uniforme. Sea  $X$  el número obtenido.

Que lo elegimos con distribución uniforme significa que para cualquier intervalo  $I \subset [0, 1]$ , postulamos que

$$P\{X \in I\} = |I|$$

donde  $|I|$  representa la medida del intervalo.

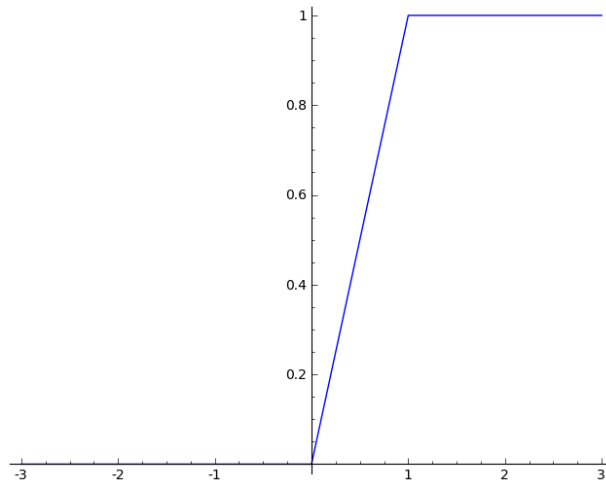


Figura 3.1: La función de distribución de una variable aleatoria con distribución uniforme en el intervalo  $[0, 1]$ .

Entonces la función de distribución de  $X$  viene dada por:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

$X$  es una variable absolutamente continua con densidad,

$$f_X(x) = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 0 & \text{si } x \notin [0, 1] \end{cases}$$

**Notación:** Notamos  $X$  se distribuye uniformemente en el intervalo  $[0, 1]$  del siguiente modo:  $X \sim \mathcal{U}(0, 1)$ .

Más generalmente si  $[a, b]$  es un intervalo de la recta, decimos que  $X$  tiene distribución uniforme en el intervalo  $[a, b]$  (Notación:  $X \sim \mathcal{U}(a, b)$ ) si para cualquier intervalo  $I \subset [a, b]$  la probabilidad de que  $X$  pertenezca a  $I$  es proporcional a la medida de  $I$ , es decir:

$$P\{X \in I\} = \frac{|I|}{b-a}$$

En este caso, la función de distribución es:

$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ (x-a)/(b-a) & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$$

y la función de densidad es,

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{si } x \notin [a, b] \end{cases}$$

**Ejemplo 3.1.3** Decimos que  $X$  tiene **distribución normal**, y lo notaremos  $X \sim N(\mu, \sigma^2)$ , si su función de densidad de probabilidad viene dada por:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

donde  $\mu, \sigma$  son dos parámetros reales con  $\sigma > 0$ . El caso  $\mu = 0, \sigma = 1$ , es decir  $N(0, 1)$ , se conoce como **distribución normal estándar**.

Si  $X \sim N(0, 1)$ , la función de distribución de  $X$  será la función:

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(t-\mu)^2/(2\sigma^2)} dt \quad (3.1)$$

Veremos en el capítulo 5 que la distribución normal resulta útil por ejemplo para aproximar la distribución binomial, del número  $S_n$  de éxitos en  $n$

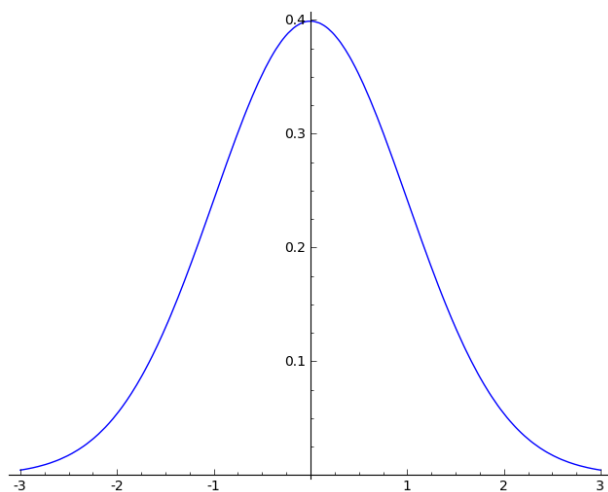


Figura 3.2: La densidad normal estándar

*ensayos de Bernoulli, cuando el número de ensayos es grande. Más generalmente, se puede usar para aproximar la suma de muchas variables aleatorias independientes cada una de las cuáles hace una pequeña contribución a la varianza de la suma (Este es el contenido del Teorema del Límite Central que veremos en dicho capítulo). Como consecuencia, esta distribución juega un papel central en estadística. Se conoce también como distribución de Laplace o de Gauss.*

### 3.1.1. Propiedades de las funciones de distribución

El siguiente lema nos dice que propiedades tienen las funciones de distribución:

**Lema 3.1.1** *Sea  $X : \Omega \rightarrow \overline{\mathbb{R}}$  una variable aleatoria y  $F = F_X$  su función de distribución. Entonces  $F$  tiene las siguientes propiedades:*

- i)  $0 \leq F(x) \leq 1$  y  $F$  es creciente.*
- ii)  $F$  es continua por la derecha.*
- iii)  $F(x_0) - \lim_{x \rightarrow x_0^-} F(x) = P\{X = x_0\}$  En particular,  $F$  es continua en  $x = x_0$  si y sólo si  $P\{X = x_0\} = 0$ .*

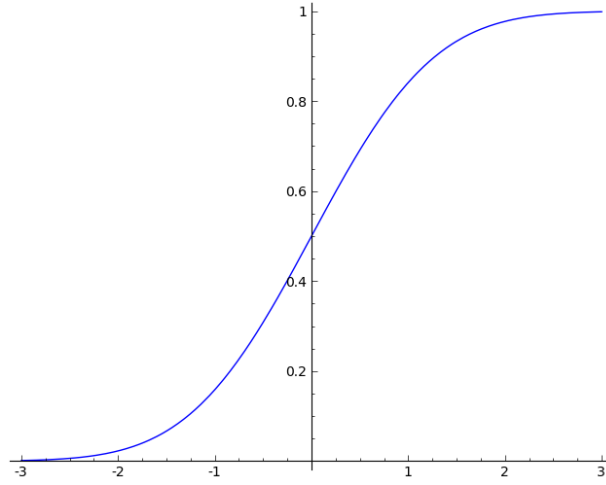


Figura 3.3: La función de distribución de una variable con distribución normal estándar

iv) Si  $X$  es finita con probabilidad 1 (o sea  $P\{X = \pm\infty\} = 0$ ), entonces:

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

$$\lim_{x \rightarrow +\infty} F(x) = 1$$

**Observación 3.1.1** La propiedad iii) significa que los saltos de una función de distribución nos indican cuando la probabilidad se concentra en un punto dado  $x_0$ , y que la magnitud del salto nos dice cuanta probabilidad se concentra en ese punto  $x_0$ .

**Prueba:** i) Que  $0 \leq F(x) \leq 1$  es obvio por ser  $F(x)$  una probabilidad. Si  $x_1 \leq x_2$  tenemos que:  $\{X \leq x_1\} \subset \{X \leq x_2\}$ , y en consecuencia  $F(x_1) \leq F(x_2)$ .

ii) Sea  $x_0 \in \mathbb{R}$  y consideremos una sucesión decreciente  $(x_n)_{n \in \mathbb{N}} > x_0$  que converja a  $x_0$ . Entonces,

$$\{X \leq x_0\} = \bigcap_{n \in \mathbb{N}} \{X \leq x_n\}$$

Es la intersección de una familia decreciente numerable de eventos. Entonces, por las propiedades de continuidad de la probabilidad:

$$P\{X \leq x_0\} = \lim_{n \rightarrow +\infty} P\{X \leq x_n\}$$

Es decir que:

$$F(x_0) = \lim_{n \rightarrow +\infty} F(x_n)$$

Y como esto vale para toda sucesión  $(x_n) > x_0$  decreciente, que converja a  $x_0$  deducimos que:

$$F(x_0) = \lim_{x \rightarrow x_0^+} F(x)$$

Es decir, que  $F$  es continua por la derecha.

iii) Análogamente, sea  $x_0 \in \mathbb{R}$  y tomemos una sucesión creciente  $(x_n)_{n \in \mathbb{N}} < x_0$  que converja a  $x_0$ . Ahora tenemos que,

$$\{X < x_0\} = \bigcup_{n \in \mathbb{N}} \{X \leq x_n\}$$

Entonces, aplicando nuevamente las propiedades de continuidad de la probabilidad:

$$P\{X < x_0\} = \lim_{n \rightarrow +\infty} P\{X \leq x_n\}$$

Es decir que:

$$P\{x < x_0\} = \lim_{n \rightarrow +\infty} F(x_n)$$

Como esto vale para toda sucesión  $(x_n)_{n \in \mathbb{N}} < x_0$  que converja a  $x_0$ , deducimos que:

$$\lim_{x \rightarrow x_0^-} F(x) = P\{X < x_0\}$$

En consecuencia,

$$F(x_0) - \lim_{x \rightarrow x_0^-} F(x) = P\{X \leq x_0\} - P\{X < x_0\} = P\{X = x_0\}$$

En particular,  $F$  será continua por la izquierda en  $x_0$  (y por lo tanto continua en  $x_0$ ) si y sólo si  $P\{X = x_0\} = 0$ .

iv) Es análoga tomando sucesiones crecientes (decrecientes) tales que  $x_n \rightarrow \pm\infty$ .  $\square$

**Observación 3.1.2** *Es posible probar que estas propiedades caracterizan a las funciones de distribución, en el sentido de que cualquier función  $F$  con estas propiedades será la función de distribución de alguna variable aleatoria  $X$ . (ver la observación 3.5.2)*

**Observación 3.1.3** *Es útil observar que como consecuencia de estas propiedades, los puntos de discontinuidad de una función de distribución son a lo sumo numerables. (Esto se prueba observando que para cada  $k$ , sólo puede haber a lo sumo  $k$  puntos donde el salto de la función de distribución sea mayor que  $1/k$ ).*

## 3.2. La integral de Riemann-Stieltjes y la definición de esperanza

La integral de Riemann-Stieltjes es una generalización de la integral de Riemann. Stieltjes observó que cualquier función creciente  $F : \mathbb{R} \rightarrow \mathbb{R}$  origina una noción de medida de intervalos,

$$m_F((a, b]) = F(b) - F(a)$$

Para las aplicaciones a la teoría de probabilidades, nos interesa el caso en que  $F$  es la función de distribución de una variable aleatoria.

Stieltjes definió la integral

$$\int_a^b \varphi(x) dF(x) \tag{3.2}$$

generalizando la definición de la integral de Riemann de la siguiente manera: sea

$$\pi : a = x_0 < x_1 < x_2 < \dots < x_n = b$$

una partición del intervalo  $[a, b]$  (Dar una partición no es otra cosa que elegir finitos puntos del intervalo en orden creciente) y elijamos puntos intermedios  $\xi_i \in (x_i, x_{i+1}]$  en cada intervalito de la partición (En realidad, estamos trabajando con particiones con puntos marcados, pero no lo haremos explícito en la notación). Consideramos entonces las sumas de Riemann-Stieltjes

$$S_\pi(\varphi, F) = \sum_{i=0}^{n-1} \varphi(\xi)(F(x_{i+1}) - F(x_i))$$

**Definición 3.2.1** Diremos que la integral (3.2) existe y toma el valor  $I \in \mathbb{R}$  si las sumas  $S_\pi(\varphi, F)$  tienden al valor  $I$  cuando la norma

$$|\pi| = \max_{0 \leq i \leq n-1} |x_{i+1} - x_i|$$

de la partición  $\pi$  tiende a cero, es decir si dado  $\varepsilon > 0$ , existe  $\delta > 0$  tal que  $|I - S_\pi(\varphi, F)| < \varepsilon$  para toda partición  $\pi$  con  $|\pi| < \delta$ .

Observemos que si  $F(x) = x$ , la integral de Riemann-Stieltjes se reduce a la integral de Riemann usual.

Algunas propiedades de la integral que son consecuencias más o menos inmediatas de las definiciones:

**Lema 3.2.1 (Linealidad)** 1. Si  $\int_a^b \varphi_1(x) dF(x)$  y  $\int_a^b \varphi_2(x) dF(x)$  existen, y  $\varphi = \lambda_1 \varphi_1 + \lambda_2 \varphi_2$  entonces,  $\int_a^b \varphi(x) dF(x)$  también existe, y tenemos que:

$$\int_a^b \varphi(x) dF(x) = \lambda_1 \int_a^b \varphi_1(x) dF(x) + \lambda_2 \int_a^b \varphi_2(x) dF(x)$$

2. Si  $\int_a^b \varphi(x) dF_1(x)$  y  $\int_a^b \varphi(x) dF_2(x)$  existen, y  $F = \lambda_1 F_1 + \lambda_2 F_2$  con  $\lambda_1, \lambda_2 \geq 0$ , entonces  $\int_a^b \varphi(x) dF$  existe, y vale que:

$$\int_a^b \varphi(x) dF(x) = \lambda_1 \int_a^b \varphi(x) dF_1(x) + \lambda_2 \int_a^b \varphi(x) dF_2(x)$$

**Lema 3.2.2 (Aditividad respecto al intervalo)** Sea  $c \in [a, b]$ . Si  $\int_a^b \varphi(x) dF(x)$  existe, entonces también existen  $\int_a^c \varphi(x) dF(x)$  y  $\int_c^b \varphi(x) dF(x)$  y se verifica:

$$\int_a^b \varphi(x) dF(x) = \int_a^c \varphi(x) dF(x) + \int_c^b \varphi(x) dF(x)$$

El siguiente teorema nos da una condición que permite garantizar la existencia de integrales de Riemann-Stieltjes:



**Teorema 3.2.1** Si  $\varphi : [a, b] \rightarrow \mathbb{R}$  es continua, y si  $F : [a, b] \rightarrow \mathbb{R}$  es creciente, entonces la integral de Riemann-Stieltjes

$$\int_a^b \varphi(x) dF(x)$$

existe

Para la prueba, véase el apéndice D.

El siguiente lema, nos dice cómo acotar una integral de Stieltjes:

**Lema 3.2.3** Supongamos que  $\int_a^b \varphi(x) dF(x)$  existe, siendo  $\varphi$  una función acotada en  $[a, b]$  y  $F$  creciente en  $[a, b]$ . Entonces,

$$\left| \int_a^b \varphi(x) dF(x) \right| \leq \left( \sup_{x \in [a, b]} |\varphi(x)| \right) (F(b) - F(a))$$

Obs: Más generalmente se puede demostrar que la integral de Riemann-Stieltjes

$$\int_a^b \varphi(x) dF(x)$$

existe si  $\varphi(x)$  es continua en  $[a, b]$  y  $F$  es de variación acotada (ya que toda función de variación acotada se puede escribir como diferencia de dos funciones crecientes). En este caso, la integral se acota del siguiente modo:

$$\left| \int_a^b \varphi(x) dF(x) \right| \leq \left( \sup_{x \in [a, b]} |\varphi(x)| \right) V_a^b(F)$$

### 3.3. La definición de Esperanza

Veamos como se aplican las integrales de Riemann-Stieltjes a la teoría de probabilidades. Para ello consideremos una variable aleatoria,  $X : \Omega \rightarrow \mathbb{R}$  no discreta y veamos como podríamos definir la esperanza de  $X$ . Supongamos por simplicidad primero que  $X$  toma valores en un cierto intervalo  $[a, b]$  de la recta.

Entonces, si tomamos una partición  $\pi$  del intervalo  $[a, b]$  (con puntos marcados como antes), podemos considerar una variable aleatoria  $X_\pi$  que aproxima a  $X$  del siguiente modo:

$$X_\pi = \xi_i \text{ si } X \in (x_i, x_{i+1}]$$

Entonces:

$$\begin{aligned} E[X_\pi] &= \sum_{i=0}^{n-1} \xi_i P\{X_\pi = \xi_i\} = \sum_{i=0}^{n-1} \xi_i P\{\xi_i < X \leq \xi_{i+1}\} \\ &= \sum_{i=0}^{n-1} \xi_i (F(x_{i+1}) - F(x_i)) \end{aligned}$$

es exactamente la suma de Riemann-Stieltjes  $S_\pi(\varphi, F)$  con  $\varphi(x) = x$ .

Entonces cuando la norma de la partición tiende a cero,  $E[X_\pi]$  tiende a la integral

$$\int_a^b x dF(x)$$

(que de acuerdo al teorema anterior siempre existe), y podemos aceptar la siguiente definición:

**Definición 3.3.1** Sea  $X$  una variable aleatoria que tome valores en un intervalo  $[a, b]$  de la recta, entonces la esperanza de  $X$  es la integral de Riemann-Stieltjes

$$E[X] = \int_a^b x dF(x) \quad (3.3)$$

siendo  $F = F_X$  su función de distribución.

Más generalmente podemos considerar la variable aleatoria  $\varphi(X)$  siendo  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  una función continua, entonces:

$$E[\varphi(X_\pi)] = \sum_{i=0}^{n-1} \varphi(\xi_i) P\{X_\pi = \xi_i\} = \sum_{i=0}^{n-1} \varphi(\xi_i) P\{\xi_i < X \leq \xi_{i+1}\}$$

$$= \sum_{i=0}^{n-1} \varphi(\xi_i)(F(x_{i+1}) - F(x_i))$$

Entonces, cuando la norma de la partición  $\pi$  tiende a cero, estas sumas convergen a la integral:

$$\int_a^b \varphi(x) dF(x)$$

y conjeturamos que

$$E[\varphi(X)] = \int_a^b \varphi(x) dF(x) \quad (3.4)$$

para toda función continua  $\varphi \in C[a, b]$  (aunque demostrar esto directamente de la definición es bastante complicado).

En particular,

$$\text{Var}(X) = E[(X - \mu)^2] = \int_a^b (x - \mu)^2 dF(x)$$

siendo  $\mu = E[X]$ .

Veamos algunos ejemplos, para familiarizarnos con esta idea:

**Ejemplo 1: Variables aleatorias discretas** Si  $X$  es una variable aleatoria discreta que solamente toma finitos valores  $v_1, v_2, \dots, v_n$  y miramos la suma  $S_\pi$  correspondiente a una partición  $\pi$  vemos que solamente contribuyen a la suma aquellos términos para los cuales  $v_j \in (x_i, x_{i+1}]$  para algún  $j$ . Refinando si es preciso la partición, podemos suponer que cada intervalito  $(x_i, x_{i+1}]$  contiene un único valor  $v_j$  a lo sumo, y en ese caso elegimos  $\xi_i = v_j$  (sino la elección de  $\xi_i$  es irrelevante). Entonces la suma de Riemann-Stieltjes para (3.3) es:

$$S_\pi = \sum_{i: v_j \in (x_i, x_{i+1}]} v_j (F(x_{i+1}) - F(x_i)) = \sum_{i: v_j \in (x_i, x_{i+1}]} v_j P\{X = v_j\}$$

que es el valor de  $E[X]$  conforme a la definición de esperanza para variables aleatorias discretas.

Otra manera de pensar esta cuenta es la siguiente: para  $x_0 \in \mathbb{R}$ , definimos la *función de Heaviside*:

$$H_{x_0}(x) = \begin{cases} 0 & \text{si } x < x_0 \\ 1 & \text{si } x \geq x_0 \end{cases}$$

$H_{x_0}$  es la función de distribución de una variable aleatoria que toma el valor  $x_0$  con probabilidad 1. Entonces tenemos:

**Lema 3.3.1** Si  $x_0 \in [a, b]$  y  $\varphi \in C[a, b]$ , entonces:

$$\int_a^b \varphi(x) dH_{x_0} = \varphi(x_0)$$

**Prueba:** En  $S_\pi(\varphi, F)$  el único término no nulo corresponde al intervalo  $[x_i, x_{i+1}]$  que contiene a  $x_0$ , en consecuencia:

$$S_\pi(\varphi, F) = \varphi(x_i)$$

y cuando  $|\pi| \rightarrow 0$ ,  $\varphi(x_i) \rightarrow \varphi(x_0)$ , por la continuidad de  $\varphi$ .  $\square$

En consecuencia si  $X$  es una función de distribución de una variable discreta que toma finitos valores  $x_1, x_2, \dots, x_n$  con probabilidad  $p_i = P\{X = x_i\}$ , tenemos que:

$$F(x) = \sum_{i=1}^n p_i H_{x_i}(x)$$

En consecuencia, por la linealidad de la integral de Riemann-Stieltjes respecto a  $F$ :

$$E[\varphi(X)] = \int_a^b \varphi(x) dF(x) = \sum_{i=1}^n p_i \int_a^b \varphi(x) dH_{x_i} = \sum_{i=1}^n p_i \varphi(x_i)$$

(donde  $a \leq x_i \leq b \forall i$ ). Este resultado coincide con la fórmula anteriormente vista para  $E[\varphi(X)]$  para variables discretas.

**Ejemplo 2: Variables aleatorias absolutamente continuas** Supongamos que  $X$  es una variable aleatoria continua, que tiene la densidad  $f(x)$ . Queremos calcular  $E[X]$ . Para ello, resultará útil el siguiente lema:

**Lema 3.3.2** Supongamos que  $F : [a, b] \rightarrow \mathbb{R}$  es una función creciente con derivada continua  $F'(x) = f(x)$ , entonces

$$\int_a^b \varphi(x) dF(x) = \int_a^b \varphi(x) f(x) dx$$

para toda función  $\varphi \in C[a, b]$ .

**Prueba:** Por el teorema del valor medio,  $F(x_{i+1}) - F(x_i) = f(\xi_i)(x_{i+1} - x_i)$  para cierto  $\xi_i \in (x_i, x_{i+1})$ . Entonces, con esta elección de los puntos intermedios, la suma  $S_\pi$  se puede escribir como

$$S_\pi = \sum_{i=0}^{n-1} \varphi(\xi_i) f(\xi_i) (x_{i+1} - x_i)$$

y vemos que cuando la norma de la partición  $\pi$  tiende a cero, tiende a la integral de Riemann

$$\int_a^b \varphi(x) f(x) dx$$

□

En particular, podemos definir la esperanza de una variable aleatoria con densidad continua  $f(x)$  por:

$$E[X] = \int_a^b x f(x) dx$$

y más generalmente,

$$E[\varphi(X)] = \int_a^b \varphi(x) f(x) dx$$

En particular:

$$\text{Var}(X) = E[(x - \mu)^2] = \int_a^b (x - \mu)^2 dx$$

siendo  $\mu = E[X]$ .

**Un ejemplo:** Si consideramos  $X$  una variable con distribución uniforme en el intervalo  $[a, b]$  entonces su densidad es:

$$f(x) = \frac{1}{b - a}$$

Con lo que

$$\mu = E(X) = \int_a^b x f(x) dx = \frac{a + b}{2}$$

y

$$\text{Var}X = \int_a^b \left( x - \frac{a + b}{2} \right)^2 f(x) dx = \frac{1}{12}(b - a)^2$$

¿Qué sucede si  $X$  no es una variable aleatoria acotada? En este caso debemos considerar integrales de Riemann-Stieltjes impropias, de la forma:

$$\int_{-\infty}^{\infty} \varphi(x) dF(x)$$

Naturalmente definimos esta integral, de la siguiente manera:

$$\int_{-\infty}^{\infty} \varphi(x) dF(x) = \lim_{a \rightarrow -\infty, b \rightarrow +\infty} \int_a^b \varphi(x) dF(x)$$

El problema es que este límite puede no existir. Si  $\varphi$  es no negativa, podemos decir que siempre existe, pero puede valer  $+\infty$ . Adoptaremos pues la siguiente definición.

**Definición 3.3.2** Sea  $X : \Omega \rightarrow \overline{\mathbb{R}}$  una variable aleatoria, y sea  $F = F_X$  su función de distribución. Diremos que  $X$  tiene esperanza finita, o que  $X$  es integrable, si

$$\int_{-\infty}^{\infty} |x| dF(x) < +\infty$$

En ese caso, definimos:

$$E[X] = \int_{-\infty}^{\infty} x dF(x) \quad (3.5)$$

Más generalmente, tenemos la fórmula<sup>3</sup>:

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) dF(x) \quad (3.6)$$

válida si

$$\int_{-\infty}^{\infty} |\varphi(x)| dF(x) < +\infty$$

análoga a la proposición 2.1.3. Y cuando  $X$  tiene una densidad continua,

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f(x) dx$$

---

<sup>3</sup>Sin embargo es complicado justificar esto directamente a partir de la definición (3.5), pues no es sencillo en general establecer cuál es la relación general entre las funciones de distribución  $F_{\varphi(X)}$  y  $F_X$ . En la observación 3.5.1 consideraremos el caso de un cambio de variable estrictamente creciente y biyectivo. Una justificación rigurosa de su validez en general se da en el apéndice B, pero utilizando herramientas de la teoría de la integral de Lebesgue.

**Ejemplo:** Supongamos que  $X$  se distribuye según la densidad normal  $N(\mu, \sigma^2)$ . Entonces, haciendo el cambio de variable  $y = \frac{x-\mu}{\sigma}$ , encontramos que

$$\begin{aligned} E[X] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/(2\sigma^2)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma y) e^{-y^2/2} dy \\ &= \mu \left[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \right] + \sigma \left[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy \right] = \mu \end{aligned}$$

[La segunda integral se anula, pues la densidad normal estándar es una función par]. Similarmente,

$$\text{Var}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-(x-\mu)^2/(2\sigma^2)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 y^2 e^{-y^2/2} dy$$

Para calcular esta integral, observamos que:

$$\left( e^{-y^2/2} \right)' = (-y) e^{-y^2/2}$$

e integramos por partes, deducimos que:

$$\text{Var}(X) = \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = \sigma^2$$

Este ejemplo aclara el significado de los parámetros de la distribución normal.

**Ejercicio:** Se dice que la variable aleatoria tiene distribución exponencial  $\text{Exp}(\lambda)$  (donde  $\lambda > 0$ ) cuando su densidad de probabilidad es

$$f_X(x) = \lambda e^{-\lambda x} I_{(0,+\infty)}(x) \quad (3.7)$$

Demostrar que entonces

$$E(X) = \frac{1}{\lambda} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

**Un ejemplo de una variable aleatoria que no es continua ni discreta:** Sea  $X$  una variable aleatoria con distribución uniforme en el intervalo  $[0, 1]$  y consideramos  $Y = \max(X, 1/2)$ , entonces:

$$Y = \begin{cases} 1/2 & \text{si } X \leq 1/2 \\ X & \text{si } X > 1/2 \end{cases}$$

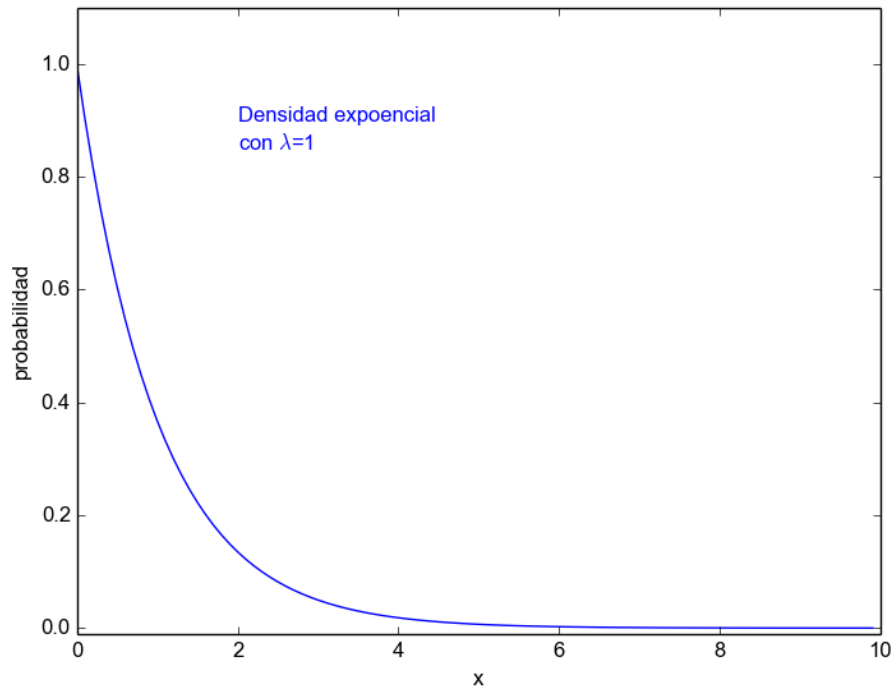


Figura 3.4: La densidad exponencial con  $\lambda = 1$  (gráfico de la función exponencial).

Calculemos la función de distribución de  $Y$ :

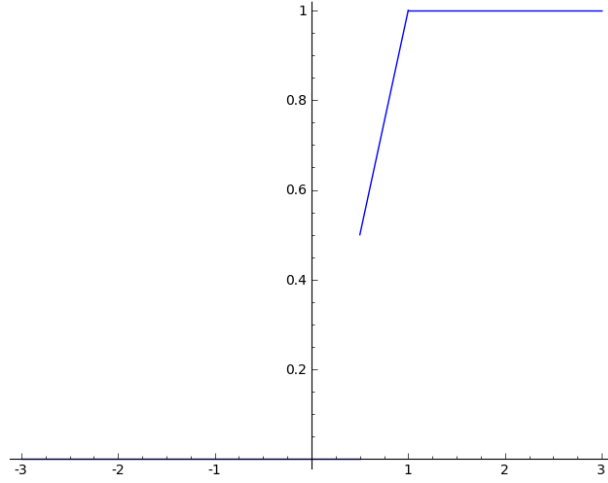
$$F_Y(x) = P\{Y \leq x\} = P\{X \leq x \wedge 1/2 \leq x\}$$

Deducimos que:

$$F_Y(x) = \begin{cases} P(\emptyset) = 0 & \text{si } x < 1/2 \\ P\{X \leq x\} = x & \text{si } 1/2 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

Deducimos que  $Y$  no es una variable discreta ya que  $F_Y$  no es una función escalera, y que tampoco  $Y$  es una variable absolutamente continua ya que  $F_Y$  no es continua.



Figura 3.5: La función de distribución  $F_Y$  en este ejemplo

Calculemos la esperanza de  $Y$ , esto puede hacerse de varias formas, por ejemplo usando la aditividad con respecto al intervalo de integración:

$$E[Y] = \int_0^1 x \, dF(x) = \int_0^{1/2} x \, dF + \int_{1/2}^1 x \, dF$$

En el intervalo cerrado  $[0, 1/2]$  la función  $F$  coincide con la función  $\frac{1}{2}H_{1/2}$  en consecuencia:

$$\int_0^{1/2} x \, dF = \frac{1}{2} \int_0^{1/2} x \, dH_{1/2} = \frac{1}{4}$$

mientras que:

$$\int_{1/2}^1 x \, dF(x) = \int_{1/2}^1 x \, dx = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$$

pues en  $[1/2, 1]$  la función  $F(x)$  tiene derivada continua  $F'(x) = 1$ . Concluimos que:

$$E[Y] = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$$

Otra manera de hacer la cuenta es considerar la función de variable real  $\varphi(x) = \max(x, 1/2)$  y utilizar la fórmula para  $E[\varphi(X)]$ :

$$E[\varphi(X)] = \int_0^1 \max(x, 1/2) dx = \int_0^{1/2} 1/2 dx + \int_{1/2}^1 x dx = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$$

**Ejercicio:** Supongamos que  $Z = \min(X, 1/2)$  donde  $X$  tiene distribución uniforme en  $[0, 1]$ . Determinar la función de distribución  $F_Z$  y la esperanza  $E(Z)$ .

### 3.4. Vectores Aleatorios

Las ideas anteriores sobre variables aleatorias continuas, pueden generalizarse para considerar vectores aleatorios.

**Definición 3.4.1** Sea  $(\Omega, \mathcal{E}, P)$  un espacio de probabilidad. Un **vector aleatorio**  $n$ -dimensional es una función  $X : \Omega \rightarrow \mathbb{R}^n$  con la propiedad de que si  $I = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_n, b_n]$  es un intervalo de  $\mathbb{R}^n$  entonces  $X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{E}$ , es decir está definida la probabilidad  $P\{X \in I\}$  de que  $X$  pertenezca a  $I$ .

**Obsevación:** Dar un vector aleatorio  $n$ -dimensional es equivalente a dar  $n$  variables aleatorias  $X_1, X_2, \dots, X_n$ .

#### Ejemplos de vectores aleatorios:

1. Un ejemplo de vector aleatorio discreto es el que consideramos al describir la distribución multinomial (ver página 53).
2. Distribución uniforme en un conjunto  $A \subset \mathbb{R}^n$  de medida positiva: si  $A$  es un conjunto de  $\mathbb{R}^n$  de medida positiva y  $X$  es un vector aleatorio  $n$ -dimensional, decimos que  $X$  se distribuye uniformemente en  $A$  si  $X$  pertenece a  $A$  con probabilidad 1, y si

$$P\{X \in B\} = \frac{m(B)}{m(A)} \quad \forall B \subset A$$

En esta definición  $A$  y  $B$  pueden ser conjuntos medibles Lebesgue cualesquiera, y  $m(A)$  denota la medida de Lebesgue de  $A$  (Quienes no hayan cursado análisis real, pueden pensar que  $A$  y  $B$  son conjuntos para los que tenga sentido calcular la medida de  $A$ , por ejemplo que  $A$  y  $B$  son abiertos de  $\mathbb{R}^2$  y  $m(A)$  representa el área de  $A$ ).

3. Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función integrable tal que  $0 \leq f(x) \leq 1$ , y

$$\int_{\mathbb{R}^n} f(x) dx = 1$$

Decimos que el vector  $X$  se distribuye según la **densidad conjunta**  $f(x)$  si para cualquier conjunto medible  $A \subset \mathbb{R}^n$ , tenemos que:

$$P\{X \in A\} = \int_A f(x) dx$$

(De nuevo, quienes no hayan cursado análisis real pueden pensar que  $f$  es integrable en el sentido de Riemann, y  $A$  es cualquier abierto de  $\mathbb{R}^n$ ).

4. Por ejemplo, una posible generalización de la distribución normal a dos dimensiones (normal bi-variada), se obtiene especificando que el vector  $(X, Y)$  se distribuye según la densidad conjunta:

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \quad (3.8)$$

Veremos más adelante que esta densidad corresponde al caso especial de dos variables aleatorias independientes con esperanza 0 y esperanza 1. Más generalmente, decimos que el vector aleatorio  $X$  tiene **distribución normal multivariada** si se distribuye según una densidad de la forma:

$$f(x) = ce^{-q(x)}$$

donde  $q(x) = x^t Ax$  es una forma cuadrática definida positiva, y  $c$  es una constante elegida de modo que la integral de  $f$  sobre todo  $\mathbb{R}^n$  dé 1. Más adelante volveremos sobre este concepto.

La noción de función de distribución puede generalizarse a vectores aleatorios.

**Definición 3.4.2** Si  $X : \Omega \rightarrow \mathbb{R}^n$  es un vector aleatorio, su **función de distribución conjunta** es la función  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  dada por:

$$F(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}$$

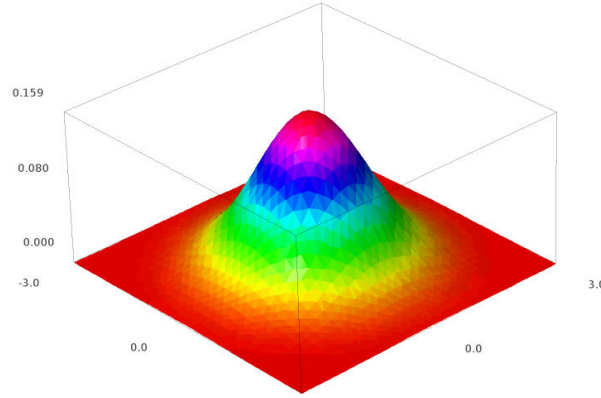


Figura 3.6: La función de densidad normal bivariada (para dos variables independientes con esperanza 0 y esperanza 1) dada por la ecuación (3.8).

Por ejemplo, si  $X$  es un vector aleatorio que se distribuye según la densidad conjunta  $f(x)$ , entonces su función de distribución conjunta es:

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) d\tilde{x}_1 d\tilde{x}_2 \dots d\tilde{x}_n$$

La noción de función de distribución resulta más complicada que en el caso de variables aleatorias unidimensionales. En el caso unidimensional, la probabilidad de que la variable  $X$  tome un valor en el intervalo  $(a, b]$  viene dada, en términos de la función de distribución  $F_X$ , por:

$$P\{X \in (a, b]\} = P\{X \leq b\} - P\{X \leq a\} = F_X(b) - F_X(a)$$

En cambio si  $(X, Y)$  es un vector aleatorio con función de distribución conjunta  $F$ , y  $R = (a, b] \times (c, d]$  es un rectángulo (semiabierto) en  $\mathbb{R}^2$ , la probabilidad de que  $(X, Y)$  tome un valor en  $R$  es (por la fórmula de inclusiones y exclusiones):

$$\begin{aligned} P\{(X, Y) \in R\} &= P\{X \leq b, Y \leq d\} - P\{X \leq a, Y \leq d\} \\ &\quad - P\{X \leq b, Y \leq c\} + P\{X \leq a, Y \leq c\} \end{aligned}$$

Es decir que:

$$P\{(X, Y) \in R\} = F(b, d) - F(a, d) - F(b, c) + F(a, c) \quad (3.9)$$

(Esta cantidad es necesariamente no negativa, esta es la generalización bidimensional del hecho de que en el caso unidimensional la función de distribución es creciente.)

Una fórmula análoga (¡pero más complicada!) es cierta para vectores aleatorios en más dimensiones. Por ello, la noción de función de distribución no resultará tan útil como lo era en el caso unidimensional (y con frecuencia resulta más cómodo pensar directamente en términos de probabilidades asignadas a rectángulos, o subconjuntos más generales de  $\mathbb{R}^n$ ).

### 3.4.1. Densidades y distribuciones marginales

Consideramos para simplificar la notación, un vector aleatorio bidimensional  $(X, Y)$ . Investiguemos qué relación existe entre la función de distribución conjunta  $F$  del vector  $(X, Y)$  y las funciones de distribución  $F_X$  y  $F_Y$  de cada variable por separado:

Notemos que:

$$F_X(x) = P\{X \leq x\} = P\{X \leq x, Y \leq +\infty\} = F(x, +\infty) = \lim_{y \rightarrow +\infty} F(x, y)$$

Similarmente,

$$F_Y(y) = \lim_{x \rightarrow +\infty} F(x, y)$$

$F_X$  y  $F_Y$  se conocen como las **funciones de distribución marginales** del vector aleatorio  $(X, Y)$ .

Consideremos ahora el caso particular, en que el vector aleatorio  $(X, Y)$  se distribuye según la densidad conjunta  $f(x, y)$ , su función de distribución será entonces:

$$F(x_0, y_0) = P\{X \leq x_0, Y \leq y_0\} = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f(x, y) \, dx \, dy$$

y en consecuencia sus funciones de distribución marginales vendrán dadas por:

$$F_X(x_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy$$

$$F_Y(y_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{y_0} f(x, y) \, dx \, dy$$

Utilizando el teorema de Fubini, podemos escribir  $F_X$  como una integral reiterada:

$$F_X(x_0) = \int_{-\infty}^{x_0} \left( \int_{-\infty}^{\infty} f(x, y) dy \right) dx$$

Esta igualdad significa que el vector aleatorio  $X$  se distribuye según la densidad:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (3.10)$$

Similarmente, el vector aleatorio  $Y$  se distribuye según la densidad:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (3.11)$$

$f_X$  y  $f_Y$  se conocen como las **densidades marginales** de probabilidad del vector aleatorio  $(X, Y)$ .

### 3.4.2. Esperanza de funciones de vectores aleatorios. Covariancia

Sea  $(X, Y)$  un vector aleatorio bidimensional, y  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  una función continua. La fórmula (3.6) para la esperanza de una función de una variable aleatoria puede generalizarse a vectores aleatorios:

$$E[\varphi(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) dF(x, y) \quad (3.12)$$

donde la integral que aparece en el segundo miembro es una integral doble de Riemann-Stieltjes.

Para definir este concepto puede procederse como en análisis II, considerando primero la integral

$$\int_a^b \int_c^d \varphi(x, y) dF(x, y) \quad (3.13)$$

en un rectángulo  $R = [a, b] \times [c, d]$  de  $\mathbb{R}^2$ . Consideramos una partición  $\pi$  del rectángulo  $R$  en rectángulos más pequeños  $R_{ij} = (x_i, x_{i+1}] \times (y_j, y_{j+1}]$ , definida por una partición  $\pi_x$  del intervalo  $[a, b]$ :

$$a = x_0 < x_1 < \dots < x_M = b$$

y otra partición  $\pi_y$  del intervalo  $[c, d]$ :

$$a = y_0 < y_1 < \dots < y_N = b$$

Elegimos puntos intermedios  $\xi_i \in [x_i, x_{i+1}]$  y  $\eta_j \in [y_j, y_{j+1}]$ , y consideramos sumas de Riemann-Stieltjes dobles:

$$S_\pi(\varphi, F) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \varphi(\xi_i, \eta_j) \Delta F(R_{ij})$$

siendo

$$\Delta F(R_{ij}) = F(x_{i+1}, y_{j+1}) - F(x_i, y_{j+1}) - F(x_{i+1}, y_j) + F(x_i, y_j)$$

que de acuerdo a la fórmula (3.9), representa la probabilidad de que el vector  $(X, Y)$  tome un valor en el rectángulo  $R_{ij}$ .

Definamos la norma  $|\pi|$  de la partición  $\pi$  como el máximo de las normas de las particiones  $\pi_x$  y  $\pi_y$ . Entonces si, cuando la norma de la partición  $\pi$  tiende a cero, las sumas  $S(\pi, F)$  convergen a un número  $I$ , diremos que la integral (3.13) existe, y que toma el valor  $I$ . Análogamente a lo que sucede en el caso unidimensional, podemos demostrar que esto sucede si  $F$  es la función de distribución de un vector aleatorio, y  $\varphi$  es continua.

La integral impropia, sobre todo el plano, que aparece en la fórmula (3.12) puede definirse como el límite de integrales sobre rectángulos:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) dF(x, y) = \lim_{a, c \rightarrow -\infty; b, d \rightarrow +\infty} \int_a^b \int_c^d \varphi(x, y) dF(x, y)$$

Para justificar intuitivamente la fórmula (3.12) podemos proceder como en el caso discreto, definiendo variables aleatorias discretas  $X_\pi$  e  $Y_\pi$  que aproximan a  $X$  e  $Y$  por:

$$X_\pi = \xi_i \text{ si } X \in (x_i, x_{i+1}]$$

$$Y_\pi = \eta_j \text{ si } Y \in (y_j, y_{j+1}]$$

y observando que:

$$E[\varphi(X_\pi, Y_\pi)] = S_\pi(\varphi, F)$$

Por lo que cuando la norma de la partición  $\pi$  tiende a cero, obtenemos formalmente la fórmula (3.12).

El caso que más nos va a interesar, es cuando el vector aleatorio  $(X, Y)$  se distribuye según una densidad conjunta  $f(x, y)$ . En este caso, como ocurría en el caso unidimensional, la esperanza de  $\varphi(X, Y)$  puede calcularse mediante una integral de Riemann ordinaria, en lugar de una integral de Riemann-Stieltjes:

$$E[\varphi(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) f(x, y) dx dy \quad (3.14)$$

Un caso importante de aplicación de las fórmulas anteriores es cuando queremos calcular la covarianza de dos variables aleatorias en el caso continuo. Recordamos que por definición:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

siendo  $\mu_X = E[X]$ ,  $\mu_Y = E[Y]$ . Entonces tomando  $\varphi(x, y) = (x - \mu_X)(y - \mu_Y)$  en las fórmulas anteriores, tenemos que:

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) dF(x, y)$$

en el caso general, y

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dx dy$$

si el vector  $(X, Y)$  admite una densidad conjunta.

**Observación 3.4.1** *Una de las propiedades más básicas de la esperanza es su linealidad. Sin embargo, es difícil justificar su validez en general partiendo de la definición 3.5, ya que la función de distribución  $F_X$  no depende linealmente de la variable  $X$ . Utilizando la fórmula (3.14), podríamos sin embargo dar una justificación de que  $E[X + Y] = E[X] + E[Y]$  para el caso en que  $X$  e  $Y$  tienen una densidad conjunta continua y esperanza finita<sup>4</sup>. En efecto, en este*

---

<sup>4</sup>Esta propiedad es válida en general, como se deduce inmediatamente de la interpretación de la esperanza como una integral de Lebesgue, ver apéndice B.



caso, tomando  $\varphi(x, y) = x + y$ . vemos que

$$\begin{aligned}
 E[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\
 &= E[X] + E[Y]
 \end{aligned}$$

### 3.4.3. Independencia

Nuestro siguiente objetivo será extender a variables no discretas la noción de independencia:

**Definición 3.4.3** *Dos variables aleatorias  $X$  e  $Y$  se dicen independientes, cuando para todo  $a < b$  y todo  $c < d$  los eventos  $\{X \in (a, b]\}$  e  $\{Y \in (c, d]\}$  son independientes. Es decir (en virtud de la definición de eventos independientes), si vale que:*

$$P\{a < X \leq b, c < Y \leq d\} = P\{a < X \leq b\}P\{c < Y \leq d\}$$

**Lema 3.4.1** *Supongamos que el vector  $(X, Y)$  admite una densidad conjunta continua  $f(x, y)$ . Entonces las variables  $X$  e  $Y$  son independientes, si y sólo si  $f$  se factoriza en la forma:*

$$f(x, y) = f_X(x)f_Y(y)$$

siendo  $f_X$  y  $f_Y$  las densidades marginales de probabilidad.

**Prueba:** Supongamos primero que  $X$  e  $Y$  son independientes, y que el vector  $(X, Y)$  se distribuye según la densidad conjunta  $f(x, y)$ . Entonces  $X$  se distribuye según la densidad marginal  $f_X$  dada por (3.10), y similarmente  $Y$  se distribuye según la densidad marginal dada por (3.11).

Entonces dado  $(x_0, y_0) \in \mathbb{R}^2$  y  $h, k > 0$ , tenemos que:

$$P\{x_0 < X \leq x_0 + h, y_0 < Y \leq y_0 + k\} = \int_{x_0}^{x_0+h} \int_{y_0}^{y_0+k} f(x, y) dx dy \quad (3.15)$$

$$P\{x_0 < X \leq x_0 + h\} = \int_{x_0}^{x_0+h} f_X(x) dx \quad (3.16)$$

$$P\{y_0 < Y \leq y_0 + k\} = \int_{y_0}^{y_0+k} f_Y(y) dy \quad (3.17)$$

En virtud de la definición (3.4.3), vemos que:

$$\begin{aligned} & \frac{P\{x_0 < X \leq x_0 + h, y_0 < Y \leq y_0 + k\}}{hk} \\ &= \frac{P\{x_0 < X \leq x_0 + h\}}{h} \frac{P\{y_0 < Y \leq y_0 + k\}}{k} \end{aligned} \quad (3.18)$$

De la expresión (3.16) cuando  $h \rightarrow 0$ , deducimos que:

$$\frac{P\{x_0 < X \leq x_0 + h\}}{h} \rightarrow f_X(x_0)$$

por el teorema fundamental del cálculo (siendo  $f_X$  continua en  $x_0$ ).

Similarmente, cuando  $k \rightarrow 0$ , (3.17) y el teorema fundamental del cálculo nos dicen que:

$$\frac{P\{y_0 < Y \leq y_0 + k\}}{k} \rightarrow f_Y(y_0)$$

Finalmente, de la expresión (3.15), por el teorema de diferenciación para integrales (generalización del teorema fundamental del cálculo), deducimos que:

$$\frac{P\{x_0 < X \leq x_0 + h, y_0 < Y \leq y_0 + k\}}{hk} \rightarrow f(x_0, y_0)$$

cuando  $h, k \rightarrow 0$ , siempre que  $f$  sea continua en el punto  $(x_0, y_0)$ .

En consecuencia, cuando  $h, k \rightarrow 0$ , a partir de la relación (3.18), obtenemos que:

$$f(x_0, y_0) = f_X(x_0)f_Y(y_0) \quad (3.19)$$

Esto prueba una de las implicaciones del teorema<sup>5</sup>

---

<sup>5</sup>Para evitar complicaciones técnicas, hemos supuesto que la densidad conjunta  $f$  es continua. No obstante, si  $f$  fuera solamente integrable, repitiendo el mismo argumento y usando el teorema de diferenciación de integrales que se ve en análisis real, obtendríamos que la relación (3.19) se verifica en casi todo punto.

Para probar la afirmación recíproca, supongamos que la densidad conjunta  $f$  puede expresarse en la forma:

$$f(x, y) = f_X(x)f_Y(y)$$

siendo  $f_X$  y  $f_Y$  dos densidades de probabilidad (Notemos que entonces,  $f_X$  y  $f_Y$  deben ser entonces necesariamente las densidades marginales dadas por (3.10 - 3.11), como se deduce integrando respecto de  $x$  y de  $y$ ).

Entonces, en virtud del teorema de Fubini,

$$\begin{aligned} P\{a < X \leq b, c < Y \leq d\} &= \int_a^b \int_c^d f(x, y) \, dx \, dy = \\ &= \left( \int_a^b f_X(x) \, dx \right) \left( \int_c^d f_Y(y) \, dy \right) = P\{a < X \leq b\}P\{c < Y \leq d\} \end{aligned}$$

por lo que se deduce que  $X$  e  $Y$  son variables aleatorias independientes.  $\square$

Notemos, que el significado de esta demostración, es que la relación (3.19), es una “expresión infinitesimal” de la definición de independencia.

Como corolario obtenemos el análogo de la proposición 2.1.6 para variables continuas<sup>6</sup>.

**Corolario 3.4.1** *Si  $X$  e  $Y$  son variables aleatorias independientes con esperanza finita, que se distribuyen según una densidad conjunta continua  $f(x, y)$  entonces  $XY$  tiene esperanza finita y se tiene que*

$$E[XY] = E[X]E[Y]$$

---

<sup>6</sup> La propiedad vale aunque  $X$  e  $Y$  no admitan una densidad conjunta continua. Una demostración se da en el apéndice C pero utilizando la integral de Lebesgue.

**Prueba:** Nuevamente usamos la fórmula (3.14), para obtener que<sup>7</sup>

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy) f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy) f_X(x) f_Y(y) dx dy \\ &= \left( \int_{-\infty}^{\infty} x f_X(x) dx \right) \left( \int_{-\infty}^{\infty} y f_Y(y) dy \right) \\ &= E[X] E[Y] \end{aligned}$$

□

### 3.4.4. Vectores aleatorios $n$ -dimensionales

Las ideas anteriores se generalizan sin dificultad a vectores aleatorios multidimensionales, pero la notación resulta más complicada. Así pues si  $X : \Omega \rightarrow \mathbb{R}^n$  es un vector aleatorio  $n$ -dimensional, que se distribuye según una densidad conjunta  $f(x) = f(x_1, x_2, \dots, x_n)$  que supongamos por simplicidad continua, tendremos que:

- La esperanza de una función  $\varphi(X)$  del vector  $X$ , donde  $\varphi : X \rightarrow \mathbb{R}$  es una función continua, se puede calcular mediante la fórmula:

$$E[\varphi(X)] = \int_{\mathbb{R}^n} \varphi(x) f(x) dx$$

- La  $k$ -ésima componente  $X_k$  del vector  $X$  ( $1 \leq k \leq n$ ) se distribuye según la densidad marginal:

$$f_{X_k}(x) = \int_{\mathbb{R}^{n-1}} f(x_1, x_2, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n) dx_1 dx_2 \dots dx_{k-1} dx_{k+1} \dots dx_n$$

---

<sup>7</sup> Para justificar rigurosamente este cálculo, hay que hacerlo primero con  $|xy|$  en lugar de  $xy$ , lo que conduce a  $E(|XY|) = E(|X|)E(|Y|)$ , con lo que se establece que la integral doble es absolutamente convergente y se justifica la aplicación del teorema de Fubini.

- Las componentes  $X_1, X_2, \dots, X_n$  del vector  $X$  se dirán mutuamente independientes si para cualquier rectángulo  $n$ -dimensional (producto de intervalos)

$$I = \prod_{k=1}^n (a_k, b_k]$$

se verifica que:

$$P\{X \in I\} = \prod_{k=1}^n P\{a_k < X_k \leq b_k\}$$

En términos de la función de distribución conjunta,  $X_1, X_2, \dots, X_n$  son mutuamente independientes si y sólo si  $f(x)$  se factoriza en la forma:

$$f(x) = f_{X_1}(x_1)f_{X_2}(x_2) \dots f_{X_n}(x_n)$$

## 3.5. Cambio de variable

### 3.5.1. Cambios de variables unidimensionales

Consideremos primero un cambio de variable de la forma  $Y = \varphi(X)$  donde  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  es una función biyectiva y estrictamente creciente.

Entonces podemos facilmente relacionar las funciones de distribución de  $X$  e  $Y$

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = P\{\varphi(X) \leq y\} \\ &= P\{X \leq \varphi^{-1}(Y)(y)\} \\ &= F_X(\varphi^{-1}(y)) \end{aligned} \tag{3.20}$$

En particular (derivando con la regla de la cadena), se deduce que si  $X$  admite una densidad de probabilidad  $f_X$  de clase  $C^1$ , vemos que  $Y$  se distribuye según la densidad:

$$f_Y(y) = f_X(\varphi^{-1}(y))[\varphi^{-1}]'(y) \tag{3.21}$$

**Ejemplo 3.5.1** Supongamos que  $X \sim N(\mu, \sigma^2)$  y hagamos un cambio de variable lineal,  $Y = aX + b$  con  $a > 0$ . Esto corresponde a elegir

$$\varphi(x) = ax + b \Rightarrow \varphi^{-1}(y) = \frac{y - b}{a}$$

Entonces según la fórmula (3.21), tenemos que

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y - b}{a}\right)$$

En particular en este ejemplo:

$$f_Y(y) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left\{-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}\right\} = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y - (a\mu + b))^2}{2(a\sigma)^2}\right\}$$

Concluimos que  $Y \sim N(a\mu + b, a^2\sigma^2)$ .

**Observación 3.5.1** Como otra aplicación, podemos dar una justificación rigurosa de la fórmula (3.6) para el caso en que  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  es biyectiva y estrictamente creciente. En efecto, en este caso, y llamamos  $Y = \varphi(X)$ , haciendo el cambio de variable  $y = \varphi(x)$  en la integral de Stieltjes y teniendo en cuenta que entonces  $F_Y(y) = F_X(x)$  por 3.20, obtenemos que:

$$E[Y] = \int_{-\infty}^{\infty} y dF_Y(y) = \int_{-\infty}^{\infty} \varphi(x) dF_X(x)$$

La situación es bastante más compleja si admitimos cambios de variables que no son monótonos o biyectivos.

Consideremos por ejemplo el cambio de variable  $Y = X^2$ . Entonces para  $z > 0$  tenemos que:

$$\begin{aligned} F_Y(y) &= P\{X^2 \leq y\} = P\{|X| \leq \sqrt{y}\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} = \\ &= P\{X \leq \sqrt{y}\} - P\{X < -\sqrt{y}\} = F_X(\sqrt{y}) - F_X(-\sqrt{y}^-) \end{aligned}$$

mientras que claramente  $F_Y(y) = 0$  si  $y < 0$ .

En particular si  $X$  es una variable absolutamente continua con densidad  $f_X$ , encontramos (derivando como antes) que:

$$f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] \quad (y > 0) \quad (3.22)$$

**Observación 3.5.2** *Una aplicación importante de los cambios de variables es la simulación de distribuciones de probabilidad. En general, una computadora sabe generar números pseudo-aleatorios, que simulan la distribución uniforme en el intervalo  $[0, 1]$ . Si queremos generar a partir de ellos números pseudo-aleatorios que simulen la distribución  $F$ , se nos plantea el problema siguiente*

*Dada una función de distribución  $F : \mathbb{R} \rightarrow [0, 1]$  (con las propiedades del lema 3.1.1) y si  $X \sim \mathcal{U}(0, 1)$ , ¿cómo podemos obtener otra variable aleatoria con distribución  $F$  ?*

Conforme a la fórmula 3.20, si  $F$  es continua (sin saltos) y estrictamente creciente, podemos tomar  $Y = F^{-1}(X)$  donde  $F^{-1}$  denota la inversa de  $F$ . Cuando  $F$  no cumple estas hipótesis, es posible hacer lo mismo, pero considerando la inversa generalizada de  $F$  definida por

$$F^{-1}(y) = \min\{x \in \mathbb{R} : F(x) \geq y\}$$

### 3.5.2. Cambios de variables $n$ -dimensionales

**Proposición 3.5.1** *Supongamos que  $X$  es un vector que se distribuye según una densidad  $f(x)$  con soporte en  $\bar{U}$  siendo  $U$  un abierto  $\mathbb{R}^n$ , y que  $\varphi : U \rightarrow V$  es un difeomorfismo  $C^1$ , donde  $V$  es otro abierto de  $\mathbb{R}^n$  entonces, si consideramos el vector aleatorio  $Y = \varphi(X)$ ,  $Y$  se distribuye en  $V$  según la densidad*

$$f(\varphi^{-1}(y)) | \det(D\varphi^{-1})(y) |$$

**Prueba:** Sea  $W \subset V$  un abierto cualquiera, entonces

$$P\{Y \in W\} = P\{X \in \varphi^{-1}(W)\} = \int_{\varphi^{-1}(W)} f(x) dx$$

En esta integral, hagamos el cambio de variable  $y = \varphi(x)$ ,  $x = \varphi^{-1}(y)$ . Entonces, según el teorema de cambio de variable

$$P\{Y \in W\} = \int_W f(\varphi^{-1}(y)) |\det D(\varphi^{-1})(y)| dy$$

Como esto vale para todo  $W \subset V$ , concluimos que  $Y$  se distribuye en  $V$  según la densidad  $f(\varphi^{-1}(y)) |\det(D\varphi^{-1})(y)|$ .  $\square$

### 3.6. Suma de variables aleatorias independientes

**Definición 3.6.1** Sean  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  funciones integrables. Definimos su **convolución**  $f * g$  de la siguiente manera:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(t) g(x - t) dt$$

Como ejemplo de la aplicación del teorema de cambio de variable, demostramos la siguiente afirmación:

**Proposición 3.6.1** Supongamos que  $X$  e  $Y$  son variables aleatorias independientes, que se distribuyen en  $\mathbb{R}$  según las densidades  $f(x)$  y  $g(x)$  respectivamente, entonces  $X + Y$  se distribuye según la densidad  $f * g(x)$ .

**Prueba:** Como  $X$  e  $Y$  son independientes,

$$(X, Y) \sim f(x)g(y)$$

Hacemos el cambio de variable lineal  $(U, V) = \varphi(X, Y) = (X + Y, Y)$ . Entonces  $(X, Y) = \varphi^{-1}(U, V) = (U - V, V)$ . Como  $\varphi$  es una transformación lineal, su diferencial coincide con ella misma. Para calcular el determinante de  $\varphi$  observamos que su matriz en la base canónica de  $\mathbb{R}^2$  es:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

En consecuencia, el determinante de  $\varphi$  es 1. Por el teorema anterior, tenemos que  $(U, V)$  que:

$$(U, V) \sim f(u - v)g(v) \text{ (densidad conjunta)}$$



Para recuperar la densidad de  $U$  (densidad marginal) debemos integrar en la variable  $v$ :

$$U \sim \int_{-\infty}^{\infty} f(u-v)g(v) dv$$

□

### Algunas Observaciones sobre la convolución:

1. La convolución es conmutativa:

$$f * g = g * f$$

También es posible probar que es asociativa:

$$(f * g) * h = f * (g * h)$$

2. Si  $f$  y  $g$  son densidades de probabilidad, entonces  $f * g$  también lo es.
3. Si  $f$  y  $g$  están soportadas en la semirrecta  $[0, +\infty)$  (es decir:  $f(t) = g(t) = 0$  si  $t < 0$ , entonces:

$$(f * g)(x) = \int_0^x f(t) g(x-t) dt$$

### 3.6.1. Suma de variables normales independientes

**Proposición 3.6.2** Si  $X \sim N(0, \sigma_1^2)$  e  $Y \sim N(0, \sigma_2^2)$  son variables aleatorias independientes, entonces  $X + Y \simeq N(0, \sigma_1^2 + \sigma_2^2)$

**Prueba:** Aplicamos la proposición 3.6.1 con

$$f(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-x^2/(2\sigma_1^2)}, \quad g(x) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-x^2/(2\sigma_2^2)}$$

Entonces  $X \sim f * g$ , donde

$$\begin{aligned} (f * g)(x) &= \int_{-\infty}^{\infty} \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-t^2/(2\sigma_1^2)} \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-(x-t)^2/(2\sigma_2^2)} dt \\ &= \frac{1}{\sigma_1 \sigma_2 2\pi} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} A(x, t) \right\} dt \end{aligned}$$

donde

$$A(x, t) := \frac{t^2}{\sigma_1^2} + \frac{(x - t)^2}{\sigma_2^2}$$

Trabajemos con esta expresión, buscando completar el cuadrado:

$$\begin{aligned} A(x, t) &= \frac{t^2}{\sigma_1^2} + \frac{x^2 - 2xt + t^2}{\sigma_2^2} \\ &= t^2 \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) - \frac{2xt}{\sigma_2^2} + \frac{x^2}{\sigma_2^2} \\ &= t^2 \frac{\sigma^2}{\sigma_1^2 \sigma_2^2} - \frac{2xt}{\sigma_2^2} + \frac{x^2}{\sigma_2^2} \end{aligned}$$

siendo  $\sigma^2 = \sigma_1^2 + \sigma_2^2$ . Luego

$$A(x, t) = \frac{\sigma^2}{\sigma_1^2 \sigma_2^2} \left[ t^2 - 2xt \frac{\sigma_1^2}{\sigma^2} \right] + \frac{x^2}{\sigma_2^2}$$

Y completando entonces el cuadrado:

$$A(x, t) = \frac{\sigma^2}{\sigma_1^2 \sigma_2^2} \left[ \left( t - x \frac{\sigma_1^2}{\sigma^2} \right)^2 - x^2 \frac{\sigma_1^4}{\sigma^4} \right] + \frac{x^2}{\sigma_2^2}$$

o sea:

$$A(x, t) = \frac{\sigma^2}{\sigma_1^2 \sigma_2^2} \left( t - x \frac{\sigma_1^2}{\sigma^2} \right)^2 + \left( \frac{1}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma^2 \sigma_2^2} \right) x^2$$

Pero

$$\frac{1}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma^2 \sigma_2^2} = \frac{\sigma^2 - \sigma_1^2}{\sigma^2 \sigma_2^2} = \frac{\sigma_2^2}{\sigma^2 \sigma_2^2} = \frac{1}{\sigma^2}$$

Con lo que nos queda finalmente que

$$A(x, t) = \frac{\sigma^2}{\sigma_1^2 \sigma_2^2} \left( t - x \frac{\sigma_1^2}{\sigma^2} \right)^2 + \frac{1}{\sigma^2} x^2$$

Sustituyendo

$$(f * g)(x) = \frac{1}{\sigma_1 \sigma_2 2\pi} \exp\left(-\frac{x^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left\{-\frac{\sigma^2}{2\sigma_1^2 \sigma_2^2} \left(t - x \frac{\sigma_1^2}{\sigma^2}\right)^2\right\} dt$$

Sólo nos falta pues calcular la integral,

$$I(x) = \int_{-\infty}^{\infty} \exp\left\{-\frac{\sigma^2}{2\sigma_1^2 \sigma_2^2} \left(t - x \frac{\sigma_1^2}{\sigma^2}\right)^2\right\} dt$$

pero haciendo el cambio de variable

$$u = t - x \frac{\sigma_1^2}{\sigma^2}$$

vemos que no depende en realidad de  $x$ , y es

$$I(x) = \int_{-\infty}^{\infty} \exp\left\{-\frac{\sigma^2}{2\sigma_1^2 \sigma_2^2} u^2\right\} du$$

Y haciendo un último cambio de variable

$$v = \frac{\sigma}{\sigma_1 \sigma_2} u$$

nos queda que

$$I(x) = \frac{\sigma_1 \sigma_2}{\sigma} \int_{-\infty}^{\infty} \exp\left\{-\frac{v^2}{2}\right\} dv = \sqrt{2\pi} \frac{\sigma_1 \sigma_2}{\sigma}$$

Reemplazando nos queda que

$$X + Y \sim (f * g)(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Es decir, que  $X + Y \sim N(0, \sigma^2)$ . □

### 3.7. Las Distribuciones Gama

**Definición 3.7.1** Definimos la función gama de Euler por

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad (\alpha > 0) \tag{3.23}$$

Introducimos también la función Beta de Euler (íntimamente relacionada con la función gama), definida para  $\alpha_1, \alpha_2 > 0$  por

$$B(\alpha_1, \alpha_2) = \int_0^1 (1-u)^{\alpha_1-1} u^{\alpha_2-1} du$$

**Proposición 3.7.1** *La función gamma tiene las siguientes propiedades:*

1.  $\Gamma(1) = 1$
2.  $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
3.  $\Gamma(k) = (k-1)!$  (En consecuencia, la función gama puede pensarse como una generalización del factorial a valores no enteros de la variable).
4.  $\Gamma(1/2) = \sqrt{\pi}$

La función gama nos será útil para definir una familia de distribuciones de probabilidad<sup>8</sup>:

**Definición 3.7.2** *Decimos que  $X$  se distribuye según la distribución gama  $\Gamma(\alpha, \lambda)$  (siendo  $\alpha, \lambda > 0$ ) si su función de densidad de probabilidad es:*

$$f_{\alpha, \lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} I_{(0, +\infty)}(x) \quad (3.24)$$

**Observación 3.7.1** *Haciendo el cambio de variable  $y = \lambda x$  en (3.23), tenemos que*

$$\frac{\Gamma(\alpha)}{\lambda^\alpha} = \int_0^\infty y^{\alpha-1} e^{-\lambda y} dy \quad (3.25)$$

Se deduce que (3.24) es efectivamente una densidad de probabilidades. Más aún esta fórmula permite calcular fácilmente los momentos de las distribuciones gama: si  $X \sim \Gamma(\alpha, \lambda)$ , entonces

$$\begin{aligned} \mu_k(X) &= E(X^k) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+k-1} e^{-\lambda x} dx \\ &= \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)} \lambda^k \alpha(\alpha + 1) \cdots (\alpha + k) \lambda^k \end{aligned}$$

---

<sup>8</sup>También tiene importantes aplicaciones en otras ramas de la matemática como la teoría de números, y aparece en numerosas fórmulas como la del volumen de una bola  $n$ -dimensional.

En particular, la esperanza y la variancia de la distribución gama son

$$E(X) = \mu_1(X) = \frac{\alpha}{\lambda} \quad (3.26)$$

y

$$Var(X) = E(X^2) - E(X)^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2} \quad (3.27)$$

**Lema 3.7.1** Si  $X \sim \Gamma(\alpha_1, \lambda)$ ,  $Y \sim \Gamma(\alpha_2, \lambda)$  y son independientes, entonces  $X + Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$ .

**Prueba:** Según la proposición 3.6.1,  $X + Y \sim f_{\alpha_1, \lambda} * f_{\alpha_2, \lambda}$ . Hemos de calcular esta convolución:

$$\begin{aligned} (f_{\alpha_1, \lambda} * f_{\alpha_2, \lambda})(x) &= \int_0^x \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} (x-t)^{\alpha_1-1} e^{-\lambda(x-t)} \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} t^{\alpha_2-1} e^{-\lambda t} dt \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \left( \int_0^x (x-t)^{\alpha_1-1} t^{\alpha_2-1} dt \right) e^{-\lambda x} \end{aligned}$$

En esta integral hacemos el cambio de variable  $u = t/x$  ( $0 \leq x \leq 1$ ). Entonces:

$$\begin{aligned} (f_{\alpha_1, \lambda} * f_{\alpha_2, \lambda})(x) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \left( \int_0^1 (x-xu)^{\alpha_1-1} (xu)^{\alpha_2-1} x du \right) e^{-\lambda x} \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1+\alpha_2-1} \left( \int_0^1 (1-u)^{\alpha_1-1} u^{\alpha_2-1} du \right) e^{-\lambda x} \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} B(\alpha_1, \alpha_2) x^{\alpha_1+\alpha_2-1} e^{-\lambda x} \end{aligned}$$

Notamos que esta es salvo la constante, la densidad gama  $f_{\alpha_1+\alpha_2, \lambda}$ , pero como la convolución de dos densidades de probabilidad es una densidad de probabilidad, y hay una única constante que hace que la integral sobre  $(0, +\infty)$  dé 1 deducimos que:

$$f_{\alpha_1, \lambda} * f_{\alpha_2, \lambda} = f_{\alpha_1+\alpha_2, \lambda} \quad (3.28)$$

Como subproducto de la demostración obtenemos que:

$$\frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} B(\alpha_1, \alpha_2) = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1 + \alpha_2)}$$

o sea

$$B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$$

□

La función beta también puede usarse para definir una familia de distribuciones: las *distribuciones beta*. Diremos que  $X \sim \beta(\alpha_1, \alpha_2)$  si se distribuye según la densidad:

$$f_X(x) = \frac{1}{B(\alpha_1, \alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1} I_{(0,1)}(x)$$

### 3.8. La Distribución Exponencial y la propiedad de Falta de Memoria

La distribución exponencial (3.7) es un modelo muy útil para distintos procesos: llamadas que llegan a una central telefónica, tiempo de duración de una lámpara, desintegración radiactiva, etc.

Por ejemplo, para fijar ideas, consideremos la desintegración radiactiva de un átomo. La hipótesis fundamental que haremos para describir este fenómeno, es la propiedad de “falta de memoria” que establece que la probabilidad de que un átomo se desintegre en un intervalo de tiempo de longitud  $\Delta t$  sólo depende de la longitud del intervalo y es independiente de la historia anterior del material.

Podemos describir con más precisión esta propiedad de la siguiente manera: Si llamamos  $T$  al tiempo en el que el átomo se desintegra,  $T$  es una variable aleatoria. La probabilidad condicional de que el átomo se desintegre en el intervalo  $(t_0, t_0 + \Delta t]$  sabiendo que no se ha desintegrado aún en tiempo  $t = t_0$ , es igual a la probabilidad de que se desintegre en el intervalo  $(0, \Delta t]$ :

$$P\{T > t_0 + \Delta t | T > t_0\} = P\{T > \Delta t\}$$

Por definición de probabilidad condicional, esto significa que:

$$\frac{P\{t < T \leq t + \Delta t\}}{P\{T > t\}} = P\{T > \Delta t\}$$

Llamemos  $F$  a la función de distribución de  $T$ , y sea  $G(t) = 1 - F(t)$ . Entonces, esta igualdad establece que:

$$G(t + \Delta t) = G(t)G(\Delta t)$$

Necesitaremos el siguiente lema:

**Lema 3.8.1** Sea  $G : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$  una función continua que satisface que:

$$G(t + s) = G(t)G(s)$$

Entonces:  $G(t) = G(0)a^t$ , siendo  $a = G(1)$ .

Volviendo a nuestro problema de la desintegración radiactiva, si ponemos  $G(1) = e^{-\lambda}$  (suponiendo  $G(0) \neq 0$ ), y observamos que  $G(0) = 1$  pues  $T > 0$  (El átomo no se desintegró aún en  $t = 0$ ), obtenemos que:

$$G(t) = e^{-\lambda t}$$

Por consiguiente la función de distribución de  $T$  es:

$$F(t) = 1 - e^{-\lambda t}$$

y derivando vemos que su densidad es

$$f(t) = \lambda e^{-\lambda t} \quad (t > 0)$$

Decimos que la variable continua  $T$  se distribuye según la densidad exponencial de parámetro  $\lambda > 0$ ,  $\text{Exp}(\lambda)$ , que introducimos en (3.7).

Supongamos ahora que tenemos un material radiactivo formado inicialmente por un grán número de átomos  $N_0$ , y llamemos  $N(t)$  a la cantidad de átomos no desintegrados hasta el instante  $t$ . Hagamos la hipótesis de que las desintegraciones de los distintos átomos son independientes. Podemos pensar que son ensayos de Bernoullí, entonces por la ley de los grandes números

$$\frac{N(t)}{N_0} \approx P\{T > t_0\}$$

y deducimos que:

$$N(t) = N_0 e^{-\lambda t} \tag{3.29}$$

Esta expresión se conoce como la ley de desintegración radiactiva de Rutherford-Soddy (1902). El valor de la constante  $\lambda$  depende de la sustancia.

Se define semivida o período de semi-desintegración  $T_{1/2}$  el tiempo en que una muestra de material radiactivo tarda en reducirse a la mitad. De la fórmula (3.29), se deduce que

$$T_{1/2} = \frac{\log 2}{\lambda}$$

La siguiente tabla muestra por ejemplo los períodos de semi-desintegración de algunos isótopos radiactivos:

Isótopo	$T_{1/2}$
Berilio-8	$10^{-16}s$
Polonio-213	$4 \times 10^{-6}s$
Aluminio-28	2.25 min
Yodo-131	8 días
Estroncio-90	28 años
Radio-226	1600 años
Carbono-14	5730 años
Rubidio-87	$5,7 \times 10^{10}$ años

**Observación 3.8.1** *Entre las distribuciones discretas, la propiedad de falta de memoria es característica de la distribución geométrica, que puede entonces considerarse como el análogo discreto de la distribución exponencial.*

### 3.8.1. Tiempos de espera y procesos de Poisson

Llamemos  $T_i$  al tiempo en que ocurre la  $i$ -ésima desintegración radiactiva, de modo que:

$$T_1 < T_2 < \dots < T_n$$

(Podemos suponer para simplificar que no hay dos desintegraciones simultáneas, ya que la probabilidad de que ello ocurra es despreciable). Notemos que:

$$T_n = T_1 + (T_2 - T_1) + (T_3 - T_2) + \dots + (T_n - T_{n-1})$$

Las variables  $T_k - T_{k-1}$  representan el tiempo entre la  $(k-1)$ -ésima desintegración y la  $k$ -ésima desintegración. Por la discusión anterior (y la propiedad



de falta de memoria),  $T_k - T_{k-1}$  tiene distribución exponencial de parámetro  $\lambda > 0$  (donde  $\lambda > 0$  es una constante que depende del material que estamos considerando).

Por otra parte, si suponemos que el tiempo que un átomo tarda en desintegrarse es independiente de lo que tardan los demás, las  $T_{k+1} - T_k$  serán variables aleatorias independientes. Entonces la variable  $T_n$  será dada por una suma de  $n$  variables aleatorias independientes, todas con distribución exponencial de parámetro  $\lambda$ .

Como  $\text{Exp}(\lambda) = \Gamma(1, \lambda)$ , deducimos que  $T_n$  tiene distribución  $\Gamma(n, \lambda)$ , es decir que se distribuye según la densidad  $g_n(t)$  dada por:

$$g_n(t) = \begin{cases} \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

Llamemos  $D(t)$  al número de desintegraciones en el intervalo  $[0, t]$ . Entonces

$$D(t_0) = n \text{ si y sólo si } T_n \leq t_0 < T_{n+1}$$

Deducimos que:

$$\{D(t_0) = n\} = \{T_n \leq t_0\} - \{T_{n+1} \leq t_0\}$$

En consecuencia,

$$P\{D(t_0) = n\} = P\{T_n \leq t_0\} - P\{T_{n+1} \leq t_0\} = \int_0^{t_0} g_n(t) dt - \int_0^{t_0} g_{n+1}(t) dt$$

Integrando por partes, tenemos que:

$$\begin{aligned} \int_0^{t_0} g_{n+1}(t) dt &= \int_0^{t_0} \frac{\lambda^{n+1}}{n!} t^n e^{-\lambda t} dt \\ &= \frac{\lambda^{n+1}}{n!} \left[ t^n \frac{e^{-\lambda t}}{(-\lambda)} \Big|_0^{t_0} - \int_0^{t_0} n t^{n-1} \frac{e^{-\lambda t}}{(-\lambda)} dt \right] \\ &= \frac{\lambda^{n+1}}{n!} t_0^n \frac{e^{-\lambda t_0}}{(-\lambda)} - 0 - \int_0^{t_0} \frac{\lambda^{n+1}}{n!} n t^{n-1} \frac{e^{-\lambda t}}{(-\lambda)} dt \\ &= -\frac{\lambda^n}{n!} t_0^n e^{-\lambda t_0} + \int_0^{t_0} \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} dt \end{aligned}$$

$$= -\frac{\lambda^n}{n!} t_0^n e^{-\lambda t_0} + \int_0^{t_0} g_n(t) dt$$

En definitiva concluimos que la distribución del número de desintegraciones viene dada por una distribución de Poisson (proceso de Poisson):

$$P\{D(t_0) = n\} = \frac{(\lambda t_0)^n}{n!} e^{-\lambda t_0}$$

Como dijimos al comienzo de la sección, aunque hemos presentado la distribución exponencial y este cálculo de los tiempos de espera como modelo de la desintegración radiactiva, este mismo modelo se puede aplicar a otros procesos donde la hipótesis de falta de memoria resulte razonable como por ejemplo la llegada de eventos a un servidor informático, o los siniestros en una compañía de seguros. Esto explica la utilización de las distribuciones exponencial y de Poisson en muchas aplicaciones de las probabilidades.

## 3.9. Algunas densidades útiles en estadística

### 3.9.1. Las densidades $\chi^2$

En esta sección veremos algunas densidades que resultan especialmente útiles en estadística. Nos proporcionarán ejemplos interesantes de las técnicas de cambio de variables.

Sea  $X \sim N(0, 1)$  una variable aleatoria con distribución normal estándar. Utilizando la fórmula (3.22), encontramos que  $Y = X^2$  se distribuye según la densidad

$$f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] = \frac{1}{2\sqrt{y}} \left[ \frac{1}{\sqrt{2\pi}} e^{-y/2} + \frac{1}{\sqrt{2\pi}} e^{-y/2} \right]$$

o sea

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} \quad (y > 0)$$

Esta densidad se conoce como la densidad  $\chi^2$  (“ji-cuadrado”) con un grado de libertad [abreviada  $\chi_1^2$ ]. Comparando con (3.24), y utilizando que  $\Gamma(1/2) = \sqrt{\pi}$ , vemos que coincide con la densidad  $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ .

Sean ahora  $X_1, X_2, \dots, X_n$  variables aleatorias independientes con distribución normal estándar, y consideremos la variable aleatoria

$$Z_n = X_1^2 + X_2^2 + \dots + X_n^2$$

¿cuál es la distribución de  $Z_n$ ? Por lo anterior cada una de las  $X_i$  se distribuye según la densidad  $\chi_1^2 = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ , y la densidad de  $Z$  será (por la independencia) la convolución de la densidad  $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$   $n$  veces con sigo misma, que por el lema 3.7.1 da la densidad  $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ . Es decir, que la densidad de  $Z_n$  será

$$f_{Z_n}(z) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2} \quad (x > 0) \quad (3.30)$$

Esta densidad se conoce como densidad  $\chi^2$  con  $n$  grados de libertad [abreviada  $\chi_n^2$ ]. Las fórmulas (3.26) y (3.27) nos dicen que si  $Z \sim \chi_n^2$ , entonces

$$E[Z_n] = n, \quad \text{Var}[Z_n] = 2n$$

### 3.9.2. Densidad del cociente de dos variables aleatorias independientes

Supongamos que  $X$  e  $Y$  son variables aleatorias continuas independientes, con densidades  $f_X$  y  $f_Y$  respectivamente. Supongamos además que  $Y$  está concentrada en la semirrecta positiva  $(0, +\infty)$ . Querramos calcular la densidad del cociente  $T = U/V$ .

La densidad conjunta del vector aleatorio  $(X, Y)$  será  $f_X(x)f_Y(y)$  como consecuencia de independencia de las variables  $X$  e  $Y$ .

Consideramos ahora el cambio de variable  $(T, V) = \varphi(X, Y)$  donde

$$(u, v) = \varphi(x, y) = (x/y, y)$$

entonces la función inversa será

$$(x, y) = \varphi^{-1}(t, v) = (tv, v)$$

Y la diferencial de  $\varphi^{-1}$  es

$$D\varphi^{-1}(t, v) = \begin{pmatrix} v & t \\ 0 & 1 \end{pmatrix}$$

de modo que el Jacobiano es  $v$ . De acuerdo a la proposición 3.5.1, encontramos que el vector  $(T, V)$  se distribuye según la densidad conjunta

$$f_X(tv)f_Y(v)v$$

e integrando respecto la variable  $v$  podemos recuperar la densidad (marginal) de  $t$  que resulta ser:

$$T \simeq \int_0^\infty f_X(tv)f_Y(v)v \, dv \quad (3.31)$$

### 3.9.3. La densidad $t$ de Student

Sea  $X$  una variable aleatoria con distribución  $\chi^2$  con  $n$  grados de libertad,  $Y$  una variable aleatoria con distribución normal estándar y supongamos que  $X$  e  $Y$  son independientes. Queremos calcular la densidad de la variable aleatoria

$$T = \frac{\sqrt{\frac{X}{n}}}{Y}$$

[El porqué esta variable aleatoria es interesante, lo veremos más adelante al desarrollar conceptos de estadística]

Ya vimos que la densidad de  $X$  viene dada por (3.30) Consideramos  $\varphi : (0, +\infty) \rightarrow (0, +\infty)$  dada por

$$\varphi(x) = \sqrt{\frac{x}{n}}$$

es un difeomorfismo cuya inversa es  $\varphi^{-1}(y) = ny^2$ .

Aplicando la fórmula de cambio de variables, encontramos que la densidad de  $Y$  es

$$\begin{aligned} f_Y(y) &= \frac{(1/2)^{n/2}}{\Gamma(n/2)} (ny^2)^{n/2-1} e^{-ny^2/2} 2ny \, I_{(0,+\infty)}(y) \\ &= \frac{2n^{n/2}}{2^{n/2}\Gamma(n/2)} y^{n-1} e^{-ny^2/2} I_{(0,+\infty)}(y) \end{aligned}$$

Utilizando la fórmula (3.31), vemos que  $T$  se distribuye según la densidad

$$\begin{aligned} f_T(t) &= \int_0^\infty f_X(tv)f_Y(v)v \, dv = \frac{2n^{n/2}}{2^{n/2}\Gamma(n/2)\sqrt{2\pi}} \int_0^\infty e^{-t^2v^2/2} v^{n-1} e^{-nv^2/2} v \, dv \\ &= \frac{2^{(1-n)/2}n^{n/2}}{\Gamma(n/2)\sqrt{\pi}} \int_0^\infty e^{-(t^2+n)v^2/2} v^n \, dv \quad (t > 0) \end{aligned}$$

Hacemos el cambio de variable  $x = \frac{v^2}{2}(t^2 + n)$ , entonces esta integral se transforma en

$$\begin{aligned}
 f_T(t) &= \frac{2^{(1-n)/2} n^{n/2}}{\Gamma(n/2) \sqrt{\pi}} \frac{1}{n + t^2} \int_0^\infty e^{-x} \left( \frac{2x}{n + t^2} \right)^{(n-1)/2} dx \\
 &= \frac{n^{n/2}}{\Gamma(n/2) \sqrt{\pi}} \frac{1}{(n + t^2)^{(n+1)/2}} \int_0^\infty e^{-x} x^{(n-1)/2} dx \\
 &= \frac{n^{n/2}}{\Gamma(n/2) \sqrt{\pi}} \Gamma\left(\frac{n+1}{2}\right) \frac{1}{(n + t^2)^{(n+1)/2}} \\
 &= \frac{1}{\Gamma(n/2) \sqrt{n\pi}} \Gamma\left(\frac{n+1}{2}\right) \frac{n^{(n+1)/2}}{(n + t^2)^{(n+1)/2}}
 \end{aligned}$$

Finalmente obtenemos

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2) \sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \quad (t > 0) \quad (3.32)$$

Esta distribución se conoce como distribución t de Student con  $n$  grados de libertad.

### 3.10. Distribución Normal Multivariada

Falta escribir esta sección.

## Capítulo 4

# Convergencia de Variables Aleatorias, y Ley Fuerte de los Grandes Números

### 4.1. Los diferentes tipos de convergencia

#### Convergencia en probabilidad

En la teoría de probabilidades se utilizan frecuentemente diferentes nociones de convergencia de una sucesión  $(X_n)_{n \in \mathbb{N}}$  de variables aleatorias.

La primera noción importante es la de convergencia en probabilidad, que aparece en el teorema de Bernoulli (ley débil de los grandes números).

**Definición 4.1.1** Sea  $(X_n)_{n \in \mathbb{N}}$  una sucesión de variables aleatorias, definidas sobre un mismo espacio de probabilidad  $(\Omega, \mathcal{E}, P)$ . Se dice que  $(X_n)$  **converge en probabilidad** a la variable  $X$  si para todo  $\varepsilon > 0$ , tenemos que

$$P\{|X - X_n| > \varepsilon\} \rightarrow 0 \text{ cuando } n \rightarrow +\infty$$

**Notación:**

$$X_n \xrightarrow{P} X$$

**Observación:** Si  $(X_n)$  converge en probabilidad a  $X$ , cualquier subsucesión de  $(X_n)$  también converge en probabilidad a  $X$ .

Veamos algunas propiedades de la convergencia en probabilidad:

**Proposición 4.1.1 (Unicidad del límite)** Si  $X_n \xrightarrow{P} X$  y  $X_n \xrightarrow{P} Y$ , entonces  $X = Y$  con probabilidad 1.

**Prueba:** Por la desigualdad triangular,

$$|X - Y| \leq |X - X_n| + |X_n - Y|$$

Entonces

$$P\{|X - Y| > \varepsilon\} \leq P\{|X - X_n| > \varepsilon/2\} + P\{|X_n - Y| > \varepsilon/2\}$$

Deducimos que para todo  $\varepsilon > 0$ ,

$$P\{|X - Y| > \varepsilon\} = 0$$

Como

$$\{X \neq Y\} = \bigcup_{n \in \mathbb{N}} \left\{ |X - Y| > \frac{1}{n} \right\}$$

Por la  $\sigma$ -subaditividad de  $P$ , deducimos que:

$$P\{X \neq Y\} \leq \sum_{n=1}^{\infty} P\left\{ |X - Y| > \frac{1}{n} \right\} = 0$$

□

**Proposición 4.1.2** Si  $X_n \xrightarrow{P} X$  y  $c \in \mathbb{R}$ , entonces  $cX_n \xrightarrow{P} cX$ .

**Prueba:** Si  $c \neq 0$ , tenemos que

$$P\{|cX_n - cX| > \varepsilon\} = P\left\{ |X_n - X| > \frac{\varepsilon}{|c|} \right\} \rightarrow 0 \text{ cuando } n \rightarrow +\infty$$

□

**Proposición 4.1.3** Si  $X_n \xrightarrow{P} X$  e  $Y_n \xrightarrow{P} Y$ , entonces  $X_n + Y_n \xrightarrow{P} X + Y$ .

**Prueba:**

$$P\{|(X + Y) - (X_n + Y_n)| > \varepsilon\} \leq P\{|X - X_n| > \varepsilon/2\} + P\{|Y - Y_n| > \varepsilon/2\}$$

□

**Lema 4.1.1** Si  $X_n \xrightarrow{P} X$ , entonces  $(X_n)$  está acotada en probabilidad, en el siguiente sentido, dado  $\varepsilon > 0$  existen  $M = M_\varepsilon$  y  $n_0 = n_0(\varepsilon)$  tales que

$$\forall n \geq n_0(\varepsilon) : P\{|X_n| > M\} < \varepsilon$$

**Prueba:** Notamos que

$$\sum_{k=1}^{\infty} P\{k-1 \leq |X| < k\} = 1$$

es una serie convergente, por consiguiente dado  $\varepsilon > 0$ , existirá un  $k_0$  tal que:

$$\sum_{k=k_0+1}^{\infty} P\{k-1 \leq |X| \leq k\} < \frac{\varepsilon}{2}$$

Es decir que:

$$P\{|X| \geq k_0\} < \frac{\varepsilon}{2}$$

De la desigualdad triangular,

$$|X_n| \leq |X_n - X| + |X|$$

Deducimos que:

$$P\{|X_n| > k_0 + \delta\} \leq P\{|X_n - X| > \delta\} + P\{|X| > k_0\}$$

y en consecuencia que

$$P\{|X_n| > k_0 + \delta\} \leq \varepsilon$$

si  $n \geq n_0(\varepsilon)$ .

Esto prueba la afirmación del lema, con  $M = k_0 + \delta$ . □

**Lema 4.1.2** Si  $X_n \xrightarrow{P} 0$  e  $Y_n$  “está acotada en probabilidad”, entonces  $X_n Y_n \xrightarrow{P} 0$ .

**Prueba:**

$$\begin{aligned} P\{|X_n Y_n| > \varepsilon\} &= P\left\{|X_n| > \frac{\varepsilon}{|Y_n|}\right\} \\ &\leq P\left\{|X_n| > \frac{\varepsilon}{|Y_n|} \wedge |Y_n| \leq M\right\} + P\left\{|X_n| > \frac{\varepsilon}{|Y_n|} \wedge |Y_n| > M\right\} \\ &\leq P\left\{|X_n| > \frac{\varepsilon}{M}\right\} + P\{|Y_n| > M\} < \varepsilon \end{aligned}$$

si  $n \geq n_0(\varepsilon)$ . □



**Corolario 4.1.1** Si  $X_n \xrightarrow{P} X$  e  $Y_n \xrightarrow{P} Y$ , entonces  $X_n Y_n \xrightarrow{P} XY$ .

**Prueba:** Utilizamos el truco habitual de “sumar y restar”:

$$XY - X_n Y_n = XY - X_n Y + X_n Y - X_n Y_n = (X - X_n)Y + X_n(Y - Y)$$

Entonces como  $X - X_n \xrightarrow{P} 0$  e  $Y$  está acotada en probabilidad, deducimos que  $(X - X_n)Y \xrightarrow{P} 0$ . Similarmente, como  $Y_n - Y \xrightarrow{P} 0$  y  $X_n$  está acotada en probabilidad (por la proposición 4.1.1, deducimos que  $X_n(Y_n - Y) \xrightarrow{P} 0$ ). Tenemos entonces que  $X_n Y_n - XY \xrightarrow{P} 0$ , y en consecuencia  $X_n Y_n \xrightarrow{P} XY$  (por la proposición 4.1.3)  $\square$

## Convergencia en Distribución

**Definición 4.1.2** Se dice que una sucesión de variables aleatorias  $X_n$  **converge en distribución** a la variable aleatoria  $X$ , si

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

en cada  $x$  en el que  $F_X$  sea continua. **Notación:**

$$X_n \xrightarrow{D} X$$

**Ejemplo 4.1.1** Supongamos que  $X_n \sim N(0, \sigma^2)$ . Entonces cuando  $\sigma \rightarrow 0$ ,  $X_n$  converge en distribución a la variable aleatoria  $X$  con  $P\{X = 0\} = 1$ , cuya distribución  $F$  (que es la función escalón de Heavside) es discontinua en cero. Este ejemplo muestra porqué resulta natural pedir que haya convergencia sólo en los puntos de continuidad de  $F$ .

**Proposición 4.1.4** Si  $X_n \xrightarrow{D} X$  y  $X_n \xrightarrow{D} Y$ , entonces  $F_X = F_Y$  ( $X$  e  $Y$  están idénticamente distribuidas)

**Prueba:**  $F_X(x) = F_Y(x)$  en cada  $x$  que sea simultáneamente punto de continuidad de  $F_X$  y  $F_Y$ . Pero  $F_X$  y  $F_Y$  son crecientes, y tienen por lo tanto a lo sumo una cantidad numerable de discontinuidades. Deducimos que  $F_X(x) = F_Y(x)$  para los  $x$  en un subconjunto denso de  $\mathbb{R}$ , y entonces para todo  $x$  ya que ambas son continuas por la derecha.  $\square$

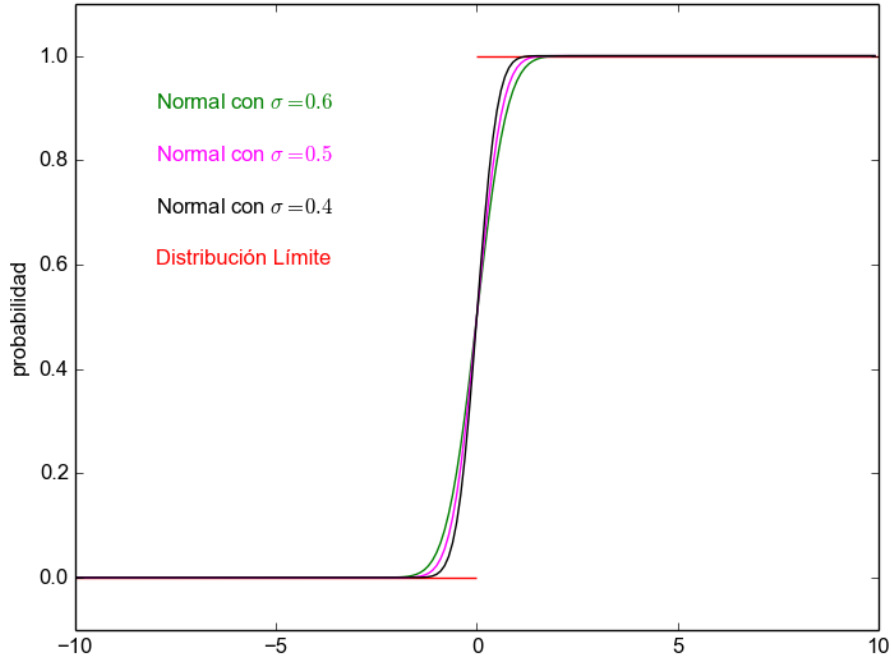


Figura 4.1: Convergencia en distribución de la densidad normal al escalón cuando  $\sigma \rightarrow 0$ .

**Proposición 4.1.5** Si  $X_n \xrightarrow{D} X$  y  $c \in \mathbb{R}$  es una constante, entonces  $cX_n \xrightarrow{D} cX$  y  $X_n + c \xrightarrow{D} X + c$ .

**Observación 4.1.1** Sin embargo, no es cierto en general que si

$$X_n \xrightarrow{D} X, Y_n \xrightarrow{D} Y \Leftrightarrow X_n + Y_n \xrightarrow{D} X + Y$$

Para comprobarlo basta tomar  $X_n, Y_n \sim N(0, \sigma^2)$  independientes,  $X \sim N(0, 1)$  y  $Y = -X$ . Entonces  $X_n + Y_n \sim N(0, 2\sigma^2)$  que no converge en distribución a cero cuando  $\sigma \rightarrow 0$ , aunque  $X + Y = 0$ . Este ejemplo patológico se explica porque la convergencia en distribución se refiere más a las distribuciones de las variables, que a las variables en sí mismas.

**Definición 4.1.3** Se dice que la sucesión  $(X_n)$  de variables aleatorias **converge casi seguramente** a la variable  $X$  si

$$P \left\{ \lim_{n \rightarrow +\infty} X_n = X \right\} = 1$$

**Notación:**

$$X_n \xrightarrow{c.s.} X$$

## 4.2. Relación entre los modos de convergencia

**Proposición 4.2.1** Si  $X_n \xrightarrow{c.s.} X$ , entonces  $X_n \xrightarrow{P} X$ .

**Prueba:** Notamos que:

$$\{\omega \in \Omega : X_n(\omega) \not\xrightarrow{c.s.} X(\omega)\} = \bigcup_{k=1}^{\infty} \bigcap_{n_0=1}^{\infty} \bigcup_{n \geq n_0} \left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| > \frac{1}{k} \right\}$$

Como  $X_n \xrightarrow{c.s.} X$ , este conjunto tiene probabilidad 0. En consecuencia, también tienen probabilidad cero los eventos (más pequeños)

$$\bigcap_{n_0=1}^{\infty} \bigcup_{n \geq n_0} \left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| > \frac{1}{k} \right\}$$

Como los eventos:

$$A_{k,n_0} = \bigcup_{n \geq n_0} \left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| > \frac{1}{k} \right\}$$

son decrecientes, deducimos (por la continuidad de la probabilidad) que:

$$\lim_{n_0 \rightarrow +\infty} P(A_{k,n_0}) = 0$$

Vale decir que si elegimos  $n_0$  suficientemente grande,  $P(A_{k,n_0}) < \delta$ , y en consecuencia

$$P \left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| > \frac{1}{k} \right\} < \delta$$

para todo  $n \geq n_0$ . Deducimos que  $X_n$  tiende en probabilidad a  $X$ .  $\square$

**Proposición 4.2.2** Sea  $(X_n)$  una sucesión de variables aleatorias tal que  $X_n \xrightarrow{P} X$ , entonces  $X_n \xrightarrow{D} X$ .

(La prueba de este teorema que aparece en versiones anteriores de este apunte debe ser corregida)

**Proposición 4.2.3** Si  $X_n \xrightarrow{D} 0$ , entonces  $X_n \xrightarrow{P} 0$ .

### 4.3. El lema de Borel-Cantelli

**Lema 4.3.1** (de Borel-Cantelli [19], [20]) Consideramos una sucesión  $(A_n)_{n \in \mathbb{N}}$  de eventos, y consideramos el evento “ocurren infinitos  $A_n$ ”, es decir:

$$A_\infty = \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A_n$$

entonces

i) Si

$$\sum_{n=1}^{\infty} P(A_n) < +\infty \quad (4.1)$$

entonces, con probabilidad 1 ocurre un número finito de tales sucesos. Es decir

$$P(A_\infty) = 0$$

ii) Si los  $A_n$  son eventos independientes, y

$$\sum_{n=1}^{\infty} P(A_n) = +\infty \quad (4.2)$$

entonces, con probabilidad 1 ocurren infinito s  $A_n$ . Es decir,

$$P(A_\infty) = 1$$

**Prueba: Demostración de i):** Dado  $\varepsilon > 0$ , teniendo en cuenta la hipótesis (4.1), podemos elegir  $k$  tal que

$$\sum_{n=k}^{\infty} P(A_n) < \varepsilon$$

Entonces, por la  $\sigma$ -subaditividad de la probabilidad:

$$P\left(\bigcup_{n \geq k} A_n\right) \leq \sum_{n=k}^{\infty} P(A_n) < \varepsilon$$

y como la probabilidad es creciente:

$$P(A_{\infty}) \leq P\left(\bigcup_{n \geq k} A_n\right) < \varepsilon$$

Como,  $\varepsilon$  es arbitrario, deducimos que:

$$P(A_{\infty}) = 0$$

**Demostración de ii):** Miremos el complemento de  $A_{\infty}$ , que es según las leyes de De Morgan:

$$A_{\infty}^c = \bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} A_n^c$$

Entonces, tenemos que:

$$P\left(\bigcap_{n=k}^l A_n^c\right) = \prod_{n=k}^l P(A_n^c) = \prod_{n=k}^l P(A_n^c)$$

ya que como los eventos  $(A_n)$  son independientes, también lo son sus complementos. Ahora utilizando la desigualdad elemental

$$1 - x \leq e^{-x} \quad x \in [0, 1],$$

tenemos que:

$$P\left(\bigcap_{n=k}^l A_n^c\right) \leq \prod_{n=k}^l e^{-P(A_n)} = \exp\left(-\sum_{n=k}^l P(A_n)\right)$$

y en consecuencia utilizando que la probabilidad es creciente, y la hipótesis (4.2), deducimos que:

$$P\left(\bigcap_{n=k}^{\infty} A_n^c\right) = 0$$

(ya que el segundo miembro de la desigualdad anterior tiende a cero cuando  $l \rightarrow \infty$ ). Entonces, por la  $\sigma$ -subaditividad de la probabilidad,

$$P(A_\infty^c) \leq \sum_{k=1}^{\infty} P\left(\bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} A_n^c\right) = 0$$

deducimos que

$$P(A_\infty) = 1$$

□

## 4.4. La ley fuerte de los grandes números

**Teorema 4.4.1** *Sea  $(X_n)_{n \in \mathbb{N}}$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con  $m_4 = E[X_n^4] < +\infty$ . Sea  $\mu = E[X_i]$  entonces*

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{c.s.} \mu$$

cuando  $n \rightarrow +\infty$ .

**Nota:** La hipótesis de que el cuarto momento  $m_4$  es finito no es necesaria para la validez de este teorema, pero facilitará enormemente la demostración. Una demostración del teorema sin esta hipótesis (ley fuerte de Kolmogorov) se da en el apéndice E.

**Prueba:** Podemos suponer que  $\mu = 0$  (cambiando sino  $X_n$  por  $X_n - \mu$ ). Notamos  $S_n = X_1 + X_2 + \dots + X_n$ . Notemos en primer lugar que las  $X_i$  tienen variancia  $\sigma^2 = E[X_i^2]$  finita.

Calculemos el cuarto momento de  $S_n$ . Por la linealidad de la esperanza tenemos que:

$$E[S_n^4] = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}]$$

Como las  $X_i$  son independientes, notamos que

$$E[X_{i_1} X_{i_2} X_{i_3} X_{i_4}] = 0$$

salvo en el caso en que los subíndices son todos iguales, o si son iguales por pares (utilizando que la esperanza del producto es el producto de las

esperanzas cuando las variables son independientes, y que la esperanza de cada variable es cero).

Ahora cuando  $i \neq j$ ,  $X_i^2$  es independiente de  $X_j^2$  en consecuencia:

$$E[X_i^2 X_j^2] = E[X_i^2] E[X_j^2] = \sigma^4$$

Cada uno de estos términos aparece  $\binom{4}{2} = 6$  veces, y hay  $\binom{n}{2} = \frac{n(n-1)}{2}$  formas de elegir 2 de las  $n$  variables aleatorias. Entonces:

$$E[S_n^4] = \sum_{i=1}^n E[X_i^4] + 6 \sum_{i,j=1, i \neq j}^n E[X_i^2 X_j^2] \leq nm_4 + 3n(n-1)\sigma^4 \leq Cn^2$$

donde  $C$  es una constante. Deducimos que:

$$E \left[ \left( \frac{S_n}{n} \right)^4 \right] \leq \frac{C}{n^2}$$

y entonces, dado  $\varepsilon > 0$ , por la desigualdad de Markov, tenemos que:

$$P \left\{ \left| \frac{S_n}{n} \right| > \varepsilon \right\} \leq \frac{C}{\varepsilon^4 n^2}$$

Como la serie

$$\sum_{n=1}^{\infty} \frac{C}{\varepsilon^4 n^2}$$

es convergente, el lema de Borel Cantelli implica que si llamamos  $A_{n,\varepsilon}$  al evento

$$A_{n,\varepsilon} = \left\{ \omega \in \Omega : \left| \frac{S_n(\omega)}{n} \right| > \varepsilon \right\}$$

entonces, con probabilidad 1 ocurren sólo finitos de los sucesos  $A_{n,\varepsilon}$ , es decir que el evento

$$A_{\infty,\varepsilon} = \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A_{n,\varepsilon}$$

tiene probabilidad cero.

Tomando  $\varepsilon = 1/m$ , con  $m \in \mathbb{N}$ , y usando la  $\sigma$  sub-aditividad de la probabilidad, vemos que el evento:

$$\begin{aligned} B &= \left\{ \omega \in \Omega : \frac{S_n(\omega)}{n} \not\rightarrow 0 \right\} = \left\{ \omega \in \Omega : \exists m \in \mathbb{N} \forall k \in \mathbb{N} \exists n \geq k : \left| \frac{S_n(\omega)}{n} \right| > \frac{1}{m} \right\} \\ &= \bigcup_{m \in \mathbb{N}} A_{\infty, 1/m} \end{aligned}$$

tiene probabilidad cero, ya que es la unión numerable de eventos de probabilidad cero. En consecuencia,  $P(B^c) = 1$ , es decir que  $S_n/n$  tiende a cero con probabilidad 1.  $\square$

#### 4.4.1. Un ejemplo: La ley fuerte de Borel para ensayos de Bernoulli

Un primer ejemplo que podemos considerar es el esquema de ensayos de Bernoulli, que consideramos en el capítulo 2. Recordamos que en este esquema, un experimento con dos posibles resultados (llamados convencionalmente éxito y fracaso) se repite infinitas veces en condiciones independientes. Llamamos  $p$  a la probabilidad del éxito.

Como antes, consideramos entonces las variables aleatorias de Bernoulli:

$$X_i = \begin{cases} 1 & \text{si el } i\text{-ésimo experimento fue un éxito} \\ 0 & \text{si el } i\text{-ésimo experimento fue un fracaso} \end{cases}$$

Entonces  $S_n$  representa la cantidad de éxitos en los  $n$  primeros ensayos, y

$$f_n = \frac{S_n}{n}$$

la frecuencia relativa de éxitos en los  $n$  primeros ensayos. La ley fuerte de los grandes números afirma entonces que

$$f_n \rightarrow p \text{ con probabilidad } 1 \quad (4.3)$$

donde llamamos  $p$  a la probabilidad del éxito (Este enunciado que se conoce como *la ley fuerte de los grandes números de Borel*, es un caso particular del teorema anterior. Notamos que la hipótesis de que las  $X_i$  tengan cuarto momento finito, se satisface trivialmente ).



¿Pero qué significa exactamente esto? ¿cuál es el espacio muestral para este experimento compuesto?. Como dijimos anteriormente, el espacio muestra podemos representarlo como

$$\Omega = \{\omega = (x_1, x_2, \dots, x_n, \dots) : \omega_i = 0 \text{ o } \omega_i = 1\} = \{0, 1\}^{\mathbb{N}}$$

donde  $\omega_i$  representará el resultado del  $i$ -ésimo ensayo. Entonces, las variables aleatorias  $X_i$  se definen sencillamente por:

$$X_i(\omega) = \omega_i$$

Para poder darle sentido a la afirmación (4.3), debemos decir cómo asignamos probabilidades en el espacio  $\Omega$ . El caso más sencillo es cuando  $p = q = 1/2$  (éxito y fracaso equiprobables).

En se caso, definamos para ello la función

$$\phi : \Omega \rightarrow [0, 1]$$

por

$$\phi(\omega) = \sum_{i=1}^{\infty} \frac{\omega_i}{2^i}$$

En otras palabras, para cada  $\omega \in \Omega$ ,  $\phi(\omega)$  será el número en  $[0, 1]$  cuyo desarrollo binario tiene por dígitos a los  $\omega_i$ .

Podemos definir entonces la sigma-álgebra  $\mathcal{E}$  como:

$$\mathcal{E} = \{E \subset \Omega : \phi(E) \text{ es un subconjunto boreliano del intervalo } [0, 1]\}$$

y la probabilidad  $P$  por

$$P(E) = m(\phi(E))$$

donde  $m$  denota la medida de Lebesgue (ver la discusión en la sección 1.6).

**Ejercicio:** Comprobar que la función  $P$  así definida asigna correctamente las probabilidades, en el sentido de que

$$P(\{\omega \in \Omega : \omega_1 = x_1, \omega_2 = x_2, \dots, \omega_n = x_n\}) = 2^{-n}$$

donde  $k = S_n(\omega)$ . En particular, las variables aleatorias  $X_1, X_2, \dots, X_n$  resultan independientes. Ayuda: notar que  $\phi(E)$  consta en este caso de una unión finita de intervalos.

Entonces, cuando  $p = 1/2$ , la afirmación (4.3) puede interpretarse equivalentemente, como la afirmación de que para casi todo número en el intervalo  $[0, 1]$ , si  $f_n$  designa la frecuencia de dígitos uno en los primeros  $n$  lugares de su desarrollo binario, se tiene que  $f_n \rightarrow 1/2$ . En esta afirmación, como es usual en la teoría de la medida, significa “salvo quizás para un conjunto de medida de Lebesgue cero”.

#### 4.4.2. Números Normales

Una generalización de la idea anterior es considerar desarrollos en otra base de numeración  $b$ , con  $b \geq 2$ . Entonces pensamos en un experimento cuyos posibles resultados son los dígitos  $0, 1, \dots, b-1$  de la base  $b$ , que consideramos equiprobables y lo repetimos infinitas veces.

$$\Omega = D^{\mathbb{N}} \text{ siendo } D = \{0, 1, \dots, b-1\}$$

Ahora definimos la función

$$\phi : \Omega \rightarrow [0, 1]$$

por

$$\phi(\omega) = \sum_{i=1}^{\infty} \frac{\omega_i}{b^i}$$

Fijamos un dígito  $d \in D$  y nos preguntamos por la frecuencia relativa de ese dígito en los primeros  $n$  lugares del número real  $x = \phi(\omega)$

$$f_n = \frac{\#\{i : 1 \leq i \leq n, \omega_i = d\}}{n}$$

que podremos escribir como antes en la forma

$$f_n = \frac{S_n}{n}$$

si definimos las variables  $X_i$  por

$$X_i = \begin{cases} 1 & \text{si } \omega_i = d \\ 0 & \text{si } \omega_i \neq d \end{cases}$$

Como antes, asignamos las probabilidades en  $\Omega$  por:

$$P(E) = m(\varphi_b(E))$$

y resulta que

$$P(\{\omega \in \Omega : \omega_1 = d_1, \omega_2 = d_2, \dots, \omega_n = d_n\}) = b^{-n}$$

$$P(\{\omega \in \Omega : X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}) = p^k q^{n-k}$$

donde  $k = S_n(\omega)$ ,  $p = 1/b$ ,  $q = 1 - 1/b$ . En particular las variables  $X_i$  son de nuevo independientes. Se deduce

$$f_n \rightarrow \frac{1}{b} \quad (4.4)$$

con probabilidad 1, o lo que es equivalente  $f_n$  tiende a cero para casi todo  $x \in [0, 1]$  (o sea: salvo para los  $x$  en un conjunto de medida cero en el sentido de Lebesgue). Los números que verifican la relación (4.4) para todo dígito  $d \in D$  fueron denominados por Borel *números (simplemente) normales* en la base  $b$ . Se deduce de lo demostrado que casi todo número es simplemente normal en la base  $b$ .

Más aún, Borel definió los *números absolutamente normales* como aquellos que son simplemente normales en cualquier base  $b \geq 2$ . Como la unión numerable de conjuntos de medida cero en el sentido de Lebesgue también tiene medida cero, se deduce el siguiente teorema:

**Teorema 4.4.2** (de Borel, [19]) *Casi todo número real del intervalo  $[0, 1]$  es absolutamente normal.*

Nota: Aunque este teorema implica que existen números absolutamente normales, su prueba no es constructiva en el sentido que no nos provee ningún ejemplo de un número absolutamente normal. El primer ejemplo fue dado por Sierpinski en 1916 [25]. Ver también [26] para una versión computable de la construcción de Sierpinski.

## 4.5. El Teorema de Helly

**Teorema 4.5.1 (Helly)** *Supongamos que  $F_n : [a, b] \rightarrow \mathbb{R}$  es una sucesión de funciones de distribución tales que  $F_n(x) \rightarrow F(x)$  en cada punto de continuidad de  $F(x)$ , entonces:*

$$\int_a^b \varphi(x) dF_n(x) \rightarrow \int_a^b \varphi(x) dF(x) \quad (4.5)$$

para toda función continua  $\varphi \in C[a, b]$ .

**Prueba:** Dado  $\varepsilon > 0$ , por el corolario D.0.1 del apéndice D (teorema de existencia para la integral de Riemman-Stieltjes; corolario sobre la convergencia uniforme respecto de la función de distribución), existirá un  $\delta > 0$  tal que:

$$\left| \int_a^b \varphi(x) dF_n(x) - S_\pi(\varphi, F_n) \right| < \varepsilon$$

para todo  $n$ , y también

$$\left| \int_a^b \varphi(x) dF(x) - S_\pi(\varphi, F) \right| < \varepsilon$$

para cualquier partición  $\pi$  de  $[a, b]$  que verifique que  $|\pi| < \delta$  (Pues  $F_n(1) - F_n(0) \leq 1$ ).

Fijemos una partición cualquiera  $\pi$  de  $[a, b]$  tal que  $|\pi| < \delta$ . Claramente podemos elegir los puntos de subdivisión de esta partición  $\pi$  para que sean puntos de continuidad de  $F$  (pues el conjunto de puntos de discontinuidad de  $F$  es a lo sumo numerable, y por lo tanto su conjunto de puntos de continuidad es denso en  $[a, b]$ ).

Entonces notamos que como hay finitos puntos en la partición, claramente tendremos que:

$$\lim_{n \rightarrow +\infty} S_\pi(\varphi, F_n) = S_\pi(\varphi, F)$$

Es decir, que dado  $\varepsilon > 0$ , existirá un  $n_0$ , tal que si  $n \geq n_0$ ,

$$|S_\pi(\varphi, F_n) - S_\pi(\varphi, F)| < \varepsilon$$

En consecuencia, si  $n \geq n_0$ ,

$$\begin{aligned} & \left| \int_a^b \varphi(x) dF_n(x) - \int_a^b \varphi(x) dF(x) \right| \\ & \leq \left| \int_a^b \varphi(x) dF_n(x) - S_\pi(\varphi, F_n) \right| \\ & \quad + |S_\pi(\varphi, F_n) - S_\pi(\varphi, F)| \\ & \quad + \left| S_\pi(\varphi, F) - \int_a^b \varphi(x) dF(x) \right| < 3\varepsilon \end{aligned}$$

□

Un resultado análogo se verifica para integrales en intervalos infinitos:

**Teorema 4.5.2** Supongamos que  $F_n : [a, b] \rightarrow \mathbb{R}$  es una sucesión de funciones de distribución tales que  $F_n(x) \rightarrow F(x)$  en cada punto de continuidad de  $F(x)$ , entonces:

$$\int_{-\infty}^{\infty} \varphi(x) dF_n(x) \rightarrow \int_{-\infty}^{\infty} \varphi(x) dF(x) \quad (4.6)$$

para toda función continua acotada  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ .

**Prueba:** Supongamos que  $|\varphi(x)| \leq M \forall x \in \mathbb{R}$ . Dado  $\varepsilon > 0$ , podemos elegir  $R > 0$  tal que:

$$F(R) - F(-R) = \int_{|x|>R} dF(x) < \frac{\varepsilon}{M}$$

y por lo tanto

$$\left| \int_{|x|>R} \varphi(x) dF_n(x) \right| < 2\varepsilon$$

Además, podemos suponer que  $R$  y  $-R$  son puntos de continuidad de  $F$ . Entonces, como  $F_n(R) \rightarrow F(R)$  y  $F_n(-R) \rightarrow F(-R)$  cuando  $n \rightarrow +\infty$ , podemos elegir  $n_1$  tal que para  $n \geq n_1$  se verifique

$$F_n(R) - F_n(-R) = \int_{|x|>R} dF_n(x) < \frac{2\varepsilon}{M}$$

y por lo tanto:

$$\left| \int_{|x|>R} \varphi(x) dF_n(x) \right| < 2\varepsilon$$

y en virtud del teorema anterior, podemos elegir un  $n_2$  tal que si  $n \geq n_2$  se verifica:

$$\left| \int_{-R}^R \varphi(x) dF_n(x) - \int_{-R}^R \varphi(x) dF(x) \right| < \varepsilon$$

Entonces, tendremos que:

$$\begin{aligned} & \left| \int_{-\infty}^{\infty} \varphi(x) dF_n(x) - \int_{-\infty}^{\infty} \varphi(x) dF(x) \right| \\ & \leq \left| \int_{-\infty}^{\infty} \varphi(x) dF_n(x) - \int_{-R}^R \varphi(x) dF(x) \right| \end{aligned}$$

$$\begin{aligned}
 & + \left| \int_{-R}^R \varphi(x) dF_n(x) - \int_{-R}^R \varphi(x) dF(x) \right| \\
 & + \left| \int_{-\infty}^{\infty} \varphi(x) dF(x) - \int_{-\infty}^{\infty} \varphi(x) dF(x) \right| < 4\varepsilon
 \end{aligned}$$

□

**Corolario 4.5.1** Si  $(X_n)$  es una sucesión de variables aleatorias tales que  $X_n \xrightarrow{D} X$ , entonces  $E[\varphi(X_n)] \rightarrow E[\varphi(X)]$  para toda función continua acotada.

## 4.6. El Recíproco del teorema de Helly

**Teorema 4.6.1** Si  $(X_n)$  es una sucesión de variables aleatorias tales que  $E[\varphi(X_n)] \rightarrow E[\varphi(X)]$  para toda función continua acotada, entonces  $X_n \xrightarrow{D} X$ .

**Prueba:** Tenemos que probar que  $F_{X_n}(x_0) \rightarrow F_X(x_0)$  cuando  $n \rightarrow +\infty$ , para cada punto de continuidad  $x_0$  de  $F_X$ . Para ello la idea es aproximar el indicador (función característica)  $I_{(-\infty, x_0]}$  del intervalo  $(-\infty, x_0]$  por funciones continuas acotadas.

Comencemos aproximando por la derecha. Para ello construimos para  $\delta > 0$ , las funciones

$$\varphi_\delta(x) = \begin{cases} 1 & \text{si } x < x_0 \\ 1 - (x - x_0)/\delta & \text{si } x_0 \leq x \leq x_0 + \delta \\ 0 & \text{si } x > x_0 + \delta \end{cases}$$

y notamos que las  $\varphi_\delta$  son continuas y acotadas. Entonces, dado  $\varepsilon > 0$ , por la hipótesis, existirá un  $n_1$  tal que si  $n \geq n_1$  tenemos que,

$$|E[\varphi_\delta(X_n)] - E[\varphi(X)]| < \frac{\varepsilon}{2}$$

Por otra parte, afirmamos que si  $\delta$  es suficientemente pequeño,

$$|E[\varphi_\delta(X)] - F_X(x_0)| < \frac{\varepsilon}{2} \tag{4.7}$$

En efecto,

$$\begin{aligned}
 |E[\varphi_\delta(X)] - F_X(x_0)| &= \left| \int_{-\infty}^{x_0} \varphi(x) dF_X(x) - \int_{-\infty}^{x_0} dF_X(x) \right| = \\
 &= \left| \int_{x_0}^{x_0+\delta} \varphi(x) dF_X(x) \right| \leq F_X(x_0 + \delta) - F_X(x_0)
 \end{aligned}$$

(acotando la integral usando el lema 3.2.3). Entonces, la afirmación se deduce de la continuidad (por la derecha)<sup>1</sup> de la función de distribución  $F_X$ .

Como consecuencia, deducimos que si  $n \geq n_0$ , tenemos que:

$$\begin{aligned} F_{X_n}(x_0) &= P\{X \leq x_0\} = E[I_{(-\infty, x_0]}(X)] \leq E[\varphi_\delta(X_n)] = \\ &E[\varphi_\delta(X)] + \frac{\varepsilon}{2} \leq F_X(x_0) + \varepsilon \end{aligned}$$

Para probar que  $F_{X_n}(x_0) \rightarrow F_X(x_0)$ , necesitamos demostrar también una desigualdad en el sentido contrario.

Para ello, aproximamos  $I_{(-\infty, x_0]}$  desde la izquierda: para ello, construimos las funciones

$$\varphi_{-\delta}(x) = \begin{cases} 1 & \text{si } x < x_0 - \delta \\ 1 - (x - (x_0 - \delta))/\delta & \text{si } x_0 - \delta \leq x \leq x_0 \\ 0 & \text{si } x > x_0 \end{cases}$$

Afirmamos que si elegimos  $\delta$  suficientemente pequeño,

$$|E[\varphi_{-\delta}(X)] - F_X(x_0)| < \frac{\varepsilon}{2} \quad (4.8)$$

ya que tenemos que:

$$\begin{aligned} |E[\varphi_{-\delta}(X)] - F_X(x_0)| &= \left| \int_{-\infty}^{x_0} \varphi_{-\delta}(x) dF_X(x) - \int_{-\infty}^{x_0} dF_X(x) \right| = \\ &= \left| \int_{x_0}^{x_0+\delta} \varphi_{-\delta}(x) dF_X(x) \right| \leq F_X(x_0) - F_X(x_0 - \delta) \end{aligned}$$

y la afirmación se deduce ahora de la continuidad de  $F_X$  en  $x_0$  (por la izquierda)

Entonces, acotando como antes tenemos que si  $n \geq n_0$ :

$$F_{X_n}(x_0) = P\{X_n \leq x_0\} \geq E[\varphi_{-\delta}(X_n)] \geq E[\varphi_{-\delta}(X)] - \frac{\varepsilon}{2} \geq F_X(x_0) - \varepsilon$$

Concluimos que si  $n \geq n_0$ ,

$$|F_{X_n}(x_0) - F_X(x_0)| < \varepsilon$$

□

**Una observación:** Con un poco más de trabajo, se puede probar que es suficiente que:

$$E[\varphi(X_n)] \rightarrow E[\varphi(X)]$$

para toda función continua  $\varphi$  con soporte compacto.

---

<sup>1</sup>Notemos que en este punto, no hemos utilizado todavía toda la fuerza de la hipótesis de que  $x_0$  es un punto de continuidad de  $F_X$ , ya que  $F_X$  siempre es continua por la derecha.

## 4.7. El Principio de Selección de Helly

Veremos en esta sección un teorema de compacidad para la convergencia en distribución.

**Teorema 4.7.1** *Supongamos que  $(F_n)_{n \in \mathbb{N}}$  es una sucesión de funciones de distribución. Entonces existe una subsucesión  $F_{n_k}$  y una función  $F : \mathbb{R} \rightarrow \mathbb{R}$  creciente y continua por la derecha, tal que*

$$\lim_{k \rightarrow +\infty} F_{n_k}(x) = F(x)$$

para todo punto de continuidad  $x$  de  $F$ .

**Observación 4.7.1** *La función límite  $F$  puede no ser una función de distribución. Por ejemplo si  $a + b + c = 1$ , y*

$$F_n(x) = aI_{[n, +\infty)}(x) + bI_{[-n, +\infty)} + cG(x)$$

donde  $G$  es alguna función de distribución, entonces

$$F_n(x) \rightarrow F(x) = b + cG(x) \text{ cuando } n \rightarrow +\infty$$

y tenemos que

$$\lim_{x \rightarrow -\infty} F(x) = b, \quad \lim_{x \rightarrow +\infty} F(x) = b + c = 1 - a$$

Luego se produce un fenómeno de “escape de masa al infinito”.

**Prueba:** Utilizando el método diagonal de Cantor (y la numerabilidad de los racionales), podemos construir una subsucesión  $F_{n_k}$  de  $F_n$  tal que

$$\lim_{k \rightarrow +\infty} F_{n_k}(q) = G(q)$$

exista para todo  $q \in \mathbb{Q}$  (es decir todo  $q$  racional).

La función  $G$  puede no ser continua por la derecha, pero si definimos

$$F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\}$$

obtenemos una función continua por la derecha pues

$$\begin{aligned} \lim_{x_n \downarrow x} F(x_n) &= \inf\{G(q) : q \in \mathbb{Q}, q > x_n \text{ para algún } n\} \\ &= \inf\{G(q) : q \in \mathbb{Q}, q > x\} = F(x) \end{aligned}$$



Para completar la prueba, consideremos un punto  $x$  de continuidad de  $F$ , y elijamos números racionales  $r_1, r_2, s$  tales que  $r_1 < r_2 < x < s$  y

$$F(x) - \varepsilon < F(r_1) \leq F(r_2) \leq F(x) \leq F(x) < F(x) + \varepsilon$$

Como  $F_{n_k}(r_2) \rightarrow G(r_2) \geq G(r_1)$  y  $F_{n_k}(s) \rightarrow G(s) \leq F(s)$ , se deduce que si  $k \geq k_0(\varepsilon)$ ,

$$F(x) - \varepsilon < F_{n_k}(r_2) \leq F_{n_k}(x) < F_{n_k}(s) < F(x) + \varepsilon$$

luego  $F_{n_k}(x) \rightarrow F(x)$ .  $\square$

La siguiente condición nos permitirá evitar la pérdida de masa en el infinito:

**Definición 4.7.1** Sea  $(F_n)$  una sucesión de funciones de distribución. Diremos que  $(F_n)$  es **ajustada**<sup>2</sup> si dado  $\varepsilon > 0$  existe  $M_\varepsilon > 0$  tal que

$$\limsup_{n \rightarrow +\infty} 1 - F_n(M_\varepsilon) + F(-M_\varepsilon) \leq \varepsilon$$

Si  $X_n$  es una sucesión de variables aleatorias con función de distribución  $F_n$ , esto es equivalente a decir que la sucesión  $(X_n)$  está acotada en probabilidad en el sentido de la proposición 4.1.1.

**Teorema 4.7.2 (Teorema de Prokhorov)** Supongamos que  $(F_n)$  es una sucesión de funciones de distribución. Entonces son equivalentes:

- i)  $(F_n)$  es ajustada.
- ii) Para cualquier subsucesión  $(F_{n_k})$  tal que

$$F_{n_k}(x) \rightarrow F(x)$$

para todo punto de continuidad de  $F$  siendo  $F$  continua por la derecha (como en el principio de selección de Helly), se tiene que  $F$  es una función de distribución, es decir que

$$F(-\infty) = 0, \quad F(+\infty) = 1 \quad (4.9)$$

**Prueba:** Supongamos primero que  $(F_n)$  es ajustada, y sea  $F_{n_k}$  una subsucesión que verifica ii). Elijamos  $r < -M_\varepsilon$  y  $s > M_\varepsilon$  puntos de continuidad de  $F$ , emtpmces

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{k \rightarrow +\infty} 1 - F_{n_k}(s) + F_{n_k}(r) \\ &\leq \limsup_{n \rightarrow +\infty} 1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon) \leq \varepsilon \end{aligned}$$

---

<sup>2</sup>**tight** en inglés

Deducimos que:

$$\limsup_{x \rightarrow +\infty} 1 - F(x) + F(-x) \leq \varepsilon$$

y como  $\varepsilon$  es arbitrario. se deduce que  $F$  que se verifica (4.9).

Para probar el recíproco, supongamos que  $(F_n)$  no es ajustada. Entonces hay un  $\varepsilon > 0$  y una subsucesión  $F_{n_k}$  tal que

$$1 - F_{n_k}(k) + F_{n_k}(-k) \geq \varepsilon$$

Utilizando el principio de selección de Helly (y pasando a una subsucesión) podemos suponer que  $F_{n_k}(x) \rightarrow F(x)$  en los puntos de continuidad de  $F$  (donde  $F$  es continua por la derecha). Sean  $r < 0 < s$  puntos de continuidad de  $F$ , entonces

$$1 - F(s) + F(r) = \lim_{k \rightarrow +\infty} 1 - F_{n_k}(s) + F_{n_k}(r) \geq \liminf_{k \rightarrow +\infty} 1 - F_{n_k}(k) + F_{n_k}(-k) \geq \varepsilon$$

Haciendo que  $s \rightarrow +\infty$  y que  $r \rightarrow +\infty$  deducimos que

$$1 - F(+\infty) + F(-\infty) \geq \varepsilon$$

Luego  $F$  no puede ser una función de distribución. □

## 4.8. Funciones Características

Para la siguiente definición, recordemos que para  $x \in \mathbb{R}$ , la función exponencial  $e^{ix}$  de exponente imaginario puro puede definirse por medio de la fórmula de Euler

$$e^{ix} = \cos x + i \sin x$$

que puede justificarse a partir de los correspondientes desarrollos de Taylor.

También notemos que podemos considerar variables aleatorias con valores complejos  $X : \Omega \rightarrow \mathbb{C}$ , en lugar de con valores reales como hemos hecho hasta ahora. Escribiendo  $X = A + Bi$  donde  $A$  y  $B$  son la parte real e imaginaria de  $X$ , no ofrece ninguna dificultad extender la definición de esperanza para ellas, escribiendo

$$E(X) = E(A) + iE(B)$$

Las propiedades de la esperanza se generalizan fácilmente para estas variables.

**Definición 4.8.1** Si  $X$  es una variable aleatoria tal que  $E(|X|)$  es finita, su función característica se define por

$$\varphi_X(t) = E[e^{itX}] \quad t \in \mathbb{R}$$

Teniendo en cuenta la definición de la esperanza, esto puede escribirse como

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} dF_X(x)$$

siendo  $F_X$  la función de distribución de  $X$ , y si  $X$  es una variable con densidad continua  $f_X(x)$  entonces

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx$$

Notemos entonces que en la teoría de probabilidades se llama *función característica* a lo que en muchos otros contextos de la matemática se conoce como transformada de Fourier. De hecho, existe toda una rama de la matemática dedicada al estudio de este tipo de transformadas, el análisis armónico. Para nosotros, será una herramienta útil para estudiar la convergencia en distribución de las variables aleatorias (ver el teorema de continuidad en la sección siguiente).

Observemos también que la función característica sólo depende de la distribución de la variable aleatoria  $X$ , por lo que tiene sentido hablar de funciones características de una determinada distribución de probabilidades  $F$ . Por eso, a veces escribiremos  $\varphi_F$  en lugar de  $\varphi_X$  para enfatizar este hecho.

**Observación 4.8.1** Si  $X$  es una variable aleatoria discreta que toma valores en  $\mathbb{N}_0$ , tenemos que

$$\varphi_X(t) = \sum_{k=0}^{\infty} e^{itk} P\{X = x_i\} = \sum_{k=0}^{\infty} (e^{it})^k P\{X = x_i\} = g_X(e^{it})$$

donde  $g_X$  es la función generatriz que introdujimos en la sección 2.4. Por ejemplo, usando esto deducimos que:

- Si  $X \sim Bi(n, p) \Rightarrow \varphi_X(t) = (p + qe^{it})^n = (1 + p(e^{it} - 1))^n$  donde  $q = 1 - p$ , por (2.3).
- Si  $X \sim \mathcal{P}(\lambda) \Rightarrow \varphi_X(t) = e^{\lambda(\exp(it)-1)}$  por (2.7).
- Si  $X \sim Ge(p) \Rightarrow \varphi_X(t) = \frac{pe^{it}}{1 - qe^{it}}$  donde  $q = 1 - p$ , por (2.8)

**Ejemplo 4.8.1** Para la distribución uniforme, la función característica puede determinarse a partir de la definición. Si  $X \sim \mathcal{U}(a, b)$ , entonces

$$\varphi_X(t) = \int_a^b e^{itx} \frac{dx}{b-a} = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

**Proposición 4.8.1** *La función característica de una variable aleatoria  $X$  con  $E(|X|) < \infty$  tiene las siguientes propiedades:*

i) *La función característica  $\varphi_X(t)$  es uniformemente continua.*

ii)

$$|\varphi_X(t)| \leq 1$$

iii)

$$\varphi_X(0) = 1$$

iv) *Si hacemos un cambio lineal de variable,  $Y = aX + b$*

$$\varphi_Y(t) = e^{itb} \varphi_X(ta)$$

**Prueba:** Falta escribir la prueba de i) ii) Es inmediata pues

$$|\varphi_X(t)| = |E(e^{itX})| \leq E(|e^{itX}|) = E(1) = 1$$

iii) También es inmediata pues

$$\varphi_X(0) = E(e^{i0}) = E(1) = 1$$

Para probar iv) notamos que

$$E(Y) = E(e^{itY}) = E(e^{it(aX+b)}) = E[e^{itaX} e^{itb}] = e^{itb} \varphi_X(ta)$$

□

**Proposición 4.8.2** *Si  $X$  e  $Y$  son variables aleatorias independientes con esperanza finita entonces*

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t)$$

**Prueba:** Como  $X$  e  $Y$  son independientes,  $e^{itX}$  y  $e^{itY}$  también lo son entonces

$$\varphi_{X+Y}(t) = E^{it(X+Y)} = E[e^{itX}] E[e^{itY}] = \varphi_X(t) \varphi_Y(t)$$

□

**Proposición 4.8.3** *Sea  $k \in \mathbb{N}$ . Si  $E(|X|^k) < \infty$ , entonces  $\varphi_X(t)$  es de clase  $C^k$  y*

$$\varphi_X^{(k)}(t) = E((iX)^k e^{itX})$$

*En particular*

$$\varphi_X^{(k)}(t) = i^k m_k(X)$$

*donde*

$$\mu_k(X) = E(X^k)$$

*es el  $k$ -ésimo momento de la variable  $X$  (respecto del origen).*

**Prueba:** Se obtiene derivando bajo el signo de esperanza. Para justificar esto, se requiere un teorema de derivación de integrales con respecto a un parámetro, que se ve en análisis real.  $\square$

**Ejemplo 4.8.2** Si  $X \sim \Gamma(\alpha, \lambda)$ , su función característica viene dada por

$$\begin{aligned}\varphi(t) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{itx} x^{\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-it)x} dx\end{aligned}$$

Usando la fórmula (3.25) (que sigue valiendo para valores complejos de  $\lambda$  con  $\operatorname{Re}(\lambda) > 0$ ) se deduce que <sup>3</sup>

$$\varphi_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{\lambda - it} = \left( \frac{\lambda}{\lambda - it} \right)^\alpha \quad (4.10)$$

**Observación 4.8.2** Cuando  $X$  es una variable aleatoria con una densidad integrable, se tiene que

$$\varphi_X(t) \rightarrow 0 \text{ cuando } |t| \rightarrow \infty$$

en virtud del lema de Riemann-Lebesgue (un resultado importante del análisis armónico). Sin embargo, esta propiedad no es cierta para variables aleatorias cualesquiera. Por ejemplo, si  $X$  es una variable aleatoria, tal que  $X = 0$  con probabilidad 1, entonces  $\varphi_X(t) \equiv 1$ .

### 4.8.1. La Función Característica de la Distribución Normal

El siguiente teorema es clave para la prueba que haremos del teorema central del límite, uno de los resultados fundamentales de la teoría de probabilidades:

**Teorema 4.8.1** Si  $X \sim N(\mu, \sigma^2)$ , entonces  $\varphi_X(t) = e^{it\mu} e^{-(\sigma t)^2/2}$

---

<sup>3</sup>Se requieren algunos conocimientos de análisis complejo para darle sentido a esta fórmula,  $z^\alpha$  se puede definir en el plano complejo menos el eje real negativo, usando la fórmula  $z^\alpha = \exp(\alpha \log(z))$  y tomando la rama principal del logaritmo.

Existen varias pruebas de este teorema. Presentaré una prueba que aprendí en el curso de V. Yohai que utiliza argumentos probabilísticos. Notemos que el teorema dice esencialmente que la densidad normal estándar es un punto fijo de la transformada de Fourier. Hay también demostraciones que utilizan argumentos de análisis complejo o de ecuaciones diferenciales. La idea de dicha prueba es usar las propiedades de invariancia de la distribución normal para obtener una ecuación funcional para la función característica buscada.

**Prueba:** Usando el resultado del ejemplo 3.5.1, vemos que basta probarlo para la variable normalizada

$$X^* = \frac{X - \mu}{\sigma}$$

que tiene distribución  $N(0, 1)$ .

Consideramos entonces dos variables aleatorias  $X, Y \sim N(0, 1)$  independientes, y sea  $Z = aX + bY$ , con  $a, b > 0$ . Tendremos entonces

$$\varphi_Z(t) = \varphi_{aX}(t)\varphi_{bY}(t) = \varphi_X(at)\varphi_Y(bt)$$

y como la función característica sólo depende de la distribución esto es igual a

$$\varphi_Z(t) = \varphi_X(at)\varphi_X(bt)$$

Por otra parte, sabemos por la proposición 3.6.2 y el ejemplo 3.5.1, que

$$Z \sim N(0, a^2 + b^2)$$

Entonces de nuevo por el ejemplo 3.5.1,

$$Z^* = \frac{Z}{\sqrt{a^2 + b^2}} \sim N(0, 1)$$

y se deduce utilizando el ítem iv) de la proposición 4.8.1 que

$$\varphi_Z(t) = \varphi_X\left(\sqrt{a^2 + b^2} t\right)$$

Comparando las dos expresiones para  $\varphi_Z(t)$  obtenemos la ecuación funcional buscada:

$$\varphi_X\left(\sqrt{a^2 + b^2} t\right) = \varphi_X(at)\varphi_X(bt)$$

En particular eligiendo  $t = 1$ , tenemos que

$$\varphi_X\left(\sqrt{a^2 + b^2}\right) = \varphi_X(a)\varphi_X(b)$$

Llamemos  $\psi(s) = \varphi_X(\sqrt{s})$ . Entonces

$$\psi(a^2 + b^2) = \psi(a^2)\psi(b^2)$$

y poniendo  $a = \alpha^2, b = \beta^2$  deducimos que

$$\psi(\alpha + \beta) = \psi(\alpha)\psi(\beta) \quad \text{para todo } \alpha, \beta \geq 0$$

(Si  $\alpha$  o  $\beta$  son cero, esto vale pues  $\varphi_X(0) = 1$ ). Entonces por el lema 3.8.1, deducimos que

$$\psi(t) = e^{tb} \quad \text{para algún } b \in \mathbb{R}$$

ya que  $\psi(0) = 1$ , y por lo tanto

$$\varphi_X(t) = e^{bt^2}$$

Para encontrar el valor de  $b$ , derivamos dos veces

$$\varphi'_X(t) = 2bt e^{bt^2}$$

$$\varphi''_X(t) = (2b + 2bt) e^{bt^2}$$

En particular,

$$\varphi''_X(0) = 2b = -\mu_2(X)$$

por la proposición 4.8.3. Pero

$$\mu_2(X) = Var(X) = 1$$

luego  $b = -1/2$ , y obtenemos que

$$\varphi_X(t) = e^{-t^2/2}$$

□

#### 4.8.2. La Fórmula de Inversión: unicidad de la función característica

Un hecho fundamental es que es posible reconstruir la distribución de probabilidades de una variable aleatoria, a partir de su función característica.

**Teorema 4.8.2** *Fórmula de inversión* Si  $X$  es una variable aleatoria, con función de distribución  $F = F_X$  y función característica  $\varphi = \varphi_X$ , y  $x$  e  $y$  son puntos de continuidad de  $F$   $x < y$  entonces

$$F(y) - F(x) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-itx} - e^{ity}}{it} \varphi(t) dt$$

**Prueba:** Véase [6], capítulo 6. □

**Observación 4.8.3** Si la función característica  $\varphi(t)$  fuera integrable, es posible pasar al límite en la integral (por el teorema de convergencia mayorada) y se obtiene

$$\frac{F(y) - F(x)}{y - x} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{ity}}{it(y - x)} \varphi(t) dt$$

Pasando nuevamente al límite cuando  $y \rightarrow x$ , y si asumimos que  $f_X$  es continua en el punto  $x$ , obtenemos

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt$$

que es la clásica fórmula de inversión para la transformada de Fourier.

**Corolario 4.8.1 (Unicidad de la función característica)** Si  $F_1$  y  $F_2$  son dos distribuciones de probabilidad, y  $\varphi_{F_1}(t) = \varphi_{F_2}(t)$  para todo  $t \in \mathbb{R}$  (es decir: sus funciones características coinciden) entonces  $F_1 = F_2$ .

**Prueba:** La fórmula de inversión implica que  $F_1(x) = F_2(x)$  si  $x$  es un punto de continuidad. Si  $x$  no lo fuera, basta observar que como los puntos de discontinuidad de  $F_1$  y  $F_2$  son a lo sumo numerables, entonces podemos elegir una sucesión  $(x_n)$  tal que  $x_n \searrow x$ , tal que  $x_n$  sea un punto de continuidad tanto de  $F_1$  como de  $F_2$ , entonces  $F_1(x_n) = F_2(x_n)$  y como  $F_1$  y  $F_2$  son continuas por la derecha, deducimos que  $F_1(x) = F_2(x)$ . □

## 4.9. El Teorema de Continuidad de Paul Levy

**Teorema 4.9.1** Sea  $(F_n)_{n \in \mathbb{N}}$  una sucesión de distribuciones de probabilidad, y sean

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x)$$

las correspondientes funciones características. Entonces



i) Si  $F_n$  converge débilmente a una distribución  $F$ , entonces

$$\varphi_n(t) \rightarrow \varphi(t) \quad \forall t \in \mathbb{R}$$

donde  $\varphi$  es la función característica de  $F$ .

ii) Recíprocamente, si

$$\varphi_n(t) \rightarrow \varphi(t) \quad \forall t \in \mathbb{R}$$

donde  $\varphi(t)$  es una función continua en  $t = 0$ , entonces existe una distribución de probabilidad  $F$  tal que  $F_n$  converge débilmente a  $F$ .

**Prueba:** La afirmación i) es una consecuencia del corolario 4.5.1 aplicado a  $\varphi(t) = e^{itx}$ .

Para probar la afirmación recíproca ii), vamos a mostrar que la sucesión de funciones de distribución  $(F_n)_{n \in \mathbb{N}}$  es ajustada. Esto será una consecuencia de la continuidad de  $\varphi(t)$  en  $t = 0$

Si  $x \in \mathbb{R}$  y  $\delta > 0$  entonces

$$1 \leq 2 \left( 1 - \frac{\sin(\delta x)}{\delta x} \right) = \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \cos(tx)) dt \quad \text{si } |\delta x| > 2$$

Podemos considerar variables aleatorias  $X_n$  con distribución  $F_n$ , tomar  $x = X_n$  y tomar esperanzas para obtener

$$\begin{aligned} P\{|\delta X_n| > 2\} &\leq \frac{1}{\delta} \int_{-\delta}^{\delta} E[1 - \cos(tX_n)] dt \\ &= \frac{1}{\delta} \int_{-\delta}^{\delta} \operatorname{Re}[E(1 - \exp(itX_n))] dt \\ &= \frac{1}{\delta} \int_{-\delta}^{\delta} \operatorname{Re}[(1 - E[\exp(itX_n)])] dt \\ &= \frac{1}{\delta} \int_{-\delta}^{\delta} \operatorname{Re}[1 - \varphi_n(t)] dt \end{aligned}$$

Como

$$|\operatorname{Re}[1 - \varphi_n(t)]| \leq |1 - \varphi_n(t)| \leq |\varphi_n(t)| \leq 2,$$

por el teorema de convergencia mayorada vemos que

$$P\{|\delta X_n| > 2\} \rightarrow \frac{1}{\delta} \int_{-\delta}^{\delta} \operatorname{Re}[1 - \varphi(t)] dt$$

Entonces, dado  $\varepsilon > 0$ , por la continuidad de  $\varphi(t)$  en  $t = 0$ , podemos encontrar un  $\delta > 0$  tal que

$$|1 - \varphi(t)| < \varepsilon \quad \text{si } |t| < \delta$$

y por lo tanto

$$P\{|\delta X_n| > 2\} \rightarrow \frac{1}{\delta} \int_{-\delta}^{\delta} |1 - \varphi(t)| dt \leq 2\varepsilon$$

Esto implica que la sucesión de funciones de distribución  $F_n$  es ajustada.

Continuemos entonces la demostración de la afirmación ii). Como  $F_n$  es ajustada, por el teorema de Prokhorov 4.7.2, existen una subsucesión  $(F_{n_k})$  y una función de distribución  $F$  tales que  $F_{n_k}(x) \rightarrow F(x)$  si  $x \in C(F)$  donde  $C(F)$  es el conjunto de puntos de continuidad de  $F$ . Entonces por la primera parte del teorema  $\varphi_{n_k}(t) \rightarrow \varphi_F(t)$ , y por la unicidad de la función característica, deducimos que  $\varphi_F = \varphi$ . Además esto implica que la sucesión  $(F_n)$  tiene un único punto de acumulación  $F$  para la convergencia en distribución (es decir: no puede haber dos subsucesiones de  $F_n$  que converjan a distribuciones distintas).

Este último hecho implica que  $F_n(x) \rightarrow F(x)$  para todo  $x \in C(F)$ . En efecto, si suponemos que no vale para algún  $x_0 \in C(F)$ , existirían un  $\varepsilon > 0$  y una subsucesión  $(F_{n_k})$  de  $(F_n)$  (no necesariamente la misma que consideramos antes), tales que

$$|F_{n_k}(x_0) - F(x_0)| > \varepsilon \tag{4.11}$$

Como  $(F_n)$  es ajustada,  $(F_{n_k})$  también lo es. Luego, de nuevo por el teorema de Prokhorov, existe una subsucesión  $(F_{n_{k_j}})$  de  $F_{n_k}$  tal que  $F_{n_{k_j}}$  converge en distribución a alguna distribución de probabilidades, que por lo que dijimos antes tiene que ser necesariamente  $F$ . Entonces  $F_{n_{k_j}}(x_0) \rightarrow F(x_0)$ , y esto contradice (4.11). Este absurdo provino de suponer que  $F_n(x_0) \not\rightarrow F(x_0)$ , por lo que  $F_{n_k}(x_0) \rightarrow F(x_0)$ . Esto vale para todo  $x_0 \in C(F)$ . □

Se deduce la siguiente caracterización de la convergencia en distribución de las funciones características:

**Corolario 4.9.1** *Sea  $(F_n)$  una sucesión de distribuciones de probabilidad y sea  $F$  otra distribución de probabilidad. Entonces  $F_n$  converge en distribución a  $F$  (es decir  $F_n(x) \rightarrow F(x)$  para todo  $x \in C(F)$ ), si y sólo si*

$$\varphi_{F_n}(t) \rightarrow \varphi_F(t) \text{ para todo } t \in \mathbb{R}$$

## Capítulo 5

# El Teorema del Límite Central

En este capítulo, presentaremos el Teorema del Límite Central, que es uno de los resultados fundamentales de la teoría de probabilidades. Informalmente, este teorema dice que la suma de un número grande de variables aleatorias independientes con varianza finita, donde la varianza de cada variable contribuye poco (en algún sentido) a la varianza total se distribuye en forma aproximadamente normal (formalizaremos esta idea más adelante). Este teorema justifica el papel central que juega la distribución normal en la estadística. Por ejemplo, los errores de medición en un experimento suelen tener una distribución normal, y esto es esperable por el teorema central del límite, si suponemos que el error de medición puede originarse en distintas fuentes independientes de error, cada una de las cuales contribuye en pequeña medida al error total.

Comenzaremos presentando una versión para la distribución binomial, conocida como el **teorema de De Moivre-Laplace**. Es históricamente la primera versión que se conoció del teorema del límite central. Y la demostraremos “a mano” utilizando la aproximación del factorial por medio de la fórmula de Stirling. Después demostraremos una versión del teorema del límite central para variables independientes y uniformemente distribuidas (con varianza finita), por medio de la técnica de las funciones características que desarrollamos en el capítulo anterior. Finalmente, haremos algunos comentarios sobre sus generalizaciones y versiones más refinadas.

### 5.1. El Teorema Local de De Moivre-Laplace

Sea  $X$  una variable aleatoria con segundo momento finito. Entonces la variable reescaldada (o “normalizada”)

$$X^* = \frac{X - E(X)}{\sqrt{\text{Var}(X)}}$$

satisface que  $E(X^*) = 0$  y  $\text{Var}(X^*) = 1$ .

Sea  $S_n$  el número de éxitos en  $n$  ensayos de Bernoulli con probabilidad  $p \in (0, 1)$ . Sabemos que  $S_n$  tiene distribución binomial:

$$P\{S_n = k\} = b(k, n, p) = \binom{n}{k} p^k q^{n-k} \quad (0 \leq k \leq n), \quad q = 1 - p$$

y que  $E[S_n] = np$ ,  $\text{Var}(S_n) = npq$ . Consideramos entonces la variable normalizada:

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \quad (5.1)$$

Nuestro objetivo es estudiar el límite de la distribución de  $S_n^*$  cuando  $n \rightarrow +\infty$ :

Comenzamos aproximando la distribución binomial, utilizando la fórmula de Stirling (ver apéndice):

$$n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n} e^{O(1/n)}$$

Obtenemos<sup>1</sup>:

**Teorema 5.1.1 (Teorema local de De Moivre-Laplace)**

$$b(k, n, p) = \frac{1}{\sqrt{2\pi npq}} e^{-x_k^2/2} (1 + \beta_{n,k})$$

donde

$$x_k = \frac{k - np}{\sqrt{npq}}$$

y para  $M \geq 0$ ,

$$\max_{|x_k| \leq M} |\beta_{n,k}| \rightarrow 0 \text{ cuando } n \rightarrow \infty \quad (5.2)$$

**Prueba:**

$$b(k, n, p) = \frac{\sqrt{2\pi} n^{n+1/2} e^{-n} e^{O(1/n)}}{\sqrt{2\pi} k^{k+1/2} e^{-k} e^{O(1/k)} \sqrt{2\pi} (n-k)^{n-k+1/2} e^{-(n-k)} e^{O(1/(n-k))}} p^k q^{n-k}$$

---

<sup>1</sup>La prueba que presentamos del teorema de De Moivre-Laplace está basada en unas notas del curso de probabilidad y estadística del profesor N. Fava.

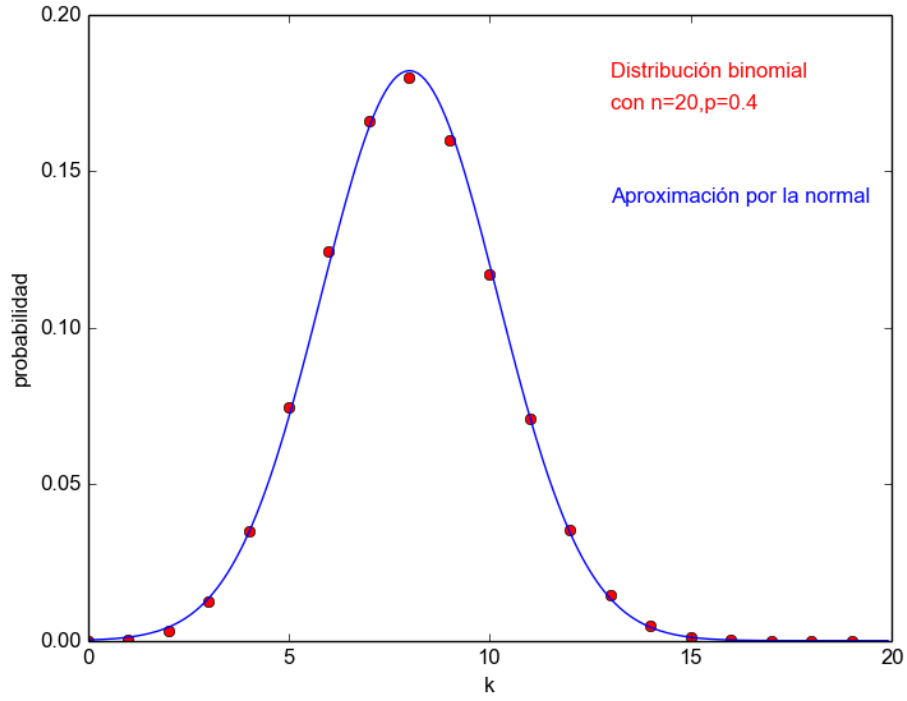


Figura 5.1: Ilustración de la bondad de la aproximación a la distribución binomial por la distribución normal dada por el teorema de local de De Moivre-Laplace, con  $n = 20$  y  $p = 0,4$ .

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} e^{O(1/n)+O(1/k)+O(1/(n-k))}$$

Notemos que:

$$k = np + x_k \sqrt{npq} = np \left(1 + x_k \sqrt{\frac{q}{np}}\right)$$

y que:

$$n - k = nq - x_k \sqrt{npq} = nq \left(1 - x_k \sqrt{\frac{p}{nq}}\right)$$

Estimaremos en forma separada el valor de cada uno de los factores a medida

que  $n \rightarrow +\infty$ :

$$\sqrt{\frac{n}{k(n-k)}} = \sqrt{\frac{n}{np \left(1 + x_k \sqrt{\frac{q}{np}}\right) nq \left(1 - x_k \sqrt{\frac{p}{nq}}\right)}} = \frac{1}{\sqrt{npq}} (1 + \alpha_{n,k})$$

donde

$$\max_{|x_k| \leq M} |\alpha_{n,k}| \rightarrow 0 \text{ cuando } n \rightarrow +\infty$$

Para estimar el segundo factor, tomamos logaritmo y hacemos uso del desarrollo de Taylor:  $\log(1+t) = t - \frac{t^2}{2} + O(t^3)$  cuando  $t \rightarrow 0$ .

En consecuencia:

$$\begin{aligned} \log \left( \frac{np}{k} \right)^k \left( \frac{nq}{n-k} \right)^{n-k} &= (-k) \log \left( \frac{k}{np} \right) - (n-k) \log \left( \frac{n-k}{nq} \right) \\ &= (-k) \log \left( 1 + x_k \sqrt{\frac{q}{np}} \right) - (n-k) \log \left( 1 - x_k \sqrt{\frac{p}{nq}} \right) \\ &= (-np - x_k \sqrt{npq}) \left\{ x_k \sqrt{\frac{q}{np}} - \frac{1}{2} x_k^2 \frac{q}{np} + O \left( \frac{1}{n^{3/2}} \right) \right\} \\ &\quad + (-nq + x_k \sqrt{npq}) \left\{ -x_k \sqrt{\frac{p}{nq}} - \frac{1}{2} x_k^2 \frac{p}{nq} + O \left( \frac{1}{n^{3/2}} \right) \right\} \\ &= -x_k \sqrt{npq} + \frac{1}{2} q x_k^2 - q x_k^2 + O \left( \frac{1}{n^{1/2}} \right) + x_k \sqrt{npq} + \frac{1}{2} p x_k^2 - p x_k^2 + O \left( \frac{1}{n^{1/2}} \right) \\ &= -\frac{1}{2} x_k^2 + O \left( \frac{1}{n^{1/2}} \right) \end{aligned}$$

Deducimos que:

$$\left( \frac{np}{k} \right)^k \left( \frac{nq}{n-k} \right)^{n-k} = e^{-x_k^2/2} \cdot e^{O(1/n^{1/2})}$$

Finalmente consideramos el término de error  $e^{O(1/n)-O(1/k)-O(1/(n-k))} = e^E$  donde

$$E = O \left( \frac{1}{n} \right) + O \left( \frac{1}{np \left( 1 + x_k \sqrt{\frac{q}{np}} \right)} \right) + O \left( \frac{1}{nq \left( 1 - x_k \sqrt{\frac{p}{nq}} \right)} \right) = O \left( \frac{1}{n} \right)$$

En consecuencia, utilizando las estimaciones que hemos obtenido para cada factor, y teniendo en cuenta que  $O(1/n^{1/2}) + O(1/n) = O(1/n^{1/2})$ , obtenemos que:

$$b(k, n, p) = \frac{1}{\sqrt{2\pi npq}} e^{-x_k^2/2} \cdot (1 + \alpha_n(x_k)) e^{O(1/n^2)}$$

Finalmente, observamos que el factor de error dado por

$$(1 + \alpha_n(x_k)) e^{O(1/n^{1/2})}$$

tiende a 1 cuando  $n \rightarrow +\infty$ , uniformemente para los  $k$  tales que  $|x_k| \leq M$ , por lo que podremos representarlo en la forma  $1 + \beta_{n,k}$  donde

$$\max_{|x_k| \leq M} |\beta_{n,k}| \rightarrow 0$$

□

**Observación 5.1.1** *La fórmula 5.2 significa que la aproximación dada por el teorema de local De Moivre-Laplace es buena en el centro de la distribución binomial, pero no en las colas de la misma. Por ejemplo, si  $n$  es grande y  $p$  es muy pequeño, como se ilustra en la figura 5.3. En esta situación es mejor la aproximación por la distribución de Poisson que discutimos en la sección 2.7. Por simetría, tampoco es buena si  $p$  está muy cerca de 1.*

## 5.2. El Teorema de De Moivre-Laplace

En este capítulo, notaremos por

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{5.3}$$

a la densidad normal estándar (que introdujimos en el ejemplo 3.1.3), y por

$$G(x) = \int_{-\infty}^x g(t) dt \tag{5.4}$$

a la correspondiente función de distribución normal (acumulada).

El siguiente teorema afirma que la distribución límite de la variable normalizada  $S_n^*$  está dada por la integral definida de  $g(x)$ :

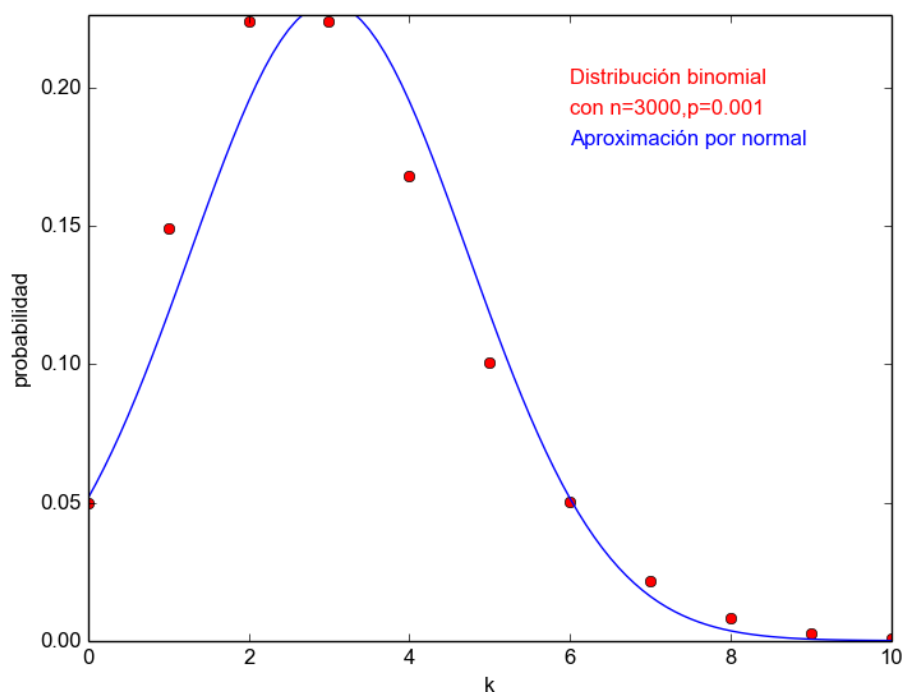


Figura 5.2: Ilustración de la bondad de la aproximación a la distribución binomial por la distribución normal dada por el teorema local de De Moivre-Laplace, con  $n = 3000$  y  $p = 0,01$ . Vemos que no resulta tan buena si  $n$  es grande y  $p$  es pequeña.

**Teorema 5.2.1 (De Moivre-Laplace)**

$$P\{a < S_n^* \leq b\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx = G(b) - G(a)$$

uniformemente en  $a$  y en  $b$  cuando  $n \rightarrow +\infty$ .

**Observación 5.2.1** De acuerdo con [21], el teorema 5.1.1 fue enunciado por De Moivre en 1754 en su trabajo *Approximatio ad Summam Terminorum Binomiali*  $(a+b)^n$  in Seriem expansi, pero sólo lo demostró para  $p = 1/2$ . La primera prueba



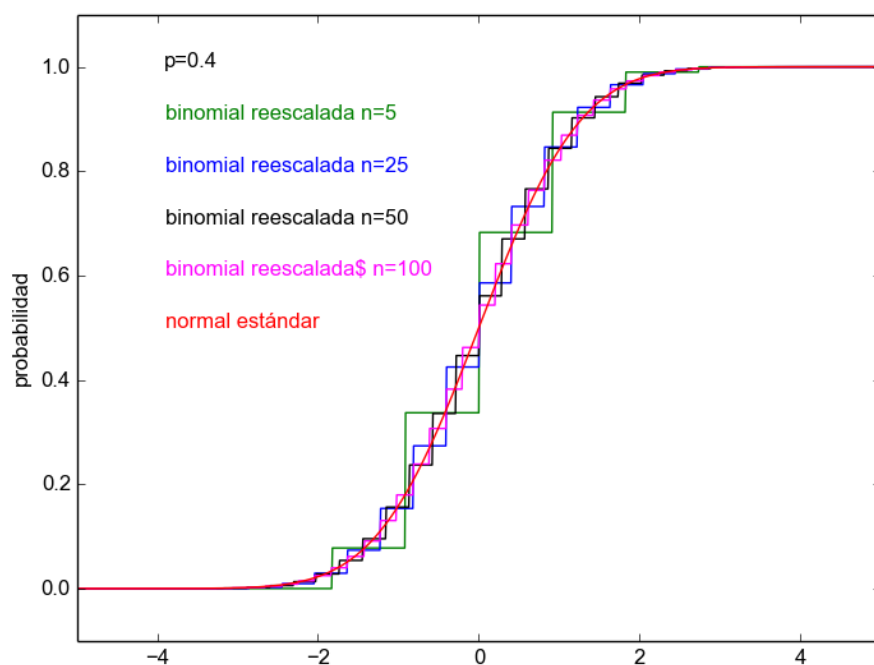


Figura 5.3: Ilustración del teorema de De Moivre-Laplace: para  $p = 0,4$  y distintos valores de  $n$ , dibujamos la función de distribución de la distribución binomial, junto con la de la normal estándar.

completa fue dada por Laplace (1795) en su libro *Théorie analytique des probabilités*. Análogamente el teorema 5.2.1 fue demostrado por De Moivre para  $p = 1/2$ , y por Laplace para cualquier  $p \in (0, 1)$ .

La idea básica de la demostración es la siguiente:

$$P_n(a, b) = P\{a < S_n^* \leq b\} = \sum_{a < x_k \leq b} b(k, n, p)$$

ya que si  $S_n^*$  toma el valor  $x_k$ , entonces  $S_n$  toma el valor  $k$ .

Los puntos  $x_k$  están cada vez más próximos a medida que  $n \rightarrow +\infty$ , ya que

$$x_{k+1} - x_k = \frac{1}{\sqrt{npq}}$$

y por el teorema anterior  $b(k, n, p) \approx g(x_k)(x_{k+1} - x_k)$  entonces,

$$P_n(a, b) = P\{a < S_n^* \leq b\} \approx \sum_{a < x_k \leq b} g(x_k)(x_{k+1} - x_k)$$

y esta es una suma de Riemann para la integral  $\int_a^b g(x) dx$ . Por lo tanto, conforme  $n \rightarrow +\infty$ , es razonable que podamos aproximar  $P_n(a, b)$  por dicha integral.

La demostración consiste en una formalización de esta idea:

**Prueba:** Dado  $\varepsilon > 0$ , elegimos  $M$  de modo que

$$G(M) - G(-M) = 1 - \varepsilon$$

y además

$$\frac{1}{M^2} < \varepsilon$$

por consiguiente:

$$G(-M) = 1 - G(M) = \varepsilon/2$$

Consideramos primero el caso en que el intervalo  $(a, b)$  está contenido en el intervalo  $(-M, M)$ . La función  $g_n$  definida por  $g_n(x) = g(x_k)$  para  $x_k < x \leq x_{k+1}$  converge uniformemente a  $g(x)$  cuando  $n \rightarrow +\infty$ , en virtud de la continuidad uniforme de  $g$ .

Denotamos por  $k_0$  el mínimo entero tal que  $a < x_{k_0}$  y sea  $k_1$  el máximo entero tal que  $x_{k_1} \leq b$ .

En virtud del teorema 5.1.1,

$$\begin{aligned} P_n(a, b) &= \sum_{a < x_k \leq b} (1 + \beta_{n,k})g(x_k)(x_{k+1} - x_k) \\ &= \sum_{a < x_k \leq b} g(x_k)(x_{k+1} - x_k) + \sum_{a < x_k \leq b} \beta_{n,k} g(x_k)(x_{k+1} - x_k) \\ &= \int_{x_{k_0}}^{x_{k_1+1}} g_n(x) dx + \sum_{a < x_k \leq b} \beta_{n,k} g(x_k)(x_{k+1} - x_k) \end{aligned}$$

En consecuencia,

$$P_n(a, b) = \int_a^b g_n(x) dx - \int_a^{x_{k_0}} g_n(x) dx + \int_b^{x_{k_1+1}} g_n(x) dx$$

$$\sum_{a < x_k \leq b} \beta_{n,k} g(x_k)(x_{k+1} - x_k)$$

o sumando y restando  $g_n$ :

$$\begin{aligned} P_n(a, b) &= \int_a^b g(x) dx + \int_a^b [g_n(x) - g(x)] dx + \int_a^{x_{k_0}} g_n(x) dx \\ &\quad + \int_b^{x_{k_1+1}} g_n(x) dx + \sum_{a < x_k \leq b} \beta_{n,k} g(x_k)(x_{k+1} - x_k) \end{aligned}$$

El segundo término de esta expresión podemos acotarlo del siguiente modo:

$$\left| \int_a^b [g_n(x) - g(x)] dx \right| \leq (b-a) \sup_{x \in [a,b]} |g_n(x) - g(x)| \leq 2M \sup_{x \in [a,b]} |g_n(x) - g(x)|$$

Además como  $g$  y por consiguiente  $g_n$  están acotadas por  $(2\pi)^{-1/2}$ , deducimos que:

$$\left| \int_a^{x_{k_0}} g_n(x) dx \right| \leq \frac{1}{\sqrt{2\pi npq}}$$

, Similarmente:

$$\left| \int_b^{x_{k_1+1}} g_n(x) dx \right| \leq \frac{1}{\sqrt{2\pi npq}}$$

Finalmente, último término podemos acotarlo del siguiente modo,

$$\begin{aligned} \left| \sum_{a < x_k \leq b} \beta_{n,k} g(x_k)(x_{k+1} - x_k) \right| &\leq \max_{|x_k| \leq M} |\beta_{n,k}| \sum_{k=k_0}^{k_1} g(x_k)(x_{k+1} - x_k) \\ &\leq \frac{1}{\sqrt{2\pi}} 2M \max_{|x_k| \leq M} |\beta_{n,k}| \rightarrow 0 \text{ cuando } n \rightarrow +\infty \end{aligned}$$

Como todas las estimaciones efectuadas, son independientes de  $a$  y  $b$ , concluimos que cuando  $n \rightarrow +\infty$ ,

$$P_n(a, b) \rightarrow \int_a^b g(x) dx$$

uniformemente en  $a$  y  $b$ . Es decir: existe un entero  $n_0 = n_0(\varepsilon)$  independiente de  $a$  y de  $b$  tal que

$$\left| P_n(a, b) - \int_a^b g(x) dx \right| < \varepsilon$$

para cualquier  $a, b \in (-M, M)$ . En particular, deducimos que:

$$\left| P_n(-M, M) - \int_{-M}^M g(x) dx \right| \leq \varepsilon$$

para  $n \geq n_0$ .

Si  $(a, b)$  no está contenido en  $(-M, M)$ , tenemos que:

$$P_n(a, b) = P_n(a, -M) + P_n(-M, M) + P_n(M, b)$$

y

$$\int_a^b g(x) dx = \int_a^{-M} g(x) dx + \int_{-M}^M g(x) dx + \int_M^b g(x) dx$$

Utilizando entonces la desigualdad triangular tenemos que:

$$\begin{aligned} \left| P_n(a, b) - \int_a^b g(x) dx \right| &\leq \left| P_n(-M, M) - \int_{-M}^M g(x) dx \right| + \\ &+ P_n(a, -M) + P_n(M, b) + \int_a^{-M} g(x) dx + \int_M^b g(x) dx \end{aligned}$$

Pero

$$\int_a^{-M} g(x) dx + \int_M^b g(x) dx \leq \int_{-\infty}^{-M} g(x) dx + \int_M^{\infty} g(x) dx = G(-M) + [1 - G(M)] < \varepsilon$$

y

$$P_n(a, -M) + P_n(M, b) \leq P\{|S_n^*| \geq M\} \leq \frac{1}{M^2} < \varepsilon$$

por la desigualdad de Tchebyshev, pues  $E(S_n^*) = 0$  y  $\text{Var}(S_n^*) = 1$  (teniendo en cuenta nuestra elección de  $M$  al comienzo de la demostración). En consecuencia,

$$\left| P_n(a, b) - \int_a^b g(x) dx \right| \leq 3\varepsilon$$

si  $n \geq n_0(\varepsilon)$  Esto concluye la demostración del teorema.  $\square$

### 5.3. Una Aplicación a la Estadística

Veremos ahora una aplicación del teorema de De Moivre-Laplace y de la distribución normal, a la estadística.

Consideremos por ejemplo, una encuesta electoral para una elección donde participan dos candidatos A y B, y supongamos que cada persona puede votar por uno de ellos (y para simplificar que no hay votos en blanco). Podemos modelizar esto utilizando la distribución binomial, para ello imaginemos un experimento aleatorio donde se elige una persona al azar y se le pregunta por quien vota. Y llamemos  $p$  a la probabilidad de que vote por A (“éxito”) y  $q = 1 - p$  a la probabilidad de que vote por B. Alternativamente, podemos pensar que tenemos una elección en la que participan varios candidatos y que nos interesa medir la intención de voto de un determinado candidato A. En este caso, consideramos el experimento aleatorio que consiste en elegir una persona al azar, preguntarle por quien vota, y hay dos resultados posibles que nos interesan: si vota por A (con probabilidad  $p$ ) o si no vota por A con probabilidad  $q=1-p$ .

Nuestro objetivo es estimar la probabilidad desconocida  $p$ . Como resulta extraordinariamente costoso y complicado preguntarle a cada votante del padrón electoral por quién piensa votar, lo que suele hacerse es elegir una **muestra**, digamos formada por  $n$  personas. Entonces, conforme a la ley de los grandes números, si llamamos  $S_n$  a la cantidad de personas de la muestra que votan por el candidato A, podemos aproximar la probabilidad desconocida  $p$  por la frecuencia:

$$f_n = \frac{S_n}{n}$$

observada en la muestra (Estamos suponiendo que las elecciones de las distintas personas pueden considerarse independientes unas de otras, de modo que la elección de  $n$  personas encuestadas, puede considerarse como realizar  $n$  ensayos de Bernoulli, y la distribución de  $S_n$  sea dada por la distribución binomial.)

Otro ejemplo análogo se da en el control de calidad en un proceso industrial. Por ejemplo, imaginemos que tenemos un lote de 10.000 lamparitas y queremos saber cuantas están falladas. Llamemos  $p$  a la probabilidad de que una lamparita elegida al azar funcione, y  $q = 1 - p$  a la probabilidad de que esté fallada. Nuevamente, sería extraordinariamente costoso probar una por una las lamparitas, por lo que se hace es elegir una muestra, y aproximar  $p$  por la frecuencia  $f_n$  observada en la muestra.

Una pregunta fundamental es entonces: ¿Cómo elegir el tamaño de la muestra?. Para ello, elegimos un margen de error  $\varepsilon$ , y un nivel de confianza  $1 - \alpha$  donde  $\varepsilon$  y  $\alpha$  son números pequeños, y nos proponemos elegir el tamaño de la muestra de modo que podamos asegurar que la probabilidad de que  $f_n$  diste de  $p$  como mucho en  $\varepsilon$  es por lo menos  $1 - \alpha$ , o sea:

$$P\{|f_n - p| \leq \varepsilon\} \geq 1 - \alpha \quad (5.5)$$

Por ejemplo: supongamos que queremos que muestra encuesta (o control de

calidad) se equivoque como mucho en un 2 % en el 95 % de las veces que realizamos la encuesta. Entonces, elegimos  $\varepsilon = 0,02$  y  $\alpha = 0,05$ .

Elegimos entonces  $x_\alpha$  de modo que:

$$G(-x_\alpha) = \frac{\alpha}{2}$$

donde  $G$  es la función de distribución normal estándar (dada por 5.4). Por la simetría de la curva normal,

$$G(x_\alpha) = 1 - \frac{\alpha}{2}$$

Llamando  $S_n^*$  a la variable normalizada dada por (5.1), por el teorema de De Moivre Laplace:

$$P\{-x_\alpha \leq S_n^* \leq x_\alpha\} \approx \frac{1}{\sqrt{2\pi}} \int_{-x_\alpha}^{x_\alpha} e^{-x^2/2} dx = G(x_\alpha) - G(-x_\alpha) = 1 - \alpha$$

si  $n$  es suficientemente grande. En consecuencia, recordando la definición de  $S_n^*$  y despejando:

$$P\{-x_\alpha \sqrt{npq} \leq S_n - np \leq x_\alpha \sqrt{npq}\} \approx 1 - \alpha$$

$$P\{np - x_\alpha \sqrt{npq} \leq S_n \leq np + x_\alpha \sqrt{npq}\} \approx 1 - \alpha$$

$$P\left\{p - x_\alpha \sqrt{\frac{pq}{n}} \leq \frac{S_n}{n} \leq p + x_\alpha \sqrt{\frac{pq}{n}}\right\} \approx 1 - \alpha$$

O sea:

$$P\left\{\left|\frac{S_n}{n} - p\right| \leq x_\alpha \sqrt{\frac{pq}{n}}\right\} \approx 1 - \alpha$$

Esta relación dice que con probabilidad  $1 - \alpha$  podemos asegurar que  $p$  está en el intervalo:

$$I_\alpha = \left[ \frac{S_n}{n} - x_\alpha \sqrt{\frac{pq}{n}}, \frac{S_n}{n} + x_\alpha \sqrt{\frac{pq}{n}} \right]$$

$I_\alpha$  se llama un **intervalo de confianza** (asintótico) para  $p$  de nivel de confianza  $1 - \alpha$ . En realidad en esta forma, esta relación no resulta todavía muy útil ya que no conocemos  $p$  y entonces tampoco conocemos el ancho del intervalo  $I_\alpha$ . Pero podemos observar que:

$$pq = p(1 - p) \leq \frac{1}{4} \quad \forall p \in [0, 1]$$

En consecuencia, podemos asegurar que

$$I_\alpha \subset \left[ \frac{S_n}{n} - x_\alpha \frac{1}{2\sqrt{n}}, \frac{S_n}{n} + x_\alpha \frac{1}{2\sqrt{n}} \right]$$

y que (si  $n$  es grande):

$$P \left\{ \left| \frac{S_n}{n} - p \right| \leq x_\alpha \frac{1}{2\sqrt{n}} \right\} \geq 1 - \alpha$$

En consecuencia, si queremos que valga la relación (5.5) debemos elegir  $n$  para que:

$$x_\alpha \frac{1}{2\sqrt{n}} \leq \varepsilon$$

o sea:

$$n \geq n_0 = \left( \frac{x_\alpha}{2\varepsilon} \right)^2$$

Esta relación nos dice cuál es el tamaño (mínimo) de la muestra que necesitamos para poder garantizar un determinado margen de error con un determinado nivel de confianza. Por ejemplo, si  $\alpha = 0,05$  y  $\varepsilon = 0,02$ , obtenemos que:  $x_\alpha = 1,96$  y  $n \geq 2401$ .

**Observación:** Notamos que cuando  $\alpha \rightarrow 0$ ,  $x_\alpha \rightarrow +\infty$  por lo que  $n_0 \rightarrow +\infty$ .

## 5.4. El Teorema del Límite Central

El siguiente teorema generaliza al de De Moivre-Laplace:

**Teorema 5.4.1 (Teorema del Límite Central, versión sencilla)** Sea  $(X_k)_{k \in \mathbb{N}} : \Omega \rightarrow \mathbb{R}$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con  $0 < \sigma^2 = \text{Var}(X_k) < +\infty$ . Sea  $\mu = E[X_k]$  (como suponemos que las  $X_k$  tienen todas la misma distribución, tendrán todas la misma esperanza y varianza). Notemos:

$$S_n = X_1 + X_2 + \dots + X_n$$

$$S_n^* = \frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n} \sigma}$$

Entonces

$$S_n^* \xrightarrow{D} N(0, 1)$$

**Observación 5.4.1** *Para comprender el significado de este teorema, observemos que si consideramos el esquema de ensayos de Bernoulli, y las  $(X_k)$  son las variables aleatorias de la sección 2.3 entonces  $S_n$  representa el número total de éxitos en  $n$  ensayos, y el teorema del límite central se reduce al teorema de De Moivre-Laplace.*

**Observación 5.4.2** *El nombre del teorema se debe a que proporciona una buena aproximación en el centro de la distribución, pero no tan buena en las colas de la misma, como vimos en la observación 5.1.1 para el caso de la distribución binomial. En inglés se denomina central limit theorem, pero por esta observación resulta más correcto traducirlo por teorema del límite central que por teorema central del límite, como muchas veces se hace.*

Para la prueba necesitamos un lema elemental sobre números complejos (que el lector fácilmente puede demostrar usando la rama principal del logaritmo).

**Lema 5.4.1** *Si  $(c_n)$  es una sucesión de números complejos tal que  $c_n \rightarrow c$ , entonces*

$$\left(1 + \frac{c_n}{n}\right)^n \rightarrow e^c$$

Pasaremos entonces a la demostración del teorema del límite central:

**Prueba:** Sin pérdida de generalidad, podemos suponer que  $\mu = 0$ , cambiando sino las  $X_k$  por las variables centradas

$$\tilde{X}_k = X_k - \mu$$

Calculemos la función característica de  $S_n^*$ . Como las  $(X_k)$  son independientes, y tienen todas la misma distribución será

$$\varphi_{S_n^*}(t) = \varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n$$

donde  $\varphi(t) = \varphi_{X_k}(t)$  para todo  $k$ . Hagamos el desarrollo de Taylor de  $\varphi(t)$  a segundo orden. Usando la proposición 4.8.3 (que relaciona los momentos de  $X_k$  con las derivadas de la función característica en  $t = 0$ ), vemos que es

$$\begin{aligned}\varphi(X_k)(t) &= 1 + \varphi'(0)t + \frac{1}{2}\varphi''(0)t^2 + t^2 e_2(t) \\ &= 1 - \frac{\sigma^2}{2}t^2 + t^2 e_2(t) \\ &= 1 + \left[-\frac{\sigma^2}{2} + e_2(t)\right] t^2\end{aligned}$$



donde

$$\lim_{t \rightarrow +\infty} e_2(t) = 0 \quad (5.6)$$

por la propiedad que tiene el resto de Taylor. Entonces:

$$\begin{aligned} \varphi_{S_n^*}(t) &= \left\{ 1 + \left[ -\frac{\sigma^2}{2} + e_2 \left( \frac{t}{\sigma\sqrt{n}} \right) \right] \left( \frac{t}{\sigma\sqrt{n}} \right)^2 \right\}^n \\ &= \left\{ 1 + \left[ -\frac{1}{2} + \frac{1}{\sigma^2} e_2 \left( \frac{t}{\sigma\sqrt{n}} \right) \right] \frac{t^2}{n} \right\}^n \end{aligned}$$

Fijado un  $t$ , si llamamos

$$c_n = \left[ -\frac{1}{2} + \frac{1}{\sigma^2} e_2 \left( \frac{t}{\sigma\sqrt{n}} \right) \right] t^2$$

como

$$c_n \rightarrow c = -\frac{t^2}{2}$$

cuando  $n \rightarrow \infty$ , por (5.6), vemos aplicando el lema que

$$\varphi_{S_n^*}(t) \rightarrow e^c = e^{-t^2/2}$$

pero esta función es justamente la función característica de la distribución normal estándar  $N(0, 1)$ . Por el corolario 4.9.1 (del teorema de continuidad de Paul Levy), se deduce que  $S_n^*$  converge en distribución a la normal estándar, como afirma el teorema.  $\square$

### 5.4.1. Aplicación a las distribuciones $\chi_n^2$

Para dar un ejemplo del teorema del límite central, consideremos nuevamente las variables

$$Z_n = X_1^2 + X_2^2 + \dots + X_n^2$$

donde las  $(X_k)$  son variables con distribución normal estándar independientes, que introdujimos en la sección 3.9.1. Entonces, por definición  $Z_n$  tiene distribución  $\chi_n^2$  y sabemos que  $E[Z_n] = n$  y  $\text{Var}(Z_n) = 2n$ . Por el teorema del límite central, para  $n$  grande, la distribución normal proporciona una buena aproximación de la distribución  $\chi_n^2$  en el sentido que las variables normalizadas

$$Z_n^* = \frac{Z_n - n}{\sqrt{2n}}$$

convergen en distribución a una normal estándar. El siguiente gráfico compara las funciones de distribución de  $Z_n^*$  con la de la distribución normal, para  $n$  grande:

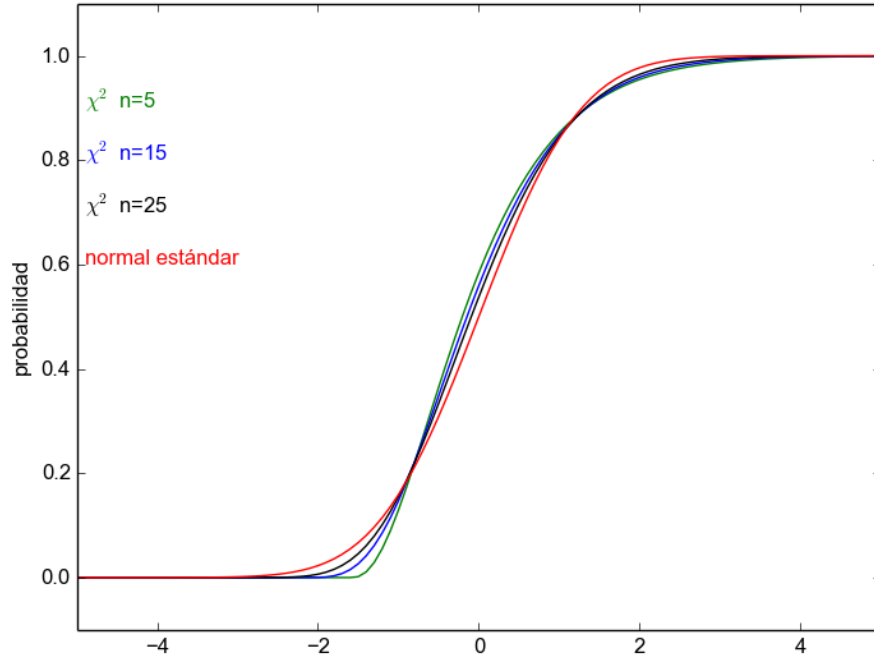


Figura 5.4: Convergencia en distribución de la distribución  $\chi_n^2$  normalizada (distribución de  $Z_n^*$ ) a la normal estándar.

## 5.5. Generalizaciones y comentarios adicionales

El teorema del límite central no está limitado al caso de variables idénticamente distribuidas. Como dijimos en la introducción, se aplica en general a sumas de variables aleatorias independientes con varianza finita, donde la varianza de cada variable contribuye (en algún sentido) a la varianza total. Una condición muy general para su validez está dada por el siguiente teorema de Lindeberg:

**Teorema 5.5.1 (Teorema Límite central de Lindeberg)** *Sea  $(X_k)_{k \in \mathbb{N}}$  una sucesión de variables aleatorias tales que  $\mu_k = E[X_k]$  y  $\sigma_k^2 = \text{Var}(X_k)$ , donde  $\sigma_k$  es finita y al menos algún  $\sigma_{k_0} > 0$ . Sean*

$$S_n = X_1 + X_2 + \dots + X_k$$

$$s_n = \sqrt{\text{Var}(S_n)} = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$$

y supongamos que se cumple la siguiente condición de Lindeberg:

$$\forall \varepsilon > 0 \lim_{\varepsilon \rightarrow 0} \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x-\mu_k| > \varepsilon s_n} (x - \mu_k)^2 dF_{X_k}(x) = 0$$

entonces si definimos

$$S_n^* = \frac{S_n - E[S_n]}{s_n} = \frac{S_n - (\mu_1 + \mu_2 + \dots + \mu_n)}{s_n}$$

tenemos que

$$S_n^* \xrightarrow{D} N(0, 1)$$

El teorema de Lindeberg implica el siguiente teorema de Lyapunov que da una condición más fuerte, pero quizás más fácil de entender:

**Teorema 5.5.2 (Teorema Límite central de Lyapunov)** Sea  $(X_k)_{k \in \mathbb{N}}$  una sucesión de variables aleatorias tales que  $\mu_k = E[X_k]$  y  $\sigma_k^2 = \text{Var}(X_k)$ , donde  $\sigma_k$  es finita y al menos algún  $\sigma_{k_0} > 0$ . Sean

$$S_n = X_1 + X_2 + \dots + X_n$$

$$s_n = \sqrt{\text{Var}(S_n)} = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$$

y supongamos que existe algún  $\delta > 0$  tal que se cumple la siguiente condición de Lyapunov:

$$\forall \varepsilon > 0 \lim_{\varepsilon \rightarrow 0} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n E[|X_k - \mu_k|^{2+\delta}] = 0$$

entonces si definimos

$$S_n^* = \frac{S_n - E[S_n]}{s_n} = \frac{S_n - (\mu_1 + \mu_2 + \dots + \mu_n)}{s_n}$$

tenemos que

$$S_n^* \xrightarrow{D} N(0, 1)$$

La demostración de estos resultados puede verse en [6] (capítulo 7). También emplea el método de las funciones características, aunque resulta mucho más técnica.

Una pregunta que podemos hacernos es ¿cuál es la velocidad de convergencia a la distribución normal en el teorema del límite central? Una respuesta es dada por el teorema de Bery-Essen<sup>2</sup> cuya versión más sencilla (correspondiente a la situación del teorema 5.4.1) es la siguiente:

**Teorema 5.5.3 (Teorema de Berry-Essen, versión sencilla)** *Si  $(X_k)$  es una sucesión de variables independientes idénticamente distribuidas, con  $E(X_k) = \mu$ ,  $E(X_k^2) = \text{Var}(X_k) = \sigma^2 > 0$  y si suponemos además que el tercer momento respecto de la media  $\mu$  de las  $X_k$*

$$\rho = E[|X_k - \mu|^3] < \infty$$

*es finito, y si definimos como antes*

$$S_n = X_1 + X_2 + \dots + X_n$$

$$S_n^* = \frac{S_n - E[S_n]}{\sqrt{\text{Var}S_n}} = \frac{S_n - n\mu}{n\sqrt{\sigma}}$$

*entonces*

$$|F_{S^*}(x) - G(x)| \leq \frac{C\rho}{\sigma^3\sqrt{n}}$$

*donde  $G$  denota la función de distribución de la normal estándar y  $C$  es una constante fija.*

También debemos mencionar que el teorema del límite central se generaliza sin dificultades esenciales a vectores aleatorios, debiendo considerar en este caso para la distribución límite a la distribución normal multivariada (ver [6], teorema 7.2). Y que existen versiones “locales” del teorema central del límite, que generalizan al teorema 5.1.1, para una discusión al respecto ver [21].

## 5.6. Una Aplicación a la Teoría de Números

Resulta sorprendente encontrar aplicaciones del teorema del límite central, en ramas de la matemática aparentemente alejadas de las probabilidades. En esta sección comentaremos brevemente una de ellas: una aplicación a la teoría de números. Esta rama de la matemática se ocupa fundamentalmente de las propiedades de los números enteros.

---

<sup>2</sup>Este teorema fue descubierto independientemente por los matemáticos Andrew C. Berry [17] y Carl-Gustav Esseen [16]. La prueba en el primero de ellos también emplea el método de las funciones características.

Comencemos con una pregunta muy básica: ¿qué quiere decir elegir un número natural al azar?. Para ello, fijado un  $N \in \mathbb{N}$  consideramos el conjunto  $\Omega_N = \{n \in \mathbb{N} : 1 \leq n \leq N\}$  como un espacio muestral discreto en el que asignamos probabilidades de acuerdo con la definición clásica de Laplace:

$$P_N(A) = \frac{\#(A)}{N} \quad A \subset \Omega_N$$

Si queremos asignar a eventos  $A \subset \Omega = \mathbb{N}$ , resulta natural entonces tomar el límite cuando  $N \rightarrow \infty$ , y definir

$$P(A) = \lim_{N \rightarrow \infty} P_N(A \cap \Omega_N) \quad A \subset \mathbb{N}$$

siempre que este límite exista

Por ejemplo: ¿cuál es la probabilidad de que un número natural elegido al azar sea par? De acuerdo a esta definición si  $D_2 = \{n \in \mathbb{N} : n \text{ es par}\}$ , entonces

$$P(D_2) = \lim_{N \rightarrow \infty} \frac{1}{N} \left[ \frac{N}{2} \right] = \lim_{N \rightarrow \infty} \frac{1}{N} \left( \frac{N}{2} + O(1) \right) = \frac{1}{2}$$

(donde los corchetes indican la parte entera de  $\frac{N}{2}$ ), que está de acuerdo con nuestra intuición. Más generalmente, si  $d \in \mathbb{N}$ , y consideramos el evento

$$D_d = \{n \in \mathbb{N} : n \text{ es divisible por } d\}$$

un argumento similar muestra que

$$P(D_d) = \frac{1}{d} \tag{5.7}$$

como esperamos<sup>3</sup>.

Sin embargo, hay que ser cuidadosos, porque esta noción de probabilidad no es  $\sigma$ -aditiva (es decir: se sale del marco de Kolmogorov en el que venimos trabajando<sup>4</sup>). Por ejemplo,  $P(\mathbb{N}) = 1$  pero

$$\mathbb{N} = \bigcup_{n \in \mathbb{N}} \{n\}$$

y  $P(\{n\}) = 0$ .

---

<sup>3</sup>Para una discusión más detallada de este concepto, ver [11]

<sup>4</sup>Sin embargo, es posible formalizarla en el contexto más general de las álgebras de probabilidad condicional propuesto por Renyi [8]

Para  $n \in \mathbb{N}$  consideremos ahora la función  $\omega(n)$  que cuenta el número de divisores primos distintos de  $n$ . Por ejemplo,

$$360 = 2^3 \times 3^2 \times 5^1 \Rightarrow \omega(360) = 3$$

Entonces se tiene el siguiente teorema:

**Teorema 5.6.1 (Teorema del límite central de Erdős–Kac,[18])** *La distribución de  $\omega(n)$  es asintóticamente normal, en el siguiente sentido:*

$$\lim_{N \rightarrow \infty} P_N \left( \left\{ n \leq N : a \leq \frac{\omega(n) - \log \log n}{\sqrt{\log \log n}} \leq b \right\} \right) = G(b) - G(a)$$

Podemos interpretar la intuición detrás de este teorema de la siguiente manera: consideremos el conjunto de los primos numerado en forma creciente

$$\mathbb{P} = \{p_1 = 2, p_2 = 3, p_3 = 5, p_4 = 7, p_5 = 11, p_6 = 13, \dots\},$$

y para cada  $k \in \mathbb{N}$  definamos la función (variable aleatoria)

$$X_k(n) = \begin{cases} 1 & \text{si } p_k \text{ divide a } n \\ 0 & \text{si no} \end{cases}$$

Las  $X_k$  se comportan como variables aleatorias independientes pues de acuerdo con 5.7:

$$P\{X_j = 1, X_k = 1\} = \frac{1}{p_j p_k} = \frac{1}{p_j} \cdot \frac{1}{p_k} = P\{X_j = 1\} \cdot P\{X_k = 1\}$$

En consecuencia como

$$\omega(n) = \sum_{k=1}^{\infty} X_k(n)$$

(Esta suma es en realidad finita para cada  $n$ , pues basta sumar los primos con  $p_k \leq n$ ), vemos que  $\omega$  se comporta como una suma de variables aleatorias independientes, y esto explica porqué el teorema del límite central se aplique a ella. Sin embargo, hacer riguroso este argumento requiere argumentos de teoría de las cribas. Una prueba relativamente sencilla aparece en [27].

# Capítulo 6

## Esperanza Condicional

### 6.1. Esperanza condicional respecto de un evento

Sea  $B$  un evento de probabilidad positiva. Recordamos que la probabilidad condicional de que ocurra el evento  $A$  sabiendo que ocurre el evento  $B$ , notada  $P(A/B)$  se define por:

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

Sea  $X : \Omega \rightarrow \overline{\mathbb{R}}$  una variable aleatoria discreta. Recordamos que la esperanza de  $X$  se define como la serie

$$E[X] = \sum_i x_i P\{X = x_i\}$$

donde  $\text{Im} = \{x_i\}$  es por hipótesis a lo sumo numerable; siempre que dicha serie sea absolutamente convergente.

En consecuencia, resulta natural definir la esperanza de  $X$  dado que ocurre el evento  $A$  de probabilidad positiva, por:

$$E[X/A] = \sum_i x_i P\{X = x_i/A\}$$

Teniendo en cuenta la definición de probabilidad condicional esto es equivalente a:

$$E[X/A] = \sum_i x_i \frac{P(\{X = x_i\} \cap A)}{P(A)} = \frac{1}{P(A)} \sum_i x_i I_A(x_i) P\{X = x_i\}$$

Es decir que:

$$E[X/A] = \frac{1}{P(A)} E[I_A X] \quad (6.1)$$

Notemos que esta fórmula puede adoptarse como definición de la esperanza condicional respecto de un evento para cualquier variable aleatoria (sea discreta o no) mientras tenga esperanza finita, y el evento  $A$  tenga probabilidad positiva.

## 6.2. Esperanzas condicionales en el caso discreto

Ahora consideremos dos variables discretas  $X, Y : \Omega \rightarrow \mathbb{R}$ . Nos proponemos definir el concepto de esperanza condicional  $E[X/Y]$  de  $X$  dada  $Y$ . Supondremos que  $X$  tiene esperanza finita.

Sean  $\{y_j\}$  los distintos valores que toma la variable  $Y$ , y notemos que los eventos  $A_j = \{\omega \in \Omega : Y(\omega) = y_j\}$  forman una partición del espacio muestral  $\Omega$ .

Si  $P\{Y = y_j\} > 0$ , podemos definir

$$E[X/Y = y_j] = E[X/A_j]$$

utilizando la definición introducida en la sección anterior.

Más explícitamente:

$$E[X/Y = y_j] = \sum_i x_i P\{X = x_i / Y = y_j\} \quad (6.2)$$

Las probabilidades  $P\{X = x_i / Y = y_j\}$  que aparecen en esta definición se llaman la distribución condicional de probabilidades de  $X$  dada  $Y$ .

Notemos que depende del valor  $y_j$  de la variable  $Y$ . En consecuencia,  $E[X/Y]$  puede considerarse como una nueva variable aleatoria. Más explícitamente, definimos  $E[X/Y] : \Omega \rightarrow \mathbb{R}$  por:

$$E[X/Y](\omega) = E[X/Y = Y(\omega)]$$

**Lema 6.2.1** *La variable aleatoria  $h(Y) = E[X/Y]$  tiene las siguientes propiedades:*



- Tiene esperanza finita.
- Para cualquier función  $f : \mathbb{R} \rightarrow \mathbb{R}$  acotada, se verifica que:

$$E[f(Y)h(Y)] = E[f(Y)X]$$

Más aún: la esperanza condicional  $E[X/Y]$  está caracterizada por estas dos propiedades. en el siguiente sentido: si  $h_1, h_2 : \mathbb{R} \rightarrow \mathbb{R}$  son dos funciones que verifican estas dos propiedades, entonces

$$P\{h_1(Y) = h_2(Y)\} = 1$$

**Prueba:** Para probar que  $h(Y)$  tiene esperanza finita, debemos mostrar que la serie

$$\sum_j h(y_j)P\{Y = y_j\}$$

donde  $(y_j)$  recorre los posibles valores que la variable  $Y$  toma con probabilidad positiva, es absolutamente convergente.

$$\begin{aligned} \sum_j |h(y_j)|P\{Y = y_j\} &= \sum_j \left| \sum_i x_i P\{X = x_i / Y = y_j\} \right| P\{Y = y_j\} \\ &\leq \sum_i \sum_j |x_i| P\{X = x_i, Y = y_j\} = E(|X|) < +\infty \end{aligned}$$

Para probar la segunda afirmación calculamos:

$$\begin{aligned} E[f(Y)h(Y)] &= \sum_j f(y_j)h(y_j)P\{Y = y_j\} \\ &= \sum_i f(y_j)P\{Y = y_j\} \sum_i x_i P\{X = x_i / Y = y_j\} \\ &= \sum_i \sum_j f(y_j)x_i P\{X = x_i, Y = y_j\} = E[f(Y)X] \end{aligned}$$

donde el reordenamiento de la serie se justifica utilizando que dicha serie converge absolutamente (dado que  $f$  es acotada).

Ahora probaremos la unicidad: supongamos que  $h_1, h_2 : \mathbb{R} \rightarrow \mathbb{R}$  son funciones que verifican las propiedades anteriores. Entonces para cualquier función  $f : \mathbb{R} \rightarrow \mathbb{R}$  acotada, tenemos que:

$$E[f(Y)h_1(Y)] = E[f(Y)h_2(Y)] = E[f(Y)X]$$

En consecuencia, si llamamos  $h = h_1 - h_2$  por la linealidad de la esperanza:

$$E[f(Y)h(Y)] = 0$$

Elijiendo  $f(t) = I_{\{y_j\}}(t)$  deducimos que:

$$h(y_j)P\{Y = y_j\} = 0$$

Por lo tanto si  $h(y_j) \neq 0$ ,  $P\{Y = y_j\} = 0$ . En consecuencia:

$$P\{h(Y) \neq 0\} = \sum_{y_j: h(y_j) \neq 0} P\{Y = y_j\} = 0$$

Es decir que:  $P\{h_1(Y) = h_2(Y)\} = 1$ . □

#### Corolario 6.2.1

$$E[E[X/Y]] = E[X]$$

(Se deduce tomando  $f \equiv 1$  en la fórmula anterior).

### 6.3. Esperanzas condicionales en el caso continuo

En el caso general, no resulta posible definir directamente  $E[X/Y]$  utilizando la fórmula (6.1) ya que usualmente el evento  $\{Y = y\}$  tiene probabilidad cero si  $Y$  es una variable continua.

**Definición 6.3.1** Sean  $X, Y : \Omega \rightarrow \overline{\mathbb{R}}$  variables aleatorias. Decimos que una variable aleatoria  $Z = h(Y)$  es una versión de la esperanza condicional  $E[X/Y]$  si se verifican las siguientes propiedades:

1.  $h(Y)$  tiene esperanza finita.
2. Para cualquier función boreliana acotada  $f : \mathbb{R} \rightarrow \mathbb{R}$  se verifica que:

$$E[f(Y)h(Y)] = E[f(Y)X]$$

Lo que hemos hecho en la sección anterior, es mostrar que la definición (6.2) proporciona una versión de la esperanza condicional en el caso discreto.

El siguiente teorema afirma que siempre existe una versión de la esperanza condicional, aunque no proporciona ninguna fórmula para calcularla. No demostraremos este teorema ya que su demostración depende de un teorema de análisis real (el teorema de Radon-Nikodym)

**Teorema 6.3.1** *Si  $X, Y : \Omega \rightarrow \overline{\mathbb{R}}$  son variables aleatorias, siempre existe una versión de la esperanza condicional  $E[X/Y]$ . Además si  $h_1(Y)$ ,  $h_2(Y)$  son dos versiones de la esperanza condicional  $E[X/Y]$ , entonces*

$$P\{h_1(Y) = h_2(Y)\} = 1$$

**Este capítulo está incompleto: faltan las propiedades de la esperanza condicional y el caso en que  $X$  e  $Y$  admiten una densidad conjunta**

# Apéndice A

## La Fórmula de Stirling

En muchas cuestiones del cálculo de probabilidades, resulta necesario disponer de una aproximación de  $n!$  para  $n$  grande. Este es el contenido de la Fórmula de Stirling:

**Teorema A.0.2 (Fórmula de Stirling)**

$$n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}$$

*Con más precisión, se tienen las desigualdades:*

$$\sqrt{2\pi} n^{n+1/2} e^{-n} < n! < \sqrt{2\pi} e^{-n} \left(1 + \frac{1}{4n}\right)$$

### A.1. La fórmula de Wallis para $\pi$

La siguiente notable fórmula expresa a  $\pi$  como un producto infinito. La utilizaremos para determinar la constante que aparece en la fórmula de Stirling:

**Teorema A.1.1 (Producto infinito de Wallis para  $\pi$ )**

$$\frac{\pi}{2} = \lim_{m \rightarrow +\infty} \left[ \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2m}{2m-1} \cdot \frac{2m}{2m+1} \right]$$

*o en forma de producto infinito*

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2m}{2m-1} \cdot \frac{2m}{2m+1} \cdots$$

Para demostrar esta fórmula, introduzcamos la cantidad

$$I_n = \int_0^{\pi/2} \text{sen}^n x \, dx$$

**Lema A.1.1** *Se verifica la relación de recurrencia:*

$$I_n = \frac{n-1}{n} I_{n-2} \quad (n \geq 2)$$

**Prueba:** Integrando por partes:

$$I_n = \int_0^{\pi/2} \text{sen}^{n-1} x (-\cos x)' dx = -\text{sen}^{n-1} x \cos x \Big|_0^{\pi/2} - \int_0^{\pi/2} (\text{sen}^{n-1} x)' (-\cos x) dx$$

Es decir:

$$I_n = \int_0^{\pi/2} (n-1) \text{sen}^{n-2} \cos^2 x \, dx = \int_0^{\pi/2} (n-1) \text{sen}^{n-2} (1 - \cos^2 x) \, dx = (n-1)[I_{n-2} - I_n]$$

En consecuencia:  $nI_n = (n-1)I_{n-2}$ , o sea:

$$I_n = \frac{n-1}{n} I_{n-2}$$

□

**Prueba de la fórmula de Wallis:**

A fin, de calcular  $I_n$  observamos que

$$I_0 = \int_0^{\pi/2} dx = \frac{\pi}{2}$$

$$I_1 = \int_0^{\pi/2} dx = 1$$

En consecuencia, podemos calcular los valores de  $I_n$  para  $n$  par o impar, respectivamente:

$$I_{2m} = \frac{2m-1}{2m} \cdot \frac{2m-3}{2m-2} \cdots \frac{5}{6} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{\pi}{2}$$

$$I_{2n+1} = \frac{2m}{2m+1} \cdot \frac{2m-2}{2m-1} \cdots \frac{8}{9} \cdot \frac{6}{7} \cdot \frac{4}{5} \cdot \frac{2}{3}$$

Podemos despejar  $\pi/2$ :

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{4}{3} \cdot \frac{6}{5} \cdot \frac{5}{7} \cdots \frac{2m}{2m-1} I_{2m}$$

y utilizando la expresión de  $I_{2m+1}$

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2m}{2m-1} \cdot \frac{2m}{2m+1} \cdot \frac{I_{2m}}{I_{2m+1}}$$

Queremos estimar el cociente  $\frac{I_{2m}}{I_{2m+1}}$ : para ello observamos que en el intervalo  $0 < x < \frac{\pi}{2}$  se tiene  $0 < \sin x < 1$ , en consecuencia  $0 < \sin^{2m+1} x < \sin^{2m-1} x$  e integrando resulta que:

$$0 \leq I_{2m+1} \leq I_{2m} \leq I_{2m-1}$$

luego

$$1 \leq \frac{I_{2m}}{I_{2m-1}} = \frac{2m+1}{2m} \cdot \frac{I_m}{I_{2m-1}} \leq \frac{2m+1}{2m} = 1 + \frac{1}{2m}$$

Por la propiedad del sandwich deducimos que  $\frac{I_{2m}}{I_{2m+1}}$  tiende a 1 cuando  $m \rightarrow +\infty$ . En consecuencia:

$$\frac{\pi}{2} = \lim_{m \rightarrow +\infty} \left[ \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2m}{2m-1} \cdot \frac{2m}{2m+1} \cdot \frac{I_{2m}}{I_{2m+1}} \right]$$

Esto completa la demostración de la fórmula de Wallis.

### A.1.1. Otra fórmula de la fórmula de Wallis

Podemos escribir el resultado anterior en la forma:

$$\frac{\pi}{2} = \lim_{m \rightarrow +\infty} \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m)^2}{3^2 \cdot 5^2 \cdot 7^2 \cdots (2m-1)^2 (2m+1)}$$

Como  $\lim_{m \rightarrow +\infty} \frac{2m+1}{2m} = 1$  obtenemos (producto de límites):

$$\frac{\pi}{2} = \lim_{m \rightarrow +\infty} \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m-2)^2}{3^2 \cdot 5^2 \cdot 7^2 \cdots (2m-1)^2} \cdot 2m$$

Tomando raíz cuadrada:

$$\sqrt{\frac{\pi}{2}} = \lim_{m \rightarrow +\infty} \frac{2 \cdot 4 \cdot 6 \cdots (2m-2)}{3 \cdot 5 \cdot 7 \cdots (2m-1)} \cdot \sqrt{2m}$$

Multiplicando el denominador y el denominador por  $2 \cdot 4 \cdot 6 \cdots (2m-2)$  resulta:

$$\begin{aligned} \sqrt{\frac{\pi}{2}} &= \lim_{m \rightarrow +\infty} \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m-2)^2}{2 \cdot 3 \cdot 5 \cdot 6 \cdot 7 \cdots (2m-1)} \cdot \sqrt{2m} \\ &= \lim_{m \rightarrow +\infty} \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m)^2}{(2m)!} \cdot \frac{\sqrt{2m}}{2m} \end{aligned}$$

$$\begin{aligned}
 &= \lim_{m \rightarrow +\infty} \frac{2^{2m} (1^2 \cdot 2^2 \cdot 3^2 \cdots m^2)}{(2m)! \sqrt{2m}} \\
 &= \lim_{m \rightarrow +\infty} \frac{2^{2m} (m!)^2}{(2m)! \sqrt{2m}}
 \end{aligned}$$

Multiplicando ambos miembros por  $\sqrt{2}$ , resulta:

**Teorema A.1.2 (Otra forma de la fórmula de Wallis)**

$$\sqrt{\pi} = \lim_{m \rightarrow +\infty} \frac{2^{2m} (m!)^2}{(2m)! \sqrt{m}}$$

## A.2. Prueba de la fórmula de Stirling

La prueba de la fórmula de Stirling, se basa en la siguiente idea: tenemos que

$$\log(n!) = \sum_{k=1}^n \log(k) \tag{A.1}$$

Cuando  $n$  es grande, es razonable que esperar que el valor de  $\log(n!)$  esté próximo del valor de la siguiente integral, que representa el área bajo la curva  $y = \log x$  (en el intervalo  $1 \leq x \leq n$ ) y que podemos calcular exactamente:

$$A_n = \int_1^n \log x \, dx = n \log n - n + 1$$

La suma en (A.1) representa una aproximación a esta integral por medio de rectángulos (sumas de Riemman). Una aproximación mejor se consigue utilizando la aproximación por medio de trapecios:

$$T_n = \sum_{k=1}^{n-1} \frac{\log(k) + \log(k+1)}{2} = \sum_{k=1}^{n-1} \log(k) + \frac{1}{2} \log n = \log(n!) - \frac{1}{2} \log n$$

Como la función  $f(x) = \log x$  es cóncava, la secante a la curva  $y = f(x)$  que une los puntos  $(k, \log(k))$  y  $(k+1, \log(k+1))$  queda por abajo de dicha curva. En consecuencia,

$$A_n \geq T_n$$

Nuestro objetivo es estimar el error  $E_n = A_n - T_n$ . Notamos que:

$$E_{k+1} - E_k = \int_k^{k+1} \log x \, dx - \frac{\log(k) + \log(k+1)}{2}$$

representa el área que queda entre la recta secante y la curva en el intervalo  $[k, k+1]$ . Como la función es cóncava,  $E_{k+1} - E_k \geq 0$ . Por otro lado el área entre la curva la secante podemos acotarla por el área entre la tangente a la curva en  $x = k + 1/2$ , es decir la recta:

$$y = T(x) = \log(k + 1/2) + \frac{1}{k + 1/2}(x - (k + 1/2))$$

y la secante (pues siendo  $f$  cóncava, tenemos que  $f(x) \leq T(x)$ ). Deducimos que:

$$E_{k+1} - E_k \leq \int_k^{k+1} T(x) \, dx - \frac{\log(k) + \log(k+1)}{2}$$

es decir:

$$\begin{aligned} E_{k+1} - E_k &\leq \log(k + 1/2) - \frac{\log(k) + \log(k+1)}{2} \\ &= \frac{1}{2} \left(1 + \frac{1}{2k}\right) - \frac{1}{2} \left(1 + \frac{1}{2(k+1/2)}\right) < \frac{1}{2} \left(1 + \frac{1}{2k}\right) - \left(1 + \frac{1}{2(k+1)}\right) \end{aligned}$$

Sumando estas igualdades para  $k = 1, 2, \dots, n-1$ , todos los términos del lado derecho se cancelan, excepto dos (serie telescópica), y como  $E_0$ , obtenemos que:

$$E_n < \frac{1}{2} \log \frac{3}{2} - \frac{1}{2} \log \left(1 + \frac{1}{2n}\right) < \frac{1}{2} \log \frac{3}{2}$$

Notamos que  $E_n$  es entonces, monótona creciente y acotada, por lo tanto  $E_n$  tiende a un límite  $E$  cuando  $n \rightarrow +\infty$ . Y la desigualdad para  $E_{k+1} - E_k$  permite estimar la diferencia  $E - E_n$ :

$$E - E_n \leq \sum_{k=n}^{\infty} (E_{k+1} - E_k) < \frac{1}{2} \left(1 + \frac{1}{2n}\right)$$

Entonces como  $A_n = T_n + E_n$ , obtenemos que:

$$\log(n!) = (n + 1/2) \log(n) - n + 1 - E_n$$

o escribiendo  $\alpha_n = e^{1-E_n}$ , y tomando exponencial:

$$n! = \alpha_n n^{n+1/2} e^{-n}$$



La sucesión  $\alpha_n$  es ahora monótona decreciente, y tiende al límite:  $\alpha = e^{1-E}$ . En consecuencia, por las estimaciones anteriores:

$$1 \leq \frac{\alpha_n}{\alpha} = e^{E-E_n} < e^{(1/2)\log(1+1/2n)} = \sqrt{1 + \frac{1}{2n}} \leq 1 + \frac{1}{2n}$$

En consecuencia, tenemos las desigualdades:

$$\alpha n^{n+1/2} e^{-n} \leq n! \leq \alpha \left(1 + \frac{1}{2n}\right) n^{n+1/2} e^{-n}$$

Nos queda determinar el valor de la constante  $\alpha$ . Para ello utilizamos la fórmula de Wallis,

$$\sqrt{\pi} = \lim_{m \rightarrow +\infty} \frac{2^{2m} (m!)^2}{(2m)! \sqrt{m}} = \lim_{n \rightarrow +\infty} \frac{\alpha_n^2}{\alpha_{2n} \sqrt{2}} = \frac{\alpha^2}{\alpha \sqrt{2}}$$

por lo que deducimos que  $\alpha = \sqrt{2\pi}$ .

## Apéndice B

# Construcción de la Integral de Lebesgue, y equivalencia de las distintas definiciones de esperanza

### Motivación

En este apéndice presentaremos una construcción de la integral de Lebesgue, que es una herramienta útil para definir esperanzas de variables aleatorias y operar con ellas (Se desarrolla en los cursos de análisis real, pero aquí presentaremos algunas nociones básicas, siempre teniendo en mente la interpretación probabilística).

Para ver porqué la integral de Stieltjes no es adecuada para muchos propósitos teóricos, consideremos la definición que hemos dado anteriormente de la esperanza de una variable aleatoria  $X$  en términos de una integral de Stieltjes:

$$E[X] = \int_{-\infty}^{+\infty} x \, dF(x)$$

siendo  $F = F_X$  su función de distribución. Esta definición es muy útil desde el punto de vista del cálculo, ya que no necesitamos conocer cuál es el espacio muestral o cuál es la función  $P$  que asigna las probabilidades. Toda la información relevante sobre la variable  $X$  está contenida en su función de distribución  $F_X$ .

Sin embargo, por ejemplo resulta complicado por ejemplo, con esta definición probar que la esperanza es lineal, ya que  $F_X$  no depende linealmente de  $X$ .

Otro ejemplo es el siguiente (tomado del libro de Barry James): Si usamos la

integral de Stieltjes, entonces la fórmula:

$$E[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x) dF(x)$$

puede no tener sentido si  $\varphi$  tiene un punto de discontinuidad en común con  $F$ . Esa es la razón por la que si utilizamos la integral de Stieltjes, debemos restringir  $\varphi$  a ser una función continua, y entonces por ejemplo  $\varphi$  no puede ser el indicador de un evento.

Por el contrario, la teoría de la integral de Lebesgue permite probar los teoremas sobre la esperanza de variables aleatorias con toda generalidad, y en forma sencilla y elegante.

Uno de los propósitos fundamentales de este apéndice es presentar una prueba de dos teoremas centrales de la teoría de Lebesgue: el teorema de convergencia monótona y el teorema de convergencia mayorada, que forman parte del programa de la asignatura Probabilidad y Estadística (para matemáticos).

Así mismo, probaremos que la definición de esperanza en términos de la integral de Stieltjes es equivalente a la que utiliza la integral de Lebesgue.

## B.1. Funciones Medibles

Consideramos un conjunto  $\Omega$  y una  $\sigma$ -álgebra  $\mathcal{M}$  de subconjuntos de  $\Omega$ . Al par  $(\Omega, \mathcal{M})$  lo llamamos *espacio medible*. A los conjuntos de  $\mathcal{M}$  los llamaremos *conjuntos medibles* (representará la clase de aquellos conjuntos a los que asignaremos medida o probabilidad).

En la interpretación probabilística,  $\Omega$  es el *espacio muestral* (conjunto de posibles resultados de un experimento aleatorio) y  $\mathcal{M}$  será la  $\sigma$ -álgebra  $\mathcal{E}$  de los eventos (aquellas partes de  $\Omega$  a las que les asignaremos probabilidad).

Las funciones con las que vamos a trabajar deberán satisfacer una condición técnica, a saber que podamos medir ciertos conjuntos asociados a la función.

**Definición B.1.1** Sea  $(\Omega, \mathcal{M})$  un espacio medible y sea  $f : \Omega \rightarrow \overline{\mathbb{R}}$  una función. Diremos que  $f$  es una **función medible** (respecto a la  $\sigma$ -álgebra  $\mathcal{M}$ ) si para todo  $\alpha \in \mathbb{R}$  el conjunto  $\{f > \alpha\} = \{\omega \in \Omega : f(\omega) > \alpha\}$  es medible, es decir pertenece a  $\mathcal{M}$ .

Si  $(\Omega, \mathcal{E}, P)$  es un espacio de probabilidad, las funciones medibles sobre  $\Omega$  (respecto a la  $\sigma$ -álgebra  $P$ ) son precisamente las *variables aleatorias* definidas sobre  $\Omega$ .

La noción de función medible puede formularse de varias maneras equivalentes. (En lo sucesivo, usaremos las notaciones abreviadas  $\{f < \alpha\} = \{\omega \in \Omega : f(\omega) < \alpha\}$ , etcétera).

**Lema B.1.1** Sea  $f : \Omega \rightarrow \overline{\mathbb{R}}$  una función. Son equivalentes:

- i)  $f$  es medible.
- ii) Para todo  $\alpha \in \mathbb{R}$ ,  $\{f \geq \alpha\}$  es medible.
- iii) Para todo  $\alpha \in \mathbb{R}$ ,  $\{f < \alpha\}$  es medible.
- iv) Para todo  $\alpha \in \mathbb{R}$ ,  $\{f \leq \alpha\}$  es medible.

**Prueba:**  $i) \Rightarrow ii)$ :

$$\{f \geq \alpha\} = \bigcap_{n \in \mathbb{N}} \{f > \alpha - 1/n\}$$

Como  $f$  es medible, cada uno de los conjuntos  $\{f > \alpha - 1/n\}$  pertenece a  $\mathcal{M}$ , y como  $\mathcal{M}$  es una  $\sigma$ -álgebra, es cerrada por intersecciones numerables. Concluimos que  $\{f \geq \alpha\} \in \mathcal{M}$ .

$ii) \Rightarrow iii)$ : Notamos que  $\{f < \alpha\} = \Omega - \{f \geq \alpha\}$ , y como  $\mathcal{M}$  es cerrada por complementos,  $\{f < \alpha\} \in \mathcal{M}$ .

$iii) \Rightarrow iv)$ : Escribimos

$$\{f \leq \alpha\} = \bigcap_{n \in \mathbb{N}} \{f < \alpha + 1/n\}$$

y utilizamos que  $\mathcal{M}$  es cerrada por intersecciones numerables.

$iv) \Rightarrow i)$ : Notamos que  $\{f > \alpha\} = \Omega - \{f \leq \alpha\}$ , y utilizamos que  $\mathcal{M}$  es cerrada por complementos.  $\square$

**Proposición B.1.1** Sean  $f, g : \Omega \rightarrow \overline{\mathbb{R}}$  funciones medibles. Entonces:  
 $\{f < g\} = \{\omega \in \Omega : f(\omega) < g(\omega)\}$  es medible.

**Prueba:** Notamos que

$$\{f < g\} = \bigcup_{q \in \mathbb{Q}} \{f < q < g\} = \bigcup_{q \in \mathbb{Q}} (\{f < q\} \cap \{q < g\})$$

y usamos que  $\mathcal{M}$  es una  $\sigma$ -álgebra y que  $\mathbb{Q}$  es numerable.  $\square$

El hecho de que la  $\sigma$ -álgebra  $\mathcal{M}$  sea cerrada por operaciones conjuntísticas numerables, tendrá como consecuencia que la clase de funciones medibles será cerrada por las operaciones algebraicas, y por las operaciones de tomar supremo o límites. Más precisamente tenemos las siguientes propiedades:

**Lema B.1.2** Sean  $f, g : \Omega \rightarrow \mathbb{R}$  funciones medibles. Entonces:

i)  $f + k$  y  $kf$  son medibles para todo  $k \in \mathbb{R}$ .

ii)  $f + g$  y  $f - g$  son medibles.

iii)  $f^2$  es medible.

iv)  $f \cdot g$  es medible,

v) Si  $g \neq 0$ ,  $f/g$  es medible.

**Prueba:** i):  $\{f + k > \alpha\} = \{f > \alpha - k\}$  Si  $k > 0$ :  $\{kf > \alpha\} = \{f > \alpha/k\}$  mientras que si  $k < 0$ :  $\{kf > \alpha\} = \{f < \alpha/k\}$

ii):  $\{f + g > \alpha\} = \{f > \alpha - g\}$  y  $\alpha - g$  es medible por i)

iii): Si  $\alpha \geq 0$ ,  $\{f^2 > \alpha\} = \{f > \sqrt{\alpha}\} \cup \{f < -\sqrt{\alpha}\}$  (sino  $\{f^2 > \alpha\} = \Omega$ ).

iv): Se deja como ejercicio (por iii) basta ver que  $1/g$  es medible)  $\square$

**Observación:** El lema se puede adaptar al caso en que  $f$  o  $g$  toman los valores  $\pm\infty$ .  $f+g$  está bien definida, salvo cuando es de la forma  $(+\infty)+(-\infty)$  o  $(-\infty)+\infty$ . Para definir  $f \cdot g$ , hay que utilizar las convenciones  $0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = 0$

**Lema B.1.3** Sea  $(f_n)_{n \in \mathbb{N}}$  una sucesión de funciones medibles. Entonces

$$\sup_{n \in \mathbb{N}} f_n(x) \quad \inf_{n \in \mathbb{N}} f_n(x)$$

$$\liminf_{n \in \mathbb{N}} f_n(x) \quad \limsup_{n \in \mathbb{N}} f_n(x)$$

son medibles.

En particular si  $f_n$  converge, entonces:

$$f(x) = \lim_{n \rightarrow +\infty} f_n(x)$$

es medible.

**Prueba:** Notamos que

$$\{\sup_{n \in \mathbb{N}} f_n(x) > \lambda\} = \bigcup_{n \in \mathbb{N}} \{f_n > \lambda\}$$

Por lo que si cada  $f_n$  es medible,  $\{f_n > \lambda\} \in \mathcal{M} \forall n \in \mathbb{N}$ , y en consecuencia como  $\mathcal{M}$  es una  $\sigma$ -álgebra,  $\{\sup_{n \in \mathbb{N}} f_n(x) > \lambda\} \in \mathcal{M}$ . Esto prueba que  $\sup_{n \in \mathbb{N}} f_n(x)$  es medible.

Del mismo modo, se prueba que  $\inf_{n \in \mathbb{N}} f_n(x)$  es medible, ya que:

$$\left\{ \inf_{n \in \mathbb{N}} f_n(x) < \lambda \right\} \bigcup_{n \in \mathbb{N}} \{f_n < \lambda\}$$

Para probar que  $\limsup f_n$  es medible, notamos que

$$\limsup f_n = \inf_k \sup_{k \geq n} f_n$$

Pero para cada  $k$ ,  $\sup_{k \geq n} f_n$  es medible por lo que ya probamos, y en consecuencia  $\limsup f_n$  es medible. De modo análogo, de que

$$\liminf f_n = \sup_k \inf_{k \geq n} f_n$$

Se deduce que  $\liminf f_n$  es medible. Finalmente notamos que si la sucesión  $(f_n)$  converge, entonces  $\lim_{n \rightarrow +\infty} f_n(x) = \liminf f_n(x) = \limsup f_n(x)$ , por lo que la función límite de las  $f_n$  es medible.  $\square$

**Definición B.1.2** Sea  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  una función. Diremos que  $\varphi$  es medible Borel si es medible con respecto a la  $\sigma$ -álgebra de Borel  $\mathcal{B}(\mathbb{R})$ , generada por los intervalos. Es decir si para todo intervalo  $(a, b]$ , su pre-imagen por  $\varphi$ ,  $\varphi^{-1}((a, b])$  es un conjunto boreliano de la recta.

**Lema B.1.4** Sean  $(\Omega, \mathcal{M})$  un espacio medible y  $f : \Omega \rightarrow \mathbb{R}$  una función. Entonces  $f$  es medible si y sólo si  $f^{-1}(B) \in \mathcal{M}$  para todo  $B \in \mathcal{B}(\mathbb{R})$ .

**Prueba:** Notamos que:

$$\mathcal{A} = \{B \subset \mathbb{R} : f^{-1}(B) \in \mathcal{M}\}$$

es una  $\sigma$ -álgebra. Si  $f$  es medible, entonces  $\mathcal{A}$  contiene a los intervalos. Por lo tanto contiene a toda la  $\sigma$ -álgebra de Borel (que es la menor  $\sigma$ -álgebra que contiene a los intervalos).  $\square$

**Corolario B.1.1** Si  $(\Omega, P)$  es un espacio medible,  $f : \Omega \rightarrow \mathbb{R}$  es medible y  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  es medible Borel, entonces  $\varphi \circ f : \Omega \rightarrow \mathbb{R}$  es medible.

**Prueba:** Sea  $B$  un boreliano de la recta, entonces  $\varphi^{-1}(B)$  es boreliano, y en consecuencia como  $f$  es medible:

$$(\varphi \circ f)^{-1}(B) = f^{-1}(\varphi^{-1}(B)) \in \mathcal{M}$$

Como esto vale para todo  $B$  boreliano, concluimos que  $\varphi \circ f$  es medible.  $\square$

**Interpretación probabilística:** Sea  $(\Omega, \mathcal{E}, P)$  un espacio de probabilidad. Si  $X : \Omega \rightarrow \mathbb{R}$  es una variable aleatoria, y  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  es medible Borel, entonces  $\varphi(X) = \varphi \circ X : \Omega \rightarrow \mathbb{R}$  es una variable aleatoria.

### B.1.1. Funciones Simples

**Definición B.1.3** Llamamos función simple a una función medible  $f : \Omega \rightarrow \mathbb{R}$  que toma un número finito de valores  $\{c_1, c_2, \dots, c_n\}$ . Podemos representarla entonces como:

$$f = \sum_{i=1}^n c_i I_{E_i} \quad (\text{B.1})$$

donde  $E_i = \{\omega \in \Omega : f(\omega) = c_i\}$ , y  $I_{A_i}$  es el **indicador** (o **función característica**<sup>1</sup>) del conjunto  $E_i$ , definido por:

$$I_{E_i}(\omega) = \begin{cases} 1 & \text{si } \omega \in E_i \\ 0 & \text{si } \omega \notin E_i \end{cases}$$

En la interpretación probabilística, las funciones simples corresponden a las variables aleatorias discretas que toman sólo un número finito de valores.

El siguiente lema de aproximación por funciones simples, será de gran utilidad para la teoría de la integral:

**Lema B.1.5** Si  $f : \Omega \rightarrow [0, +\infty]$  es una función medible no negativa, entonces existe una sucesión  $\varphi_n(x)$  de funciones simples no negativas tales que

$$\lim_{n \rightarrow +\infty} \varphi_n(x) = f(x) \quad \forall x \in \Omega$$

**Prueba:** Para cada  $n \in \mathbb{N}$ , definimos:

$$\varphi_n(x) = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} I_{E_{n,i}}(x) + n I_{F_n}(x)$$

siendo

$$E_{n,i} = \left\{ x \in \Omega : \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n} \right\}$$

$$F_n = \{x \in \Omega : f(x) \geq n\}$$

Es decir que:

$$\varphi_n(x) = \begin{cases} \frac{i-1}{2^n} & \text{si } \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n} \\ n & \text{si } f(x) \geq n \end{cases}$$

Se prueba que  $\varphi_n(x)$  tiene las propiedades del enunciado. □

---

<sup>1</sup>En la teoría de probabilidades el nombre de función característica suele usarse para otra cosa, por eso preferimos en estas notas el de indicador. A veces se usa la notación  $\chi_E$  en lugar de  $I_E$

## B.2. Integral de Funciones Simples

Consideramos ahora un espacio de medida  $(\Omega, \mathcal{M}, \mu)$  es decir un espacio medible, donde además está definida una medida ( $\sigma$ -aditiva)  $\mu : \mathcal{M} \rightarrow [0, +\infty]$ .

Si  $f : \Omega \rightarrow \mathbb{R}$  es una función simple, representada por (B.1) definimos su integral de la siguiente manera:

$$\int_{\Omega} f \, d\mu = \sum_{i=1} c_i \mu(A_i)$$

En la interpretación probabilística, tenemos un espacio de probabilidad  $(\Omega, \mathcal{E}, P)$  donde la probabilidad no es otra cosa que una medida que asigna a todo el espacio  $\Omega$  medida 1 (o sea:  $P(\Omega) = 1$ ).

Entonces la definición de integral de una función simple, no es otra cosa que nuestra definición de esperanza de una variable aleatoria discreta, escrita en el lenguaje de la teoría de la medida. Es decir, que si  $X : \Omega \rightarrow \mathbb{R}$  es una variable aleatoria discreta, entonces

$$E[X] = \int_{\Omega} X \, dP$$

La integral de las funciones simples, tiene las siguientes propiedades: (que se demuestran exactamente como las propiedades de la esperanza de variables aleatorias discretas)

**Proposición B.2.1**    1. *linealidad: Si  $f$  y  $g$  son funciones simples:*

$$\int_{\Omega} (f + g) \, d\mu = \int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu$$

*Si  $f$  es una función simple, y  $k$  una constante:*

$$\int_{\Omega} (kf) \, d\mu = k \int_{\Omega} f \, d\mu$$

2. *Monotonía: si  $f$  y  $g$  son funciones simples y  $f \leq g$ , entonces:*

$$\int_{\Omega} f \, d\mu \leq \int_{\Omega} g \, d\mu$$

3. *Si  $f$  es una función simple, entonces*

$$\left| \int_{\Omega} f \, d\mu \right| \leq \int_{\Omega} |f| \, d\mu$$



## B.3. Integral de funciones no negativas

**Definición B.3.1** Sea  $(\Omega, \mathcal{M}, \mu)$  un espacio de medida, y  $f : \Omega \rightarrow [0, +\infty]$  una función medible no negativa. Definimos la integral de  $f$  de la siguiente manera:

$$\int_{\Omega} f d\mu = \sup \left\{ \int_{\Omega} \varphi d\mu : 0 \leq \varphi \leq f, \varphi \text{ simple} \right\}$$

Una consecuencia inmediata de la definición es la siguiente:

**Proposición B.3.1** Si  $f, g : \Omega \rightarrow [0, +\infty]$  son funciones simples no negativas tales que  $f \leq g$ , entonces

$$\int_{\Omega} f(x) d\mu \leq \int_{\Omega} g(x) d\mu$$

**Definición B.3.2** Si  $A \in \mathcal{M}$  es un conjunto medible, y  $f : \Omega \rightarrow [0, +\infty]$  es una función medible no negativa, definimos la integral de  $f$  sobre  $E$  como:

$$\int_{\Omega} f d\mu = \int_{\Omega} f \cdot I_A d\mu$$

**Lema B.3.1** Sea  $\varphi$  una función simple no negativa. Entonces la función  $\lambda = \lambda_{\varphi} : \mathcal{M} \rightarrow [0, +\infty]$  definida por:

$$\lambda(A) = \int_A \varphi d\mu$$

es una medida

**Prueba:** Supongamos que un conjunto medible  $A$  se representa como una unión disjunta numerable de una sucesión  $(A_n)_{n \in \mathbb{N}}$  de conjuntos medibles:

$$A = \bigcup_{n \in \mathbb{N}} A_n$$

Queremos probar que:

$$\lambda(A) = \sum_{n=1}^{\infty} \lambda(A_n)$$

Como  $\varphi$  es una función simple, podremos representarla en la forma

$$\varphi = \sum_{i=1}^N c_i I_{E_i}$$

siendo  $E_i$  conjuntos medibles disjuntos.

Notamos que  $\varphi(x)I_{A_n}(x)$  es una función simple, que toma el valor  $c_i$  en el conjunto  $A_n \cap E_i$ , es decir que su representación canónica es:

$$\varphi(x)I_{A_n}(x) = \sum_{i=1}^N c_i I_{E_i \cap A_n}$$

En consecuencia,

$$\lambda(A_n) = \sum_{i=1}^N c_i \mu(E_i \cap A_n)$$

Y por lo tanto

$$\sum_{n=1}^{\infty} \lambda(A_n) = \sum_{n=1}^{\infty} \sum_{i=1}^N c_i \mu(E_i \cap A_n)$$

Como en esta suma doble los términos  $\mu(E_i \cap A_n)$  son no negativos, da lo mismo efectuar la suma en cualquier orden. En consecuencia,

$$\sum_{n=1}^{\infty} \lambda(A_n) = \sum_{i=1}^N \sum_{n=1}^{\infty} c_i \mu(E_i \cap A_n) = \sum_{i=1}^N c_i \sum_{n=1}^{\infty} \mu(E_i \cap A_n)$$

Ahora notamos que:

$$E_i \cap A = \bigcup_{n \in \mathbb{N}} (E_i \cap A_n)$$

siendo esta unión disjunta. En consecuencia, como  $\mu$  es una medida,

$$\mu(E_i \cap A) = \sum_{n=1}^{\infty} \mu(E_i \cap A_n)$$

y concluimos que:

$$\sum_{n=1}^{\infty} \lambda(A_n) = \sum_{i=1}^N c_i \mu(E_i \cap A) = \int_{\Omega} \varphi(x) I_A(x) d\mu = \int_A \varphi(x) d\mu$$

□

**Teorema B.3.1 (Teorema de la Convergencia Monótona )** <sup>2</sup> Sea  $f_n(x) : \Omega \rightarrow [0, +\infty]$  una sucesión creciente (o sea:  $f_n(x) \leq f_{n+1}(x)$ ) de funciones medibles no negativas. Entonces,

$$\int_{\Omega} \lim_{n \rightarrow +\infty} f_n(x) d\mu = \lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu$$

---

<sup>2</sup>También conocido como teorema de Beppo Levi.

**Prueba:** Sea

$$f(x) = \lim_{n \rightarrow +\infty} f_n(x)$$

Por la monotonía de la integral es claro que:

$$\int_{\Omega} f_n(x) d\mu \leq \int_{\Omega} f(x) d\mu$$

Y por lo tanto que:

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu \leq \int_{\Omega} f(x) d\mu$$

Por otra parte, sea  $\varphi$  una función simple tal que  $\varphi \leq f$ . Dado  $\alpha \in (0, 1)$ , consideramos los conjuntos (medibles)

$$A_n = \{x \in \Omega : f_n(x) \geq \alpha \varphi(x)\}$$

Entonces la sucesión  $(A_n)_{n \in \mathbb{N}}$  es monótona creciente (o sea  $A_n \subset A_{n+1}$ ) y

$$\Omega = \bigcup_{n \in \mathbb{N}} A_n$$

Además la función  $\lambda_{\varphi}$  definida en el lema anterior, es una medida, por lo tanto:

$$\lambda(\Omega) = \lim_{n \rightarrow +\infty} \lambda(A_n)$$

es decir,

$$\lim_{n \rightarrow +\infty} \int_{A_n} \varphi(x) d\mu = \int_{\Omega} \varphi(x) d\mu$$

Por otra parte, para cada  $n \in \mathbb{N}$ ,

$$\alpha \int_{A_n} \varphi(x) d\mu \leq \int_{A_n} f_n(x) d\mu \leq \int_{\Omega} f_n(x) d\mu$$

De modo que,

$$\alpha \int_{\Omega} \varphi(x) d\mu = \alpha \lim_{n \rightarrow +\infty} \int_{A_n} \varphi(x) d\mu \leq \lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu$$

Haciendo tender  $\alpha$  a 1 deducimos que:

$$\int_{\Omega} \varphi(x) d\mu \leq \lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu$$

y por lo tanto como esto vale para toda función simple  $\varphi$  con  $0 \leq \varphi \leq f$ , por la definición de integral, deducimos que:

$$\int_{\Omega} f(x) d\mu \leq \lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu$$

□

**Proposición B.3.2 (Linealidad de la integral)** Si  $f, g : \Omega \rightarrow [0, +\infty]$  son funciones medibles no negativas y  $\lambda_1, \lambda_2 \geq 0$  son números reales no negativos, entonces:

$$\int_{\Omega} [\lambda_1 f(x) + \lambda_2 g(x)] d\mu = \lambda_1 \int_{\Omega} f(x) d\mu + \lambda_2 \int_{\Omega} g(x) d\mu$$

**Prueba:** Utilizamos el lema de aproximación por funciones simples: sabemos que existen una sucesión creciente  $(f_n(x))$  de funciones simples que converge a  $f(x)$ , y una sucesión creciente  $(g_n(x))$  de funciones simples que converge a  $g(x)$ . Entonces por la linealidad de la integral de funciones simples,

$$\int_{\Omega} [\lambda_1 f_n(x) + \lambda_2 g_n(x)] d\mu = \lambda_1 \int_{\Omega} f_n(x) d\mu + \lambda_2 \int_{\Omega} g_n(x) d\mu$$

Y el teorema de convergencia monótona implica entonces que:

$$\int_{\Omega} [\lambda_1 f(x) + \lambda_2 g(x)] d\mu = \lambda_1 \int_{\Omega} f(x) d\mu + \lambda_2 \int_{\Omega} g(x) d\mu$$

□

**Teorema B.3.2 (Lema de Fatou)** Sea  $f_n : \mathcal{M} \rightarrow [0, +\infty]$  una sucesión de funciones medibles no negativas. Entonces:

$$\int_{\Omega} \liminf_{n \rightarrow +\infty} f_n(x) d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu$$

**Prueba:** Llamemos

$$f(x) = \liminf_{n \rightarrow +\infty} f_n(x) = \sup_{k \in \mathbb{N}} \left( \inf_{n \geq k} f_n(x) \right)$$

y consideremos la sucesión creciente de funciones no negativas:

$$g_k(x) = \inf_{n \geq k} f_n(x)$$

Entonces por el teorema de convergencia monótona:

$$\int_{\Omega} f(x) d\mu = \int_{\Omega} \lim_{k \rightarrow +\infty} g_k(x) d\mu = \lim_{k \rightarrow +\infty} \int_{\Omega} g_k(x) d\mu \quad (\text{B.2})$$

Por otra parte si  $n \geq k$ , tenemos que

$$\int_{\Omega} g_k(x) d\mu \leq \int_{\Omega} f_n(x) d\mu$$

y en consecuencia:

$$\int_{\Omega} g_k(x) d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu$$

Y por lo tanto:

$$\lim_{k \rightarrow +\infty} \int_{\Omega} g_k(x) d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu$$

En consecuencia utilizando (B.2), deducimos que:

$$\int_{\Omega} f(x) d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu$$

□

## B.4. Funciones Integrables

Si  $f : \Omega \rightarrow \overline{\mathbb{R}}$  es una función medible, hacemos la descomposición:

$$f = f^+ - f^- \tag{B.3}$$

como diferencia de dos funciones medibles no negativas, siendo

$$f^+(x) = \begin{cases} f(x) & \text{si } f(x) \geq 0 \\ 0 & \text{si } f(x) < 0 \end{cases}$$

y

$$f^-(x) = \begin{cases} 0 & \text{si } f(x) \geq 0 \\ -f(x) & \text{si } f(x) < 0 \end{cases}$$

Notamos que:

$$|f| = f^+ + f^-$$

**Definición B.4.1** Diremos que una función medible  $f : \Omega \rightarrow \overline{\mathbb{R}}$  es integrable si son finitas las integrales

$$\int_{\Omega} f^+(x) d\mu$$

y

$$\int_{\Omega} f^-(x) d\mu$$

En ese caso, definimos la integral de  $f$  con respecto a  $\mu$  en el espacio  $\Omega$  por:

$$\int_{\Omega} f(x) d\mu = \int_{\Omega} f^+(x) d\mu + \int_{\Omega} f^-(x) d\mu$$

Observación: De la definición de función integrable, deducimos que  $f$  es integrable si y sólo si

$$\int_{\Omega} |f(x)| d\mu < +\infty$$

Además:

$$\left| \int_{\Omega} f(x) d\mu \right| \leq \int_{\Omega} |f(x)| d\mu$$

**Proposición B.4.1 (Linealidad de la integral)** Si  $f, g : \Omega \rightarrow \overline{\mathbb{R}}$  son funciones integrables y  $\lambda_1, \lambda_2$  son números reales, entonces  $\lambda_1 f + \lambda_2 g$  es integrable, y se tiene que:

$$\int_{\Omega} [\lambda_1 f(x) + \lambda_2 g(x)] d\mu = \lambda_1 \int_{\Omega} f(x) d\mu + \lambda_2 \int_{\Omega} g(x) d\mu$$

**Prueba:** Primero probaremos que es posible sacar escalares de la integral: En efecto si  $\lambda > 0$ , tenemos que:

$$(\lambda f)^+ = \lambda f^+$$

$$(\lambda f)^- = \lambda f^-$$

Entonces es claro por la definición y la linealidad de la integral para funciones no negativas, que si  $f$  es integrable,  $\lambda f$  también lo es y se verifica que:

$$\begin{aligned} \int_{\Omega} \lambda f d\mu &= \int_{\Omega} (\lambda f)^+ d\mu - \int_{\Omega} (\lambda f)^- d\mu = \\ &= \lambda \int_{\Omega} f^+ d\mu - \lambda \int_{\Omega} f^- d\mu \\ &= \lambda \int_{\Omega} f d\mu \end{aligned}$$

Si  $\lambda < 0$ , notamos que:

$$(\lambda f)^+ = (-\lambda) f^-$$

$$(\lambda f)^- = (-\lambda) f^+$$

y de nuevo, vemos usando la definición y la linealidad de la integral para funciones no negativas, que si  $f$  es integrable,  $\lambda f$  también lo es y se verifica que:

$$\begin{aligned} \int_{\Omega} \lambda f d\mu &= \int_{\Omega} (\lambda f)^+ d\mu - \int_{\Omega} (\lambda f)^- d\mu = \\ &= -\lambda \int_{\Omega} f^- d\mu + \lambda \int_{\Omega} f^+ d\mu \end{aligned}$$

$$= \lambda \int_{\Omega} f \, d\mu$$

(El caso  $\lambda = 0$  es trivial porque la integral de la función nula da 0).

Ahora probaremos que la integral distribuye la suma: Para ello notamos que (B.3) proporciona una escritura de  $f$  como diferencia de dos funciones no negativas. Pero que si tenemos otra escritura de  $f$  como diferencia de dos funciones medibles no negativas:

$$f = f_1 - f_2$$

Entonces de  $f^+ - f^- = f_1 - f_2$ , deducimos  $f^+ + f_2 = f_1 + f^-$ , entonces por la linealidad de la integral para funciones no negativas:

$$\int_{\Omega} f^+ \, d\mu + \int_{\Omega} f_2 \, d\mu = \int_{\Omega} f_1 \, d\mu + \int_{\Omega} f^- \, d\mu$$

En consecuencia,

$$\int_{\Omega} f \, d\mu = \int_{\Omega} f_1 \, d\mu - \int_{\Omega} f_2 \, d\mu$$

Vale decir que si en lugar de (B.3), utilizáramos cualquier otra descomposición de  $f$  como diferencia de funciones medibles no negativas obtendríamos el mismo valor de la integral.

Hecha esta observación, notamos que

$$f + g = f^+ - f^- + g^+ - g^- = (f^+ + g^+) - (f^- + g^-)$$

y que esta última expresión proporciona una escritura de  $f + g$  como diferencia de funciones no negativas. En consecuencia, por la observación anterior, y la linealidad de la integral para funciones no negativas:

$$\begin{aligned} \int_{\Omega} (f + g) \, d\mu &= \int_{\Omega} (f^+ + g^+) \, d\mu - \int_{\Omega} (f^- + g^-) \, d\mu = \\ &= \int_{\Omega} f^+ \, d\mu + \int_{\Omega} g^+ \, d\mu - \int_{\Omega} f^- \, d\mu - \int_{\Omega} g^- \, d\mu = \\ &= \int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu \end{aligned}$$

□

**Teorema B.4.1 (De convergencia mayorada, de Lebesgue)** Sea  $f_n(x) : \Omega \rightarrow \overline{\mathbb{R}}$  una sucesión de funciones integrables, que converge puntualmente a una función  $f(x)$

$$f(x) = \lim_{n \rightarrow +\infty} f_n(x)$$

y tal que existe una función integrable  $g$  de modo que  $|f_n(x)| \leq g$  (en casi todo punto con respecto a la medida  $\mu$ ). Entonces

$$\lim_{n \rightarrow +\infty} \int_{\Omega} |f_n(x) - f(x)| d\mu = 0$$

En particular,

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu = \int_{\Omega} f(x) d\mu$$

**Prueba:** Sea  $h_n(x)$  la sucesión de funciones medibles no negativas, definida por:

$$h_n(x) = 2g(x) - |f_n(x) - f(x)|$$

Entonces, por el lema de Fatou,

$$\begin{aligned} 2 \int_{\Omega} g(x) d\mu &= \int_{\Omega} \lim_{n \rightarrow +\infty} h_n(x) d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\Omega} f_n(x) d\mu \\ 2 \int_{\Omega} g(x) d\mu - \limsup_{n \rightarrow +\infty} \int_{\Omega} |f_n(x) - f(x)| d\mu & \end{aligned}$$

En consecuencia,

$$\limsup_{n \rightarrow +\infty} \int_{\Omega} |f_n(x) - f(x)| d\mu = 0$$

Entonces,

$$\left| \int_{\Omega} f_n(x) d\mu - \int_{\Omega} f(x) d\mu \right| \leq \int_{\Omega} |f_n(x) - f(x)| d\mu \rightarrow 0 \text{ cuando } n \rightarrow \infty$$

□

## B.5. Equivalencia de las distintas definiciones de Esperanza

Sean como antes  $(\Omega, \mathcal{E}, P)$  un espacio de probabilidad y  $X : \Omega \rightarrow \overline{\mathbb{R}}$  una variable aleatoria. La esperanza de  $X$  no es otra cosa que su integral de Lebesgue respecto a la medida  $P$ :

$$E[X] = \int_{\Omega} X d\mu$$

A la variable aleatoria  $X$  le podemos asociar la medida  $\mu_X$  (o probabilidad), definida para los conjuntos borelianos de la recta por:

$$\mu_X(B) = P(X^{-1}(B))$$



$\mu_X$  se llama la distribución de probabilidades de  $X$ . Notamos que  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ , donde  $\mathcal{B}(\mathbb{R})$  denota la  $\sigma$ -álgebra de Borel de la recta, es un espacio de probabilidad.

El siguiente lema afirma que es posible transformar las integrales respecto a  $P$ , en integrales respecto a  $\mu_X$ . Por consiguiente  $\mu_X$  contiene toda la información sobre  $X$  que es necesaria para calcular la esperanza de  $X$ , o más generalmente, de una función  $\varphi(X)$  de  $X$ .

**Lema B.5.1** *Sea  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  una función medible Borel. Entonces se tiene que*

$$E[\varphi(X)] = \int_{\Omega} \varphi(X) d\mu = \int_{\mathbb{R}} \varphi(x) d\mu_X$$

*en el siguiente sentido.*

1. *Si  $\varphi$  es no negativa, la fórmula vale sin restricciones. (Notar que estas integrales siempre existen, aunque pueden ser infinitas)*
2. *Si  $\varphi$  es cualquiera, entonces  $\varphi(X)$  es integrable con respecto a  $P$  si y sólo si  $\varphi(x)$  lo es con respecto a  $\mu_X$  y en este caso es válida dicha fórmula.*

**Prueba:** Primero consideramos el caso en que  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  es una función boreliana simple, entonces:

$$\varphi(x) = \sum_{i=1}^n c_i I_{B_i}(x)$$

para ciertos conjuntos  $B_i \subset \mathbb{R}$  borelianos, de modo que:

$$\int_{\mathbb{R}} \varphi(x) d\mu_X = \sum_{i=1}^n c_i \mu_X(B_i)$$

Por otra parte, notamos que  $\varphi(X) : \mathcal{M} \rightarrow \mathbb{R}$  es una función simple que toma el valor  $c_i$  en el conjunto  $X^{-1}(B_i)$ , de modo que:

$$\int_{\Omega} \varphi(X) dP = \sum_{i=1}^n c_i P(X^{-1}(B_i))$$

Dado que por definición de  $\mu_X$ ,  $\mu_X(B_i) = P(X^{-1}(B_i))$ , ambas integrales coinciden.

Sea ahora  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  una función boreliana no negativa. Y consideramos una sucesión creciente de funciones borelianas simples  $\varphi_n : \mathbb{R} \rightarrow \mathbb{R}$  que converge a  $\varphi$  en forma creciente. Dado que para cada  $n \in \mathbb{N}$  tenemos que:

$$\int_{\Omega} \varphi_n(X) dP = \int_{\mathbb{R}} \varphi_n(x) d\mu_X$$

El teorema de convergencia monótona, implica que:

$$\int_{\Omega} \varphi(X) dP = \int_{\mathbb{R}} \varphi(x) d\mu_X$$

Finalmente, consideremos una función boreliana  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  cualquiera. Como  $|\varphi|$  es no negativa, ya sabemos que:

$$\int_{\Omega} |\varphi(X)| dP = \int_{\mathbb{R}} |\varphi(x)| d\mu_X$$

En consecuencia,  $\varphi(X)$  es integrable con respecto a  $P$  si y sólo si  $\varphi(x)$  lo es con respecto a  $\mu_X$ .

Finalmente, hagamos uso de la descomposición:

$$\varphi(x) = \varphi^+(x) - \varphi^-(x)$$

Entonces como  $\varphi^+$  y  $\varphi^-$  son no negativas, tenemos que:

$$\int_{\Omega} \varphi^+(X) dP = \int_{\mathbb{R}} \varphi^+(x) d\mu_X$$

y que:

$$\int_{\Omega} \varphi^-(X) d\mu = \int_{\mathbb{R}} \varphi^-(x) d\mu_X$$

La linealidad de la integral implica entonces que:

$$\int_{\Omega} \varphi(X) dP = \int_{\mathbb{R}} \varphi(x) d\mu_X$$

□

Anteriormente definimos la esperanza utilizando integrales de Stieltjes respecto a la función de distribución de  $X$ . El siguiente teorema afirma que la definición de esperanza que dimos anteriormente coincide con la nueva definición.

En la demostración, utilizaremos la notación:

$$\int_a^b \varphi(x) d\mu_X = \int_{[a,b]} \varphi(x) d\mu_X$$

**Teorema B.5.1** Sea  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  una función continua. Entonces se tiene que:

$$\int_a^b \varphi(x) d\mu_X = \int_a^b \varphi(x) dF(x)$$

en el siguiente sentido:

1. Si  $\varphi$  tiene soporte en un intervalo  $[a, b]$  entonces, la fórmula es válida, y ambos miembros son finitos.
2. Si  $\varphi$  es no negativa, la fórmula es válida sin restricciones (aunque ambas integrales pueden ser infinitas)
3. Si  $\varphi$  es de signo arbitrario, entonces  $\varphi(x)$  es integrable con respecto a  $\mu_X$  si y sólo si

$$\int_{-\infty}^{\infty} |\varphi(x)| dF(x) < +\infty$$

y en este caso, también es válida dicha fórmula.

**Prueba:** Supongamos primero que  $\varphi$  tiene soporte en un intervalo cerrado  $[a, b]$ . Consideremos una partición  $\pi : a = x_0 < x_1 < \dots < x_n = b$  del intervalo y elijamos puntos intermedios  $\xi_i \in (x_i, x_{i+1})$ .

Definamos la función simple  $\varphi_\pi : [a, b] \rightarrow \mathbb{R}$  dada por:

$$\varphi_\pi(x) = \xi_i \text{ si } x \in (x_i, x_{i+1}]$$

Entonces:

$$S(\pi, F) = \sum_{i=1}^{n-1} \varphi(x_i) [F(x_{i+1}) - F(x_i)] = \sum_{i=1}^{n-1} \varphi(\xi_i) \mu_X((x_i, x_{i+1}]) = \int_{\Omega} \varphi_\pi(x) d\mu$$

Ahora bien, como  $\varphi$  es uniformemente continua en  $[a, b]$ , deducimos que  $\varphi_\pi$  converge uniformemente a  $\varphi$  en  $[a, b]$  cuando la norma de la partición  $\pi$  tiende a cero. En efecto, dado  $\varepsilon > 0$ , sea  $\delta > 0$  el que corresponde a  $\varepsilon$  por la continuidad uniforme de  $\varphi$  en  $[a, b]$ . Entonces, si  $x \in (x_i, x_{i+1}]$ ,

$$|\varphi_\pi(x) - \varphi(x)| = |\varphi(\xi_i) - \varphi(x)| < \varepsilon$$

si  $|x_{i+1} - x_i| < \delta$ .

Deducimos que:

$$\lim_{|\pi| \rightarrow 0} \int_a^b \varphi_\pi(x) d\mu_X = \int_a^b \varphi(x) d\mu_X$$

ya que

$$\left| \int_{\Omega} \varphi_\pi(x) d\mu_X - \int_{\Omega} \varphi(x) d\mu_X \right| \leq \int_a^b |\varphi_\pi(x) - \varphi(x)| d\mu_X < \varepsilon \mu_X([a, b]) \leq \varepsilon$$

Por definición de integral de Stieltjes esto dice que la integral

$$\int_a^b \varphi(x) dF(x)$$

existe, y coincide con

$$\int_a^b \varphi(x) d\mu_X$$

Para el caso general, en el que  $\varphi$  no tiene soporte compacto, consideremos cualquier sucesión decreciente  $(a_n)_{n \in \mathbb{N}}$  tal que  $a_n \rightarrow -\infty$ , y cualquier sucesión creciente  $(b_n)_{n \in \mathbb{N}}$  tal que  $b_n \rightarrow +\infty$ , y observemos que

$$\int_{a_n}^{b_n} \varphi(x) d\mu_X = \int_{\mathbb{R}} \varphi(x) I_{[a_n, b_n]}(x) d\mu(x) \rightarrow \int_{\mathbb{R}} \varphi(x) d\mu_X$$

Por el teorema de convergencia monótona aplicado a  $\varphi(x) I_{[a_n, b_n]}$ , si  $\varphi$  es no negativa. En consecuencia,

$$\int_{\mathbb{R}} \varphi(x) d\mu_X = \int_{-\infty}^{\infty} \varphi(x) dF(x) \quad (\text{B.4})$$

vale siempre que  $\varphi(x)$  sea no negativa.

Cuando  $\varphi$  tiene cualquier signo, observamos primero que

$$\int_{\mathbb{R}} |\varphi(x)| d\mu = \int_{-\infty}^{\infty} |\varphi(x)| dF(x)$$

Lo que en particular, dice que  $|\varphi(x)|$  es integrable con respecto a  $\mu_X$  si y sólo si:

$$\int_{-\infty}^{\infty} |\varphi(x)| dF(x) < +\infty$$

Si esto sucede, podemos aplicar el teorema de convergencia mayorada a la sucesión  $\varphi(x) I_{[a_n, b_n]}$  (que claramente está mayorada por  $|\varphi(x)|$ ), y deducir que la fórmula (B.4) es cierta, también en este caso.  $\square$

### B.5.1. Vectores Aleatorios

Las ideas anteriores pueden generalizarse fácilmente a vectores aleatorios. Si  $(\Omega, \mathcal{E}, P)$  es un espacio de probabilidad, un vector aleatorio no es otra cosa que una función medible  $\Omega : X \rightarrow \mathbb{R}^n$ .

Podemos definir la distribución de probabilidades de  $X$  como la medida  $\mu_X$ , definida en la  $\sigma$ -álgebra de Borel de  $\mathbb{R}^n$  por:

$$\mu_X(B) = P(X^{-1}(B))$$

Y si  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  es una función medible Borel, entonces tendremos la fórmula (generalización del lema B.5.1):

$$E[\varphi(X)] = \int_{\Omega} \varphi(X) = \int_{\mathbb{R}^n} \varphi(x) d\mu_X$$

# Apéndice C

## Independencia

En este apéndice utilizaremos las herramientas de la teoría de la medida para probar algunas propiedades de las variables aleatorias independientes.

### C.1. El teorema $\pi - \lambda$ de Dynkin

Para la prueba de algunos teoremas de la teoría de probabilidades (y de la teoría de la medida) se necesita un resultado técnico conocido como el teorema  $\pi - \lambda$  de Dynkin. Para enunciarlo, necesitamos algunas definiciones previas:

**Definición C.1.1** Sea  $\Omega$  un conjunto. Una clase  $\mathcal{P}$  de subconjuntos de  $\Omega$  se llamará un  $\pi$ -sistema si es cerrado bajo intersecciones finitas, o sea si  $A, B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$ .

**Definición C.1.2** Una clase  $\mathcal{L}$  de subconjuntos de  $\Omega$  se llama un  $\lambda$ -sistema si verifica las siguientes propiedades:

$$\lambda_1) \Omega \in \mathcal{L}$$

$$\lambda_2) A \in \mathcal{L} \Rightarrow A^c = \Omega - A \in \mathcal{L}$$

$$\lambda_3) \text{ Si } (A_n) \text{ es una familia numerable } \underline{\text{disjunta}} \text{ y } A_n \in \mathcal{L}, \text{ entonces } \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}$$

**Obs:** Debido a la condición de que los conjuntos sean disjuntos en la condición  $\lambda_3$ ), la definición  $\lambda$ -sistema es mucho más débil que la de  $\sigma$ -álgebra. Toda  $\sigma$ -álgebra es un  $\lambda$ -sistema pero la recíproca no es válida.

### Algunas propiedades de los $\lambda$ -sistemas

- $\emptyset \in \mathcal{L}$
- Si  $A \subset B$ , y  $A, B \in \mathcal{L} \rightarrow B - A \in \mathcal{L}$ .  
**Prueba:**  $B - A = B \cap A^c = (B^c \cup A)^c$  y  $B^c \cap A = \emptyset$ .
- $\mathcal{L}$  es cerrado por uniones numerables crecientes. Si  $A_n \in \mathcal{L} \forall n \in \mathbb{N}$ , y  $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$ , entonces  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}$ .  
**Prueba:**

$$\bigcup_{n \in \mathbb{N}} A_n = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \dots \cup (A_n - A_{n-1}) \cup \dots$$

- Si  $\mathcal{L}$  es a la vez un  $\lambda$ -sistema y un  $\pi$ -sistema, entonces  $\mathcal{L}$  es una  $\sigma$ -álgebra.

**Notación:** Si  $\mathcal{P}$  es una familia de partes de  $\Omega$ , notamos por  $\sigma(\mathcal{P})$  la  $\sigma$ -álgebra generada por  $\mathcal{P}$ .

**Teorema C.1.1 (Teorema  $\pi - \lambda$  de Dynkin)** Si  $\mathcal{P}$  es un  $\pi$ -sistema,  $\mathcal{L}$  es un  $\lambda$ -sistema, y  $\mathcal{P} \subset \mathcal{L}$  entonces  $\sigma(\mathcal{P}) \subset \mathcal{L}$ .

**Prueba:** Sea  $\mathcal{L}_0$  el  $\lambda$ -sistema generado por  $\mathcal{P}$ , esto es la intersección de todos los  $\lambda$ -sistemas que contienen a  $\mathcal{P}$  (que es a su vez un  $\lambda$ -sistema). Notamos que en particular  $\mathcal{L}_0 \subset \mathcal{L}$ . Afirmamos que  $\mathcal{L}_0$  es un  $\pi$ -sistema. Para probar que  $\mathcal{L}_0$  es un  $\pi$ -sistema, procedemos del siguiente modo: dado  $A \in \mathcal{L}$ , definimos

$$\mathcal{L}_A = \{B \subset \Omega : A \cap B \in \mathcal{L}_0\}$$

**Afirmación 1:** Si  $A \in \mathcal{L}_0$ , entonces  $\mathcal{L}_A$  es un  $\lambda$ -sistema.

- $A \cap \Omega = A \in \mathcal{L}_0$  por hipótesis, luego  $\Omega \in \mathcal{L}_A$ .
- Si  $B_1, B_2 \in \mathcal{L}_A$  y  $B_1 \subset B_2$ , entonces por definición  $A \cap B_1, A \cap B_2 \in \mathcal{L}_0$ . Ahora como  $\mathcal{L}_0$  es un  $\lambda$ -sistema y  $A \cap B_1 \subset A \cap B_2$ , tenemos que  $A \cap B_1 - A \cap B_2 = A \cap (B_1 - B_2) \in \mathcal{L}_0$ . En consecuencia,  $B_1 - B_2 \in \mathcal{L}_A$ .
- Si  $(B_n)$  es una familia disjunta de conjuntos de  $\mathcal{L}_A$  entonces  $A \cap B_n$  es una familia disjunta de conjuntos de  $\mathcal{L}_0$ , y como

$$A \cap \left( \bigcup_{n \in \mathbb{N}} B_n \right) = \bigcap_{n \in \mathbb{N}} (A \cap B_n) \in \mathcal{L}_0$$

entonces

$$\bigcup_{n \in \mathbb{N}} B_n \in \mathcal{L}_A$$

**Afirmación 2:** Si  $A \in \mathcal{P}$ , entonces  $\mathcal{L}_0 \subset \mathcal{L}_A$ .

Si  $A \in \mathcal{P}$ , entonces para cualquier  $B \in \mathcal{P}$  tenemos que  $A \cap B \in \mathcal{P}$ , ya que  $\mathcal{P}$  es por hipótesis un  $\pi$ -sistema. Deducimos que  $\mathcal{P} \subset \mathcal{L}_A$ . Luego por la afirmación 1,  $\mathcal{L}_A$  es un  $\lambda$ -sistema que contiene a  $\mathcal{P}$ , lo cual por la definición de  $\mathcal{L}_0$  implica que  $\mathcal{L}_0 \subset \mathcal{L}_A$ .

**Afirmación 3:** Si  $C \in \mathcal{L}_0$ , entonces  $\mathcal{L}_0 \subset \mathcal{L}_C$ .

Para todo  $A \in \mathcal{P}$ , por la afirmación 2, tenemos que  $\mathcal{L}_A \subset \mathcal{L}_0$ . Luego si  $C \in \mathcal{L}_0$ , entonces  $C \in \mathcal{L}_A$ , que por simetría de la definición implica que  $A \in \mathcal{L}_C$ . Como esto vale para todo  $A \in \mathcal{P}$ , deducimos que  $\mathcal{P} \subset \mathcal{L}_C$ .

Por la afirmación 1, deducimos que  $\mathcal{L}_C$  es un  $\lambda$ -sistema que contiene a  $\mathcal{P}$ , lo que por la definición de  $\mathcal{L}_0$ , implica que  $\mathcal{L}_0 \subset \mathcal{L}_C$ .

Finalmente sean  $D, E \in \mathcal{L}_0$ . Entonces por la afirmación 3,  $D \in \mathcal{L}_0 \subset \mathcal{L}_E$ . En consecuencia por definición de  $\mathcal{L}_E$ ,  $D \cap E \in \mathcal{L}_0$ . Concluimos que  $\mathcal{L}_0$  es un  $\pi$ -sistema.

**Conclusión de la prueba:** Como  $\mathcal{L}_0$  es a la vez un  $\pi$ -sistema, y un  $\lambda$ -sistema, es una  $\sigma$ -álgebra. Como contiene a  $\mathcal{P}$ , deducimos que  $\sigma(\mathcal{P}) \subset \mathcal{L}_0$ . Y entonces, como  $\mathcal{L}_0 \subset \mathcal{L}$ , concluimos que  $\sigma(\mathcal{P}) \subset \mathcal{L}$ .  $\square$

## C.2. Variables independientes

Si  $X$  e  $Y$  son dos variables aleatorias, recordamos que  $X$  e  $Y$  se dicen independientes si para cualquier par de intervalos  $(a, b]$  y  $(c, d]$  de la recta, los eventos  $\{X \in (a, b]\}$  y  $\{Y \in (c, d]\}$  son independientes, es decir que:

$$P\{(X, Y) \in (a, b] \times (c, d]\} = P\{X \in (a, b]\} \times P\{Y \in (c, d]\}$$

Podemos interpretar esta fórmula como:

$$\mu_{(X,Y)}((a, b] \times (c, d]) = \mu_X((a, b])\mu_Y((c, d])$$

El siguiente lema afirma que una fórmula análoga es válida si sustituimos los intervalos por conjuntos borelianos de la recta:

**Lema C.2.1** Sean  $X$  e  $Y$  dos variables aleatorias. Entonces  $X$  e  $Y$  son independientes si y sólo si:

$$P\{(X, Y) \in B_1 \times B_2\} = P\{X \in B_1\} \cdot P\{Y \in B_2\}$$

para cualquier par  $B_1, B_2$  de conjuntos borelianos de la recta.

**Prueba:** Fijemos primero  $B_1$ , como siendo un intervalo  $(a, b]$  de la recta, y consideremos la familia

$$\mathcal{L}_1 = \{B \subset \mathbb{R} : P\{(X, Y) \in (a, b] \times B\} = P\{X \in (a, b]\} \cdot P\{Y \in B\}\}$$

Afirmamos que  $\mathcal{A}_1$  es un  $\lambda$ -sistema de subconjuntos de  $\mathbb{R}$ . Chequeamos las tres condiciones de la definición:

$\lambda_1)$   $\mathbb{R} \in \mathcal{L}_1$ :

$$P\{(X, Y) \in (a, b] \times \mathbb{R}\} = P\{X \in (a, b]\} = P\{X \in (a, b]\} \cdot P\{Y \in \mathbb{R}\}$$

ya que  $P\{Y \in \mathbb{R}\} = 1$ .

$\lambda_2)$   $B \in \mathcal{L}_1 \Rightarrow B^c = \mathbb{R} - B \in \mathcal{L}_1$

En efecto,

$$\begin{aligned} P\{(X, Y) \in (a, b] \times B^c\} &= P\{(X, Y) \in (a, b] \times \mathbb{R}\} - P\{(X, Y) \in (a, b] \times B\} \\ &= P\{X \in (a, b]\} - P\{X \in (a, b]\}P\{Y \in B\} \\ &= P\{X \in (a, b]\}(1 - P\{Y \in B\}) \\ &= P\{X \in (a, b]\}P\{Y \in B^c\} \end{aligned}$$

$\lambda_3)$  Si  $(B_n)$  es una familia numerable disjunta y  $B_n \in \mathcal{L}_1$ , entonces  $B = \bigcup_{n \in \mathbb{N}} B_n \in \mathcal{L}_1$

En efecto, utilizando que los  $B_n$  son disjuntos, tenemos que:

$$\begin{aligned} P\{(X, Y) \in (a, b] \times B\} &= P\{(X, Y) \in \bigcup_{n \in \mathbb{N}} ((a, b] \times B_n)\} \\ &= \sum_{n \in \mathbb{N}} P\{(X, Y) \in (a, b] \times B_n\} \\ &= \sum_{n \in \mathbb{N}} P\{X \in (a, b]\}P\{Y \in B_n\} \\ &= P\{X \in (a, b]\} \left( \sum_{n \in \mathbb{N}} P\{Y \in B_n\} \right) \\ &= P\{X \in (a, b]\}P\{Y \in B\} \end{aligned}$$

Notemos que no es posible probar que  $\mathcal{L}_1$  sea una  $\sigma$ -álgebra, pues este argumento no funciona si los  $B_n$  no fueran disjuntos.



Por otra parte la familia  $\mathcal{P}$  de los intervalos semiabiertos de la recta (contando como intervalo semiabierto al conjunto vacío  $(a, a] = \emptyset$  es un  $\pi$ -sistema, y por la definición de variables aleatorias independientes,  $\mathcal{P} \subset \mathcal{L}_1$ .

El teorema  $\pi - \lambda$  nos permite concluir entonces que  $\sigma(\mathcal{P}) \subset \mathcal{L}_1$ , es decir: que la  $\sigma$ -álgebra  $\mathcal{B}(\mathbb{R})$  de los borelianos de la recta, está contenida en  $\mathcal{L}_1$ . Entonces, hemos probado que la fórmula del enunciado, se verifica cuando  $B_1$  es un intervalo semiabierto y  $B_2$  un boreliano arbitrario.

Ahora, repetimos el argumento, fijando la otra variable. Para ello consideramos la familia:

$$\mathcal{L}_2 = \{B \subset \mathbb{R} : P\{(X, Y) \in B \times B_2\} = P\{X \in B\} \cdot P\{Y \in B_2\} : \forall B \in \mathcal{B}(\mathbb{R})\}$$

Repitiendo el argumento anterior, podemos probar que  $\mathcal{L}_2$  es un  $\lambda$ -sistema, y por lo anteriormente probado,  $\mathcal{L}_2$  contiene a la clase  $\mathcal{P}$  de los intervalos semiabiertos. Nuevamente, por el teorema  $\pi - \lambda$ ,  $\mathcal{L}_2$  contiene a los borelianos. Pero esto significa precisamente, que la fórmula del enunciado es válida para  $B_1, B_2$  borelianos arbitrarios de la recta.

□

**Corolario C.2.1** Sean  $X, Y$  variables aleatorias independientes, y sean  $\varphi_1, \varphi_2 : \mathbb{R} \rightarrow \mathbb{R}$  funciones medibles Borel. Entonces:  $\varphi_1(X)$  y  $\varphi_2(Y)$  son variables aleatorias independientes.

Estos resultados se generalizan a varias variables independientes.

### C.3. Esperanza del producto de variables independientes

A modo de ilustración de la utilidad de los teoremas de paso al límite en la integral, demostraremos la siguiente propiedad:

**Teorema C.3.1** Si  $X$  e  $Y$  son variables aleatorias independientes con esperanza finita (esto es, integrables) entonces

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

**Prueba:** Hacemos uso una vez más del método de aproximación por funciones simples. Supongamos pues primero que  $X$  e  $Y$  son no negativas, y sean  $(X_n)$  e  $(Y_n)$  variables aleatorias simples (discretas) tales que  $X_n$  converja a  $X$  en forma

creciente, e  $Y_n$  converja en forma creciente a  $Y$ , dadas por la construcción del lema B.1.5. Notamos que como consecuencia de dicha construcción, si  $X$  e  $Y$  son independientes,  $X_n$  e  $Y_n$  resultan independientes. En consecuencia, como ya probamos que el resultado es cierto para variables discretas (proposición 2.1.6), tenemos que

$$E[X_n \cdot Y_n] = E[X_n] \cdot E[Y_n]$$

Ahora, en virtud del teorema de convergencia monótona,

$$E[X_n] \rightarrow E[X]$$

$$E[Y_n] \rightarrow E[Y]$$

$$E[X_n \cdot Y_n] \rightarrow E[X \cdot Y]$$

Luego,

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Esto establece el resultado para funciones no negativas. En el caso general, hacemos uso, una vez más de la descomposición:

$$X = X^+ - X^-$$

$$Y = Y^+ - Y^-$$

Entonces,

$$\begin{aligned} E[X \cdot Y] &= E[(X^+ - X^-)(Y^+ - Y^-)] = E[X^+Y^+ - X^-Y^+ - X^+Y^- + X^-Y^-] = \\ &= E[X^+Y^+] - E[X^-Y^+] - E[X^+Y^-] + E[X^-Y^-] \end{aligned}$$

Pero como  $X$  e  $Y$  son independientes  $X^+$ ,  $X^-$  son independientes de  $Y^+$ ,  $Y^-$  respectivamente; en consecuencia:

$$\begin{aligned} E[X \cdot Y] &= E[X^+]E[Y^+] - E[X^-]E[Y^+] - E[X^+]E[Y^-] + E[X^-]E[Y^-] = \\ &= (E[X^+] - E[X^-])(E[Y^+] - E[Y^-]) = E[X]E[Y] \end{aligned}$$

□

La prueba de este teorema ilustra como los teoremas de paso al límite resultan útiles para generalizar las propiedades que conocemos para variables discretas, al caso de variables aleatorias continuas.

# Apéndice D

## Existencia de las Integrales de Riemann-Stieltjes

En esta apéndice, presentaremos una prueba del siguiente resultado fundamental de la teoría de la integral de Riemann-Stieltjes:

**Teorema D.0.2** *Si  $F$  es una función creciente en un intervalo cerrado  $[a, b]$  de la recta, y  $\varphi$  es una función continua en  $[a, b]$ , entonces la integral de Riemann-Stieltjes*

$$\int_a^b \varphi(x) dF(x)$$

*existe*

Recordamos que esta integral, se define como el límite conforme la norma  $|\pi|$  de la partición tiende a cero, de las sumas:

$$S_\pi(\varphi.F) = \sum_{i=0}^{n-1} \varphi(\xi_i)(F(x_{i+1}) - F(x_i))$$

donde  $\pi : a = x_0 < x_1 < \dots x_n = b$  es una partición de  $[a, b]$  y  $\xi_i \in [x_i, x_{i+1}]$  es un punto intermedio.

Estas sumas son poco manejables para nuestros propósitos pues dependen de los puntos intermedios  $\xi_i$  variables. Por ello, las reemplazamos por sumas superiores e inferiores que són de más fácil manejo:

Para cada  $i$  ( $0 \leq i \leq n - 1$ ), notamos:

$$m_i = \inf_{x \in [x_i, x_{i+1}]} \varphi(x)$$

$$M_i = \sup_{x \in [x_i, x_{i+1}]} \varphi(x)$$

y consideramos las sumas superiores  $U_\pi$  y las sumas inferiores  $L_\pi$  definidas por:

$$L_\pi(\varphi, F) = \sum_{i=0}^{n-1} M_i(F(x_{i+1}) - F(x_i))$$

$$U_\pi(\varphi, F) = \sum_{i=0}^{n-1} M_i(F(x_{i+1}) - F(x_i))$$

Es claro entonces que:

$$L_\pi(\varphi, F) \leq S_\pi(\varphi, F) \leq U_\pi(\varphi, F)$$

Las sumas superiores e inferiores, tienen la siguiente propiedad importante (de monotonía): Si  $\pi'$  es un refinamiento de  $\pi$ , entonces

$$L_{\pi'}(\varphi, F) \geq L_\pi(\varphi, F)$$

$$U_{\pi'}(\varphi, F) \leq U_\pi(\varphi, F)$$

(Las sumas superiores decrecen al afinar la partición, mientras que las inferiores crecen.)

Para demostrarla, es fácil observar que se verifica si  $\pi'$  es una partición obtenida de  $\pi$  agregando un punto. Por inducción, se obtiene el caso general, ya que si  $\pi'$  es un refinamiento de  $\pi$ , ello significa que se obtiene de  $\pi$  agregando finitos puntos.

De esta observación, se deduce lo siguiente: toda suma superior es mayor que cualquier suma inferior. Es decir que si  $\pi$  y  $\pi'$  son dos particiones arbitrarias, siempre se verifica que:

$$L_\pi(\varphi, F) \leq U_{\pi'}(\varphi, F)$$

Para demostrar esta afirmación, es suficiente notar que la partición  $\pi'' = \pi \cup \pi'$  es un refinamiento común <sup>1</sup>

Entonces, utilizando la propiedad de monotonía,

$$L_\pi(\varphi, F) \leq L_{\pi''} \leq U_{\pi''} \leq U_{\pi'}$$

---

<sup>1</sup>Es esta propiedad de las particiones, de que dos particiones siempre tienen un refinamiento común, hace de las particiones un *conjunto dirigido*. Así pues,  $S_\pi(\varphi, F)$  es una *red* que converge a la integral de Stieltjes.

**Lema D.0.1** Dado  $\varepsilon > 0$ , existe  $\delta > 0$  tal que si  $|\pi| < \delta$ , tenemos que

$$0 \leq U_\pi(\varphi, F) - L_\pi(\varphi, F) < \varepsilon$$

**Prueba:** Dado  $\varepsilon > 0$ , como  $\varphi$  es uniformemente continua en  $[a, b]$ , existirá un  $\delta > 0$  tal que si  $|x - y| < \delta$  con  $x, y \in [a, b]$ , se tiene que  $|\varphi(x) - \varphi(y)| < \varepsilon$ . Entonces, si  $\pi$  es cualquier partición de  $[a, b]$  tal que  $|\pi| < \delta$ , tendremos que:

$$\begin{aligned} U_\pi(\varphi, F) - L_\pi(\varphi, F) &= \sum_{i=0}^{n-1} (M_i - m_i)(F(x_{i+1}) - F(x_i)) \\ &\leq \sum_{i=0}^{n-1} \varepsilon(F(x_{i+1}) - F(x_i)) \leq \varepsilon(F(b) - F(a)) \end{aligned}$$

□

Hechas estas observaiones, estamos en condiciones de demostrar el teorema, para ello comencemos eligiendo una sucesión  $(\pi_n)$  de particiones de  $[a, b]$  de modo que  $\pi_{n+1}$  sea un refinamiento de  $\pi_n$ , y que  $|\pi_n| \rightarrow 0$ . Por ejemplo, podemos elegir como  $\pi_n$  la partición uniforme de  $[a, b]$  en  $2^n$  partes de igual longitud.

Entonces, por la propiedad de monotonía la sucesión de sumas inferiores  $L_{\pi_n}(\varphi, F)$  será monótona creciente, y además está acotada pues

$$L_{\pi_n} \leq \left( \sup_{x \in [a, b]} \varphi(x) \right) (F(b) - F(a))$$

En consecuencia, existe el límite

$$I = \lim_{n \rightarrow +\infty} L_{\pi_n}(\varphi, F)$$

En virtud del lema, también tendremos que:

$$I = \lim_{n \rightarrow +\infty} U_{\pi_n}(\varphi, F)$$

Dado  $\varepsilon > 0$ , sea  $\delta > 0$  el que corresponde a  $\varepsilon$  de acuerdo al lema, y elijamos  $n$  tal que  $|\pi_n| < \delta$ , y

$$|L_{\pi_n} - I| < \varepsilon$$

$$|U_{\pi_n} - I| < \varepsilon$$

Afirmamos entonces que:

$$|S_\pi(\varphi, F) - I| < 2\varepsilon$$

En efecto,

$$\begin{aligned} S_\pi(\varphi, F) - I &\leq U_\pi(\varphi, F) - U_{\pi_n} + U_{\pi_n} - I \\ &\leq U_\pi(\varphi, F) - L_\pi(\varphi, F) + \varepsilon < 2\varepsilon \end{aligned}$$

Similarmente,

$$\begin{aligned} S_\pi(\varphi, F) - I &\geq L_\pi(\varphi, F) - L_{\pi_n} + L_{\pi_n} - I \\ &\geq L_\pi(\varphi, F) - U_\pi(\varphi, F) - \varepsilon > -2\varepsilon \end{aligned}$$

En consecuencia,

$$\lim_{|\delta| \rightarrow 0} S_\pi(\varphi, F) = I$$

Una observación adicional nos será útil para demostrar el teorema de Helly sobre paso al límite en la integral de Stieltjes: este  $\delta$  sólo depende de la continuidad uniforme de  $\varphi$  y de la magnitud de la variación  $F(b) - F(a)$  de  $F$  en  $[a, b]$  (La partición  $\pi_n$  sólo juega un rol auxiliar en el argumento, pero  $\delta$  es independiente de  $n$  y por lo tanto de  $F$  mientras  $F(b) - F(a)$  permanezca acotado). Esto nos proporciona el siguiente corolario (sobre convergencia uniforme de la integral de Stieltjes respecto de la función  $F$ ):

**Corolario D.0.1** *Sea  $\varphi \in C[a, b]$ . Dados  $\varepsilon > 0$  y  $C > 0$ , existe un  $\delta > 0$  (que depende de  $\varepsilon > 0$  y  $C$  pero es independiente de  $F$ ) tal que si  $F$  es cualquier función  $F : [a, b] \rightarrow \mathbb{R}$  creciente tal que*

$$F(b) - F(a) \leq C$$

*entonces*

$$\left| \int_a^b \varphi(x) dF(x) - S_\pi(\varphi, F) \right| < \varepsilon$$

# Apéndice E

## Las Leyes Fuertes de Kolmogorov

En este apéndice expondremos la demostración de la ley fuerte de los grandes números de Kolmogorov.

### E.1. La Desigualdad de Kolmogorov

La desigualdad de Kolmogorov es una generalización de la desigualdad de Tchebyshev:

**Proposición E.1.1 (Desigualdad de Kolmogorov)** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias independientes tales que  $E[X_k] = 0$  y  $Var(X_k) < +\infty$  para  $k = 1, 2, \dots, n$ . Entonces para todo  $\lambda > 0$ ,

$$P\left\{\max_{1 \leq k \leq n} |S_k| \geq \lambda\right\} \leq \frac{1}{\lambda^2} Var(S_n) = \frac{1}{\lambda^2} \sum_{k=1}^n Var(X_k)$$

donde  $S_k = X_1 + X_2 + \dots + X_n$ .

**Prueba:** Consideremos el evento:

$$A = \left\{\max_{1 \leq k \leq n} S_k^2 \geq \lambda^2\right\}$$

Queremos obtener una cota para  $P(A)$ . Para ello lo descomponemos en eventos disjuntos, de acuerdo a cual es la primera vez que  $S_k^2 \geq \lambda^2$ :

$$A_1 = \{S_1^2 \geq \lambda^2\}$$

$$A_2 = \{S_1^2 < \lambda, S_2^2 \geq \lambda^2\}$$

y en general:

$$A_k = \{S_1^2 < \lambda^2, S_2^2 < \lambda^2, \dots, S_{k-1}^2 < \lambda^2, S_k^2 \geq \lambda^2\}$$

Entonces los  $A_k$  son disjuntos dos a dos, y

$$A = \bigcup_{k \in N} A_k$$

Luego,

$$I_A = \sum_{k=1}^n I_{A_k}$$

$$S_n^2 \geq S_n^2 I_A = \sum_{k=1}^n S_n^2 I_{A_k}$$

y tomando esperanza:

$$E[S_n^2] \geq \sum_{k=1}^n E[S_n^2 I_{A_k}] \quad (\text{E.1})$$

Nos gustaria sustituir  $S_n$  por  $S_k$  en esta sumatoria. Para ello, notamos que:

$$S_n^2 = (S_n - S_k + S_k)^2 = (S_n - S_k)^2 + 2S_k(S_n - S_k) + S_k^2 \geq 2S_k(S_n - S_k) + S_k^2$$

Multiplicando por  $I_{A_k}$  y tomando esperanza tenemos que:

$$E[S_n^2 I_{A_k}] \geq E[S_k^2 I_{A_k}] + 2E[S_k(S_n - S_k) I_{A_k}]$$

Observamos ahora que  $S_k I_{A_k}$  y  $S_n - S_k$  son independientes (pues  $S_k I_{A_k}$  depende de  $X_1, X_2, \dots, X_k$  y  $S_n - S_k$  depende de  $X_{k+1}, X_{k+2}, \dots, X_n$ ). En consecuencia:

$$E[S_k(S_n - S_k) I_{A_k}] = E[S_k I_{A_k}] E[S_n - S_k] = 0$$

pues  $E[S_n] = E[S_k] = 0$ . En consecuencia:

$$E[S_n^2 I_{A_k}] \geq E[S_k^2 I_{A_k}]$$

Ahora en  $A_k$ ,  $S_k^2 \geq \lambda^2$ . En consecuencia,

$$E[S_n^2 I_{A_k}] \geq E[\lambda^2 I_{A_k}] = \lambda^2 P(A_k)$$



Sustituyendo este resultado en la desigualdad (E.1), tenemos que:

$$E[S_n^2] \geq \lambda^2 \sum_{k=1}^n P(A_k) = \lambda^2 P(A)$$

Luego

$$P(A) \leq \frac{1}{\lambda^2} E[S_n^2] = \frac{1}{\lambda^2} \sum_{k=1}^n \text{Var}(X_k)$$

□

## E.2. La Ley Fuerte de los Grandes Números

### E.2.1. La Primera Ley Fuerte de Kolmogorov

**Teorema E.2.1 (Primera ley fuerte de Kolmogorov)** Sea  $(X_n)_{n \in \mathbb{N}}$  una sucesión de variables aleatorias independientes con esperanza finita, y supongamos que:

$$\sum_{n=1}^{\infty} \frac{\text{Var}(X_n)}{n^2} < +\infty \quad (\text{E.2})$$

Entonces  $(X_n)_{n \in \mathbb{N}}$  verifica la ley fuerte de los grandes números, es decir:

$$\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{E(X_1) + E(X_2) + \dots + E(X_n)}{n} \xrightarrow{\text{c.s.}} 0$$

**Prueba:** Podemos suponer sin pérdida de generalidad que  $E[X_n] = 0 \ \forall \ n \in \mathbb{N}$  (Sino cambiamos  $X_n$  por  $X_n - E[X_n]$ ). Queremos probar que:

$$\frac{S_n}{n} \xrightarrow{\text{c.s.}} 0$$

donde  $S_n = X_1 + X_2 + \dots + X_n$ . Definamos las “variables maximales diádicas”:

$$M_n = \max_{2^n < k \leq 2^{n+1}} \frac{|S_k|}{k}$$

Basta probar que  $M_n \rightarrow 0$  casi seguramente.

Vamos a probar esto en dos etapas:

**Etapla 1:** Probaremos que

$$\sum_{n=1} P \left\{ M_n > \frac{1}{m} \right\} < +\infty$$

para  $m = 1, 2, \dots$ , utilizando la desigualdad de Kolmogorov.

**Etapla 2:** Probaremos que  $M_n \rightarrow 0$  casi seguramente, utilizando el lema de Borel-Cantelli.

**Etapla 1:** Para probar la primera afirmación notamos que:

$$P \left\{ \max_{2^n < k \leq 2^{n+1}} \frac{|S_k|}{k} > \frac{1}{m} \right\} \leq P \left\{ \max_{2^n < k \leq 2^{n+1}} |S_k| > \frac{2^n}{m} \right\}$$

(ya que dividir por  $2^n$  en lugar de  $k$  agranda el máximo)

$$\leq P \left\{ \max_{1 \leq k \leq 2^{n+1}} |S_k| > \frac{2^n}{m} \right\} \leq \left( \frac{m}{2^n} \right)^2 \sum_{k=1}^{2^{n+1}} \text{Var}(X_k)$$

Definamos el evento  $A_{m,n} = \{M_n \geq \frac{1}{m}\}$ . Entonces

$$\sum_{n=1}^{\infty} P(A_{m,n}) \leq \sum_{n=1}^{\infty} \left( \frac{m^2}{4^n} \sum_{k=1}^{2^{n+1}} \text{Var}(X_k) \right)$$

Cambiando el orden de la suma deducimos que:

$$\begin{aligned} \sum_{n=1}^{\infty} P(A_{m,n}) &\leq m^2 \sum_{k=1}^{\infty} \left( \sum_{n: 2^{n+1} \geq k} \frac{\text{Var}(X_k)}{4^n} \right) \\ &= m^2 \sum_{k=1}^{\infty} \text{Var}(X_k) \left( \sum_{n: 2^{n+1} \geq k} \frac{1}{4^n} \right) \end{aligned}$$

Ahora bien, sumando la serie geométrica:

$$\sum_{n=j}^{\infty} \frac{1}{4^n} = \frac{4}{3} \frac{1}{4^j}$$

En consecuencia:

$$\sum_{n: 2^{n+1} \geq k} \frac{1}{4^n} = \sum_{n=j(k)}^{\infty} \frac{1}{4^n}$$

donde  $j(k)$  cumple:

$$2^{j(k)} < k \leq 2^{j(k)+1}$$

En consecuencia:

$$\sum_{n: 2^{n+1} \geq k} \frac{1}{4^n} = \frac{4}{3} \frac{1}{4^{j(k)}} \leq \frac{4}{3} \frac{4}{k^2} = \frac{16}{3k^2}$$

(pues  $2^{j(k)} \geq \frac{k}{2}$ ).

Por lo que sustituyendo, concluimos que:

$$\sum_{n=1}^{\infty} P(A_{m,n}) \leq \frac{16m^2}{3} \sum_{k=1}^{\infty} \frac{\text{Var}(X_k)}{k^2} < +\infty$$

por la hipótesis.

**Etapla 2:** Por el lema de Borel-Cantelli, concluimos que, fijado  $m$  con probabilidad 1, sólo ocurren finitos de los eventos  $A_{n,m}$ . Vale decir que si

$$A_{m,\infty} = \left\{ \omega \in \Omega : M_n(\omega) \geq \frac{1}{m} \text{ para infinitos } n \right\} = \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A_{m,n}$$

entonces  $P(A_{m,\infty}) = 0$ . Y entonces si consideramos el evento:

$$A = \{ \omega \in \Omega : M_n(\omega) \not\rightarrow 0 \} = \bigcup_{m \in \mathbb{N}} \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A_{m,n}$$

por la  $\sigma$ -aditividad, tenemos que:  $P(A) = 0$ . Concluimos que  $M_n \rightarrow 0$  con probabilidad 1.  $\square$

**Corolario E.2.1** *La ley fuerte de los grandes números,*

$$\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{E(X_1) + E(X_2) + \dots + E(X_n)}{n} \xrightarrow{c.s.} 0$$

es válida para toda sucesión  $(X_n)_{n \rightarrow \infty}$  de variables aleatorias independientes y uniformemente acotadas.

**Prueba:** Supongamos que  $|X_n| \leq c$ . Entonces  $\text{Var}(X_n) \leq E[X_n^2] \leq c^2$ , y entonces la hipótesis (E.2) es satisfecha.  $\square$

### E.2.2. Algunos Lemas Preparatorios

Nuestro siguiente objetivo será probar que la ley fuerte de los grandes números es válida sin la restricción de acotación uniforme. Para ello necesitaremos algunos lemas preparatorios:

**Lema E.2.1 (Criterio de Integrabilidad)** Sea  $X : \Omega \rightarrow \overline{\mathbb{R}}$  una variable aleatoria. Entonces  $E[|X|] < +\infty$  (“ $X$  es integrable”) si y sólo si

$$\sum_{n=1}^{\infty} P\{|X| > n\} < +\infty$$

**Prueba:** Pongamos

$$A_0 = \{\omega \in \Omega : X(\omega) = 0\}$$

$$A_n = \{\omega \in \Omega : n-1 < |X| \leq n\}$$

$$A_{\infty} = \{\omega \in \Omega : X(\omega) = \pm\infty\}$$

Los eventos  $A_n$  (con  $n \in \mathbb{N} \cup \{\infty\}$ ) forman una partición del espacio  $\Omega$ . Notemos así mismo que bajo cualquiera de las dos condiciones del enunciado  $X$  es finita con probabilidad 1, es decir  $A_{\infty}$  tiene probabilidad cero. En consecuencia, por la  $\sigma$ -aditividad de la integral (de Lebesgue) respecto del conjunto <sup>1</sup>:

$$E[|X|] = \sum_{n=0}^{\infty} \int_{A_n} |X| dP$$

y por lo tanto:

$$\sum_{n=1}^{\infty} \int_{A_n} (n-1) dP \leq E[|X|] \leq \sum_{n=1}^{\infty} \int_{A_n} n dP$$

(Notamos que el término correspondiente a  $n = 0$  se anula). Es decir que:

$$\sum_{n=1}^{\infty} (n-1)P(A_n) \leq E[|X|] \leq \sum_{n=1}^{\infty} nP(A_n)$$

o sea, teniendo en cuenta que los  $A_n$  forman una partición (y que por lo tanto sus probabilidades suman 1):

$$\left( \sum_{n=1}^{\infty} nP(A_n) \right) - 1 \leq E[|X|] \leq \sum_{n=1}^{\infty} nP(A_n)$$

---

<sup>1</sup>Aquí presentamos una prueba usando la integral Lebesgue. Son posibles pruebas alternativas, por ej. usando la integral de Stieltjes. Ver Barry James

Deducimos pues que:

$$E[|X|] < +\infty \Leftrightarrow \sum_{n=1}^{\infty} nP(A_n) < +\infty$$

Para escribir esto de otra forma (y obtener la conclusión del enunciado), introduzcamos los eventos:

$$B_n = \{\omega \in \Omega : |X(\omega)| > n\}$$

Entonces  $A_n = B_{n-1} - B_n$  y como  $B_n \subset B_{n-1}$  deducimos que:

$$P(A_n) = P(B_{n-1}) - P(B_n)$$

En consecuencia,

$$E[|X|] < +\infty \Leftrightarrow \sum_{n=1}^{\infty} n \{P(B_{n-1}) - P(B_n)\} < +\infty \quad (\text{E.3})$$

Ahora notamos que “sumando por partes”:

$$\begin{aligned} \sum_{n=1}^N n \{P(B_{n-1}) - P(B_n)\} &= 1(P(B_0) - P(B_1)) + 2(P(B_1) - P(B_2)) + \dots \\ &+ N(P(B_{N-1}) - P(B_N)) = P(B_0) + P(B_1) + P(B_2) + \dots + P(B_{N-1}) - NP(B_N) \end{aligned}$$

Es decir que:

$$\sum_{n=1}^N n \{P(B_{n-1}) - P(B_n)\} = \sum_{n=0}^{N-1} P(B_n) - NP(B_N) \quad (\text{E.4})$$

Ahora probaremos el enunciado: Si  $E[|X|]$  es finita, por la desigualdad de Markov:

$$P(B_N) \leq \frac{1}{N} E[|X|]$$

En consecuencia, de (E.4) y (E.3), deducimos que la serie de términos no negativos:

$$\sum_{n=1}^{\infty} P(B_n)$$

tiene sumas parciales acotadas, y es por lo tanto convergente. Esto prueba una de las implicaciones del enunciado. Para probar la otra, supongamos que dicha serie es convergente. Entonces, por (E.4):

$$\sum_{n=1}^N n \{P(B_{n-1}) - P(B_n)\} \leq \sum_{n=1}^{N-1} P(B_0)$$

y en consecuencia por (E.3),  $E[|X|] < +\infty$ . □

**Lema E.2.2** Sea  $X$  una variable aleatoria con esperanza finita, y pongamos para cada  $n$ ,  $A_n = \{\omega \in \Omega : -n \leq |X(\omega)| \leq n\}$ . Entonces:

$$K = \sum_{n=1}^{\infty} \frac{1}{n^2} E[X^2 I_{A_n}] < +\infty$$

**Prueba:** Necesitamos la siguiente propiedad:

$$\sum_{n=j}^{\infty} \frac{1}{n^2} \leq \frac{2}{j} \quad (\text{E.5})$$

Para establecer esta fórmula, notemos que para cada  $n \in N$ :

$$\frac{1}{n^2} \leq \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}$$

En consecuencia, sumando esta serie telescópica, obtenemos que:

$$\begin{aligned} \sum_{n=j}^{\infty} \frac{1}{n^2} &= \frac{1}{j^2} + \sum_{n=j+1}^{\infty} \frac{1}{n^2} \leq \frac{1}{j^2} + \sum_{n=j+1}^{\infty} \left( \frac{1}{n-1} - \frac{1}{n} \right) \\ &= \frac{1}{j^2} + \frac{1}{j} < \frac{2}{j} \end{aligned}$$

Volviendo a la prueba del lema, para cada  $j \in N$ , consideramos el evento:

$$B_j = \{\omega \in \Omega : j-1 < |X(\omega)| \leq j\}$$

y

$$B_0 = \{\omega \in \Omega : X(\omega) = 0\}$$

Entonces:

$$A_n = \bigcup_{j=0}^n B_j \text{ (unión disjunta)}$$

En consecuencia:

$$E[X^2 I_{A_n}] = \sum_{j=0}^n E[X^2 I_{B_j}]$$

y por lo tanto:

$$K = \sum_{n=1}^{\infty} \frac{1}{n^2} E[X^2 I_{A_n}] = \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=0}^n E[X^2 I_{B_j}]$$

Cambiando el orden de la suma (cosa que está permitida, ya que es una serie de términos no negativos):

$$K = \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} \frac{1}{n^2} E[X^2 I_{B_j}]$$

Utilizando entonces la propiedad (E.5), vemos que:

$$K \leq \sum_{j=1}^{\infty} \frac{2}{j} E[X^2 I_{B_j}]$$

Ahora bien, cuando ocurre el evento  $B_j$ ,  $X^2 \leq j|X|$ . Deducimos que,

$$K \leq 2 \sum_{j=1}^{\infty} E[|X| I_{B_j}] \leq 2E[|X|] < +\infty$$

ya que los eventos  $(B_j)$  forman una partición de  $\Omega$ . □

### E.2.3. La Segunda Ley Fuerte de Kolmogorov

**Teorema E.2.2** Sea  $(X_n)_{n \in \mathbb{N}}$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con  $E[|X_i|] < +\infty$ . Sea  $\mu = E[X_i]$  entonces

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{c.s.} \mu$$

cuando  $n \rightarrow +\infty$ .

La prueba se basa en el *método de truncamiento*. Definimos unas nuevas variables aleatorias  $Y_n$  por:

$$Y_n = \begin{cases} X_n & \text{si } |X_n| \leq n \\ 0 & \text{si } |X_n| > n \end{cases}$$

**Lema E.2.3** Supongamos que se cumplen las hipótesis del teorema E.2.2 Las variables truncadas  $Y_n$  tienen las siguientes propiedades:

i)

$$\lim_{n \rightarrow +\infty} E[Y_n] = \mu$$

ii)

$$\sum_{n=1}^{\infty} \frac{\text{Var}(Y_n)}{n^2} < +\infty$$

iii) Con probabilidad 1, dado  $\omega \in \Omega$  existe un  $n_0 = n_0(\omega)$  tal que  $X_n(\omega) = Y_n(\omega)$  para  $n \geq n_0$ .

**Prueba: i):** Como las  $X_n$  son idénticamente distribuidas:

$$E[Y_n] = E[X_n I_{\{|X_n| \leq 1\}}] = E[X_1 I_{\{|X_1| \leq 1\}}]$$

Ahora bien la secuencia de variables aleatorias:  $X_1 I_{\{|X_1| \leq 1\}}$  está acotada por  $|X_1|$ :

$$|X_1 I_{\{|X_1| \leq 1\}}| \leq |X_1|$$

que es integrable por hipótesis. En consecuencia, por el teorema de convergencia mayorada:

$$E[Y_n] \rightarrow E[X_1] = \mu$$

**ii):** Nuevamente, como las  $X_n$  son idénticamente distribuidas

$$\text{Var}(Y_n) = \text{Var}(X_1 I_{\{|X_1| \leq n\}})$$

y la conclusión se sigue del lema E.2.2 pues  $X_1$  es integrable.

**iii):** Consideramos el evento

$$A = \{\omega \in \Omega : \exists n_0 = n_0(\omega) \text{ tal que } \forall n \geq n_0 : X_n(\omega) = Y_n(\omega)\}$$

Queremos ver que  $P(A) = 1$ . Para ello consideramos los eventos,

$$A_n = \{\omega \in \Omega : X_n(\omega) \neq Y_n(\omega)\}$$

Entonces:

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P\{X_n \neq Y_n\} = \sum_{n=1}^{\infty} P\{|X_n| > n\} = \sum_{n=1}^{\infty} P\{|X_1| > n\} < +\infty$$

por el criterio de integrabilidad (lema E.2.1). En consecuencia, por el lema de Borel-Cantelli, con probabilidad 1, sólo ocurre un número finito de los sucesos  $A_n$ , es decir que  $P(A) = 1$ .  $\square$

**Corolario E.2.2** Si consideramos el evento

$$B = \left\{ \omega \in \Omega : \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n |X_k(\omega) - Y_k(\omega)| = 0 \right\}$$

tenemos que  $P(B) = 1$



En efecto, como  $A \subset B$  (donde  $A$  es el evento definido en la prueba anterior), y  $P(A) = 1$  deducimos que  $P(B) = 1$ .

Necesitaremos también un lema (ejercicio) de análisis I:

**Lema E.2.4** Sea  $(\mu_k)_{k \in \mathbb{N}}$  una sucesión de números reales tales que  $\mu_k \rightarrow \mu$  cuando  $k \rightarrow +\infty$ , y pongamos  $z_n = \frac{1}{n} \sum_{k=1}^n \nu_k$  entonces  $z_n \rightarrow \mu$  cuando  $n \rightarrow +\infty$ .

Podemos ahora concluir la prueba de la segunda ley fuerte de Kolmogorov (teorema E.2.2): consideramos el evento

$$C = \left\{ \omega \in \Omega : \frac{X_1(\omega) + X_2(\omega) + \dots + X_n(\omega)}{n} \rightarrow \mu \text{ cuando } n \rightarrow +\infty \right\}$$

Y consideramos también el evento:

$$D = \left\{ \omega \in \Omega : \frac{Y_1(\omega) + Y_2(\omega) + \dots + Y_n(\omega)}{n} - \bar{\mu} \rightarrow 0 \text{ cuando } n \rightarrow +\infty \right\}$$

siendo  $\mu_k = E(Y_k)$  y  $\bar{\mu} = -\frac{\mu_1 + \mu_2 + \dots + \mu_n}{n}$ .

En virtud del lema E.2.3, ii), vemos que las variables truncadas  $Y_n$  verifican las hipótesis de la primera ley fuerte de Kolmogorov (teorema E.2.1), en consecuencia  $P(D) = 1$ . Ahora bien, en virtud del lema E.2.4:

$$\frac{\mu_1 + \mu_2 + \dots + \mu_n}{n} \rightarrow \mu$$

y en consecuencia:  $B \cap D \subset C$ . Pero como,  $P(B) = P(D) = 1$ , deducimos que  $P(C) = 1$ .

Esto concluye la prueba de la segunda ley fuerte de Kolmogorov.

**Nota:** Una demostración alternativa del teorema (E.2.2), que no depende de la desigualdad de Kolmogorov, se da en el artículo de N. Etemadi [15].

# Bibliografía

- [1] R. Ash, *Basic Probability Theory* <http://www.math.uiuc.edu/~r-ash/BPT.html>
- [2] R. Ash, *Lectures on Statistics Lecutres on Statistics* <http://www.math.uiuc.edu/~r-ash/Stat.html>
- [3] Durrett, Richard. Probability:theory and examples. Duxbury Press, Second edition, 1996.
- [4] Feller, William. An introduction to probability theory and its applications, J. Wiley. 1978.
- [5] García Álvarez, Miguel Angel. Introducción a la Teoría de la Probabilidad (Primero y Segundo curso). Fondo de Cultura Económica, 2005.
- [6] James, Barry. Probabilidade: um curso em nível intermediário (2 ed.), IMPA. 2002.
- [7] A. Mood, F. Graybill. Introducción a la Teoría de la Estadística. Aguilar. 1969.
- [8] Renyi, Alfred . Teoría de Probabilidades. Reverté 1978.
- [9] Ross, Sheldon. A first course in Probability. 1994-1998.
- [10] Rozanov, Yu. Procesos Aleatorios. Editorial Mir, 1973.
- [11] Luis A. Santaló, *La Probabilidad y sus aplicaciones*. Ed. Iberoamericana, Buenos Aires. (1955).
- [12] Yohai, Victor. Notas del curso Probabilidades y Estadística. <http://mate.dm.uba.ar/~vyohai/Notas%20de%20Probabilidades.pdf>

**Libros avanzados sobre probabilidad: (para quienes hayan cursado análisis real y quieran profundizar en estos temas)**

- [13] Patrick Billingsley. Probability and Measure. John Wiley & Sons. (1979)

**Artículos elementales sobre probabilidad:**

- [14] Henryk Gzyl, José Luis Palacios. The Weierstrass Aproximation theorem and Large Desviations. American Mathematical Monthly, August- Sepetember (1997).
- [15] N. Etemadi, *An elementary proof of the strong law of large numbers*. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete. (Probability Theory and Related Topics). February 1981, Volume 55, Issue 1, pp 119-122.

**Otros artículos originales citados**

- [16] Carl-Gustav Esseen, *On the Liapunoff limit of error in the theory of probability*. Arkiv för matematik, astronomi och fysik A28: 1–19.(1942).
- [17] Andrew C.Berry, *The Accuracy of the Gaussian Approximation to the Sum of Independent Variates*. Transactions of the American Mathematical Society 49 (1): 122–136, (1941)
- [18] Erdős, Paul; Kac, Mark .*The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions*. American Journal of Mathematics 62 (1/4): 738–742, (1940)

**Referencias de Interés Histórico**

- [19] E. Borel, *Les probabilités dénombrables et leurs applications arithmetiques*. Rend. Circ. Mat. Palermo (2) 27 (1909) pp. 247–271.
- [20] F.P. Cantelli, *Sulla probabilità come limite della frequenza*, Atti Accad. Naz. Lincei 26:1 (1917) pp.39–45.
- [21] D. R. McDonald, *The Local Limit Theorem: A Historical Perspective*, Journal of the Royal Statistical Society, Vol. 4, No. 2, pp. 73-86.

**Libros de Análisis Real:**

- [22] R. Wheeden, A. Zygmund. Measure and Integral. Marcel Dekker Inc. 1977.
- [23] A.N. Kolmogorov, S.V. Fomín. Elementos de la teoría de funciones y del análisis funcional. Editorial MIR, Moscú, (1975).

**Otra bibliografía consultada para la elaboración de estas notas:**

- [24] R. Courant, F. John. Introducción al cálculo y al análisis matemático. Ed. Limusa (1985).

**Otros artículos sobre temas mencioandos en estas notas:**

- [25] M.W. Sierpinski. *Démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et détermination effective d'un tel nombre.* Bull. Soc. Math. France, 45 (1917), pp. 127–132.
- [26] V. Becher, S. Figueira, *An example of a computable absolutely normal number.* Theoretical Computer Science. Volume 270, Issues 1–2, 6 January (2002), pp. 947–958.
- [27] Andrew Granville, K. Soundararajan, *Sieving and the Erdős-Kac theorem.* <http://arxiv.org/abs/math/0606039>