

Curso intermedio de **PROBABILIDAD**

Luis Rincón
Departamento de Matemáticas
Facultad de Ciencias UNAM
Circuito Exterior de CU
04510 México DF

Junio 2006

El presente texto corresponde a la versión de junio de 2006.
Este material se encuentra en permanente actualización y corrección.
La última versión disponible en formato pdf puede obtenerse en
<http://www.matematicas.unam.mx/lars>

Prefacio

El presente texto está dirigido a estudiantes de mitad de carrera de las licenciaturas de matemáticas, actuaría, y áreas afines. Contiene el material básico para un segundo curso de probabilidad, y tiene como origen las notas de clase del curso semestral de Probabilidad II, que he impartido en la Facultad de Ciencias de la UNAM.

Este volumen trata temas de probabilidad a un nivel intermedio. El énfasis de este segundo curso se centra en la formalización de algunos conceptos estudiados en un primer curso de probabilidad, y en el estudio de vectores aleatorios y sus varios conceptos relacionados. El lector puede comprobar que se hace poco énfasis en las aplicaciones, y que la exposición se concentra mayormente en el desarrollo matemático. La objetivo es que después de este curso, el estudiante pueda continuar con facilidad con un curso de estadística matemática, de procesos estocásticos, o tal vez un curso avanzado de probabilidad o de teoría de la medida, teniendo como elementos básicos los conceptos teóricos aquí desarrollados. En particular se incluye un capítulo sobre esperanza condicional, cuyo uso y aplicación es cada vez más frecuente. También se incluye un capítulo sobre distribuciones muestrales y estadísticas de orden, con aplicaciones inmediatas en temas de la estadística matemática.

El texto contiene una gran cantidad de ejercicios, la mayoría de los cuales son de tipo mecánico, algunos de ellos son muy sencillos de modo que el término *ejercicios* me parece justo y adecuado. La intención es la de crear confianza y soltura por parte del alumno en el manejo de los conceptos y notación involucrados. El número de ejercicios excede lo que normalmente puede realizarse en un semestre, y el objetivo que siempre tuve en mente estos años fue el tener un número suficiente de ejercicios para presentar algunos en clase, dejar otros para trabajo en casa, y asignar algunos otros para preguntas de examen, usando material ligeramente distinto cada semestre para evitar repeticiones. Los ejercicios se encuentran regularmente al final de cada capítulo, y se han numerado de manera consecutiva a lo largo del curso.

La presentación del material mantiene la estructura de las notas de clase, y creo que será particularmente útil al estudiante con poco tiempo para leer párrafos completos, y quien sólo busca una definición, un resultado, un ejemplo, un ejercicio, o tal vez orientación breve acerca de un concepto. En este sentido, el libro contiene tablas a manera de resumen, y los enunciados están enmarcados para su fácil localización. También he intentado que la notación fuera lo más simple y mínima posible. Personalmente me gustan los libros con imágenes y diagramas, y he buscado plasmar ese gusto en este texto. Este material fue escrito en \LaTeX , y las gráficas fueron elaboradas usando el paquete *pstricks*.

Al final del texto aparece una lista de referencias que me permito sugerir al lector consultar para profundizar y a veces precisar en algunos temas. Algunos de estos textos no han sido referenciados explícitamente pero aparecen en la lista por que en algún momento he obtenido inspiración de ellos.

Agradezco sinceramente a todas aquellas personas, alumnos y profesores, quienes a

través de sus comentarios y sugerencias, han contribuido al mejoramiento de este texto. Cualquier corrección o comentario acerca de este trabajo será muy bien recibido en el correo electrónico que aparece abajo. Por último, me parece importante mencionar que este texto ha sido posible, en gran medida, al excelente ambiente de trabajo y de libertad académica que he tenido la fortuna de encontrar en el Departamento de Matemáticas de la Facultad de Ciencias de la UNAM. Gracias a todos por su confianza y apoyo.

Luis Rincón
Junio 2006
Ciudad Universitaria UNAM
`lars@ciencias.unam.mx`

Contenido

| | |
|--|------------|
| 1. Espacios de probabilidad | 5 |
| 1.1. Espacios de probabilidad | 5 |
| 1.2. σ -álgebras | 6 |
| 1.3. Medidas de probabilidad | 18 |
| 1.4. Independencia de eventos | 27 |
| 1.5. Lema de Borel-Cantelli | 28 |
| 1.6. Ejercicios | 30 |
| 2. Variables aleatorias | 40 |
| 2.1. Variables aleatorias | 40 |
| 2.2. Función de distribución | 47 |
| 2.3. Tipos de variables aleatorias | 51 |
| 2.4. Integral de Riemann-Stieltjes | 55 |
| 2.5. Características numéricas | 57 |
| 2.6. Distribuciones discretas | 62 |
| 2.7. Distribuciones continuas | 68 |
| 2.8. Ejercicios | 73 |
| 3. Vectores aleatorios | 95 |
| 3.1. Vectores aleatorios | 95 |
| 3.2. Distribución conjunta | 96 |
| 3.3. Densidad conjunta | 100 |
| 3.4. Distribución marginal | 104 |
| 3.5. Distribución condicional | 105 |
| 3.6. Independencia | 106 |
| 3.7. Esperanza de una función de un vector aleatorio | 109 |
| 3.8. Covarianza | 111 |
| 3.9. Coeficiente de correlación | 113 |
| 3.10. Esperanza y varianza de un vector aleatorio | 116 |
| 3.11. Distribuciones multivariadas discretas | 117 |
| 3.12. Distribuciones multivariadas continuas | 119 |
| 3.13. Ejercicios | 120 |
| 4. Esperanza condicional | 137 |
| 4.1. Esperanza condicional | 137 |
| 4.2. Varianza condicional | 139 |
| 4.3. Ejercicios | 140 |

| | |
|---|------------|
| 5. Transformaciones | 142 |
| 5.1. Transformación de una variable aleatoria | 142 |
| 5.2. Transformación de un vector aleatorio | 146 |
| 5.3. Ejercicios | 155 |
| 6. Distribuciones muestrales y estadísticas de orden | 161 |
| 6.1. Distribuciones muestrales | 162 |
| 6.2. Estadísticas de orden | 169 |
| 6.3. Ejercicios | 175 |
| 7. Convergencia | 182 |
| 7.1. Convergencia puntual | 182 |
| 7.2. Convergencia casi segura | 183 |
| 7.3. Convergencia en probabilidad | 185 |
| 7.4. Convergencia en media | 185 |
| 7.5. Convergencia en media cuadrática | 186 |
| 7.6. Convergencia en distribución | 186 |
| 7.7. Relaciones generales entre los tipos de convergencia | 187 |
| 7.8. Dos resultados importantes de convergencia | 192 |
| 7.9. Ejercicios | 194 |
| 8. Funciones generadoras | 197 |
| 8.1. Función generadora de probabilidad | 197 |
| 8.2. Función generadora de momentos | 200 |
| 8.3. Función característica | 205 |
| 8.4. Ejercicios | 214 |
| 9. Teoremas límite | 221 |
| 9.1. Desigualdad de Markov | 221 |
| 9.2. Desigualdad de Chebyshev | 222 |
| 9.3. Ley de los grandes números | 224 |
| 9.4. Teorema central del límite | 226 |
| 9.5. Ejercicios | 228 |
| A. Distribuciones de probabilidad | 232 |
| B. Formulario | 238 |
| B.1. El alfabeto griego | 238 |
| B.2. Notación | 238 |
| B.3. Límite superior e inferior | 238 |
| B.4. Imagen inversa | 239 |
| B.5. Función indicadora | 240 |
| B.6. Tabla de la distribución normal estándar | 241 |

Capítulo 1

Espacios de probabilidad

La teoría de la probabilidad es la parte de las matemáticas que se encarga del estudio de los fenómenos o *experimentos aleatorios*. Se entiende por experimento aleatorio todo aquel experimento tal que cuando se le repite bajo las mismas condiciones iniciales, el resultado que se obtiene no siempre es el mismo. A menudo y por muy diversas razones es necesario aceptar que no es posible predecir el resultado de un experimento particular aún cuando se le haya efectuado con anterioridad varias veces bajo las mismas condiciones iniciales, y en consecuencia se considera aleatorio. Bajo estas circunstancias, la teoría de la probabilidad tiene el objetivo de modelar matemáticamente cualquier experimento aleatorio de interés.

1.1. Espacios de probabilidad

El modelo matemático creado durante el primer tercio del siglo *XX* para estudiar los experimentos aleatorios es el así llamado *espacio de probabilidad*. Este modelo consiste de una terna ordenada, denotada usualmente por (Ω, \mathcal{F}, P) , en donde Ω es un conjunto arbitrario, \mathcal{F} es una σ -álgebra de subconjuntos de Ω , y P es una medida de probabilidad definida sobre \mathcal{F} . Explicamos a continuación brevemente cada uno de estos elementos.

Espacio muestral. El conjunto Ω es llamado *espacio muestral* o *espacio muestra*, y tiene como objetivo agrupar a todos los posibles resultados del experimento aleatorio en cuestión. No es imprescindible darle esta interpretación al conjunto Ω , y matemáticamente se le considera entonces como un conjunto arbitrario.

σ -álgebra. Una clase o colección no vacía \mathcal{F} de subconjuntos de Ω es una σ -álgebra si es cerrada bajo las operaciones de tomar complementos y uniones numerables. A los elementos de una σ -álgebra se les llama *eventos* o *conjuntos medibles*¹. En particular, un evento es *simple* si consta de a lo más un elemento de Ω , y es *compuesto* cuando consta de dos o más elementos de Ω .

¹Debido a su uso extendido, se usa el término *medible*, aunque tal vez lo correcto sea decir *mesurable*.

Medida de probabilidad. Una función P definida sobre una σ -álgebra \mathcal{F} y con valores en el intervalo $[0, 1]$ es una *medida de probabilidad* si $P(\Omega) = 1$, es no negativa, y es σ -aditiva, es decir,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n),$$

cuando $A_1, A_2, \dots \in \mathcal{F}$ son ajenos dos a dos, esto es, $A_i \cap A_j = \emptyset$ para valores de i y j distintos. El número $P(A)$ representa una forma de medir la *posibilidad* de observar la ocurrencia del evento A , al efectuar una vez el experimento aleatorio. Tenemos entonces formalmente la siguiente definición.

Definición (Espacio de probabilidad). Un espacio de probabilidad es una terna (Ω, \mathcal{F}, P) , en donde Ω es un conjunto arbitrario, \mathcal{F} es una σ -álgebra de subconjuntos de Ω , y P es una medida de probabilidad definida sobre \mathcal{F} .

En este primer capítulo se estudian con más detalle los conceptos de σ -álgebra y medida de probabilidad. Empecemos con el primero.

1.2. σ -álgebras

En esta sección se estudia el concepto de σ -álgebra y se define la mínima σ -álgebra generada por una colección arbitraria. Recordemos nuevamente la definición de esta estructura.

Definición (σ -álgebra). Una colección \mathcal{F} de subconjuntos de Ω es una σ -álgebra si cumple las siguientes condiciones:

1. $\Omega \in \mathcal{F}$.
2. Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$.
3. Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

A la pareja (Ω, \mathcal{F}) se le llama espacio medible y a los elementos de \mathcal{F} se les llama eventos o conjuntos medibles.

En palabras, una σ -álgebra es una colección de subconjuntos de Ω que es no vacía y cerrada bajo las operaciones de tomar complemento y efectuar uniones infinitas numerables. En probabilidad elemental el conjunto Ω denota el *espacio muestral* o conjunto de posibles resultados de un experimento aleatorio, y los elementos de \mathcal{F} representan eventos en el experimento aleatorio. Una σ -álgebra es entonces una

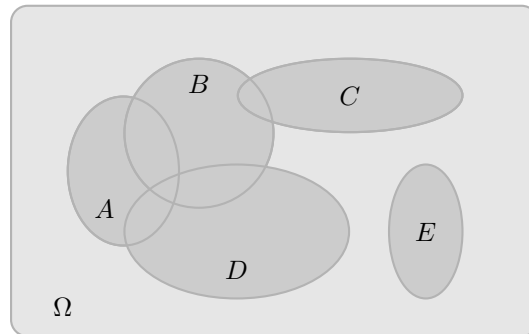
estructura que nos permite agrupar ciertos subconjuntos de Ω de interés, aquellos a los cuales se desea calcular su probabilidad, y esta estructura constituye el dominio de definición de una medida de probabilidad. A menudo no pueden definirse medidas de probabilidad sobre colecciones de subconjuntos más grandes o naturales, como podría ser 2^Ω , la *teoría de la medida* garantiza que por lo menos el concepto de medida de probabilidad, con los axiomas mencionados antes, puede obtenerse sobre σ -álgebras, y por ello es que estudiamos estas estructuras. En general existen varias σ -álgebras que pueden asociarse a un conjunto cualquiera no vacío Ω como se muestra a continuación.

Ejemplo Sea Ω un conjunto cualquiera no vacío. Las siguientes colecciones son σ -álgebras de subconjuntos de Ω .

- a) $\mathcal{F}_1 = \{\emptyset, \Omega\}$.
- b) $\mathcal{F}_2 = \{\emptyset, A, A^c, \Omega\}$, en donde $A \subseteq \Omega$.
- c) $\mathcal{F}_3 = 2^\Omega$, conjunto potencia.

o

Es fácil verificar que las tres condiciones de la definición de σ -álgebra se cumplen para cada caso en el ejemplo anterior. La σ -álgebra del inciso (a) es la σ -álgebra más pequeña que podemos asociar a un conjunto cualquiera Ω , y la σ -álgebra del inciso (c) es la más grande. En el siguiente diagrama puede observarse gráficamente una σ -álgebra como una colección de subconjuntos de Ω .



Una σ -álgebra es una colección $\mathcal{F} = \{A, B, C, D, E, \dots\}$ de subconjuntos de Ω que es no vacía y es cerrada bajo complementos y uniones numerables.

Ejemplo Sean A y B subconjuntos de Ω tales que $A \subseteq B$. La colección

$$\mathcal{F} = \{\emptyset, A, B, A^c, B^c, B - A, (B - A)^c, \Omega\}$$

es una σ -álgebra de subconjuntos de Ω que contiene explícitamente a los conjuntos A y B . Esto puede verificarse directamente con la ayuda de un diagrama de Venn.

◦

En la sección de ejercicios se pueden encontrar algunos otros ejemplos de σ -álgebras. El uso de la letra \mathcal{F} para denotar una σ -álgebra proviene del nombre en inglés *field* que significa *campo*. A menudo se usa también el término σ -campo en lugar de σ -álgebra. Observe con cuidado el uso y significado de los símbolos de contención y pertenencia: $A \subseteq \Omega$ y $A \in \mathcal{F}$. Demostraremos a continuación algunas otras propiedades que satisface cualquier σ -álgebra.

Proposición. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω . Entonces

1. $\emptyset \in \mathcal{F}$.
2. Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.
3. Si $A, B \in \mathcal{F}$, entonces $A - B \in \mathcal{F}$.
4. Si $A, B \in \mathcal{F}$, entonces $A \Delta B \in \mathcal{F}$.

Demostración. (1) Como $\Omega \in \mathcal{F}$ y \mathcal{F} es una colección cerrada bajo complementos entonces $\Omega^c = \emptyset \in \mathcal{F}$. (2) Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $A_1^c, A_2^c, \dots \in \mathcal{F}$. Por lo tanto $\bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F}$. Tomando complementos y usando las leyes de De Morgan se obtiene el resultado. Las proposiciones (3) y (4) se siguen de lo demostrado antes y de las definiciones $A - B = A \cap B^c$, y $A \Delta B = (A - B) \cup (B - A)$. \square

La proposición anterior establece entonces que las σ -álgebras son estructuras también cerradas bajo las operaciones de diferencia e intersecciones numerables. En la sección de ejercicios pueden encontrarse algunas otras definiciones de σ -álgebra equivalentes a la que hemos enunciado, y que involucran las operaciones de la proposición anterior. Una operación de particular importancia es aquella en la que se intersectan dos σ -álgebras produciendo una nueva σ -álgebra. Este es el contenido del siguiente resultado.

Proposición. La intersección de dos σ -álgebras es una σ -álgebra.

Demostración. Sean \mathcal{F}_1 y \mathcal{F}_2 dos σ -álgebras de subconjuntos de Ω . Entonces $\mathcal{F}_1 \cap \mathcal{F}_2$ es aquella colección de subconjuntos de Ω cuyos elementos pertenecen tanto a \mathcal{F}_1 como a \mathcal{F}_2 . Demostraremos que $\mathcal{F}_1 \cap \mathcal{F}_2$ es una σ -álgebra. (1) Como \mathcal{F}_1 y \mathcal{F}_2

son σ -álgebras entonces $\Omega \in \mathcal{F}_1$ y $\Omega \in \mathcal{F}_2$. Por lo tanto $\Omega \in \mathcal{F}_1 \cap \mathcal{F}_2$. (2) Sea A un elemento en $\mathcal{F}_1 \cap \mathcal{F}_2$. Entonces $A \in \mathcal{F}_1$ y $A \in \mathcal{F}_2$. Por lo tanto $A^c \in \mathcal{F}_1$ y $A^c \in \mathcal{F}_2$, es decir, $A^c \in \mathcal{F}_1 \cap \mathcal{F}_2$. (3) Sea A_1, A_2, \dots una sucesión de elementos en $\mathcal{F}_1 \cap \mathcal{F}_2$. Entonces $A_1, A_2, \dots \in \mathcal{F}_1$ y $A_1, A_2, \dots \in \mathcal{F}_2$. Por lo tanto $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_1$ y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_2$, es decir, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_1 \cap \mathcal{F}_2$. \square

Hemos entonces comprobado que si \mathcal{F}_1 y \mathcal{F}_2 son dos σ -álgebras de un mismo conjunto Ω , entonces $\mathcal{F}_1 \cap \mathcal{F}_2$ es nuevamente una σ -álgebra de subconjuntos de Ω , naturalmente más pequeña que \mathcal{F}_1 y \mathcal{F}_2 en el sentido $\mathcal{F}_1 \cap \mathcal{F}_2 \subseteq \mathcal{F}_1, \mathcal{F}_2$. La siguiente pregunta consiste en verificar si la unión de dos σ -álgebras produce nuevamente una σ -álgebra. En este caso la respuesta es negativa. En general no es cierto que la unión de dos σ -álgebras produce una nueva σ -álgebra. Véanse por ejemplo los Ejercicios 10 y 11 a este respecto. Por otro lado se puede extender la validez de la proposición recién demostrada a intersecciones más generales como indica el siguiente resultado.

Proposición. La intersección finita, infinita numerable o bien arbitraria de σ -álgebras es nuevamente una σ -álgebra.

Demostración. Sea T un conjunto arbitrario distinto del vacío. Suponga que para cada t en T se tiene una σ -álgebra \mathcal{F}_t de subconjuntos de Ω . Sea $\mathcal{F} = \bigcap_{t \in T} \mathcal{F}_t$. Siguiendo los mismos pasos que en la demostración anterior es fácil probar que \mathcal{F} es una σ -álgebra. Observe que como T es un conjunto arbitrario, la σ -álgebra \mathcal{F} es efectivamente una intersección arbitraria de σ -álgebras. \square

El resultado anterior garantiza que la siguiente definición tiene sentido.

Definición (σ -álgebra generada). Sea \mathcal{C} una colección no vacía de subconjuntos de Ω . La σ -álgebra generada por \mathcal{C} , denotada por $\sigma(\mathcal{C})$, es la colección

$$\sigma(\mathcal{C}) = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ es } \sigma\text{-álgebra y } \mathcal{C} \subseteq \mathcal{F} \}.$$

Es decir, la colección $\sigma(\mathcal{C})$ es la intersección de todas aquellas σ -álgebras que contienen a \mathcal{C} . Por la proposición anterior sabemos que $\sigma(\mathcal{C})$ es una σ -álgebra. A $\sigma(\mathcal{C})$ también se le llama *mínima σ -álgebra generada* por \mathcal{C} , y el adjetivo *mínima* es claro a partir del hecho de que es la σ -álgebra más pequeña que contiene a la colección \mathcal{C} . Es decir, si \mathcal{F} es una σ -álgebra que contiene a \mathcal{C} entonces forzosamente $\sigma(\mathcal{C}) \subseteq \mathcal{F}$. Observe que $\mathcal{C} \subseteq \sigma(\mathcal{C})$ pues a la colección \mathcal{C} se le han añadido posiblemente algunos otros subconjuntos para convertirla en la σ -álgebra $\sigma(\mathcal{C})$.

Ejemplo Sean $A, B \subseteq \Omega$ con $A \cap B = \emptyset$. Defina la colección $\mathcal{C} = \{A, B\}$. En general esta colección no es una σ -álgebra pero podemos añadirle algunos subconjuntos de Ω para encontrar la σ -álgebra generada por \mathcal{C} . Esto es

$$\sigma(\mathcal{C}) = \{\emptyset, A, B, (A \cup B)^c, A \cup B, A^c, B^c, \Omega\}.$$

No es difícil verificar que ésta es la mínima σ -álgebra que contiene a la colección \mathcal{C} . ◻

Los siguientes dos resultados son proposiciones sencillas y naturales acerca de σ -álgebras generadas. Las demostraciones son cortas pero requieren algunos momentos de reflexión en una primera lectura.

Proposición. Sean \mathcal{C}_1 y \mathcal{C}_2 dos colecciones de subconjuntos de Ω tales que $\mathcal{C}_1 \subseteq \mathcal{C}_2$. Entonces $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$.

Demostración. Claramente $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \sigma(\mathcal{C}_2)$. Entonces $\sigma(\mathcal{C}_2)$ es una σ -álgebra que contiene a la colección \mathcal{C}_1 . Por lo tanto $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$. ◻

Proposición. Si \mathcal{F} es una σ -álgebra, entonces $\sigma(\mathcal{F}) = \mathcal{F}$.

Demostración. Sabemos que $\mathcal{F} \subseteq \sigma(\mathcal{F})$. Como \mathcal{F} es una σ -álgebra que contiene a \mathcal{F} , entonces $\sigma(\mathcal{F}) \subseteq \mathcal{F}$. Esto demuestra la igualdad. ◻

Otras estructuras de subconjuntos

En esta sección se presentan los conceptos de álgebra y semi-álgebra, y su relación con σ -álgebras. No estudiaremos estas estructuras con detalle pero las mencionamos porque desempeñan un papel importante en la construcción y extensión de medidas de probabilidad.

Definición (álgebra). Una colección \mathcal{F} de subconjuntos de Ω es una álgebra si cumple las siguientes condiciones:

1. $\Omega \in \mathcal{F}$.
2. Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$.
3. Si $A_1, \dots, A_n \in \mathcal{F}$, entonces $\bigcup_{k=1}^n A_k \in \mathcal{F}$.

La diferencia entre una álgebra y una σ -álgebra estriba en que para la primera se pide que sea una colección cerrada bajo uniones finitas mientras que la segunda es una colección cerrada bajo uniones infinitas numerables. Claramente toda σ -álgebra es una álgebra.

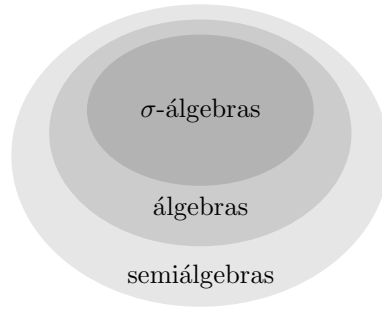
Definición (semiálgebra) Una colección \mathcal{F} de subconjuntos de Ω es una semiálgebra si cumple las siguientes condiciones:

1. $\Omega \in \mathcal{F}$.
2. Si $A, B \in \mathcal{F}$, entonces $A \cap B \in \mathcal{F}$.
3. Si $A, A_1 \in \mathcal{F}$ son tales que $A_1 \subseteq A$, entonces existen $A_2, \dots, A_n \in \mathcal{F}$ tales que

$$A = \bigcup_{k=1}^n A_k,$$

en donde A_1, A_2, \dots, A_n son ajenos dos a dos.

Los conceptos de σ -álgebra, álgebra y semiálgebra están relacionados como se muestra en la siguiente figura. En la sección de ejercicios se pide demostrar las implicaciones y no implicaciones que se obtienen de este diagrama.



Relación general entre σ -álgebras, álgebras y semiálgebras.

En la siguiente sección se estudia un ejemplo importante de una σ -álgebra de subconjuntos de los números reales: la σ -álgebra de Borel.

Conjuntos de Borel

Considere la colección de todos los intervalos abiertos (a, b) de \mathbb{R} , en donde $a \leq b$. A la mínima σ -álgebra generada por esta colección se le llama σ -álgebra de Borel de \mathbb{R} , y se le denota por $\mathcal{B}(\mathbb{R})$.

Definición (σ -álgebra de Borel de \mathbb{R}). $\mathcal{B}(\mathbb{R}) = \sigma \{(a, b) \subseteq \mathbb{R} : a \leq b\}$.

A los elementos de $\mathcal{B}(\mathbb{R})$ se les llama *conjuntos de Borel*, *Borelianos* o *conjuntos Borel medibles*. De esta forma se puede asociar la σ -álgebra $\mathcal{B}(\mathbb{R})$ al conjunto de números reales, y obtener así el espacio medible $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Se muestran a continuación algunos elementos explícitos de la σ -álgebra $\mathcal{B}(\mathbb{R})$.

Proposición. Para cualesquiera números reales $a \leq b$, los intervalos

$$[a, b], (a, \infty), (-\infty, b), [a, b), (a, b], \{a\},$$

son todos elementos de $\mathcal{B}(\mathbb{R})$.

Demostración. Primeramente observe que los intervalos cerrados $[a, b]$ son conjuntos Borelianos, pues podemos escribirlos en términos de una intersección numerable de intervalos abiertos de la siguiente forma

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right).$$

Observe que cada elemento de la intersección anterior es un conjunto Boreliano. Siendo $\mathcal{B}(\mathbb{R})$ una σ -álgebra, la intersección infinita es un elemento de $\mathcal{B}(\mathbb{R})$. De esta forma se concluye que cada intervalo cerrado es un elemento de $\mathcal{B}(\mathbb{R})$. Así mismo tenemos que

$$(a, \infty) = \bigcup_{n=1}^{\infty} (a, a + n) \in \mathcal{B}(\mathbb{R}),$$

y

$$(-\infty, b) = \bigcup_{n=1}^{\infty} (b - n, b) \in \mathcal{B}(\mathbb{R}).$$

Por lo tanto

$$[a, \infty) = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, \infty\right) \in \mathcal{B}(\mathbb{R}),$$

y

$$(-\infty, b] = \bigcap_{n=1}^{\infty} \left(-\infty, b + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R}).$$

De forma análoga se puede hacer ver que los intervalos semiabiertos de la forma $[a, b)$ y $(a, b]$ son conjuntos Borelianos. Los conjuntos que constan de un solo número también son conjuntos Borelianos pues

$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n}\right).$$

□

Complementos, intersecciones y uniones numerables de estos conjuntos son todos ellos Borelianos. Además de la definición enunciada, existen otras formas equivalentes de generar a los conjuntos Borelianos. Este es el contenido de la siguiente proposición.

Proposición. Las siguientes σ -álgebras son todas idénticas a $\mathcal{B}(\mathbb{R})$.

- a) $\sigma\{[a, b] : a \leq b\}$.
- b) $\sigma\{(a, b] : a \leq b\}$.
- c) $\sigma\{[a, b) : a \leq b\}$.
- d) $\sigma\{(a, \infty) : a \in \mathbb{R}\}$.
- e) $\sigma\{(-\infty, b) : b \in \mathbb{R}\}$.

Demostación. Se prueba únicamente el primer inciso. El resto de ellos se demuestra usando el mismo procedimiento. Para demostrar que $\mathcal{B}(\mathbb{R}) = \sigma\{[a, b] : a \leq b\}$ se verifican ambas contenciones. Claramente $[a, b] \in \mathcal{B}(\mathbb{R})$, por lo tanto $\{[a, b] : a \leq b\} \subseteq \mathcal{B}(\mathbb{R})$. Entonces

$$\sigma\{[a, b] : a \leq b\} \subseteq \mathcal{B}(\mathbb{R}).$$

Ahora se demuestra la contención contraria. Sabemos que $(a, b) \in \sigma\{[a, b] : a \leq b\}$ pues $(a, b) = \bigcup_{n=1}^{\infty} [a + \frac{1}{n}, b - \frac{1}{n}]$. Entonces

$$\{(a, b) : a \leq b\} \subseteq \sigma\{[a, b] : a \leq b\}.$$

Por lo tanto $\mathcal{B}(\mathbb{R}) = \sigma\{[a, b] : a \leq b\}$. □

De manera equivalente se puede definir a $\mathcal{B}(\mathbb{R})$ como la mínima σ -álgebra generada por una colección más grande, aquella de todos los subconjuntos abiertos de \mathbb{R} . En ambos casos la σ -álgebra generada es $\mathcal{B}(\mathbb{R})$.

Es natural preguntarse si la colección $\mathcal{B}(\mathbb{R})$ contiene a todos los subconjuntos de \mathbb{R} . La respuesta es negativa, es decir, puede demostrarse que existe un subconjunto de \mathbb{R} que no pertenece a la colección $\mathcal{B}(\mathbb{R})$. La construcción del tal conjunto no es sencilla, y puede obtenerse indirectamente de la siguiente forma: la colección $\mathcal{B}(\mathbb{R})$ está contenida en una clase más amplia llamada la colección de conjuntos *Lebesgue medibles* de \mathbb{R} , y se demuestra que existen subconjuntos de \mathbb{R} que no son Lebesgue medibles, y por tanto tampoco Borel medibles. Los detalles de estas afirmaciones pueden encontrarse en textos de teoría de la medida, como por ejemplo [3].

Es posible también considerar la σ -álgebra de conjuntos de Borel restringidos a una porción de los números reales como se indica a continuación.

Definición. Sea $A \in \mathcal{B}(\mathbb{R})$. La σ -álgebra de Borel de A , denotada por $\mathcal{B}(A)$ o $A \cap \mathcal{B}(\mathbb{R})$, se define como sigue

$$\mathcal{B}(A) = \{A \cap B : B \in \mathcal{B}(\mathbb{R})\}.$$

No es difícil comprobar que $\mathcal{B}(A)$ es efectivamente una σ -álgebra de subconjuntos de A . Observe que el nuevo conjunto total es A y no \mathbb{R} . El concepto de σ -álgebra de Borel de \mathbb{R} puede extenderse a dimensiones mayores de la siguiente forma. Considere la colección \mathcal{C} de todas los rectángulos abiertos de \mathbb{R}^2 , es decir,

$$\mathcal{C} = \{(a, b) \times (c, d) : a \leq b, c \leq d\}.$$

Se definen los conjuntos de Borel de \mathbb{R}^2 como los elementos de la mínima σ -álgebra generada por la colección \mathcal{C} , es decir, $\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{C})$. De manera equivalente se puede definir $\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}))$. En forma análoga se define $\mathcal{B}(\mathbb{R}^n)$ usando productos cartesianos de intervalos.

Definición (σ -álgebra de Borel de \mathbb{R}^n). $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R}))$.

En general el producto cartesiano de dos σ -álgebras no es una σ -álgebra de subconjuntos del espacio producto, de modo que debe anteponerse la operación σ a tal colección para convertirla en una σ -álgebra.

Sucesiones de eventos

En esta sección se estudia el concepto de convergencia de una sucesión infinita de eventos. Para enunciar tal concepto necesitaremos antes las definiciones de límite superior y límite inferior para conjuntos que se establecen a continuación. Estas definiciones son análogas al caso de sucesiones numéricas como puede consultarse en el apéndice al final del texto.

Definición (Límite superior e inferior). Para una sucesión de eventos $\{A_n : n \in \mathbb{N}\}$, se define el límite superior y el límite inferior como sigue:

$$\begin{aligned} 1. \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k. \\ 2. \liminf_{n \rightarrow \infty} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \end{aligned}$$

Tanto el límite superior como el límite inferior son operaciones bien definidas, es decir, el resultado siempre existe y es único. En cada caso, el conjunto resultante es siempre un evento, es decir, un conjunto medible. Es sencillo también comprobar que

$$\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n.$$

Tampoco es difícil verificar que un elemento pertenece al evento $\limsup_{n \rightarrow \infty} A_n$ si y sólo si pertenece a una infinidad² de elementos de la sucesión. Por otro lado un elemento pertenece al evento $\liminf_{n \rightarrow \infty} A_n$ si, y sólo si, pertenece a todos los elementos de la sucesión excepto un número finito de ellos. Con estos elementos podemos ahora establecer la definición de convergencia de una sucesión de eventos.

²En textos de habla inglesa a menudo se escribe $\limsup_{n \rightarrow \infty} A_n = (A_n \text{ i.o.})$, en donde i.o. significa *infinitely often*.

Definición (Convergencia de eventos). Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Si existe un evento A tal que

$$\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = A,$$

entonces se dice que la sucesión converge al evento A , y se escribe

$$\lim_{n \rightarrow \infty} A_n = A.$$

Para calcular el posible límite de una sucesión de eventos debemos entonces calcular el límite superior y el límite inferior, y cuando el resultado de ambas operaciones coincida en el mismo evento, entonces a tal resultado común se le llama el límite de la sucesión. Por supuesto que no todas las sucesiones de eventos convergen. Mostramos a continuación que en particular toda sucesión monótona es convergente. Más adelante presentaremos algunos ejemplos concretos de sucesiones de eventos y en la sección de ejercicios se encuentran algunos otros.

Proposición. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión monótona de eventos.

1. Si $A_1 \subseteq A_2 \subseteq \dots$, entonces $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$.
2. Si $A_1 \supseteq A_2 \supseteq \dots$, entonces $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Demostración. (1) Como la sucesión es creciente, entonces

$$\begin{aligned} \bigcup_{k=n}^{\infty} A_k &= \bigcup_{k=1}^{\infty} A_k, \\ \text{y } \bigcap_{k=n}^{\infty} A_k &= A_n. \end{aligned}$$

Por lo tanto

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k, \\ \text{y } \liminf_{n \rightarrow \infty} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} A_n. \end{aligned}$$

(2) La demostración es completamente análoga al inciso anterior. En este caso como

la sucesión es decreciente se tiene que

$$\bigcap_{k=n}^{\infty} A_k = \bigcap_{k=1}^{\infty} A_k,$$

y

$$\bigcup_{k=n}^{\infty} A_k = A_n.$$

Entonces

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} A_n,$$

y

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} A_k = \bigcap_{k=1}^{\infty} A_k.$$

□

Ejemplo Para cada número natural n defina $A_n = [-1/n, 0]$ si n es impar, y $A_n = [0, 1/n]$ si n es par. Entonces $\lim_{n \rightarrow \infty} A_n = \{0\}$ pues

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} [-1/n, 1/n] = \{0\},$$

y

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \{0\} = \{0\}.$$

○

El siguiente resultado establece que a partir de una sucesión de eventos puede construirse otra sucesión cuyos elementos son ajenos dos a dos y cuya unión es la unión de la sucesión original. Este procedimiento de separación será de utilidad más adelante.

Proposición. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Defina

$$B_1 = A_1,$$

y

$$B_n = A_n - \bigcup_{k=1}^{n-1} A_k, \quad \text{para } n \geq 2.$$

Entonces $\{B_n : n \in \mathbb{N}\}$ es una sucesión de eventos con las siguientes propiedades:

1. $B_n \subseteq A_n$.
2. $B_n \cap B_m = \emptyset$, si $n \neq m$.
3. $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$.

Demostración. El inciso (1) es evidente a partir de la definición de B_n . Para demostrar (2) suponga $n < m$, entonces

$$\begin{aligned}
 B_n \cap B_m &= \left(A_n - \bigcup_{k=1}^{n-1} A_k \right) \cap \left(A_m - \bigcup_{k=1}^{m-1} A_k \right) \\
 &= \left(A_n \cap \bigcap_{k=1}^{n-1} A_k^c \right) \cap \left(A_m \cap \bigcap_{k=1}^{m-1} A_k^c \right) \\
 &\subseteq A_n \cap A_n^c \\
 &= \emptyset.
 \end{aligned}$$

Ahora se demuestra (3) considerando cada contención por separado. Como cada B_n está contenido en A_n entonces el lado izquierdo de (3) es efectivamente un subconjunto del lado derecho. Por el contrario, sea x un elemento en $\bigcup_{n=1}^{\infty} A_n$. Entonces existe un índice n tal que $x \in A_n$. Sea n_0 el primer índice tal que $x \in A_{n_0}$ y $x \notin A_j$ para $1 \leq j \leq n_0 - 1$. Entonces $x \in A_{n_0} - \bigcup_{n=1}^{n_0-1} A_n = B_{n_0}$. Por lo tanto x pertenece a $\bigcup_{n=1}^{\infty} B_n$. \square

1.3. Medidas de probabilidad

En esta sección y en lo que resta del presente capítulo se estudian algunas propiedades de las medidas de probabilidad. Empezaremos por recordar nuevamente la definición de este concepto.

Definición (Medida de probabilidad). Sea (Ω, \mathcal{F}) un espacio medible. Una medida de probabilidad es una función $P : \mathcal{F} \rightarrow [0, 1]$ que satisface

1. $P(\Omega) = 1$.
2. $P(A) \geq 0$, para cualquier $A \in \mathcal{F}$.
3. Si $A_1, A_2, \dots \in \mathcal{F}$ son ajenos dos a dos, esto es, $A_n \cap A_m = \emptyset$ para $n \neq m$, entonces $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$.

Entonces toda función P definida sobre una σ -álgebra \mathcal{F} , con valores en el intervalo $[0, 1]$ y que cumpla los tres postulados anteriores se le llama *medida de probabilidad*. Estos axiomas fueron establecidos por Kolmogorov en 1933. En particular, la tercera propiedad se conoce con el nombre de *σ -aditividad*.



Andrey Nikolaevich Kolmogorov (Rusia 1903–1987).
Fuente: Archivo MacTutor, Universidad de St. Andrews.

Veamos algunos ejemplos de medidas de probabilidad.

Ejemplo (Probabilidad clásica). Considere un experimento aleatorio con espacio muestral un conjunto finito Ω . Asocie al conjunto Ω la σ -álgebra el conjunto potencia 2^Ω . Para cualquier $A \subseteq \Omega$ defina

$$P(A) = \frac{\#A}{\#\Omega}.$$

Entonces P es una medida de probabilidad, y es llamada *probabilidad clásica*. De acuerdo a esta definición, para calcular la probabilidad de un evento es necesario entonces conocer su cardinalidad. De esta forma de calcular probabilidades surgen muchos y muy variados problemas de conteo, algunos de los cuales pueden ser complicados de resolver. ◦

Ejemplo. Considere un experimento aleatorio con espacio muestral el conjunto de números naturales \mathbb{N} . Asocie al conjunto \mathbb{N} la σ -álgebra el conjunto potencia $2^\mathbb{N}$. Para cualquier $A \subseteq \mathbb{N}$ defina

$$P(A) = \sum_{n \in A} \frac{1}{2^n}.$$

No es difícil verificar que P es efectivamente una medida de probabilidad. ◦

Ejemplo. Considere el espacio medible $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Sea $f : \mathbb{R} \rightarrow \mathbb{R}$ una función no negativa y continua, tal que su integral sobre el intervalo $(-\infty, \infty)$ es uno. Para cualquier A en $\mathcal{B}(\mathbb{R})$ defina

$$P(A) = \int_A f(x) dx.$$

Se puede demostrar que P es una medida de probabilidad. ◦

Ejemplo (Probabilidad geométrica). Sea $\Omega \subseteq \mathbb{R}^2$ una región tal que su área es positiva y finita. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω para los cuales el concepto de área esté bien definido. Para cada A en \mathcal{F} defina

$$P(A) = \frac{\text{Área}(A)}{\text{Área}(\Omega)}.$$

La función P , que resulta ser una medida de probabilidad, es llamada *probabilidad geométrica*, y puede definirse este concepto en espacios de dimensión mayor. \circ

En la siguiente sección estudiaremos algunas propiedades generales que cumple toda medida de probabilidad, y a lo largo del texto estudiaremos varios modelos particulares para calcular probabilidades.

Propiedades elementales

A partir de los postulados enunciados en la sección anterior es posible demostrar una extensa serie de propiedades que cumplen todas las medidas de probabilidad. En esta sección se estudian algunas propiedades elementales y más adelante se demuestran otras propiedades más avanzadas.

Proposición. Sea P una medida de probabilidad. Entonces

1. $P(\emptyset) = 0$.
2. Si $A_1, A_2, \dots, A_n \in \mathcal{F}$ son ajenos dos a dos, entonces

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k).$$

3. $P(A^c) = 1 - P(A)$.
4. Si $A \subseteq B$, entonces $P(B - A) = P(B) - P(A)$.
5. Si $A \subseteq B$, entonces $P(A) \leq P(B)$.
6. $0 \leq P(A) \leq 1$.
7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
8. $P(A \cup B) \leq P(A) + P(B)$.

Demostración. Como $\emptyset = \emptyset \cup \emptyset \cup \dots$, por la σ -aditividad se tiene que

$$P(\emptyset) = \sum_{n=1}^{\infty} P(\emptyset),$$

lo cual sucede únicamente cuando $P(\emptyset) = 0$. Para la propiedad (2) se toma $A_{n+1} = A_{n+2} = \dots = \emptyset$, y la igualdad se obtiene al aplicar la σ -aditividad y la propiedad anterior. Para la propiedad (3) expresamos a Ω como la unión disjunta $A \cup A^c$. Aplicamos P y obtenemos la igualdad requerida. Para demostrar (4) escribimos

$B = A \cup (B - A)$. Aplicando P obtenemos $P(B) - P(A) = P(B - A)$. Como la probabilidad de cualquier evento es un número no negativo, de la igualdad anterior obtenemos también la propiedad (5). La primera desigualdad de la propiedad (6) es el segundo axioma, y la segunda es consecuencia de la propiedad (5) cuando $B = \Omega$ y el primer axioma. Finalmente para demostrar (7) descomponemos el evento $A \cup B$ como la siguiente unión de tres eventos disjuntos dos a dos: $A \cup B = (A - B) \cup (A \cap B) \cup (B - A) = (A - A \cap B) \cup (A \cap B) \cup (B - A \cap B)$. Por lo tanto $P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$. La propiedad (8) es consecuencia de la anterior y el segundo axioma. \square

La propiedad (2) establece que las probabilidades son funciones *finitamente aditivas*, y la propiedad (5) que son funciones monótonas. La desigualdad (8) dice que las probabilidades son funciones *finitamente subaditivas*. Veamos algunas otras propiedades de las medidas de probabilidad.

Proposición (Desigualdades de Boole). Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Entonces

1. $P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$.
2. $P\left(\bigcap_{n=1}^{\infty} A_n\right) \geq 1 - \sum_{n=1}^{\infty} P(A_n^c)$.

Demostración. Para la primera desigualdad tome $B_1 = A_1$, y para $n \geq 2$ defina

$$B_n = A_n - \bigcup_{k=1}^{n-1} A_k.$$

Entonces $\{B_n : n \in \mathbb{N}\}$ es una sucesión de eventos disjuntos dos a dos tales que $B_n \subseteq A_n$ y $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Esto es consecuencia de la Proposición 1.2 de la página 17. Por lo tanto

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &\leq \sum_{n=1}^{\infty} P(A_n). \end{aligned}$$

La segunda desigualdad se sigue de la primera tomando complementos. \square

Proposición. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos.

1. Si $P(A_n) = 1$ para toda n , entonces $P(\bigcap_{n=1}^{\infty} A_n) = 1$.
2. Si $P(A_n) = 1$ para alguna n , entonces $P(\bigcup_{n=1}^{\infty} A_n) = 1$.
3. Si $P(A_n) = 0$ para alguna n , entonces $P(\bigcap_{n=1}^{\infty} A_n) = 0$.
4. Si $P(A_n) = 0$ para toda n , entonces $P(\bigcup_{n=1}^{\infty} A_n) = 0$.

Demostración. (1) Por las leyes de De Morgan y la desigualdad de Boole,

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1 - P\left(\bigcup_{n=1}^{\infty} A_n^c\right) \\ &\geq 1 - \sum_{n=1}^{\infty} P(A_n^c) \\ &= 1. \end{aligned}$$

(2) Como $A_n \subseteq \bigcup_{n=1}^{\infty} A_n$, se tiene que $1 = P(A_n) \leq P\left(\bigcup_{n=1}^{\infty} A_n\right)$.

(3) Como $\bigcap_{n=1}^{\infty} A_n \subseteq A_n$, entonces $P\left(\bigcap_{n=1}^{\infty} A_n\right) \leq P(A_n) = 0$.

(4) Por la desigualdad de Boole, $P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) = 0$. □

Las propiedades (1) y (4) de la proposición anterior pueden interpretarse de la siguiente forma. Intersectar dos eventos produce en general un evento más pequeño o por lo menos no mayor a los intersectandos. Sin embargo la propiedad (1) establece que la intersección, aún infinita, de eventos con probabilidad uno produce un evento con probabilidad todavía uno. Análogamente, unir dos eventos produce en general un evento mayor, pero por la propiedad (4), la unión, aún infinita, de eventos con probabilidad cero tiene probabilidad que se mantiene en cero.

Continuidad

Ahora demostraremos que las medidas de probabilidad son funciones continuas. Primero se prueba este resultado para dos tipos de sucesiones particulares, aquellas que son monótonas crecientes o decrecientes, y después se prueba en general. Empezaremos con el caso de sucesiones crecientes.

Proposición. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión no decreciente de eventos, esto es, $A_1 \subseteq A_2 \subseteq \dots$. Entonces

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Demostración. Como $A_n \subseteq A_{n+1}$ tenemos que $P(A_n) \leq P(A_{n+1})$. Por lo tanto la sucesión numérica $\{P(A_n) : n \in \mathbb{N}\}$ es no decreciente y acotada superiormente por uno. Entonces el límite de esta sucesión existe y el lado derecho de la igualdad tiene sentido. Defina los eventos

$$\begin{aligned} B_1 &= A_1, \\ \text{y } B_n &= A_n - A_{n-1}, \quad \text{para } n \geq 2. \end{aligned}$$

La sucesión $\{B_n : n \in \mathbb{N}\}$ es una colección de eventos disjuntos dos a dos, y es tal que

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n.$$

Por lo tanto

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &= P(B_1) + \sum_{n=2}^{\infty} P(B_n) \\ &= P(A_1) + \sum_{n=2}^{\infty} P(A_n - A_{n-1}) \\ &= P(A_1) + \sum_{n=2}^{\infty} P(A_n) - P(A_{n-1}) \\ &= P(A_1) + \lim_{m \rightarrow \infty} \sum_{n=2}^m P(A_n) - P(A_{n-1}) \\ &= P(A_1) + \lim_{m \rightarrow \infty} P(A_m) - P(A_1) \\ &= \lim_{m \rightarrow \infty} P(A_m). \end{aligned}$$

□

Las medidas de probabilidad también son continuas respecto de sucesiones no crecientes de eventos. Esta afirmación es el contenido del siguiente resultado que se demuestra a partir de la proposición anterior.

Proposición. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión no creciente de eventos, esto es, $A_1 \supseteq A_2 \supseteq \dots$. Entonces

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Demostración. Observe que si $A_n \supseteq A_{n+1}$ entonces $A_n^c \subseteq A_{n+1}^c$. Por la proposición anterior,

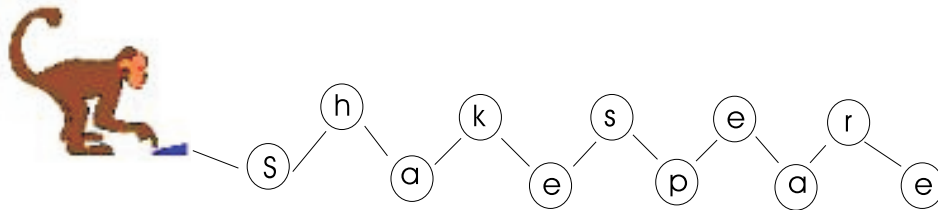
$$P\left(\bigcup_{n=1}^{\infty} A_n^c\right) = \lim_{n \rightarrow \infty} P(A_n^c).$$

Aplicando las leyes de De Morgan,

$$1 - P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} (1 - P(A_n)),$$

de donde se sigue inmediatamente el resultado. \square

Ejemplo (El problema del mono). Un mono escribe caracteres al azar en una máquina de escribir. ¿Cuál es la probabilidad de que eventualmente obtenga exactamente, y sin ningún error, las obras completas de Shakespeare?



Mono escribiendo al azar.

Demostramos a continuación que la probabilidad de este raro evento es uno. Imagine entonces que un mono escribe caracteres al azar en una máquina de escribir, y que lo hace de manera continua generando una sucesión lineal de caracteres. Sea m el total de caracteres disponibles, y sea N el total de caracteres de los que constan las obras completas de Shakespeare. Segmentamos el arreglo lineal de caracteres generados por el mono en bloques disjuntos de N caracteres, uno después de otro, y observamos si algún bloque contiene las obras de Shakespeare. Por ejemplo,

$$\underbrace{Xku \dots aTs}_N \underbrace{hwW \dots pZq}_N Ot \dots$$

Para cada número natural k defina el evento A_k correspondiente a que el k -ésimo bloque contiene exactamente y sin error alguno las obras completas de Shakespeare. Observe que los eventos A_k son independientes pues los bloques no se superponen, además $P(A_k) = (1/m)^N = p$. Defina $B_k = A_1^c \cap \dots \cap A_k^c$, que indica el evento de que el mono no obtenga éxito en los primeros k bloques. Observe que $B_{k+1} \subseteq B_k$, es decir la sucesión es decreciente, por lo tanto

$$\lim_{k \rightarrow \infty} B_k = \bigcap_{k=1}^{\infty} B_k,$$

en donde el evento $\bigcap_{k=1}^{\infty} B_k$ se interpreta como aquel en el que el mono nunca tiene éxito. Entonces

$$P\left(\bigcap_{k=1}^{\infty} B_k\right) = \lim_{k \rightarrow \infty} P(B_k) = \lim_{k \rightarrow \infty} (1-p)^k = 0.$$

Por lo tanto la probabilidad de que eventualmente el mono obtenga éxito es uno. \circ

Ahora se enuncia un resultado más fuerte. La siguiente proposición establece que las medidas de probabilidad son funciones continuas. Esta propiedad es muy útil pues permite el cálculo de probabilidades en procedimientos límite, y se encuentra siempre presente de manera implícita en toda la teoría que se desarrolla más adelante.

Proposición (Continuidad de la probabilidad). Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos convergente al evento A . Entonces

$$\lim_{n \rightarrow \infty} P(A_n) = P(A).$$

Demostración. La prueba se basa en las siguientes dos desigualdades:

- a) $\limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n).$
- b) $P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n).$

Como la sucesión de eventos $\{A_n : n \in \mathbb{N}\}$ es convergente al evento A entonces

$$\limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = A.$$

Se sigue entonces de las desigualdades (a) y (b) que

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(A_n) &\leq P(\limsup_{n \rightarrow \infty} A_n) \\ &= P(A) \\ &= P(\liminf_{n \rightarrow \infty} A_n) \\ &\leq \liminf_{n \rightarrow \infty} P(A_n). \end{aligned}$$

De donde se concluye el resultado. Nos concentraremos ahora en demostrar las desigualdades enunciadas. (a) Como $A_n \subseteq \bigcup_{k=n}^{\infty} A_k$, entonces

$$P(A_n) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right),$$

en donde $\{\bigcup_{k=n}^{\infty} A_k : n \in \mathbb{N}\}$ es una sucesión decreciente de eventos. Tomando el límite superior se obtiene

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(A_n) &\leq \limsup_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \\ &= P\left(\lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k\right) \\ &= P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \\ &= P(\limsup_{n \rightarrow \infty} A_n). \end{aligned}$$

(b) Como $\bigcap_{k=n}^{\infty} A_k \subseteq A_n$, entonces

$$P\left(\bigcap_{k=n}^{\infty} A_k\right) \leq P(A_n),$$

en donde $\{\bigcap_{k=n}^{\infty} A_k : n \in \mathbb{N}\}$ es una sucesión creciente de eventos. Tomando el límite inferior se obtiene

$$\begin{aligned} \liminf_{n \rightarrow \infty} P(A_n) &\geq \liminf_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\ &= P\left(\lim_{n \rightarrow \infty} \bigcap_{k=n}^{\infty} A_k\right) \\ &= P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) \\ &= P(\liminf_{n \rightarrow \infty} A_n). \end{aligned}$$

□

Ejemplo. Se lanza un dado equilibrado una infinidad de veces. Sea el evento $A = \{2, 4, 6\}$, y sea A_n el evento correspondiente a obtener el evento A en cada uno de los

primeros n lanzamientos del dado. Entonces claramente $A_n \supseteq A_{n+1}$ para cualquier n en \mathbb{N} . Por lo tanto

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Entonces

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= P\left(\lim_{n \rightarrow \infty} A_n\right) \\ &= \lim_{n \rightarrow \infty} P(A_n) \\ &= \lim_{n \rightarrow \infty} \frac{1}{2^n} \\ &= 0. \end{aligned}$$

El evento $\bigcap_{n=1}^{\infty} A_n$ se interpreta como aquel resultado en el que siempre se obtiene un número par en una sucesión infinita de lanzamientos. Hemos demostrado que la probabilidad de tal evento es cero. En consecuencia la probabilidad de que eventualmente aparezca un número impar es uno. Observe que el argumento presentado funciona de la misma forma cuando el evento A es cualquier subconjunto propio de Ω distinto del vacío. Por ejemplo, si $A = \{1, 2, 3, 4, 5\}$, entonces la probabilidad de nunca obtener “6” es cero. Por lo tanto, con probabilidad uno, cada una de las caras del dado aparecerá eventualmente. \circ

1.4. Independencia de eventos

En esta sección se define el concepto importante de independencia de eventos. Este es un concepto central en la teoría de la probabilidad y uno de sus rasgos distintivos. De manera natural la independencia aparecerá con frecuencia a lo largo del texto a partir de ahora, y ayudará a simplificar el cálculo de probabilidades. La definición matemática es la siguiente.

Definición (Independencia de dos eventos). Dos eventos A y B son independientes, y se escribe $A \perp B$, cuando

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

Aceptar la hipótesis de que dos eventos son independientes es una cuestión de apreciación por parte del observador. Puede interpretarse en el sentido de que la ocurrencia de uno de los eventos no proporciona información que modifique la probabilidad de ocurrencia del segundo evento. Contrario a alguna primera concepción intuitiva errónea, el hecho de que dos eventos sean independientes no implica que ellos sean

ajenos. La proposición contraria tampoco es válida, dos eventos ajenos no necesariamente son independientes. La definición de independencia puede extenderse a colecciones finitas e incluso infinitas de eventos del siguiente modo.

Definición (Independencia de n eventos). Los eventos A_1, \dots, A_n son independientes si se cumplen todas y cada una de las siguientes condiciones:

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad i, j \text{ distintos.} \quad (1.1)$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k), \quad i, j, k \text{ distintos.} \quad (1.2)$$

$$\vdots$$

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n).$$

Más generalmente, una colección infinita de eventos es independiente si cualquier subcolección finita lo es.

Observe que según la definición anterior, se necesitan verificar o suponer varias condiciones para que n eventos sean independientes entre sí. De hecho el número total de igualdades a demostrar es $2^n - n - 1$. En la siguiente sección usaremos el hecho de que si A_1, \dots, A_n son independientes, entonces los eventos A_1^c, \dots, A_n^c también son independientes. Es posible además demostrar que la independencia dos a dos, igualdad (1.1), no implica en general la independencia tres a tres, igualdad (1.2), ni viceversa. También se tiene la noción de independencia entre dos colecciones de eventos, cuya definición es la siguiente.

Definición (Independencia de σ -álgebras). Dos sub- σ -álgebras \mathcal{F}_1 y \mathcal{F}_2 son independientes si para cada A en \mathcal{F}_1 y cada B en \mathcal{F}_2 se cumple

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

1.5. Lema de Borel-Cantelli

Concluimos este capítulo con el enunciado y demostración del famoso lema de Borel-Cantelli. El objetivo es demostrar este resultado y con ello poner en práctica algunas propiedades de las medidas de probabilidad, aunque también lo usaremos para demostrar la ley fuerte de los grandes números en la última parte del curso.

Proposición (Lema de Borel-Cantelli). Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos, y defina $A = \limsup_{n \rightarrow \infty} A_n$.

1. Si $\sum_{n=1}^{\infty} P(A_n) < \infty$, entonces $P(A) = 0$.
2. Si A_1, A_2, \dots son independientes y $\sum_{n=1}^{\infty} P(A_n) = \infty$, entonces $P(A) = 1$.

Demostración. (1) Para cada n en \mathbb{N} ,

$$P(A) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k).$$

Como $\sum_{n=1}^{\infty} P(A_n) < \infty$, el lado derecho tiende a cero cuando n tiende a infinito. Esto implica que $P(A) = 0$. (2) Es suficiente demostrar que para todo número natural n se cumple la igualdad $P(\bigcup_{k=n}^{\infty} A_k) = 1$, pues la intersección numerable de eventos con probabilidad uno tiene probabilidad uno. Para cada $m > n$,

$$\begin{aligned} 1 - P\left(\bigcup_{k=n}^{\infty} A_k\right) &\leq 1 - P\left(\bigcup_{k=n}^m A_k\right) \\ &= P\left(\bigcap_{k=n}^m A_k^c\right) \\ &= \prod_{k=n}^m [1 - P(A_k)] \\ &\leq \exp\left(-\sum_{k=n}^m P(A_k)\right). \end{aligned}$$

Para obtener la última expresión se usa la desigualdad: $1 - x \leq e^{-x}$, válida para cualquier número real x . Como $\sum_{n=1}^{\infty} P(A_n) = \infty$, el lado derecho tiende a cero cuando m tiende a infinito. Por lo tanto $P(\bigcup_{k=n}^{\infty} A_k) = 1$ para cualquier valor de n y entonces $P(A) = 1$. \square

Ejemplo (El problema del mono, nuevamente). El problema de encontrar la probabilidad de que un mono que escribe caracteres al azar eventualmente escriba las obras completas de Shakespeare puede resolverse también usando el lema de Borel-Cantelli. Suponga que N es el total de caracteres de los que constan las obras completas de Shakespeare y considere nuevamente la división por bloques de longitud N ,

$$\underbrace{x_1, \dots, x_N}_{A_1}, \underbrace{x_{N+1}, \dots, x_{2N}}_{A_2}, \dots$$

El evento A_k se define nuevamente como aquel en el que el mono tiene éxito en el k -ésimo bloque. Si nuevamente m denota el total de caracteres disponibles, entonces $P(A_k) = (1/m)^N$. Entonces claramente la sucesión A_1, A_2, \dots constituye una

sucesión de eventos independientes tales que $\sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{\infty} (1/m)^N = \infty$. Entonces por la segunda parte del lema de Borel-Cantelli, $P(\limsup_{k \rightarrow \infty} A_k) = 1$. Ahora sólo hay que recordar que el evento $\limsup_{n \rightarrow \infty} A_n$ corresponde a aquel en el que una infinidad de eventos A_k ocurren. Es decir, con probabilidad uno, el mono tiene, no una, sino ¡una infinidad de éxitos!. \circ

1.6. Ejercicios

σ -álgebras

1. *Definición alternativa de σ -álgebra.* Demuestre que \mathcal{F} es una σ -álgebra de subconjuntos de Ω si y sólo si satisface las siguientes propiedades:

a) $\emptyset \in \mathcal{F}$.

b) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.

c) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

2. *Definición alternativa de σ -álgebra.* Demuestre que \mathcal{F} es una σ -álgebra de subconjuntos de Ω si y sólo si satisface las siguientes propiedades:

a) $\Omega \in \mathcal{F}$.

b) $A, B \in \mathcal{F} \Rightarrow A - B \in \mathcal{F}$.

c) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

3. Sean A_1, A_2, \dots, A_n eventos de un espacio muestral Ω . Demuestre que el conjunto de elementos de Ω que pertenecen a exactamente k de estos eventos es un evento, $1 \leq k \leq n$.
4. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω . Demuestre que la colección

$$\mathcal{F}^c = \{F^c : F \in \mathcal{F}\}$$

es una σ -álgebra. Compruebe además que $\mathcal{F}^c = \mathcal{F}$.

5. Sea $\Omega = \{a, b, c, d\}$, y sean $A = \{a, b\}$ y $B = \{b, c\}$. Defina la colección $\mathcal{C} = \{A, B\}$. Claramente \mathcal{C} no es una σ -álgebra. Encuentre $\sigma(\mathcal{C})$.
6. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω y sea A un elemento de \mathcal{F} . Demuestre que la colección $\{A \cap F : F \in \mathcal{F}\}$ es una σ -álgebra de subconjuntos de A . Se usan los símbolos \mathcal{F}_A ó $A \cap \mathcal{F}$ para denotar a esta colección.
7. Sea Ω un conjunto no numerable. Demuestre que la siguiente colección es una σ -álgebra:

$$\mathcal{F} = \{A \subseteq \Omega : A \text{ o } A^c \text{ es finito o numerable}\}.$$

8. Sean Ω_1 y Ω_2 dos conjuntos arbitrarios, y sea $X : \Omega_1 \rightarrow \Omega_2$ una función en donde $(\Omega_2, \mathcal{F}_2)$ es un espacio medible. Demuestre que la siguiente colección es una σ -álgebra de subconjuntos de Ω_1 :

$$X^{-1}\mathcal{F}_2 = \{X^{-1}F : F \in \mathcal{F}_2\}.$$

9. ¿Es la diferencia de dos σ -álgebras una σ -álgebra?
10. Sean \mathcal{F}_1 y \mathcal{F}_2 dos σ -álgebras de subconjuntos de Ω . Demuestre que $\mathcal{F}_1 \cup \mathcal{F}_2$ no necesariamente es una σ -álgebra. Para ello considere el espacio $\Omega = \{1, 2, 3\}$ con $\mathcal{F}_1 = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ y $\mathcal{F}_2 = \{\emptyset, \{1, 2\}, \{3\}, \Omega\}$.
11. Sean \mathcal{F}_1 y \mathcal{F}_2 dos σ -álgebras de subconjuntos de Ω tales que $\mathcal{F}_1 \subseteq \mathcal{F}_2$. Demuestre que $\mathcal{F}_1 \cup \mathcal{F}_2$ es una σ -álgebra.
12. Sea T un conjunto arbitrario distinto del vacío. Suponga que para cada t en T se tiene una σ -álgebra \mathcal{F}_t de subconjuntos de Ω . Demuestre con detalle que $\bigcap_{t \in T} \mathcal{F}_t$ es una σ -álgebra.
13. Sea \mathcal{C} una colección de subconjuntos de Ω . Demuestre que $\sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$.
14. Sean $A, B \subseteq \Omega$ arbitrarios. Demuestre que la cardinalidad de $\sigma\{A, B\}$ es a lo sumo 16.
15. Sean $A, B \subseteq \Omega$ arbitrarios. Encuentre explícitamente todos los elementos de $\sigma\{A, B\}$. Por el ejercicio anterior, el total de elementos en $\sigma\{A, B\}$, en el caso más general, es 16.
16. Sea $\{A_1, \dots, A_n\}$ una partición finita de Ω , es decir, la unión de todos estos conjuntos es igual a Ω , y la intersección de cualesquiera dos de ellos es vacía. Demuestre que la cardinalidad de $\sigma\{A_1, \dots, A_n\}$ es 2^n .
17. Sea $\{A, B, C\}$ una partición de Ω . Encuentre explícitamente los ocho elementos de $\sigma\{A, B, C\}$.
18. Sea \mathcal{C} una colección de subconjuntos de Ω . Diga falso o verdadero justificando en cada caso: $\mathcal{C} \subseteq \sigma(\mathcal{C}) \subseteq 2^\Omega$.
19. Demuestre que 2^Ω es una σ -álgebra de subconjuntos de Ω y que no existe una σ -álgebra de subconjuntos de Ω que sea más grande.
20. Demuestre que toda σ -álgebra de un espacio muestral finito contiene un número par de elementos.
21. Sea Ω un conjunto, \mathcal{F} una σ -álgebra de subconjuntos de Ω y A un evento. De cada una de las dos expresiones siguientes determine la que es notacionalmente correcta. Explique su respuesta.
- $\Omega \in \mathcal{F}$ ó $\Omega \subseteq \mathcal{F}$.
 - $A \in \Omega$ ó $A \subseteq \Omega$.
 - $\emptyset \in \mathcal{F}$ ó $\emptyset \subseteq \mathcal{F}$.
 - $A \in \mathcal{F}$ ó $A \subseteq \mathcal{F}$.

σ -álgebras, álgebras y semiálgebras

22. *Definición alternativa de álgebra.* Demuestre que \mathcal{F} es una álgebra de subconjuntos de Ω si, y sólo si, cumple las siguientes condiciones:

- a) $\Omega \in \mathcal{F}$.
- b) $A, B \in \mathcal{F} \Rightarrow A - B \in \mathcal{F}$.

23. Demuestre que

$$\mathcal{F} \text{ es } \sigma\text{-álgebra} \Rightarrow \mathcal{F} \text{ es álgebra} \Rightarrow \mathcal{F} \text{ es semiálgebra.}$$

24. *álgebra $\not\Rightarrow$ σ -álgebra.* Sea $\Omega = (0, 1]$ y defina la colección \mathcal{F} de subconjuntos de la forma

$$\bigcup_{i=1}^n (a_i, b_i],$$

en donde $(a_i, b_i] \subseteq (0, 1]$ con $(a_i, b_i] \cap (a_j, b_j] = \emptyset$ para $i \neq j$ y $n \in \mathbb{N}$. Demuestre que \mathcal{F} es una álgebra pero no una σ -álgebra.

25. Mediante un contraejemplo demuestre que no toda semiálgebra es una álgebra.

Conjuntos de Borel

26. Demuestre que \mathbb{N} , \mathbb{Z} y \mathbb{Q} son elementos de $\mathcal{B}(\mathbb{R})$.

27. Demuestre que el conjunto de números irracionales es un conjunto de Borel de \mathbb{R} .

28. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(a, b] : a \leq b\}$.

29. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{[a, b) : a \leq b\}$.

30. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(a, \infty) : a \in \mathbb{R}\}$.

31. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{[a, \infty) : a \in \mathbb{R}\}$.

32. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(-\infty, b) : b \in \mathbb{R}\}$.

33. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(-\infty, b] : b \in \mathbb{R}\}$.

34. Sea $A \in \mathcal{B}(\mathbb{R})$. Demuestre que $\mathcal{B}(A)$ es efectivamente una σ -álgebra de subconjuntos de A .

35. Diga falso o verdadero. Justifique su respuesta.

a) $\sigma\left\{\left(\frac{1}{n+1}, \frac{1}{n}\right] : n \in \mathbb{N}\right\} = \mathcal{B}(0, 1]$.

b) $\sigma\left\{\left(0, \frac{1}{n}\right] : n \in \mathbb{N}\right\} = \mathcal{B}(0, 1]$.

c) $\sigma\left\{\left(\frac{1}{n+1}, \frac{1}{n}\right] : n \in \mathbb{N}\right\} = \sigma\left\{\left(0, \frac{1}{n}\right] : n \in \mathbb{N}\right\}$.

36. Demuestre que $\mathcal{B}(\mathbb{R}^2) = \sigma\{[a, b] \times [c, d] : a \leq b, c \leq d\}$.

37. *σ -álgebra producto.* Demuestre que el producto cartesiano de dos σ -álgebras no es necesariamente σ -álgebra. Esto es, suponga que $(\Omega_1, \mathcal{F}_1)$ y $(\Omega_2, \mathcal{F}_2)$ son dos espacios medibles. Mediante un ejemplo muestre que $\mathcal{F}_1 \times \mathcal{F}_2$ no necesariamente es una σ -álgebra de subconjuntos del espacio producto $\Omega_1 \times \Omega_2$. Se define entonces la *σ -álgebra producto* de la forma siguiente:

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{F}_1 \times \mathcal{F}_2).$$

Sucesiones de eventos

38. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Demuestre que

- $\limsup_{n \rightarrow \infty} A_n$ es un evento.
- $\liminf_{n \rightarrow \infty} A_n$ es un evento.
- $\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n$.

39. Demuestre que

- $\limsup_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ para una infinidad de valores de } n\}$.
- $\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ para toda } n \text{ excepto un número finito de ellas}\}$.

40. Suponga $A_n \subseteq B_n$ para cada n en \mathbb{N} . Demuestre que

- $\limsup_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} B_n$.
- $\liminf_{n \rightarrow \infty} A_n \subseteq \liminf_{n \rightarrow \infty} B_n$.
- $\limsup_{n \rightarrow \infty} A_n \subseteq \liminf_{n \rightarrow \infty} B_n$.

41. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Demuestre que

- $(\liminf_{n \rightarrow \infty} A_n)^c = \limsup_{n \rightarrow \infty} A_n^c$.
- $(\limsup_{n \rightarrow \infty} A_n)^c = \liminf_{n \rightarrow \infty} A_n^c$.
- $P(\liminf_{n \rightarrow \infty} A_n) = 1 - P(\limsup_{n \rightarrow \infty} A_n^c)$.
- $P(\limsup_{n \rightarrow \infty} A_n) = 1 - P(\liminf_{n \rightarrow \infty} A_n^c)$.

42. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Demuestre que

- $\lim_{n \rightarrow \infty} A_n = A \iff \lim_{n \rightarrow \infty} A_n^c = A^c$.
- $\lim_{n \rightarrow \infty} 1_{A_n} = 1_A \iff \lim_{n \rightarrow \infty} 1_{A_n^c} = 1_{A^c}$.

El símbolo 1_A denota la función indicadora del conjunto A . Véase el final del texto para la definición y algunas propiedades de esta función.

43. Sea $\{a_n : n \in \mathbb{N}\}$ una sucesión de números no negativos convergente al número $a \geq 0$. Sea $A_n = [0, a_n]$. Calcule $\liminf_{n \rightarrow \infty} A_n$ y $\limsup_{n \rightarrow \infty} A_n$.

44. Determine si cada una de las siguientes sucesiones de conjuntos es convergente:

a) $A_n = (1/n, 2 + (-1)^n)$.

b) $A_n = \{(x, y) : x^2 + y^2 \leq (1 + 1/n)^n\}$.

c) $A_n = \{(x, y) : x^2 + y^2 \leq 2 + \text{sen}(n\pi/2)\}$.

45. Demuestre que las siguientes sucesiones de eventos no son convergentes:

a) $A_n = \emptyset$ si n es impar, y $A_n = \Omega$ si n es par.

b) $A_n = (0, 1 + (-1/2)^n)$.

46. Suponga que $\lim_{n \rightarrow \infty} A_n = A$, y $\lim_{n \rightarrow \infty} B_n = B$. Determine si la siguiente sucesión es convergente:

$$C_n = \begin{cases} A_n & \text{si } n \text{ es impar,} \\ B_n & \text{si } n \text{ es par.} \end{cases}$$

47. Sean A y B dos eventos. Determine en cada caso si la sucesión es convergente:

a) $A_n = \begin{cases} A & \text{si } n \text{ es impar,} \\ A^c & \text{si } n \text{ es par.} \end{cases}$

b) $A_n = \begin{cases} A & \text{si } n \text{ es impar,} \\ B & \text{si } n \text{ es par.} \end{cases}$

48. Suponga que $\lim_{n \rightarrow \infty} A_n = A$. Demuestre que para cualquier evento B ,

a) $\lim_{n \rightarrow \infty} (A_n \cap B) = A \cap B$.

b) $\lim_{n \rightarrow \infty} (A_n \cup B) = A \cup B$.

c) $\lim_{n \rightarrow \infty} (A_n - B) = A - B$.

d) $\lim_{n \rightarrow \infty} (A_n \triangle B) = A \triangle B$.

49. Suponga que $\lim_{n \rightarrow \infty} A_n = A$ y $\lim_{n \rightarrow \infty} B_n = B$. Diga falso o verdadero. Demuestre en cada caso.

a) $\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} (A_n \cap B_m) = A \cap B$.

b) $\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} (A_n \cup B_m) = A \cup B$.

c) $\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} (A_n - B_m) = A - B$.

d) $\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} (A_n \triangle B_m) = A \triangle B$.

50. Suponga que $\lim_{n \rightarrow \infty} A_n = A$ y $\lim_{n \rightarrow \infty} B_n = B$. Diga falso o verdadero. Demuestre en cada caso.

a) $\lim_{n \rightarrow \infty} (A_n \cap B_n) = A \cap B$.

b) $\lim_{n \rightarrow \infty} (A_n \cup B_n) = A \cup B$.

c) $\lim_{n \rightarrow \infty} (A_n - B_n) = A - B$.

d) $\lim_{n \rightarrow \infty} (A_n \triangle B_n) = A \triangle B$.

Medidas de probabilidad

51. Determine completamente un espacio de probabilidad (Ω, \mathcal{F}, P) para el experimento aleatorio de
- lanzar una moneda equilibrada.
 - lanzar un dado equilibrado.
 - escoger al azar un número real dentro del intervalo unitario $[0, 1]$.
 - extraer dos bolas de una urna en donde hay dos bolas blancas y dos negras.
 - lanzar una moneda honesta hasta obtener las dos caras.

52. Sea $\{x_n : n \in \mathbb{N}\}$ una sucesión de números reales. Sea $\{a_n : n \in \mathbb{N}\}$ otra sucesión de números reales no negativos tal que $\sum_{n=1}^{\infty} a_n = 1$. Demuestre que la función $P : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ definida de la siguiente forma es una medida de probabilidad.

$$P(A) = \sum_{n=1}^{\infty} a_n \cdot \mathbf{1}_{\{n : x_n \in A\}}(n).$$

53. Sean P y Q dos medidas de probabilidad definidas sobre una misma σ -álgebra. Demuestre que $\alpha P + (1 - \alpha)Q$ es una medida de probabilidad para cada α en $[0, 1]$.
54. Sea P una medida de probabilidad. Determine si las siguientes funciones también son medidas de probabilidad: a) $1 - P$. b) $(1 + P)/2$. c) P^2 . d) $|P|$.
55. Considere el espacio medible $(\mathbb{N}, 2^{\mathbb{N}})$. Demuestre en cada caso que P es una medida de probabilidad. Para cada $A \in 2^{\mathbb{N}}$ defina:

$$a) P(A) = \sum_{n \in A} 2/3^n.$$

$$b) P(A) = \sum_{n \in A} 1/2^n.$$

56. Sea $\Omega = \{1, 2, \dots, n\}$, y considere el espacio medible $(\Omega, 2^{\Omega})$. Investigue en cada caso si P es una medida de probabilidad. Para cada $A \in 2^{\Omega}$ defina:

$$a) P(A) = \sum_{k \in A} \frac{2k}{n(n+1)}.$$

$$b) P(A) = \prod_{k \in A} \left(1 - \frac{1}{k}\right).$$

57. Considere el espacio medible $((0, 1), \mathcal{B}(0, 1))$. Demuestre en cada caso que P es una medida de probabilidad. Para cada $A \in \mathcal{B}(0, 1)$ defina:

$$a) P(A) = \int_A 2x \, dx.$$

$$b) P(A) = \int_A \frac{3}{2} \sqrt{x} \, dx.$$

58. *Probabilidad condicional.* Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sea B un evento con probabilidad estrictamente positiva. Demuestre que la probabilidad condicional definida para cada A en \mathcal{F} como sigue:

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

es una medida de probabilidad. En consecuencia toda propiedad válida para $P(\cdot)$ es también válida para $P(\cdot|B)$.

59. Sea P una medida de probabilidad, y sean $P_1(\cdot) = P(\cdot|B)$ y $P_2(\cdot) = P(\cdot|C)$. Demuestre que $P_2(A) = P(A|B \cap C)$.
60. Sea P una medida de probabilidad. Demuestre que la colección $\{A \in \mathcal{F} : P(A) = 0 \text{ ó } P(A) = 1\}$ es una sub σ -álgebra de \mathcal{F} .

Propiedades elementales

61. Demuestre que $P(\emptyset) = 0$, sin usar $P(\Omega) = 1$.
62. Demuestre que $P(A \cap B) - P(A)P(B) = P(A^c)P(B) - P(A^c \cap B)$.
63. Demuestre que

$$P(A \cap B) \leq \min\{P(A), P(B)\} \leq P(A) \leq \max\{P(A), P(B)\} \leq P(A \cup B).$$

64. Demuestre que

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C). \end{aligned}$$

65. Demuestre que

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \\ &\quad - \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n). \end{aligned}$$

66. Demuestre que

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cup A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cup A_j \cup A_k) \\ &\quad - \dots + (-1)^{n+1} P(A_1 \cup \dots \cup A_n). \end{aligned}$$

67. Demuestre que $P\left(\bigcap_{k=1}^n A_k\right) \geq 1 - \sum_{k=1}^n P(A_k^c)$.

68. Demuestre que

$$0 \leq P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B) \leq 2.$$

69. Diga falso o verdadero. Demuestre en cada caso.

a) $P(B - A) = P(B) - P(A)$.

b) $P(A \cup B) = P(A - B) + P(B - A)$.

c) $P(A) > 0 \Rightarrow P(A \cup B) > 0$.

d) $P(A) > 0 \Rightarrow P(A \cap B) > 0$.

e) $P(A) < 1 \Rightarrow P(A \cup B) < 1$.

f) $P(A) < 1 \Rightarrow P(A \cap B) < 1$.

g) $P(A) = 0 \Rightarrow P(A \cup B) = 0$.

h) $P(A) = 0 \Rightarrow P(A \cap B) = 0$.

i) $P(A \cup B) = 0 \Rightarrow P(A) = 0$.

j) $P(A \cap B) = 0 \Rightarrow P(A) = 0$.

k) $P(A) = 1 \Rightarrow P(A \cup B) = 1$.

l) $P(A) = 1 \Rightarrow P(A \cap B) = 1$.

m) $P(A \cup B) = 1 \Rightarrow P(A) = 1$.

n) $P(A \cap B) = 1 \Rightarrow P(A) = 1$.

70. Diga falso o verdadero. Demuestre en cada caso.

a) $P(A \cap B) \leq P(A)P(B)$.

b) $P(A|B) < P(A)$.

c) $P(A|B) > P(A) \Rightarrow P(B|A) > P(B)$.

71. *Teorema de probabilidad total.* Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sea $\{A_1, A_2, \dots\}$ una partición de Ω tal que cada elemento de la partición es un evento con probabilidad estrictamente positiva. Demuestre que para cualquier evento B ,

$$P(B) = \sum_{n=1}^{\infty} P(B|A_n)P(A_n).$$

72. Se lanza una moneda tantas veces como indica un dado previamente lanzado. Calcule la probabilidad de que:

a) se obtengan ambas caras de la moneda igual número de veces.

b) se obtenga una misma cara siempre.

73. *Teorema de Bayes.* Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sea $\{A_1, A_2, \dots\}$ una partición de Ω tal que cada elemento de la partición es un evento con probabilidad estrictamente positiva. Demuestre que para cualquier evento B tal que $P(B) > 0$, y para cualquier $m \geq 1$ fijo,

$$P(A_m | B) = \frac{P(B | A_m)P(A_m)}{\sum_{n=1}^{\infty} P(B | A_n)P(A_n)}.$$

74. *Regla del producto.* Demuestre que

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap \dots \cap A_{n-1}).$$

75. *Desigualdad de Bonferroni.* Demuestre que

$$P\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j).$$

76. *Desigualdad de Kounias.* Demuestre que

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \min_j \left\{ \sum_{i=1}^n P(A_i) - \sum_{\substack{i=1 \\ i \neq j}}^n P(A_i \cap A_j) \right\}.$$

Continuidad

77. Se lanza una moneda honesta una infinidad de veces. Demuestre que la probabilidad de que eventualmente cada una de las dos caras aparezca es uno.
78. Se lanza un dado equilibrado una infinidad de veces. Demuestre que la probabilidad de que eventualmente cada una de las seis caras aparezca es uno.
79. Sea A un evento con probabilidad estrictamente positiva. Demuestre que si se efectúa una infinidad de ensayos independientes del experimento aleatorio, la probabilidad de que nunca ocurra el evento A es cero.

Independencia de eventos

80. Demuestre que los eventos A y B son independientes si, y sólo si,
- A y B^c lo son.
 - A^c y B lo son.
 - A^c y B^c lo son.
81. Demuestre que los eventos A_1, \dots, A_n son independientes si, y sólo si, A_1^c, \dots, A_n^c también lo son.
82. Encuentre tres eventos A, B, C tales que

- a) sean independientes dos a dos pero no independientes tres a tres.
 b) sean independientes tres a tres pero no independientes dos a dos.
83. Demuestre que un evento A es independiente consigo mismo si, y sólo si, su probabilidad es cero o uno.
84. Demuestre que un evento que tiene probabilidad cero o uno, es independiente de cualquier otro evento.
85. Mediante un contraejemplo demuestre que
- a) En general, A, B independientes $\not\Rightarrow A, B$ ajenos.
 b) En general, A, B ajenos $\not\Rightarrow A, B$ independientes.
86. Diga falso o verdadero. Demuestre o proporcione un contraejemplo.
- a) $A \perp A$.
 b) $A \perp A^c$.
 c) $A \perp \emptyset$.
 d) $A \perp \Omega$.
87. ¿Es la independencia de dos eventos una relación de equivalencia?
88. Sean A_1, \dots, A_n independientes. Demuestre que

$$P\left(\bigcup_{k=1}^n A_k\right) = 1 - \prod_{k=1}^n [1 - P(A_k)].$$

89. Sea A_1, A_2, \dots una sucesión infinita de eventos. Defina

$$B_n = \bigcup_{k=n}^{\infty} A_k \quad \text{y} \quad C_n = \bigcap_{k=n}^{\infty} A_k.$$

Demuestre que si B_n y C_n son independientes para cada n entonces $\limsup A_n$ y $\liminf A_n$ también son independientes. En particular, cuando $\lim_{n \rightarrow \infty} A_n = A$, entonces A tiene probabilidad cero o uno.

90. Sean A y B independientes. Demuestre que $\sigma\{A\}$ y $\sigma\{B\}$ son independientes.

Lema de Borel-Cantelli

91. Se lanza una moneda honesta una infinidad de veces. Use el lema de Borel-Cantelli para demostrar que la probabilidad de que cada cara aparezca una infinidad de veces es uno.
92. Se lanza un dado equilibrado una infinidad de veces. Demuestre que con probabilidad uno cada una de las seis caras aparece una infinidad de veces.
93. Sea A un evento con probabilidad positiva. Use el lema de Borel-Cantelli para demostrar que si se efectúa una infinidad de ensayos independientes del experimento aleatorio, la probabilidad de que ocurra una infinidad de veces el evento A , es uno.

Capítulo 2

Variables aleatorias

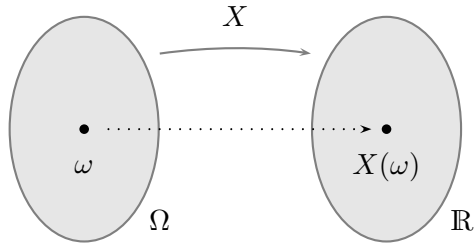
En este capítulo se estudian los conceptos de variable aleatoria, función de distribución, función de densidad y esperanza. Se estudian también algunas distribuciones de probabilidad de variables aleatorias discretas y continuas particulares. A partir de ahora y en el resto del curso consideraremos como elemento base un espacio de probabilidad (Ω, \mathcal{F}, P) .

2.1. Variables aleatorias

El concepto de variable aleatoria es fundamental en la teoría de la probabilidad. Una vez que enunciemos su definición, el término aparecerá con mucha frecuencia a lo largo del curso.

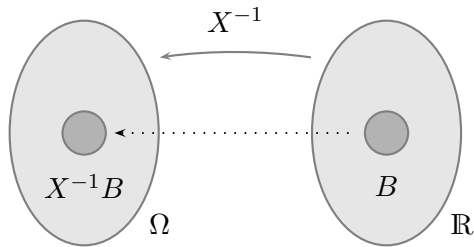
Definición (Variable aleatoria). Una variable aleatoria es una función $X : \Omega \rightarrow \mathbb{R}$ tal que para cualquier conjunto Boreliano B , se cumple que el conjunto $X^{-1}B$ es un elemento de \mathcal{F} .

Gráficamente una variable aleatoria puede representarse de la siguiente forma:



Una variable aleatoria es una función de Ω en \mathbb{R} .

Esto es, una variable aleatoria (v.a.) es una función de Ω en \mathbb{R} tal que la imagen inversa de cualquier conjunto Boreliano es un elemento de la σ -álgebra del espacio de probabilidad. Esta condición se conoce como *medibilidad* en teoría de la medida, y se dice entonces que dicha función es *medible* respecto de las σ -álgebras \mathcal{F} y $\mathcal{B}(\mathbb{R})$. En un apéndice al final del texto aparece una sección que contiene una discusión breve del concepto de imagen inversa de una función, que para el caso de variables aleatorias puede ilustrarse gráficamente como indica la siguiente figura:



La imagen inversa de un conjunto de Borel.

Explicamos a continuación la razón técnica por la cual se le pide a una función $X : \Omega \rightarrow \mathbb{R}$ que cumpla la condición de medibilidad. Recordemos que P es una medida de probabilidad definida sobre el espacio medible (Ω, \mathcal{F}) . Si X es una variable aleatoria, entonces podemos trasladar la medida de probabilidad P al espacio medible $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ del siguiente modo: Si B es un conjunto Boreliano definimos $P_X(B) = P(X^{-1}B)$, lo cual es posible pues el conjunto $X^{-1}B$ es un elemento de \mathcal{F} , dominio de definición de P . La función $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ resulta ser una medida de probabilidad, y se le llama por tanto la *medida de probabilidad inducida* por la variable aleatoria X . De este modo se construye el espacio de probabilidad $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.

Si B es un conjunto Boreliano, se usan los símbolos $X^{-1}B$ y $(X \in B)$ para denotar el conjunto $\{\omega \in \Omega : X(\omega) \in B\}$. Por ejemplo, el conjunto $\{\omega \in \Omega : X(\omega) \in [0, \infty)\}$ puede ser denotado por $X^{-1}[0, \infty)$ o $(X \in [0, \infty))$, o simplemente por $(X \geq 0)$, incluyendo los paréntesis. Veamos otro ejemplo, si (a, b) es un intervalo de la recta

real, se puede usar el símbolo $X^{-1}(a, b)$, o $(X \in (a, b))$, o bien $(a < X < b)$ para denotar el conjunto $\{\omega \in \Omega : X(\omega) \in (a, b)\}$. Para hacer la escritura más corta, a menudo se omite el argumento ω de una v.a. X y se omite también el término *variable aleatoria* para X asumiendo, en la mayoría de las veces, que lo es.

Para comprobar que una función $X : \Omega \rightarrow \mathbb{R}$ es realmente una variable aleatoria, la definición requiere verificar la condición $X^{-1}B \in \mathcal{F}$ para cualquier conjunto Boreliano B . En muy pocos casos tal condición puede comprobarse de manera tan general. La siguiente proposición establece que no es necesario demostrar la condición de medibilidad para cualquier conjunto Boreliano B , sino que es suficiente tomar intervalos de la forma $(-\infty, x]$, para cada x en \mathbb{R} . Este resultado, como uno puede imaginar, es de suma utilidad y lo usaremos con frecuencia en el resto del capítulo.

Proposición. Una función $X : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria si, y sólo si, para cada x en \mathbb{R} ,

$$X^{-1}(-\infty, x] \in \mathcal{F}.$$

Demostración.

(\Rightarrow) Si X es variable aleatoria, entonces claramente se cumple que para cualquier número real x el conjunto $X^{-1}(-\infty, x]$ es un elemento de \mathcal{F} .

(\Leftarrow) Ahora suponga que para cada real x , el conjunto $X^{-1}(-\infty, x]$ es un elemento de \mathcal{F} . Sean \mathcal{B} y \mathcal{C} las colecciones

$$\begin{aligned} \mathcal{B} &= \{B \in \mathcal{B}(\mathbb{R}) : X^{-1}B \in \mathcal{F}\}, \\ \text{y } \mathcal{C} &= \{(-\infty, x] : x \in \mathbb{R}\}. \end{aligned}$$

Entonces claramente $\mathcal{C} \subseteq \mathcal{B} \subseteq \mathcal{B}(\mathbb{R})$. La primera contención es por hipótesis, y la segunda es por definición de la colección \mathcal{B} . Suponga por un momento que \mathcal{B} es una σ -álgebra de subconjuntos de \mathbb{R} . Entonces \mathcal{B} es una σ -álgebra que contiene a \mathcal{C} . Por lo tanto $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}) \subseteq \mathcal{B}$. Esto implica que $\mathcal{B} = \mathcal{B}(\mathbb{R})$, y entonces X es variable aleatoria. Resta entonces hacer ver que \mathcal{B} es efectivamente una σ -álgebra.

- a) Primeramente tenemos que $\mathbb{R} \in \mathcal{B}$, pues $\mathbb{R} \in \mathcal{B}(\mathbb{R})$ y $X^{-1}\mathbb{R} = \Omega \in \mathcal{F}$.
- b) Sea $B \in \mathcal{B}$. Entonces $B \in \mathcal{B}(\mathbb{R})$ y $X^{-1}B \in \mathcal{F}$. Por lo tanto $B^c \in \mathcal{B}(\mathbb{R})$ y $X^{-1}B^c = (X^{-1}B)^c \in \mathcal{F}$. Es decir, $B^c \in \mathcal{B}$.
- c) Sea B_1, B_2, \dots una sucesión en \mathcal{B} . Es decir, para cada número natural n , $B_n \in \mathcal{B}(\mathbb{R})$ y $X^{-1}B_n \in \mathcal{F}$. Entonces $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}(\mathbb{R})$ y $\bigcup_{n=1}^{\infty} X^{-1}B_n = X^{-1} \bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$. Es decir, $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$.

□

Además de la condición anterior para demostrar que una función es variable aleatoria, existen otras condiciones igualmente equivalentes y útiles. Por ejemplo, X es variable aleatoria si para cada x en \mathbb{R} , $X^{-1}(-\infty, x) \in \mathcal{F}$, o $X^{-1}(x, \infty) \in \mathcal{F}$, o $X^{-1}[x, \infty) \in \mathcal{F}$. Cualquiera de estas condiciones es necesaria y suficiente para que X sea variable aleatoria. También la condición $X^{-1}(a, b) \in \mathcal{F}$ para cualquier intervalo (a, b) de \mathbb{R} es necesaria y suficiente para que X sea variable aleatoria. La demostración de todas estas aseveraciones es completamente análoga al caso demostrado arriba y se pide desarrollar los detalles en la sección de ejercicios.

Considere ahora los espacios medibles (Ω, \mathcal{F}) y $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si X es una función de Ω en \mathbb{R} entonces se denota por $\sigma(X)$ a la mínima σ -álgebra de subconjuntos de Ω respecto de la cual X es variable aleatoria. Es decir,

$$\sigma(X) = \{X^{-1}B : B \in \mathcal{B}(\mathbb{R})\}.$$

Es sencillo probar que tal colección de imágenes inversas es efectivamente una σ -álgebra, y claramente X es variable aleatoria si y sólo si $\sigma(X) \subseteq \mathcal{F}$. En particular, una función $g : \mathbb{R} \rightarrow \mathbb{R}$ es *Borel medible* si $g^{-1}B \in \mathcal{B}(\mathbb{R})$, para cada B en $\mathcal{B}(\mathbb{R})$.

A continuación se demuestra que algunas operaciones básicas entre variables aleatorias producen nuevas variables aleatorias. Suponga entonces que (Ω, \mathcal{F}, P) es un espacio de probabilidad dado. Todas las variables aleatorias que se consideran a continuación están definidas sobre este mismo espacio de probabilidad.

Proposición. La función constante $X = c$ es una v.a.

Demostración. Sea B un elemento cualquiera de $\mathcal{B}(\mathbb{R})$. Para la función constante $X = c$ se tiene que $X^{-1}B = \Omega$ si $c \in B$, y $X^{-1}B = \emptyset$ si $c \notin B$. En ambos casos el conjunto $X^{-1}B$ es un elemento de \mathcal{F} , por lo tanto X es v.a. □

Proposición. Si X es v.a. y c es una constante, entonces cX es v.a.

Demostración. Comprobaremos que para cada número real x , la imagen inversa del conjunto $(-\infty, x]$, bajo la función cX , es un elemento de \mathcal{F} . Tenemos tres casos: Si $c > 0$, entonces el conjunto $(cX \leq x) = (X \leq x/c)$ es un elemento de \mathcal{F} , pues X es v.a. Si $c < 0$ entonces nuevamente el conjunto $(cX \leq x) = (X \geq x/c)$ es un elemento de \mathcal{F} pues X es v.a. Finalmente si $c = 0$ entonces es claro que $cX = 0$ es v.a. por la proposición anterior. □

Proposición. Si X y Y son v.a.s, entonces $X + Y$ es v.a.

Demostración. Probaremos que para cada número real x , el conjunto $(X + Y > x)$ es un elemento de \mathcal{F} . Para ello usaremos la igualdad

$$(X + Y > x) = \bigcup_{r \in \mathbb{Q}} (X > r) \cap (Y > x - r). \quad (2.1)$$

Es claro que a partir de esta igualdad se concluye que el conjunto $(X + Y > x)$ es un elemento de \mathcal{F} , pues tanto X como Y son variables aleatorias, y la operación de unión involucrada es numerable. Resta entonces demostrar (2.1).

- (\subseteq) Sea ω en Ω tal que $X(\omega) + Y(\omega) > x$. Entonces $X(\omega) > x - Y(\omega)$. Como los números racionales son un conjunto denso en \mathbb{R} , tenemos que existe un número racional r tal que $X(\omega) > r > x - Y(\omega)$. Por lo tanto $X(\omega) > r$ y $Y(\omega) > x - r$. De aquí se desprende que ω es un elemento del lado derecho.
- (\supseteq) Sea ahora ω un elemento de $\bigcup_{r \in \mathbb{Q}} (X > r) \cap (Y > x - r)$. Entonces existe un número racional r_0 tal que $X(\omega) > r_0$ y $Y(\omega) > x - r_0$. Sumando obtenemos $X(\omega) + Y(\omega) > x$, y por lo tanto ω es un elemento del lado izquierdo.

□

Proposición. Si X y Y son v.a.s, entonces XY es v.a.

Demostración. Suponga primero el caso particular $X = Y$. Entonces necesitamos probar que para todo número real x , el conjunto $(X^2 \leq x)$ es un elemento de \mathcal{F} . Pero esto es cierto pues $(X^2 \leq x) = \emptyset$ si $x < 0$, y $(X^2 \leq x) = (-\sqrt{x} \leq X \leq \sqrt{x})$ si $x \geq 0$. En ambos casos, $(X^2)^{-1}(-\infty, x]$ es un elemento de \mathcal{F} . Para el caso general, $X \neq Y$, usamos la fórmula

$$XY = \frac{1}{4} [(X + Y)^2 - (X - Y)^2].$$

Por lo demostrado antes, el producto XY es efectivamente una v.a. □

Como consecuencia de la proposición anterior se cumple que si multiplicamos X por sí misma n veces, entonces X^n es variable aleatoria. Por lo tanto toda función polinomial de una variable aleatoria es también variable aleatoria.

Proposición. Sean X y Y v.a.s con $Y \neq 0$. Entonces X/Y es v.a.

Demostración. Como el producto de v.a.s es nuevamente una v.a., es suficiente demostrar que $1/Y$ es v.a. Para cualquier número real $y > 0$ tenemos que

$$\begin{aligned} \left(\frac{1}{Y} \leq y\right) &= \left(\frac{1}{Y} \leq y, Y > 0\right) \cup \left(\frac{1}{Y} \leq y, Y < 0\right) \\ &= \left(Y \geq \frac{1}{y}, Y > 0\right) \cup \left(Y \leq \frac{1}{y}, Y < 0\right) \\ &= \left(Y \geq \frac{1}{y}\right) \cup (Y < 0), \end{aligned}$$

que es un elemento de \mathcal{F} puesto que Y es v.a. Por otro lado, si $y < 0$ tenemos que

$$\begin{aligned} \left(\frac{1}{Y} \leq y\right) &= \left(\frac{1}{Y} \leq y, Y > 0\right) \cup \left(\frac{1}{Y} \leq y, Y < 0\right) \\ &= \left(Y \leq \frac{1}{y}, Y > 0\right) \cup \left(Y \geq \frac{1}{y}, Y < 0\right) \\ &= \emptyset \cup \left(Y \geq \frac{1}{y}, Y < 0\right) \\ &= \left(\frac{1}{y} \leq Y < 0\right). \end{aligned}$$

Nuevamente vemos que este conjunto es un elemento de \mathcal{F} , puesto que Y es v.a. Finalmente cuando $y = 0$ obtenemos una vez mas un elemento de \mathcal{F} pues

$$\begin{aligned} \left(\frac{1}{Y} \leq 0\right) &= \left(\frac{1}{Y} \leq 0, Y > 0\right) \cup \left(\frac{1}{Y} \leq 0, Y < 0\right) \\ &= \emptyset \cup (Y < 0) \\ &= (Y < 0). \end{aligned}$$

□

Proposición. Si X y Y son variables aleatorias, entonces $\max\{X, Y\}$ y $\min\{X, Y\}$ también lo son.

Demostración. Para cualquier número real x ,

$$(\max\{X, Y\} \leq x) = (X \leq x, Y \leq x) = (X \leq x) \cap (Y \leq x).$$

Análogamente,

$$(\min\{X, Y\} \geq x) = (X \geq x, Y \geq x) = (X \geq x) \cap (Y \geq x).$$

En ambos casos los conjuntos del lado derecho son elementos de \mathcal{F} .

□

Como consecuencia de la proposición anterior se obtiene que tanto $X^+ = \max\{0, X\}$ como $X^- = -\min\{0, X\}$ son variables aleatorias.

Proposición. Si X es v.a., entonces $|X|$ es v.a.

Demostración. Si $x \geq 0$ entonces $(|X| \leq x) = (-x \leq X \leq x)$, y si $x < 0$ entonces $(|X| \leq x) = \emptyset \in \mathcal{F}$, de modo que $|X|$ es v.a. Alternativamente se puede escribir $|X| = X^+ + X^-$, y por lo expuesto anteriormente concluir que $|X|$ es v.a. \square

Se muestra a continuación que en general el recíproco de la proposición anterior es falso, esto es, si $X : \Omega \rightarrow \mathbb{R}$ es una función tal que $|X|$ es v.a., entonces no necesariamente X es v.a.

Ejemplo Considere el espacio muestral $\Omega = \{-1, 0, 1\}$ junto con la σ -álgebra $\mathcal{F} = \{\emptyset, \{0\}, \{-1, 1\}, \Omega\}$. Sea $X : \Omega \rightarrow \mathbb{R}$ la función identidad $X(\omega) = \omega$. Entonces $|X|$ es v.a. pues para cualquier conjunto Boreliano B ,

$$|X|^{-1}B = \begin{cases} \Omega & \text{si } 0, 1 \in B, \\ \{-1, 1\} & \text{si } 0 \notin B \text{ y } 1 \in B, \\ \{0\} & \text{si } 0 \in B \text{ y } 1 \notin B, \\ \emptyset & \text{si } 0, 1 \notin B. \end{cases}$$

Es decir, $|X|^{-1}B$ es un elemento de \mathcal{F} . Sin embargo X no es variable aleatoria pues $X^{-1}\{-1\} = \{-1\}$ no es un elemento de \mathcal{F} . \circ

Ahora consideraremos algunas operaciones límite en sucesiones infinitas de variables aleatorias.

Proposición. Sea $\{X_n : n \in \mathbb{N}\}$ una sucesión de v.a.s tales que para cada ω en Ω , $\sup_{n \geq 0} \{X_n(\omega)\}$ e $\inf_{n \geq 0} \{X_n(\omega)\}$ son finitos. Entonces $\sup_{n \geq 0} \{X_n\}$ e $\inf_{n \geq 0} \{X_n\}$ son v.a.s

Demostración. Este resultado se sigue directamente de las siguientes igualdades. Para cualquier número real x ,

$$\begin{aligned} \left(\sup_{n \geq 0} X_n \leq x \right) &= \bigcap_{n=1}^{\infty} (X_n \leq x) \in \mathcal{F}, \\ \text{e } \left(\inf_{n \geq 0} X_n \geq x \right) &= \bigcap_{n=1}^{\infty} (X_n \geq x) \in \mathcal{F}. \end{aligned}$$

\square

Proposición. Sea $\{X_n : n \in \mathbb{N}\}$ una sucesión de v.a.s tales que para cada ω en Ω , $\limsup_{n \rightarrow \infty} X_n(\omega)$ y $\liminf_{n \rightarrow \infty} X_n(\omega)$ son finitos. Entonces $\limsup_{n \rightarrow \infty} X_n$ y $\liminf_{n \rightarrow \infty} X_n$ son v.a.s.

Demostración. Esto es consecuencia de la proposición anterior pues

$$\begin{aligned} \limsup_{n \rightarrow \infty} X_n &= \inf_k (\sup_{n \geq k} X_n), \\ \text{y } \liminf_{n \rightarrow \infty} X_n &= \sup_k (\inf_{n \geq k} X_n). \end{aligned}$$

□

Proposición. Sea $\{X_n : n \in \mathbb{N}\}$ es una sucesión de v.a.s tales que $\lim_{n \rightarrow \infty} X_n(\omega)$ existe y es finito para cada $\omega \in \Omega$. Entonces $\lim_{n \rightarrow \infty} X_n$ es v.a.

Demostración. Como $\lim_{n \rightarrow \infty} X_n$ existe, los límites $\limsup_{n \rightarrow \infty} X_n$ y $\liminf_{n \rightarrow \infty} X_n$ coinciden. Entonces por lo demostrado antes, $\lim_{n \rightarrow \infty} X_n$ es variable aleatoria. □

2.2. Función de distribución

Toda variable aleatoria tiene asociada una función llamada *función de distribución*. En esta sección se define este importante concepto y se demuestran algunas de sus propiedades.

Definición (Función de distribución). La función de distribución de una variable aleatoria X es la función $F(x) : \mathbb{R} \rightarrow [0, 1]$, definida como sigue

$$F(x) = P(X \leq x).$$

Cuando sea necesario especificar la variable aleatoria en cuestión se escribe $F_X(x)$, pero en general se omite el subíndice X cuando no haya posibilidad de confusión. El argumento de la función es la letra minúscula x que puede tomar cualquier valor real. Por razones obvias a esta función se le conoce también con el nombre de *función de acumulación de probabilidad* o *función de probabilidad acumulada*. Observe que la función de distribución de una variable aleatoria está definida sobre la totalidad del

conjunto de números reales, y siendo una probabilidad, toma valores en el intervalo $[0, 1]$. La función de distribución es importante pues, como se ilustrará más adelante, contiene ella toda la información de la variable aleatoria y la correspondiente medida de probabilidad. Veamos ahora algunas propiedades básicas de esta función, en una de estas propiedades aparece la expresión $F(x+)$, que significa el límite por la derecha de la función F en el punto x .

Proposición. Sea $F(x)$ la función de distribución de una variable aleatoria. Entonces

1. $\lim_{x \rightarrow +\infty} F(x) = 1$.
2. $\lim_{x \rightarrow -\infty} F(x) = 0$.
3. Si $x_1 \leq x_2$, entonces $F(x_1) \leq F(x_2)$.
4. $F(x)$ es continua por la derecha, es decir, $F(x+) = F(x)$.

Demostración. (1) Sea $\{x_n : n \in \mathbb{N}\}$ una sucesión cualquiera de números reales creciente a infinito, y sean los eventos $A_n = (X \leq x_n)$. Entonces $\{A_n : n \in \mathbb{N}\}$ es una sucesión de eventos creciente cuyo límite es Ω . Por la propiedad de continuidad

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(A_n) = P(\Omega) = 1.$$

Dado que \mathbb{R} es un espacio métrico, lo anterior implica que $F(x)$ converge a uno cuando x tiende a infinito. (2) Ahora sea $\{x_n : n \in \mathbb{N}\}$ una sucesión cualquiera de números reales decreciente a menos infinito, y sean los eventos $A_n = (X \leq x_n)$. Entonces $\{A_n : n \in \mathbb{N}\}$ es una sucesión de eventos decreciente al conjunto vacío. Por la propiedad de continuidad

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(A_n) = P(\emptyset) = 0.$$

Por lo tanto, $F(x)$ converge a cero cuando x tiende a menos infinito. (3) Para $x_1 \leq x_2$,

$$\begin{aligned} F(x_1) &\leq F(x_1) + P(x_1 < X \leq x_2) \\ &= P[(X \leq x_1) \cup (x_1 < X \leq x_2)] \\ &= P(X \leq x_2) \\ &= F(x_2). \end{aligned}$$

(4) Sea $\{x_n : n \in \mathbb{N}\}$ una sucesión cualquiera de números reales no negativos y decreciente a cero. Entonces

$$F(x + x_n) = F(x) + P(x < X \leq x + x_n),$$

en donde $A_n = (x < X \leq x + x_n)$ es una sucesión de eventos decreciente al conjunto vacío. Por lo tanto $\lim_{n \rightarrow \infty} F(x + x_n) = F(x)$. Es decir $F(x+) = F(x)$. \square

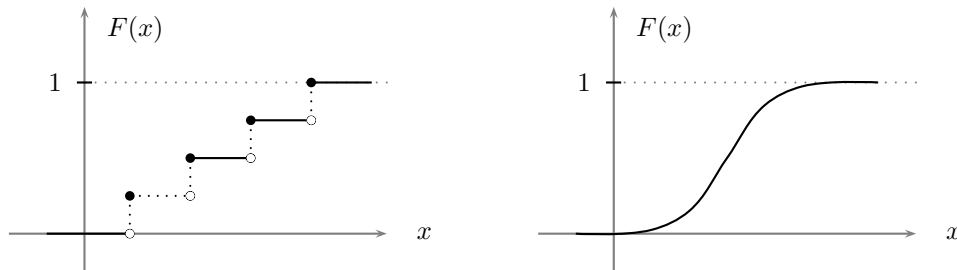
El recíproco de la proposición anterior es válido y justifica la importancia de la función de distribución. Se enuncia a continuación este interesante resultado cuya demostración omitiremos y puede encontrarse por ejemplo en [12].

Proposición. Sea $F : \mathbb{R} \rightarrow [0, 1]$ una función que satisface las cuatro propiedades de la proposición anterior. Entonces existe un espacio de probabilidad y una variable aleatoria cuya función de distribución es F .

Como consecuencia tenemos la siguiente definición general, no haciendo referencia a variables aleatorias ni a espacios de probabilidad particulares.

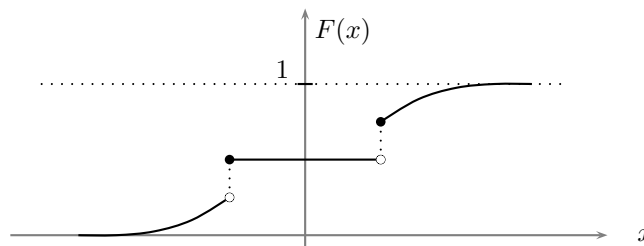
Definición (Función de distribución). Una función $F(x) : \mathbb{R} \rightarrow [0, 1]$ es llamada función de distribución si cumple las cuatro propiedades anteriores.

A continuación se presentan algunos ejemplos gráficos de funciones de distribución.



Ejemplos gráficos de funciones de distribución.

También pueden presentarse situaciones como la que se muestra a continuación.



Otro ejemplo de función de distribución.

Se demuestran ahora algunas otras propiedades que establecen la forma de calcular probabilidades usando la función de distribución. La expresión $F(x-)$ significa el límite por la izquierda de la función F en el punto x .

Proposición. Para cualquier número x y para cualesquiera números reales $a \leq b$,

1. $P(X < x) = F(x-)$.
2. $P(X = x) = F(x) - F(x-)$.
3. $P(X \in (a, b]) = F(b) - F(a)$.
4. $P(X \in [a, b]) = F(b) - F(a-)$.
5. $P(X \in (a, b)) = F(b-) - F(a)$.
6. $P(X \in [a, b)) = F(b-) - F(a-)$.

Demostración. (1) Sea $\{x_n : n \in \mathbb{N}\}$ una sucesión cualquiera de números reales no negativos y decreciente a cero. Sea A_n el evento $(X \leq a - x_n)$. Entonces $\{A_n : n \in \mathbb{N}\}$ es una sucesión de eventos decreciente al evento $(X < a)$. Por la propiedad de continuidad

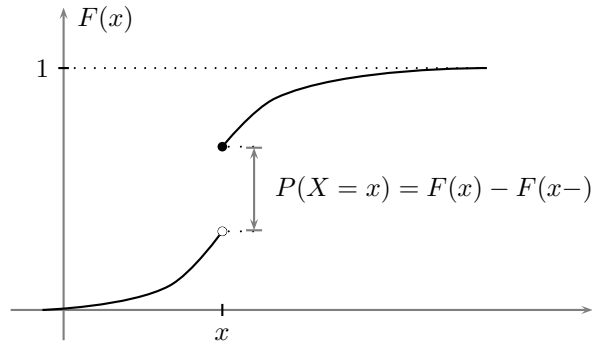
$$\begin{aligned} P(X < a) &= \lim_{n \rightarrow \infty} P(A_n) \\ &= \lim_{n \rightarrow \infty} F(a - x_n) \\ &= F(a-). \end{aligned}$$

Para (2) simplemente se escribe

$$\begin{aligned} P(X = x) &= P(X \leq x) - P(X < x) \\ &= F(x) - F(x-). \end{aligned}$$

Las igualdades (3),(4),(5) y (6) se siguen directamente de (1) y (2). □

Observe que como $F(x)$ es una función no decreciente y continua por la derecha, la probabilidad $P(X = x) = F(x) - F(x-)$ representa el tamaño del salto o discontinuidad de la función de distribución en el punto x , como se muestra en la siguiente figura.



La probabilidad $P(X = x)$ es el tamaño del salto de la función F en el punto x .

En consecuencia, cuando $F(x)$ es una función continua y para $a < b$,

$$\begin{aligned}
 F(b) - F(a) &= P(X \in (a, b]) \\
 &= P(X \in [a, b]) \\
 &= P(X \in (a, b)) \\
 &= P(X \in [a, b)).
 \end{aligned}$$

Es decir, cuando $F(x)$ es una función continua, incluir o excluir los extremos de un intervalo no afecta el cálculo de la probabilidad de dicho intervalo. Por lo tanto, para cualquier número x , $P(X = x) = 0$. Finalizamos esta sección con un resultado interesante cuya prueba es sorprendentemente simple.

Proposición. Toda función de distribución tiene a lo sumo un número numerable de discontinuidades.

Demostración. Sea D el conjunto de puntos de discontinuidad de una función de distribución $F(x)$. Para cada número natural n defina los subconjuntos

$$D_n = \left\{ x \in D : \frac{1}{n+1} < F(x) - F(x-) \leq \frac{1}{n} \right\}.$$

Cada conjunto D_n tiene a lo sumo n elementos. Como $D = \bigcup_{n=1}^{\infty} D_n$, se concluye que D es numerable. \square

2.3. Tipos de variables aleatorias

Las variables aleatorias se clasifican en varios tipos dependiendo del conjunto de valores que éstas toman. Al menos existen dos tipos: discretas y continuas. La definición es la siguiente.

Definición (Variable aleatoria discreta). La variable aleatoria X se llama discreta si su correspondiente función de distribución $F(x)$ es una función constante por pedazos. Sean x_1, x_2, \dots los puntos de discontinuidad de $F(x)$. En cada uno de estos puntos el tamaño de la discontinuidad es $P(X = x_i) = F(x_i) - F(x_i-) > 0$. A la función $f(x)$ que indica estos incrementos se le llama función de probabilidad de X , y se define como sigue

$$f(x) = \begin{cases} P(X = x) & \text{si } x = x_1, x_2, \dots \\ 0 & \text{otro caso.} \end{cases} \quad (2.2)$$

En este caso discreto la función $f(x)$ siempre existe, y se le llama también *función de masa de probabilidad* o simplemente *función de probabilidad* de la variable aleatoria X . Aunque también se acostumbra usar el término: función de densidad, como una analogía con el caso de variables aleatorias continuas mencionado más adelante. Cuando sea necesario especificarlo se escribe $f_X(x)$ en lugar de $f(x)$. Observe que $f(x)$ es una función no negativa que suma uno en el sentido que $\sum_i f(x_i) = 1$. Recíprocamente, toda función de la forma (2.2) que cumpla estas dos propiedades se le llama función de probabilidad, sin que haya necesariamente una variable aleatoria de por medio. Es posible reconstruir la función de distribución a partir de la función de probabilidad mediante la relación

$$F(x) = \sum_{x_i \leq x} f(x_i).$$

Definición (Variable aleatoria continua). La variable aleatoria X se llama continua si su correspondiente función de distribución $F(x)$ es una función continua. Cuando existe una función no negativa e integrable f tal que para cualquier valor de x ,

$$F(x) = \int_{-\infty}^x f(u) du, \quad (2.3)$$

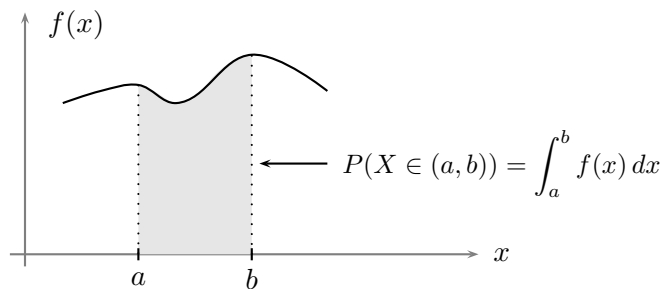
entonces se dice que X es absolutamente continua. En tal caso a la función $f(x)$ se le llama función de densidad de X .

Pueden construirse ejemplos de variables aleatorias continuas que no tienen función de densidad, es decir, que no existe una función f no negativa e integrable que cumpla (2.3) para cualquier número real x . En tales situaciones se dice que la distribución es *singular*.

Aún cuando exista una función no negativa e integrable f que cumpla (2.3), ésta puede no ser única, pues basta modificarla en un punto para que sea ligeramente distinta pero aún así seguir cumpliendo (2.3). A pesar de ello, nos referiremos a la función de densidad como si ésta fuera única, y ello se justifica por el hecho de

que las probabilidades son las mismas, ya sea usando una función de densidad o modificaciones de ella que cumpla (2.3).

Es claro que la función de densidad de una variable aleatoria absolutamente continua es no negativa y su integral sobre toda la recta real es uno. Recíprocamente, toda función $f(x)$ no negativa que integre uno en \mathbb{R} se llama *función de densidad*. Si X es absolutamente continua con función de distribución $F(x)$ y función de densidad continua $f(x)$, entonces el teorema fundamental del cálculo establece que, a partir de (2.3), $F'(x) = f(x)$. Además, la probabilidad de que X tome un valor en el intervalo (a, b) es el área bajo la función de densidad sobre dicho intervalo. Esto se ilustra en la siguiente figura, la probabilidad es la misma si se incluyen o excluyen los extremos del intervalo.



La probabilidad como un área.

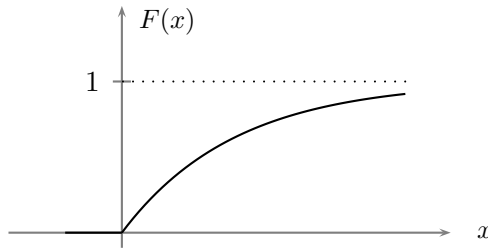
Definición (Variable aleatoria mixta). Una variable aleatoria que no es discreta ni continua se llama variable aleatoria mixta.

Un ejemplo de este tipo de variables se presenta a continuación.

Ejemplo (Una variable aleatoria que no es discreta ni continua). Sea X una variable aleatoria con función de distribución

$$F(x) = \begin{cases} 1 - e^{-x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0, \end{cases}$$

cuya gráfica es

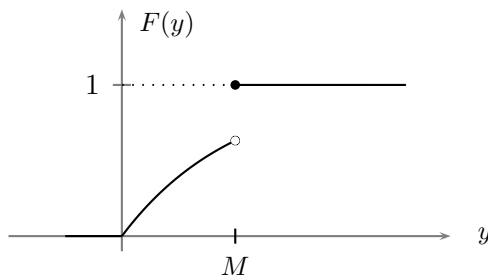


Función de distribución de X .

Como $F(x)$ es continua, entonces X es una variable aleatoria continua. Sea $Y = X \wedge M$, con $M > 0$ constante. Observe que Y está acotada superiormente por la constante M . La función de distribución de Y es

$$F(y) = \begin{cases} 1 & \text{si } y \geq M, \\ 1 - e^{-x} & \text{si } 0 < y < M, \\ 0 & \text{si } x \leq 0, \end{cases}$$

con gráfica



Función de distribución de Y .

Esta función no es constante por pedazos pues es creciente en el intervalo $(0, M)$, y tampoco es continua pues tiene una discontinuidad en $y = M$. Por lo tanto Y es una variable aleatoria que no es discreta ni continua. \circ

Proposición. Toda función de distribución $F(x)$ se puede escribir como una combinación lineal convexa de una función de distribución discreta y otra continua, es decir, admite la siguiente representación:

$$F(x) = \alpha F^d(x) + (1 - \alpha) F^c(x),$$

en donde $\alpha \in [0, 1]$, $F^d(x)$ es una función de distribución discreta, y $F^c(x)$ es una función de distribución continua.

Igualdad de variables aleatorias

Dos variables aleatorias X y Y son estrictamente iguales si para cada ω en Ω se cumple $X(\omega) = Y(\omega)$. Existen, sin embargo, otras formas más débiles de igualdad:

Definición (Igualdad de variables aleatorias). Se dice que dos variables aleatorias X y Y son

- a) iguales casi seguramente, y se escribe $X = Y$ c.s., o bien $X \stackrel{c.s.}{=} Y$, si se cumple que $P(X = Y) = 1$. Más generalmente, un evento ocurre casi seguramente si su probabilidad es uno.
- b) iguales en distribución, y se escribe $X \stackrel{d}{=} Y$, si sus correspondientes funciones de distribución coinciden, es decir, para cada x en \mathbb{R} , $F_X(x) = F_Y(x)$.

Es interesante observar que la igualdad casi segura es más fuerte que la igualdad en distribución, es decir, si X y Y son iguales casi seguramente, entonces son iguales en distribución. Sin embargo, si X y Y tienen la misma distribución, entonces no necesariamente son iguales casi seguramente, al respecto véase el Ejercicio 147. A menos que se indique lo contrario, cuando aparezca una expresión de igualdad entre variables aleatorias, se considera que la igualdad es válida en el sentido casi seguro.

2.4. Integral de Riemann-Stieltjes

En esta sección se define la integral de Riemann-Stieltjes, la cual tiene la forma

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

en donde las funciones $h(x)$ y $F(x)$ deben cumplir ciertas condiciones para que la integral tenga sentido y esté bien definida. Esta integral es una generalización de la integral usual de Riemann. Al integrando $h(x)$ se le pide inicialmente que sea una función acotada en el intervalo $(a, b]$, aunque después se omitirá esta condición. A la función integradora $F(x)$ se le pide que sea continua por la derecha, monótona no decreciente y tal que $F(\infty) - F(-\infty) < M$, para algún número $M > 0$. Observe que $F(x)$ debe cumplir propiedades casi idénticas a las de una función de distribución, y de hecho la notación es la misma. Esto no es coincidencia pues usaremos las funciones de distribución como funciones integradoras.

Presentamos a continuación la definición de la integral de Riemann-Stieltjes bajo las condiciones arriba señaladas. En [12] puede encontrarse una exposición más completa y rigurosa de esta integral. Nuestro objetivo en esta sección es simplemente presentar la definición y mencionar algunas propiedades. Sea $\{a = x_0 < x_1 < \dots <$

$x_n = b\}$ una partición finita del intervalo $(a, b]$, y defina

$$\begin{aligned}\bar{h}(x_i) &= \sup \{h(x) : x_{i-1} < x \leq x_i\}, \\ \text{y } \underline{h}(x_i) &= \inf \{h(x) : x_{i-1} < x \leq x_i\}.\end{aligned}$$

Se define la *suma superior e inferior* de Riemann-Stieltjes como sigue

$$\begin{aligned}\bar{S}_n &= \sum_{i=1}^n \bar{h}(x_i)[F(x_i) - F(x_{i-1})], \\ \underline{S}_n &= \sum_{i=1}^n \underline{h}(x_i)[F(x_i) - F(x_{i-1})].\end{aligned}$$

Ahora se hace n tender a infinito de tal forma que la longitud $\max\{|x_i - x_{i-1}| : 1 \leq i \leq n\}$ tienda a cero. Si sucede que

$$-\infty < \lim_{n \rightarrow \infty} \underline{S}_n = \lim_{n \rightarrow \infty} \bar{S}_n < \infty,$$

entonces a este valor común se le denota por

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

y se le llama la *integral de Riemann-Stieltjes* de la función $h(x)$, respecto de la función $F(x)$, sobre el intervalo $(a, b]$. Cuando la función $h(x)$ no es acotada se define la función auxiliar

$$h_N(x) = \begin{cases} -N & \text{si } h(x) < -N, \\ h(x) & \text{si } |h(x)| \leq N, \\ N & \text{si } h(x) > N. \end{cases}$$

Y entonces

$$\int_a^b h(x) dF(x) = \lim_{N \rightarrow \infty} \int_a^b h_N(x) dF(x),$$

cuando este límite existe. Se puede extender la definición de esta integral de la siguiente forma

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

cuando el límite del lado derecho exista.

La integral de Riemann-Stieltjes tiene muchas propiedades semejantes a la integral de Riemann. Enunciaremos a continuación algunas de ellas. Primeramente es lineal tanto en el integrando como en el integrador, es decir, si α es constante entonces

$$\begin{aligned}\int_a^b (\alpha h_1(x) + h_2(x)) dF(x) &= \alpha \int_a^b h_1(x) dF(x) + \int_a^b h_2(x) dF(x), \\ \text{y } \int_a^b h(x) d(\alpha F_1(x) + F_2(x)) &= \alpha \int_a^b h(x) dF_1(x) + \int_a^b h(x) dF_2(x).\end{aligned}$$

Cuando $h(x)$ tiene primera derivada continua se cumple la fórmula

$$\int_a^b h(x) dF(x) = h(b)F(b) - h(a)F(a) - \int_a^b F(x)h'(x) dx.$$

De particular importancia en la teoría de la probabilidad son los siguientes dos casos particulares. Cuando $F(x)$ es diferenciable entonces se tiene la igualdad

$$\int_a^b h(x) dF(x) = \int_a^b h(x)F'(x) dx.$$

De modo que integrar respecto de una función de distribución absolutamente continua se reduce a efectuar una integral de Riemann. El otro caso interesante ocurre cuando $h(x)$ es continua y $F(x)$ es constante excepto en los puntos x_1, x_2, \dots en donde la función tiene saltos positivos de tamaño $p(x_1), p(x_2), \dots$ respectivamente. En este caso y suponiendo convergencia,

$$\int_a^b h(x) dF(x) = \sum_{i=1}^{\infty} h(x_i)p(x_i).$$

Por lo tanto integrar respecto de la función de distribución de una variable aleatoria discreta se reduce a efectuar una suma. Finalmente enunciamos la propiedad que ilustra el hecho de que la integral de Riemann es un caso particular de la integral de Riemann-Stieltjes. Cuando $F(x) = x$ se cumple

$$\int_a^b h(x) dF(x) = \int_a^b h(x) dx.$$

2.5. Características numéricas

Se estudian a continuación algunas características numéricas asociadas a variables aleatorias. En particular, se definen los conceptos de esperanza, varianza y más generalmente los momentos de una variable aleatoria. Para ello haremos uso de la integral de Riemann-Stieltjes.

Esperanza

La esperanza de una variable aleatoria es un número que representa el promedio ponderado de los posible valores que toma la variable aleatoria, y se calcula como se indica a continuación.

Definición (Esperanza). Sea X con función de distribución $F(x)$. La esperanza de X , denotada por $E(X)$, se define como el número

$$E(X) = \int_{-\infty}^{\infty} x dF(x),$$

cuando esta integral sea absolutamente convergente, es decir, cuando

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

y en tal caso se dice que X es integrable.

A la esperanza se le conoce también con el nombre de: *media*, *valor esperado*, *valor promedio* o *valor medio*, y en general se usa la letra griega μ (mu) para denotarla. Cuando X es discreta con función de probabilidad $f(x)$, su esperanza, si existe, se calcula como sigue

$$E(X) = \sum_x x f(x).$$

Cuando X es absolutamente continua con función de densidad $f(x)$, entonces su esperanza, si existe, es

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

La integral o suma arriba mencionados pueden no existir y en ese caso se dice que la variable aleatoria no tiene esperanza finita. El Ejercicio 157 en la página 81 contiene algunos ejemplos que ilustran esta situación.

Ejemplo. Sea X discreta con valores en el conjunto $\{1, 2, \dots\}$, y con función de probabilidad $f(x) = P(X = x) = 1/2^x$, para $x \geq 1$. Entonces

$$E(X) = \sum_{x=1}^{\infty} x f(x) = \sum_{x=1}^{\infty} \frac{x}{2^x} = 2.$$

o

Ejemplo. Sea X continua con función de densidad $f(x) = 2x$, para $0 < x < 1$. Entonces

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x \cdot 2x dx = \frac{2}{3}.$$

o

Con frecuencia surge el problema de calcular esperanzas de funciones de variables aleatorias, es decir, si X es una variable aleatoria y $g : \mathbb{R} \rightarrow \mathbb{R}$ es una función Borel medible, entonces $g(X)$ es una variable aleatoria y el problema es encontrar

su esperanza. Usando directamente la definición, la esperanza de $g(X)$ se calcula del siguiente modo:

$$E[g(X)] = \int_{-\infty}^{\infty} x dF_{g(X)}(x),$$

pero ello requiere encontrar primero la distribución de $g(X)$, lo cual puede no ser fácil. Afortunadamente se cuenta con el siguiente resultado que establece una forma muy conveniente de calcular la esperanza de $g(X)$, sin conocer su distribución, pero suponiendo conocida la distribución de X .

Teorema (Esperanza de una función de una v.a.) Sea X con función de distribución $F_X(x)$, y sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función Borel medible tal que $g(X)$ tiene esperanza finita. Entonces

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x).$$

La demostración de este resultado en general no es sencilla y la omitiremos, aunque un camino cómodo que puede adoptarse es aceptar la fórmula anterior como la definición de la esperanza de $g(X)$. En particular, cuando la función g es la identidad, se recupera la definición básica de esperanza de una variable aleatoria. Por otro lado, cuando X es discreta, la demostración del teorema resulta no ser complicada, y en el Ejercicio 152 se pide hacer los detalles. Se establecen a continuación algunas propiedades de la esperanza.

Proposición (Propiedades de la esperanza). Sean X y Y con esperanza finita, y sea c una constante. Entonces

1. $E(c) = c$.
2. $E(cX) = cE(X)$.
3. Si $X \geq 0$, entonces $E(X) \geq 0$.
4. Si $X \leq Y$, entonces $E(X) \leq E(Y)$.
5. $E(X + Y) = E(X) + E(Y)$.

Las demostraciones de las primeras cuatro propiedades son sencillas pues se siguen directamente de la definición. La última propiedad es fácilmente demostrable en el caso discreto, y ello se ha dejado como ejercicio. Esta propiedad, en el caso general, será demostrada más adelante.

Proposición. Sea X con función de distribución $F(x)$, la cual admite la descomposición

$$F(x) = \alpha F^d(x) + (1 - \alpha)F^c(x),$$

en donde $\alpha \in [0, 1]$, $F^d(x)$ es una función de distribución discreta, y $F^c(x)$ es una función de distribución continua. Sea X_d con distribución $F^d(x)$, y sea X_c con distribución $F^c(x)$. Entonces X tiene esperanza finita si, y sólo si, tanto X_d como X_c tienen esperanza finita, y en tal caso,

$$E(X) = \alpha E(X_d) + (1 - \alpha)E(X_c).$$

Este resultado es fácil de demostrar usando la propiedad de linealidad de la integral de Riemann-Stieltjes respecto de la función integradora.

Varianza

La varianza de una variable aleatoria es una medida del grado de dispersión de los diferentes valores tomados por la variable aleatoria. Su definición es la siguiente.

Definición (Varianza). La varianza de X , denotada por $\text{Var}(X)$, se define como el número no negativo

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

cuando esta esperanza existe.

Cuando X es discreta con función de probabilidad $f(x)$ y esperanza finita μ , la varianza de X , cuando existe, se calcula como sigue

$$\text{Var}(X) = \sum_x (x - \mu)^2 f(x).$$

Cuando X es absolutamente continua con función de densidad $f(x)$ y esperanza finita μ , entonces la varianza de X , cuando existe, es

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

La varianza se denota regularmente por el símbolo σ^2 (sigma cuadrada). A la raíz cuadrada positiva de $\text{Var}(X)$ se le llama *desviación estándar*, y se le denota naturalmente por σ . Nuevamente hay casos en los que la varianza no es finita, y en

en esas situaciones se dice que la variable aleatoria no tiene varianza. Observe que para calcular $\text{Var}(X)$ se necesita conocer primero $E(X)$. Enunciamos a continuación algunas propiedades de la varianza.

Proposición (Propiedades de la varianza). Sean X y Y con varianza finita, y sea c una constante. Entonces

1. $\text{Var}(X) \geq 0$.
2. $\text{Var}(c) = 0$.
3. $\text{Var}(cX) = c^2\text{Var}(X)$.
4. $\text{Var}(X + c) = \text{Var}(X)$.
5. $\text{Var}(X) = E(X^2) - E^2(X)$.
6. En general, $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$.

La demostración de estas propiedades es sencilla pues todas ellas, excepto la última, se siguen directamente de la definición y de la propiedad lineal de la esperanza. Para la última propiedad puede tomarse $Y = X$, con $\text{Var}(X) \neq 0$, y verificarse la no igualdad. Otras propiedades de la varianza aparecen más adelante.

Momentos

Los momentos de una variable aleatoria son números que representan algunas características de la distribución de probabilidad asociada. Bajo ciertas condiciones el conjunto de momentos determinan de manera única a la distribución de probabilidad.

Definición (Momentos). Sea X una variable aleatoria con esperanza μ y sea n un número natural. Cuando existe, el número

1. $E(X^n)$ es el n -ésimo momento de X .
2. $E|X|^n$ es el n -ésimo momento absoluto de X .
3. $E[(X - \mu)^n]$ es el n -ésimo momento central de X .
4. $E|X - \mu|^n$ es el n -ésimo momento central absoluto de X .
5. $E[X(X - 1) \cdots (X - n + 1)]$ es el n -ésimo momento factorial de X .

Observe que el primer momento de X es $E(X)$, y el segundo momento central es $\text{Var}(X)$. En algunos textos al n -ésimo momento de X se le denota por μ'_n , mientras que el n -ésimo momento central es μ_n . En el Capítulo 8 sobre funciones generadoras se estudian ciertas funciones asociadas a las distribuciones de probabilidad, y a través de las cuales los momentos de una variable aleatoria pueden ser encontrados, cuando existen, de manera más eficiente.

Bajo ciertas condiciones los momentos de una variable aleatoria determinan la distribución de probabilidad de la misma. Por ejemplo, si X es tal que $E(X), E(X^2), \dots$ son todos finitos y si se cumple que la serie

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n)$$

es absolutamente convergente para algún $t > 0$, entonces la sucesión de momentos determina de manera única a la distribución de X . Las condiciones enunciadas para la determinación de la distribución de probabilidad son suficientes pero no necesarias.

Mediana

La mediana de una variable aleatoria es una medida de tendencia central que permite, cuando existe, dividir en dos partes iguales a la distribución de probabilidad sobre la recta numérica.

Definición (Mediana). El número m es una mediana de la variable X , o de su distribución, si se cumplen las siguientes dos desigualdades:

$$\begin{aligned} P(X \leq m) &\geq 1/2, \\ \text{y } P(X \geq m) &\geq 1/2. \end{aligned}$$

La mediana no es única y puede no existir, por ejemplo, si X es una variable aleatoria discreta tal que $P(X = 1) = p$, y $P(X = 0) = 1 - p$, con $p = 1/2$, entonces cualquier m en el intervalo $(0, 1)$ es una mediana, en cambio, cuando $p \neq 1/2$, no existe una mediana.

2.6. Distribuciones discretas

En esta sección se estudian algunas distribuciones discretas de probabilidad de uso común. Estas distribuciones son ejemplos particulares de medidas de probabilidad concentradas en un conjunto discreto de números reales. Se presentan estos ejemplos

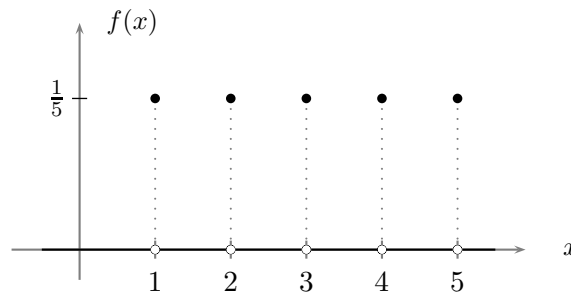
sin hacer mayor énfasis en las aplicaciones de los modelos. En el Apéndice A, al final del libro, aparecen algunas otras distribuciones de probabilidad.

Distribución uniforme discreta

La variable aleatoria X tiene una distribución uniforme sobre el conjunto $\{x_1, \dots, x_n\}$ si la probabilidad de que X tome cualquiera de estos valores es $1/n$. Esta distribución surge en espacios de probabilidad equiprobables, esto es, en situaciones en donde se tienen n resultados diferentes y todos ellos tienen la misma probabilidad de ocurrir. Los juegos de lotería justos son un ejemplo donde puede aplicarse esta distribución. Se escribe $X \sim \text{unif}\{x_1, \dots, x_n\}$, y su función de probabilidad es

$$f(x) = \begin{cases} \frac{1}{n} & \text{si } x = x_1, \dots, x_n, \\ 0 & \text{otro caso.} \end{cases}$$

Gráficamente



Función de probabilidad $\text{unif}\{1, 2, 3, 4, 5\}$.

Es fácil ver que

$$E(X) = \frac{1}{n} \sum_{i=1}^n x_i,$$

y

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - E(X))^2.$$

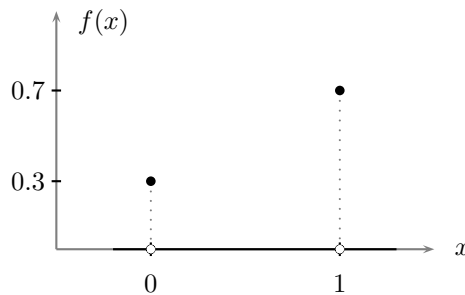
Distribución Bernoulli

Un *ensayo Bernoulli* es un experimento aleatorio con únicamente dos posibles resultados, llamados genéricamente *éxito* y *fracaso*, y con probabilidades respectivas p y

$1 - p$. Se define la variable aleatoria X como aquella función que lleva el resultado éxito al número 1, y el resultado fracaso al número 0. Entonces se dice que X tiene una distribución Bernoulli con parámetro $p \in (0, 1)$. Se escribe $X \sim \text{Ber}(p)$ y la correspondiente función de probabilidad es

$$f(x) = \begin{cases} 1 - p & \text{si } x = 0, \\ p & \text{si } x = 1, \\ 0 & \text{otro caso,} \end{cases}$$

cuya gráfica es



Función de probabilidad $\text{Ber}(p)$ con $p=0.7$.

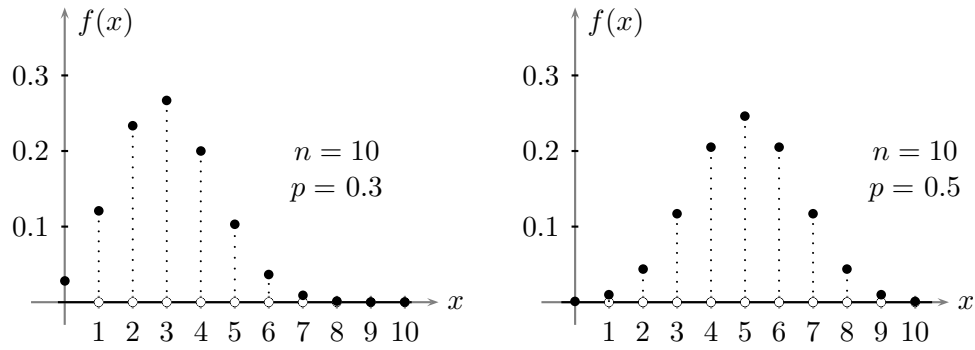
Es sencillo verificar que $E(X) = p$, y $\text{Var}(X) = p(1 - p)$. En particular, si A es un evento con probabilidad p , entonces la función indicadora 1_A es una variable aleatoria con distribución $\text{Ber}(p)$.

Distribución binomial

Suponga que se realizan n ensayos independientes Bernoulli en donde la probabilidad de éxito en cada uno de ellos es $p \in (0, 1)$. Si denotamos por E el resultado éxito y por F el resultado fracaso, entonces el espacio muestral consiste de todas las posibles sucesiones de longitud n de caracteres E y F. Usando el principio multiplicativo, es fácil ver que el conjunto Ω tiene 2^n elementos. Si ahora se define la variable aleatoria X como el número de éxitos en cada una de estas sucesiones, entonces X toma los valores $0, 1, \dots, n$, y se dice que X tiene una distribución binomial con parámetros n y p . Se escribe $X \sim \text{bin}(n, p)$, y su función de probabilidad es

$$f(x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & \text{si } x = 0, 1, \dots, n, \\ 0 & \text{otro caso.} \end{cases}$$

En las siguientes gráficas se muestra el comportamiento de esta función.



Función de probabilidad $\text{bin}(n, p)$.

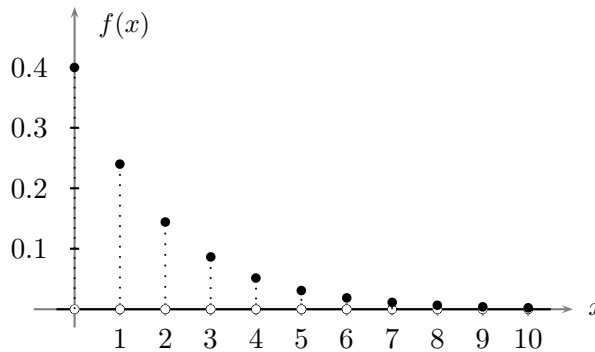
Se puede demostrar que $E(X) = np$, y $\text{Var}(X) = np(1 - p)$.

Distribución geométrica

Suponga que se tiene una sucesión infinita de ensayos independientes Bernoulli en donde la probabilidad de éxito en cada uno de ellos es $p \in (0, 1)$. Se define X como el número de fracasos antes de obtener el primer éxito. Se dice entonces que X tiene una distribución geométrica con parámetro p . Se escribe $X \sim \text{geo}(p)$, y su función de probabilidad es

$$f(x) = \begin{cases} p(1 - p)^x & \text{si } x = 0, 1, \dots \\ 0 & \text{otro caso,} \end{cases}$$

cuya gráfica es de la siguiente forma:



Función de probabilidad $\text{geo}(p)$ con $p=0.4$.

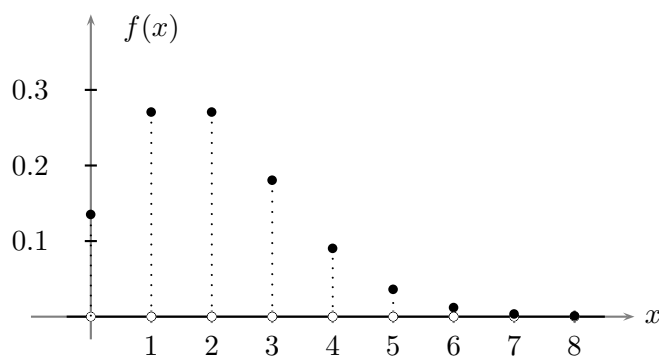
Para esta distribución se puede demostrar que $E(X) = (1 - p)/p$, y $\text{Var}(X) = (1 - p)/p^2$. En algunos textos se define también la distribución geométrica como el número de ensayos, y no el de fracasos, antes del primer éxito. En tal caso, la función de probabilidad es $f(x) = p(1 - p)^{x-1}$ para $x = 1, 2, \dots$. La media es entonces $1/p$ y la varianza es como antes.

Distribución Poisson

La variable aleatoria discreta X tiene una distribución Poisson con parámetro $\lambda > 0$, y se escribe $X \sim \text{Poisson}(\lambda)$ si su función de probabilidad es

$$f(x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & \text{si } x = 0, 1, \dots \\ 0 & \text{otro caso.} \end{cases}$$

Esta distribución fue descubierta por Simeón Denis Poisson en 1873 como límite de la distribución binomial, véase el Ejercicio 217. La gráfica de la función de probabilidad Poisson es de la siguiente forma:



Función de probabilidad Poisson(λ) con $\lambda = 2$.

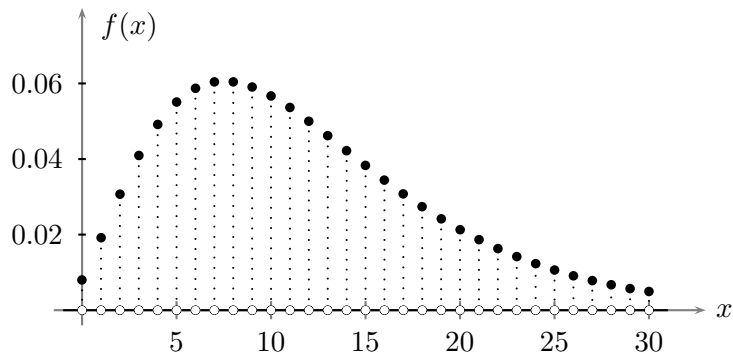
Puede demostrarse que $E(X) = \lambda$, y $\text{Var}(X) = \lambda$.

Distribución binomial negativa

Suponga una sucesión infinita de ensayos independientes Bernoulli en donde la probabilidad de éxito en cada ensayo es $p \in (0, 1)$. Sea X el número de fracasos antes de obtener el r -ésimo éxito. Se dice entonces que X tiene una distribución binomial negativa con parámetros r y p . Se escribe $X \sim \text{bin neg}(r, p)$, y su función de probabilidad es

$$f(x) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x & \text{si } x = 0, 1, \dots \\ 0 & \text{otro caso.} \end{cases}$$

Para $r = 3$ y $p = 0.2$, esta función tiene la siguiente forma.



Función de probabilidad bin neg(r, p) con $r = 3$ y $p = 0.2$.

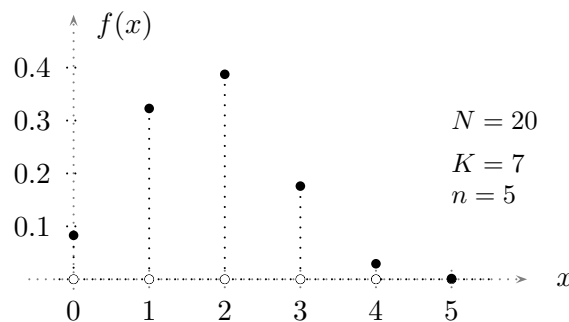
Es claro que esta distribución es una generalización de la distribución geométrica, la cual se obtiene cuando $r = 1$. Se puede demostrar que $E(X) = r(1 - p)/p$, y $\text{Var}(X) = r(1 - p)/p^2$.

Distribución hipergeométrica

Suponga que se tiene un conjunto de N objetos de los cuales K son de una primera clase, y $N - K$ son de una segunda clase. Suponga que de este conjunto se toma una muestra de tamaño n , sin reemplazo y en donde el orden de los objetos seleccionados no importa. Se define X como el número de objetos de la primera clase contenidos en la muestra seleccionada. Entonces X puede tomar los valores $0, 1, 2, \dots, n$, suponiendo $n \leq K$. Decimos que X tiene una distribución hipergeométrica con parámetros N, K y n . Se escribe $X \sim \text{hipergeo}(N, K, n)$, y su función de probabilidad es

$$f(x) = \begin{cases} \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} & \text{si } x = 0, 1, \dots, n, \\ 0 & \text{otro caso.} \end{cases}$$

Gráficamente:



Función de probabilidad hipergeo(N, K, n).

Es posible comprobar que

$$E(X) = n \frac{K}{N},$$

y

$$\text{Var}(X) = n \frac{K}{N} \frac{N - K}{N} \frac{N - n}{N - 1}.$$

2.7. Distribuciones continuas

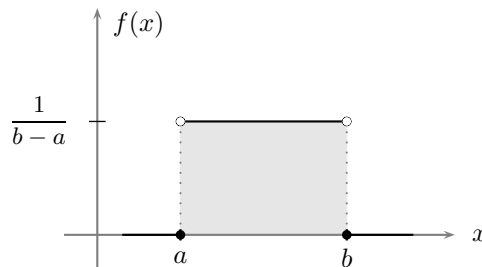
Ahora se estudian algunas distribuciones de probabilidad de variables aleatorias continuas. Algunas otras distribuciones continuas que surgen en la estadística serán estudiadas en el Capítulo 5. Recuerde que estos modelos son ejemplos particulares de medidas de probabilidad continuas sobre el conjunto de números reales, y que, en general, no se hace énfasis en sus aplicaciones.

Distribución uniforme continua

La variable aleatoria X tiene distribución uniforme en el intervalo (a, b) y se escribe $X \sim \text{unif}(a, b)$, cuando su función de densidad es

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a, b), \\ 0 & \text{otro caso.} \end{cases}$$

Gráficamente:



Función de densidad $\text{unif}(a, b)$.

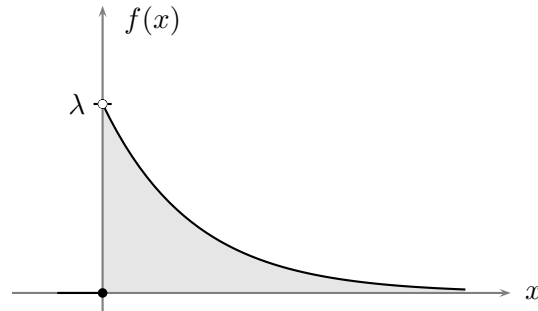
En este caso es inmediato verificar que $E(X) = (a + b)/2$, y $\text{Var}(X) = (b - a)^2/12$.

Distribución exponencial

La variable continua X tiene una distribución exponencial con parámetro $\lambda > 0$ y se escribe $X \sim \text{exp}(\lambda)$ cuando tiene función de densidad

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

cuya gráfica es



Función de densidad $\exp(\lambda)$.

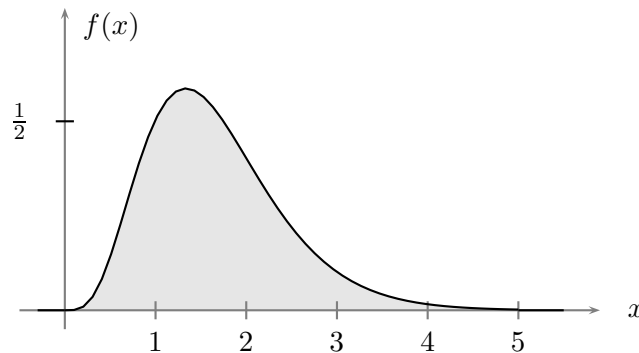
Para esta distribución es muy sencillo verificar que $E(X) = 1/\lambda$, y $\text{Var}(X) = 1/\lambda^2$.

Distribución gama

La variable aleatoria continua X tiene distribución gama con parámetros $n > 0$ y $\lambda > 0$ si su función de densidad es

$$f(x) = \begin{cases} \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

La gráfica de esta función se muestra a continuación.



Función de densidad $\text{gama}(n, \lambda)$ con $n = 5$ y $\lambda = 3$.

En tal caso se escribe $X \sim \text{gama}(n, \lambda)$. El término $\Gamma(n)$ es la *función gama* definida como sigue

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt,$$

para valores de n tal que la integral es convergente. Esta función satisface las siguientes propiedades:

- a) $\Gamma(n+1) = n\Gamma(n)$.
- b) $\Gamma(n+1) = n!$ para n entero positivo.
- c) $\Gamma(2) = \Gamma(1) = 1$.
- d) $\Gamma(1/2) = \sqrt{\pi}$.

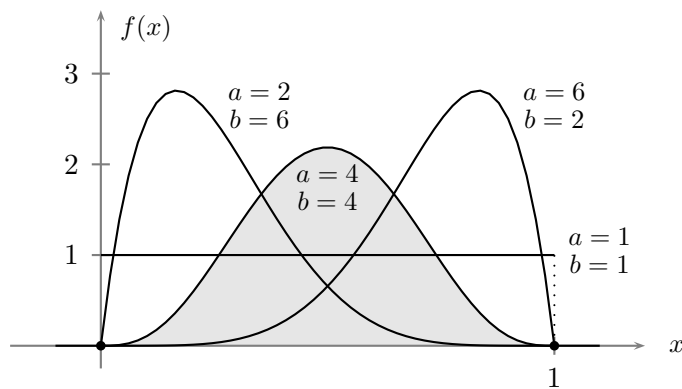
Observe que cuando $n = 1$, la distribución $\text{gama}(n, \lambda)$ se reduce a la distribución $\text{exp}(\lambda)$. Resolviendo un par de integrales se puede demostrar que $E(X) = n/\lambda$, y $\text{Var}(X) = n/\lambda^2$.

Distribución beta

La variable continua X tiene distribución beta con parámetros $a > 0$ y $b > 0$, y se escribe $X \sim \text{beta}(a, b)$ cuando su función de densidad es

$$f(x) = \begin{cases} \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} & \text{si } 0 < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

En la siguiente gráfica se ilustra la forma de esta función para varios valores de los parámetros.



Función de densidad $\text{beta}(a, b)$.

El término $B(a, b)$ se conoce como la *función beta* y se define para $a > 0$ y $b > 0$ como sigue

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx.$$

Esta función satisface las siguientes propiedades.

- a) $B(a, b) = B(b, a)$.
- b) $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Para la distribución beta(a, b) se tiene que

$$E(X) = \frac{a}{a+b},$$

y

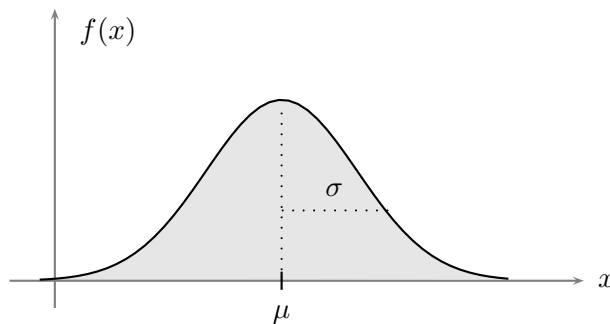
$$\text{Var}(X) = \frac{ab}{(a+b+1)(a+b)^2}.$$

Distribución normal

Esta es posiblemente la distribución de probabilidad de mayor importancia. Se dice que la variable aleatoria continua X tiene una distribución normal o Gaussiana si su función de densidad es

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

en donde $\mu \in \mathbb{R}$ y $\sigma^2 > 0$ son dos parámetros. En este caso se escribe $X \sim N(\mu, \sigma^2)$. No es difícil demostrar que $E(X) = \mu$, y $\text{Var}(X) = \sigma^2$. La gráfica de la función de densidad normal aparece en la siguiente figura:



Función de densidad $N(\mu, \sigma^2)$.

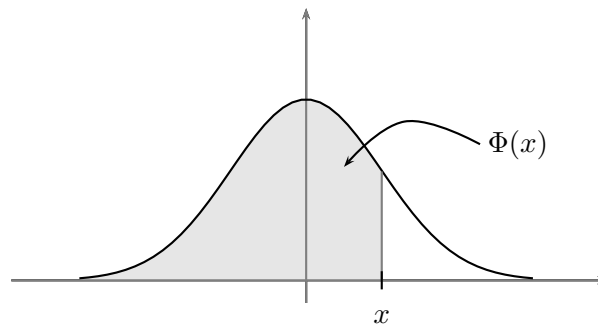
En particular se dice que X tiene una distribución normal *estándar* si $\mu = 0$ y $\sigma^2 = 1$. En este caso particular, la función de densidad se reduce a la expresión más sencilla

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Es posible transformar una variable aleatoria normal no estándar en una estándar mediante la siguiente operación llamada *estandarización*. La demostración de este resultado es elemental y se deja como ejercicio.

Proposición $X \sim N(\mu, \sigma^2) \Leftrightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$.

Comúnmente se usa la letra Z para denotar una variable aleatoria con distribución normal estándar. En particular la función $\Phi(x)$ denota la función de distribución de una variable aleatoria normal estándar, es decir, $\Phi(x) = P(Z \leq x)$. Gráficamente:



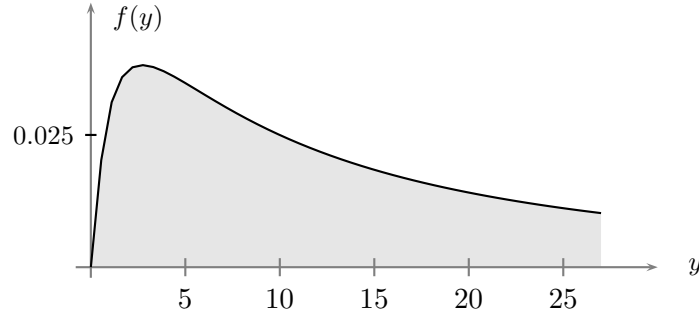
Área cubierta por la función de distribución $\Phi(x) = P(Z \leq x)$.

Distribución log normal

Si X tiene distribución $N(\mu, \sigma^2)$, entonces $Y = e^X$ tiene una distribución log normal (μ, σ^2) y su función de densidad es:

$$f(y) = \begin{cases} \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln y - \mu)^2}{2\sigma^2}\right] & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

La gráfica de esta función es la siguiente:



Función de densidad log normal(μ, σ^2) con $\mu = 3$ y $\sigma^2 = 2$.

Se puede demostrar que

$$E(Y) = \exp(\mu + \sigma^2/2),$$

$$y \quad \text{Var}(Y) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2).$$

Otras distribuciones continuas de interés se encuentran en el capítulo sobre distribuciones muestrales.

2.8. Ejercicios

Variables aleatorias

94. Demuestre que la función identidad $X(\omega) = \omega$ no es variable aleatoria cuando $\Omega = \{1, 2, 3\}$ y $\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$.
95. Sea $\Omega = \{-1, 0, 1\}$ y $\mathcal{F} = \{\emptyset, \{0\}, \{-1, 1\}, \Omega\}$. Considere la función identidad $X(\omega) = \omega$. Demuestre que X^2 es variable aleatoria pero X no lo es.
96. Considere el espacio medible (Ω, \mathcal{F}) , con $\mathcal{F} = \{\emptyset, \Omega\}$. Demuestre que $X : \Omega \rightarrow \mathbb{R}$ es variable aleatoria si, y sólo si, X es constante.
97. Sea (Ω, \mathcal{F}) un espacio medible tal que $\mathcal{F} = \{\emptyset, \Omega, A, A^c\}$ con $A \subseteq \Omega$. Demuestre que toda función medible $X : \Omega \rightarrow \mathbb{R}$ es constante en A y en A^c . Por lo tanto toda función medible respecto de esta σ -álgebra toma a lo sumo dos valores distintos. El siguiente ejercicio generaliza este resultado.
98. Sea A_1, \dots, A_n una partición finita de Ω , y considere el espacio medible (Ω, \mathcal{F}) , con $\mathcal{F} = \sigma\{A_1, \dots, A_n\}$. Demuestre que $X : \Omega \rightarrow \mathbb{R}$ es variable aleatoria si, y sólo si, X es constante en cada elemento de la partición. En consecuencia, X toma a lo sumo n valores distintos.
99. Demuestre que X es variable aleatoria si, y sólo si, $X^{-1}(-\infty, x) \in \mathcal{F}$ para cada número real x .
100. Demuestre que X es variable aleatoria si, y sólo si, $X^{-1}[x, \infty) \in \mathcal{F}$ para cada número real x .

101. Demuestre que X es variable aleatoria si, y sólo si, $X^{-1}(x, \infty) \in \mathcal{F}$ para cada número real x .
102. Demuestre que X es variable aleatoria si, y sólo si, $X^{-1}(a, b) \in \mathcal{F}$ para cada intervalo (a, b) de \mathbb{R} .
103. Sea c una constante y X una v.a. Demuestre directamente que las siguientes funciones también son variables aleatorias: cX , $X + c$, $X \vee c$, $X \wedge c$.
104. Demuestre directamente que la diferencia de dos variables aleatorias es variable aleatoria.
105. Sea X una variable aleatoria. Demuestre que la parte entera de X , denotada por $\lfloor X \rfloor$, es una variable aleatoria discreta, es decir, toma un número numerable de valores.
106. Demuestre que el conjunto de v.a.s definidas sobre un espacio de probabilidad es un espacio vectorial con las operaciones usuales de suma y producto por escalares.
107. Sean X y Y variables aleatorias. Demuestre directamente que tanto $X \vee Y$ como $X \wedge Y$ son variables aleatorias.
108. Demuestre directamente que si X es variable aleatoria, entonces también lo son X^n y $2X^3 - 5X$.
109. Demuestre que X es variable aleatoria si, y sólo si, tanto $X^+ = \max\{0, X\}$ como $X^- = -\min\{0, X\}$, lo son.
110. Sea $A \subseteq \Omega$. Demuestre que la función indicadora¹ $1_A : \Omega \rightarrow \mathbb{R}$ es variable aleatoria si, y sólo si, el conjunto A es medible.
111. Sean $A, B \subseteq \Omega$. Diga falso o verdadero. Demuestre en cada caso.
- a) A, B medibles $\Rightarrow 1_A + 1_B$ es v.a.
- b) $1_A + 1_B$ es v.a. $\Rightarrow A, B$ son medibles.
112. Sean A, B subconjuntos disjuntos de Ω y sean a, b dos números reales distintos. Demuestre que

$$a1_A + b1_B \text{ es v.a.} \Leftrightarrow A, B \text{ son medibles.}$$

Una de estas implicaciones resulta falsa cuando se omite la condición de que los números a y b son distintos. ¿Cuál de ellas es?.

113. Sean A_1, \dots, A_n subconjuntos disjuntos de Ω , y sean a_1, \dots, a_n constantes distintas. Demuestre que

$$\sum_{i=1}^n a_i 1_{A_i} \text{ es v.a.} \Leftrightarrow A_1, \dots, A_n \text{ son medibles.}$$

¹Véase el final del texto para la definición y algunas propiedades de la función indicadora.

114. Sean A y B dos eventos, y sean 1_A y 1_B las correspondientes funciones indicadoras. Directamente de la definición demuestre que las funciones $1_A + 1_B$, $1_A - 1_B$ y $1_A \cdot 1_B$ son variables aleatorias.
115. Sean X y Y dos variables aleatorias. Demuestre que los conjuntos $(X = Y)$, $(X \leq Y)$, $(X > Y)$ y $(X \neq Y)$ son eventos. Sugerencia: Proceda como en la fórmula (2.1) de la página 44.
116. Sean X , Y y Z tres variables aleatorias. Demuestre que los conjuntos $(X = Y = Z)$, y $(X < Y < Z)$ son eventos.
117. Sea X una variable aleatoria y $g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ una función Borel medible. Demuestre que $g(X) = g \circ X : \Omega \rightarrow \mathbb{R}$ es también una variable aleatoria. Sugerencia: Demuestre que la colección $\mathcal{B} = \{B \in \mathcal{B}(\mathbb{R}) : g^{-1}B \in \mathcal{B}(\mathbb{R})\}$ coincide con $\mathcal{B}(\mathbb{R})$ usando los siguientes dos resultados: (1) Dada una función continua de \mathbb{R} en \mathbb{R} , la imagen inversa de un conjunto abierto es nuevamente un conjunto abierto. (2) Todo conjunto abierto de \mathbb{R} distinto del vacío puede expresarse como una unión numerable de intervalos abiertos.
118. Sea X una variable aleatoria. Demuestre que las funciones: e^X , $\text{sen } X$, y $\text{cos } X$ son variables aleatorias.
119. Sea $X : \Omega \rightarrow \mathbb{R}$ una función. Proporcione un ejemplo en el que X^2 sea variable aleatoria pero $|X|$ no lo sea.
120. Sean X_1, \dots, X_n variables aleatorias. Demuestre que

$$a) \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{es v.a.}$$

$$b) S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{es v.a.}$$

121. Sea X una variable aleatoria, y sean $a < b$ dos constantes. Demuestre que las siguientes funciones son variables aleatorias.

$$a) Y = \begin{cases} X & \text{si } X < a, \\ a & \text{si } X \geq a. \end{cases}$$

$$b) Y = \begin{cases} a & \text{si } X < a, \\ X & \text{si } a \leq X \leq b, \\ b & \text{si } X > b, \end{cases}$$

$$c) Y = \begin{cases} X & \text{si } |X| \leq a, \\ 0 & \text{si } |X| > a. \end{cases}$$

122. Defina la *función signo* como sigue

$$\text{signo}(x) = \begin{cases} +1 & \text{si } x > 0, \\ -1 & \text{si } x < 0, \\ 0 & \text{si } x = 0. \end{cases}$$

Demuestre que si X es variable aleatoria, entonces $\text{signo}(X)$ también lo es.

123. Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sea $X : \Omega \rightarrow \mathbb{R}$ una función. Demuestre que la colección $\{X^{-1}B : B \in \mathcal{B}(\mathbb{R})\}$ es una sub σ -álgebra de \mathcal{F} si, y sólo si, X es variable aleatoria.
124. Sea X una variable aleatoria con valores en el conjunto $\{0, 1, 2, \dots\}$. Sea $(X)_{10}$ el valor de X módulo 10. Demuestre que $(X)_{10}$ es también variable aleatoria.
125. Sean $(\Omega_1, \mathcal{F}_1)$ y $(\Omega_2, \mathcal{F}_2)$ dos espacios medibles, y sea $X : \Omega_1 \rightarrow \Omega_2$ una *función medible*, es decir, para cualquier A en \mathcal{F}_2 se cumple que $X^{-1}A \in \mathcal{F}_1$. Suponga que $P : \mathcal{F}_1 \rightarrow [0, 1]$ es una medida de probabilidad. Demuestre que

$$P \circ X^{-1} : \mathcal{F}_2 \rightarrow [0, 1]$$

es también una medida de probabilidad. A la medida $P \circ X^{-1}$ se le llama *medida de probabilidad inducida* por X .

126. Sea X una variable aleatoria discreta. Demuestre que la probabilidad de que X tome un valor finito es uno.

Función de distribución

127. Demuestre que las siguientes funciones son de distribución.

- a) $F(x) = 1 - e^{-x}$, para $x > 0$.
- b) $F(x) = 1 - (1 + x)e^{-x}$, para $x > 0$.
- c) $F(x) = \begin{cases} 0 & \text{si } x < -1, \\ (x + 1)/2 & \text{si } x \in [-1, 1], \\ 1 & \text{si } x > 1. \end{cases}$

128. Investigue si las siguientes funciones son de distribución.

- a) $F(x) = x$, para $x \in \mathbb{R}$.
- b) $F(x) = 1 - e^{-x^2}$, para $x > 0$.
- c) $F(x) = e^{-1/x}$, para $x > 0$.
- d) $F(x) = \frac{e^x}{1 + e^x}$, para $x \in \mathbb{R}$.
- e) $F(x) = \frac{e^x}{e^x + e^{-x}}$, para $x \in \mathbb{R}$.

129. Sean $F(x)$ y $G(x)$ dos funciones de distribución. Determine si las siguientes funciones son de distribución.

- a) $aF(x) + (1 - a)G(x)$, con $0 \leq a \leq 1$.
- b) $F(x) + G(x)$.
- c) $F(x)G(x)$.

130. Sea X con función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < 2, \\ 1 - \frac{4}{x^2} & \text{si } x \geq 2. \end{cases}$$

Grafique $F(x)$ y demuestre que es una función de distribución. Calcule además $P(X \leq 4)$, $P(X > 1)$, $P(4 < X < 6)$ y $P(X = 2)$.

131. Sea X con función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 0.2 & \text{si } 0 \leq x < 1, \\ 0.5 & \text{si } 1 \leq x < 3, \\ 0.9 & \text{si } 3 \leq x < 4, \\ 1 & \text{si } x \geq 4. \end{cases}$$

Grafique $F(x)$ y demuestre que es una función de distribución. Calcule además $P(X \leq 1)$, $P(X = 1)$, $P(0 < X < 3)$, $P(X = 4)$ y $P(X \geq 3)$.

132. Sea X con función de distribución $F(x)$. Demuestre que

- a) $P(X < x) = F(x-)$.
- b) $P(X = x) = F(x) - F(x-)$.
- c) $P(X > x) = 1 - F(x)$.

133. En la escuela rusa de probabilidad se define la función de distribución de una variable aleatoria X como $G(x) = P(X < x)$. Observe el signo “ $<$ ” en lugar de “ \leq ”. Demuestre que esta función cumple todas las propiedades de una función de distribución, excepto que ahora la continuidad es por la izquierda.

134. Sea $F(x)$ una función de distribución continua. Demuestre que para cualquier entero $n \geq 1$, las siguientes funciones también son de distribución.

- a) $G(x) = [F(x)]^n$.
- b) $G(x) = 1 - [1 - F(x)]^n$.

135. Sea X con función de distribución $F(x)$. Diga falso o verdadero. Demuestre en cada caso.

- a) $F(x) = P(X < x) + P(X = x)$.
- b) $1 - F(x) = P(X \geq x)$.
- c) $1 - P(X < x) - P(X > x) = P(X = x)$.

136. Encuentre $F_Y(y)$ en términos de $F_X(x)$ cuando

- a) $Y = aX + b$, con a, b constantes.
- b) $Y = e^X$.
- c) $Y = e^{-X}$.
- d) $Y = X^2$.
- e) $Y = X^+ = \max\{0, X\}$.
- f) $Y = X^- = -\min\{0, X\}$.
- g) $Y = |X|$.
- h) $Y = -X$.
- i) $Y = \text{sen } X$.

137. Sea X con función de distribución $F_X(x)$, y sean $a < b$ dos constantes. Calcule la función de distribución de Y en términos de la función de distribución de X , y muestre gráficamente el comportamiento de $F_Y(y)$ en los puntos a y b .

$$a) Y = \begin{cases} X & \text{si } X < a, \\ a & \text{si } X \geq a. \end{cases}$$

$$b) Y = \begin{cases} a & \text{si } X < a, \\ X & \text{si } a \leq X \leq b, \\ b & \text{si } X > b. \end{cases}$$

$$c) Y = \begin{cases} X & \text{si } |X| \leq a, \\ 0 & \text{si } |X| > a. \end{cases}$$

138. Sean $F(x)$ y $G(x)$ dos funciones de distribución continuas y estrictamente crecientes. Demuestre que

$$a) \text{ si } F(x) \geq G(x), \text{ entonces } F^{-1}(y) \leq G^{-1}(y).$$

$$b) \text{ si } X \text{ tiene función de distribución } F(x), \text{ entonces } Y = G^{-1}(F(X)) \text{ tiene función de distribución } G(x).$$

$$c) \text{ si } F(x) \geq G(x), \text{ entonces existen variables aleatorias } X \text{ y } Y \text{ cuyas funciones de distribución son } F(x) \text{ y } G(x) \text{ respectivamente, y son tales que } X \leq Y. \text{ Sugerencia: Use el inciso anterior.}$$

139. Sea X con función de distribución $F(x)$. Demuestre que $F(x)$ es continua en $x = x_0$ si, y sólo si, $P(X = x_0) = 0$.

Tipos de variables aleatorias

140. Encuentre la constante c que hace a $f(x)$ una función de probabilidad.

$$a) f(x) = \frac{c}{x(x+1)}, \quad \text{para } x = 1, 2, \dots$$

$$b) f(x) = ce^{-x}, \quad \text{para } x = 1, 2, \dots$$

$$c) f(x) = \frac{c}{x!}, \quad \text{para } x = 1, 2, \dots$$

141. Encuentre la constante c que hace a $f(x)$ una función de densidad.

$$a) f(x) = cx^2, \quad \text{para } 0 < x < 1.$$

$$b) f(x) = cxe^{-2x^2}, \quad \text{para } x > 0.$$

$$c) f(x) = cx^{-2}, \quad \text{para } x > 1.$$

$$d) f(x) = \frac{ce^x}{(1+e^x)^2}, \quad \text{para } x \in \mathbb{R}.$$

$$e) f(x) = cx(1-x), \quad \text{para } 0 < x < 1.$$

$$f) f(x) = \frac{c}{\sqrt{1-x^2}}, \quad \text{para } 0 < x < 1.$$

$$g) f(x) = \frac{c}{1+x^2}, \quad \text{para } x \in \mathbb{R}.$$

142. Demuestre que las siguientes funciones son de densidad. Encuentre la correspondiente función de distribución y demuestre que satisface las propiedades de toda función de distribución. Grafique ambas funciones.

a) $f(x) = 2x$ para $x \in [0, 1]$.

b) $f(x) = \frac{3}{2}x^2$, para $x \in [-1, 1]$.

c) $f(x) = 1 - \frac{1}{2}x$, para $x \in [0, 2]$.

d) $f(x) = \frac{2}{m^2}x$, para $x \in [0, m]$, con $m > 0$.

e) $f(x) = \frac{1}{(1-x)^2}$, para $x \in [0, 1/2]$.

f) $f(x) = \frac{1}{2}e^{|x|}$, para $x \in \mathbb{R}$.

143. Demuestre que las siguientes funciones son de distribución. Encuentre la correspondiente función de densidad y compruebe que efectivamente es una función de densidad. Grafique ambas funciones.

a) $F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 & \text{si } x \geq 0. \end{cases}$

b) $F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ x & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$

c) $F(x) = \frac{e^x}{1 + e^x}$.

d) $F(x) = \frac{1}{2} \int_{-\infty}^x e^{-|u|} du$.

144. Sea $f(x)$ una función de densidad y sea c una constante. Demuestre que $f(x+c)$ es también una función de densidad.

145. Diga falso o verdadero. Demuestre en cada caso.

a) Toda función de densidad es acotada.

b) Toda función de distribución es acotada.

146. Sea X absolutamente continua y sea $Y = aX + b$ con a y b constantes. Demuestre que si $a \neq 0$, entonces

$$f_Y(y) = \frac{1}{|a|} f_X((y-b)/a).$$

Igualdad de variables aleatorias

147. Demuestre que

a) si $X \stackrel{c.s.}{=} Y$, entonces $X \stackrel{d}{=} Y$.

b) si $X \stackrel{d}{=} Y$, entonces no necesariamente $X \stackrel{c.s.}{=} Y$. Como sugerencia considere X tal que $P(X = -1) = P(X = 1) = 1/2$, y defina $Y = -X$.

148. Demuestre que la igualdad casi segura de variables aleatorias es una relación de equivalencia. ¿Cumple tal propiedad la igualdad en distribución?

Integral de Riemann-Stieltjes

149. Sea X una variable aleatoria con función de distribución F , y sea a cualquier número real. Demuestre que

$$\int_{-\infty}^{\infty} 1_{\{a\}}(x) dF(x) = P(X = a).$$

150. Sea X una variable aleatoria con función de distribución F , y sea $(a, b) \subseteq \mathbb{R}$. Demuestre que

$$\int_{-\infty}^{\infty} 1_{(a,b)}(x) dF(x) = P(a < X < b).$$

151. Sea F una función de distribución absolutamente continua. Demuestre que para cualesquiera números naturales n y m

$$\int_{-\infty}^{\infty} F^n(x) dF^m(x) = \frac{m}{n+m}.$$

Esperanza

152. *Esperanza de una función de una v.a. discreta.* Sea X discreta con valores en el conjunto $\{x_1, x_2, \dots\}$, y sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función Borel medible tal que $g(X)$ tiene esperanza finita. Demuestre que

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i)P(X = x_i).$$

153. Calcule la esperanza de X cuya función de probabilidad o de densidad es

a) $f(x) = 1/5$, para $x = -2, -1, 0, 1, 2$.

b) $f(x) = e^{-1}/x!$, para $x = 0, 1, 2, \dots$

c) $f(x) = |x|$, para $-1 < x < 1$.

d) $f(x) = \frac{1}{2}e^{-|x|}$, para $x \in \mathbb{R}$.

154. Calcule la esperanza de la variable aleatoria X cuya función de distribución es

$$F(x) = \begin{cases} 0 & \text{si } x < 1, \\ 1 - \frac{1}{2}e^{-x} & \text{si } x \geq 1. \end{cases}$$

155. Sean X y Y con esperanza finita, y sea c una constante. Demuestre que

a) $E(c) = c$.

- b) $E(cX) = cE(X)$.
 c) $E(X + c) = E(X) + c$.
 d) Si $X \geq 0$, entonces $E(X) \geq 0$.
 e) Si $X \leq Y$, entonces $E(X) \leq E(Y)$.
 f) $|E(X)| \leq E|X|$.
156. Sean X y Y discretas ambas con esperanza finita. Demuestre directamente que $E(X + Y) = E(X) + E(Y)$.

157. Demuestre que no existe la esperanza de X cuando su función de probabilidad o de densidad es

a) $f(x) = \frac{1}{x(x+1)}$, para $x = 1, 2, \dots$

b) $f(x) = \frac{3}{\pi^2 x^2}$, para $x \in \mathbb{Z} \setminus \{0\}$.

c) $f(x) = 1/x^2$, para $x > 1$.

d) $f(x) = \frac{1}{\pi(1+x^2)}$, para $x \in \mathbb{R}$.

158. *La paradoja de San Petersburgo*. Un juego consiste en lanzar una moneda equilibrada repetidas veces hasta que una de las caras, seleccionada previamente, aparezca por primera vez. Si un jugador lanza la moneda y requiere de n lanzamientos para que se cumpla la condición, entonces recibe 2^n unidades monetarias. ¿Cuál debe ser el pago inicial justo para ingresar a este juego?
159. Sea $\{A_1, A_2, \dots\}$ una colección de eventos que forman una partición de Ω tal que cada elemento de la partición tiene probabilidad estrictamente positiva. Sea X una variable aleatoria discreta con esperanza finita. Para cualquier evento A con probabilidad positiva defina

$$E(X|A) = \sum_x xP(X = x | A).$$

Demuestre que

$$E(X) = \sum_{i=1}^{\infty} E(X | A_i)P(A_i).$$

160. Demuestre que

a) $E(X \wedge Y) \leq E(X) \wedge E(Y) \leq E(X)$.

b) $E(X \vee Y) \geq E(X) \vee E(Y) \geq E(X)$.

161. Sea $X > 0$, discreta y con esperanza finita. Demuestre directamente que

$$E(X)E(1/X) \geq 1.$$

Este resultado puede ser demostrado usando la desigualdad de Jensen, pero en este ejercicio se pide obtener el resultado sin usar Jensen.

162. Sea X discreta con valores $0 \leq x_1 \leq x_2 \leq \dots \leq x_k$. Demuestre que

- a) $\lim_{n \rightarrow \infty} \frac{E(X^{n+1})}{E(X^n)} = x_k,$
 b) $\lim_{n \rightarrow \infty} \sqrt[n]{E(X^n)} = x_1.$

163. Sea X discreta con valores $0, 1, \dots$ y con esperanza finita. Demuestre que

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=0}^{\infty} P(X > n).$$

Ahora use esta fórmula para demostrar que

- a) si X tiene distribución $\text{geo}(p)$, entonces $E(X) = (1 - p)/p$.
 b) si X tiene distribución $\text{Poisson}(\lambda)$, entonces $E(X) = \lambda$.
164. Sea $X \geq 0$ con esperanza finita, y suponga que para algún $p \in (0, 1)$, se cumple la desigualdad $P(X \geq k) \leq p^k$, para cada $k = 0, 1, \dots$. Demuestre que

$$E(X) \leq \frac{1}{1 - p}.$$

165. Sea $X \geq 0$ con esperanza finita, y para cada número natural n defina el evento $A_n = (n - 1 \leq X < n)$. Demuestre que

$$\sum_{n=1}^{\infty} (n - 1)1_{A_n} \leq X < \sum_{n=1}^{\infty} n1_{A_n}.$$

Ahora demuestre las desigualdades

$$\sum_{n=1}^{\infty} P(X \geq n) \leq E(X) < 1 + \sum_{n=1}^{\infty} P(X \geq n).$$

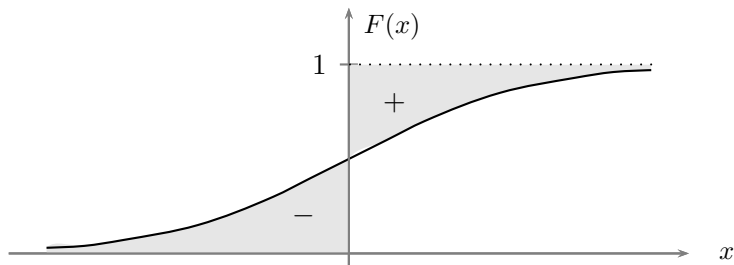
166. Sea X con función de distribución $F(x)$, y con esperanza finita. Demuestre que

- a) $\lim_{x \rightarrow \infty} x[1 - F(x)] = 0.$
 b) $\lim_{x \rightarrow -\infty} xF(x) = 0.$

167. Sea X con función de distribución $F(x)$, y con esperanza finita. Demuestre que

$$E(X) = \int_0^{\infty} [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx.$$

Gráficamente estas integrales pueden interpretarse como indica la siguiente figura:



La esperanza como la diferencia de dos áreas.

Use esta fórmula para demostrar que

- a) si X tiene distribución $\exp(\lambda)$, entonces $E(X) = 1/\lambda$.
 - b) si X tiene distribución $\text{gama}(n, \lambda)$, entonces $E(X) = n/\lambda$.
168. Sea X con función de distribución continua $F(x)$, y con esperanza finita μ . Demuestre que

$$\int_{-\infty}^{\mu} F(x)dx = \int_{\mu}^{\infty} [1 - F(x)]dx.$$

169. Demuestre que la condición $E(X) = 0$ no implica que X sea simétrica alrededor de cero. Considere el ejemplo $P(X = -1) = 1/2$, $P(X = 0) = 1/8$, $P(X = 1) = 1/4$ y $P(X = 2) = 1/8$. ¿Puede construir un ejemplo de una distribución continua con esperanza cero, que no sea simétrica?

Varianza

170. Calcule la varianza de X cuya función de probabilidad o de densidad es
- a) $f(x) = 1/5$, para $x = -2, -1, 0, 1, 2$.
 - b) $f(x) = e^{-1}/x!$, para $x = 0, 1, 2, \dots$
 - c) $f(x) = |x|$, para $-1 < x < 1$.
 - d) $f(x) = \frac{1}{2}e^{-|x|}$, para $x \in \mathbb{R}$.
171. Sean X y Y con varianza finita y sea c una constante. Demuestre las siguientes propiedades de la varianza.
- a) $\text{Var}(X) \geq 0$.
 - b) $\text{Var}(c) = 0$.
 - c) $\text{Var}(cX) = c^2\text{Var}(X)$.
 - d) $\text{Var}(X + c) = \text{Var}(X)$.
 - e) $\text{Var}(X) = E(X^2) - E^2(X)$.
172. Sea X con valores en $[a, b]$. Demuestre que
- a) $a \leq E(X) \leq b$.
 - b) $0 \leq \text{Var}(X) \leq (b - a)^2/4$.

173. *Minimización del error cuadrático medio.* Sea X con segundo momento finito. A la función $g(u) = E[(X - u)^2]$ se le conoce como error cuadrático medio. Demuestre que $g(u)$ se minimiza cuando $u = E(X)$. En consecuencia, para cualquier valor real de u ,

$$\text{Var}(X) \leq E[(X - u)^2].$$

174. Sea X con varianza finita y sea c una constante. Demuestre que

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

175. Sea X con media μ y varianza σ^2 . Demuestre que $E|X - \mu| \leq \sigma$. Sugerencia: $\text{Var}(|X - \mu|) \geq 0$.

176. Diga falso o verdadero. Demuestre en cada caso.

- a) Si $X \leq Y$, entonces $\text{Var}(X) \leq \text{Var}(Y)$.
- b) $\text{Var}(X) \leq E(X^2)$.
- c) $E^2(X) \leq E(X^2)$.

177. Sea $X \geq 0$ con varianza finita, y sea $M \geq 0$ una constante fija. Diga si las siguientes afirmaciones son falsas o verdaderas, demuestre en cada caso.

- a) $E(X \wedge M) \leq E(X) \leq E(X \vee M)$.
- b) $\text{Var}(X \wedge M) \leq \text{Var}(X)$.
- c) $\text{Var}(X \vee M) \leq \text{Var}(X)$.

178. Sean X y Y con varianza finita. Diga si las siguientes desigualdades son falsas o verdaderas, demuestre en cada caso.

$$\text{Var}(X \wedge Y) \leq \text{Var}(X) \leq \text{Var}(X \vee Y).$$

179. Sea X con varianza finita, y sea c una constante cualquiera. Diga si las siguientes afirmaciones son falsas o verdaderas, demuestre en cada caso.

- a) $\text{Var}(X + c) = \text{Var}(X - c)$.
- b) $\text{Var}(|X|) \leq \text{Var}(X)$.
- c) $\text{Var}(|X - c|) \leq \text{Var}(X)$.

Momentos

180. Calcule el n -ésimo momento de X cuya función de probabilidad o de densidad es

- a) $f(x) = 1/5$ para $x = -2, -1, 0, 1, 2$.
- b) $f(x) = e^{-1}/x!$ para $x = 0, 1, 2, \dots$
- c) $f(x) = |x|$ para $-1 < x < 1$.
- d) $f(x) = \frac{1}{2}e^{-|x|}$ para $x \in \mathbb{R}$.

181. Sea X tal que $E|X|^n < \infty$ para algún natural n . Demuestre que para cualquier número natural de $m \leq n$, se cumple

$$E|X|^m \leq E|X|^n.$$

Este resultado establece entonces que si el n -ésimo momento de una variable aleatoria es finito, entonces todos los momentos anteriores a n también son finitos. Sugerencia:

$$|X|^m = |X|^m \cdot 1_{(|X| \leq 1)} + |X|^m \cdot 1_{(|X| > 1)}.$$

182. Sea X con distribución simétrica alrededor de $x = 0$, y con cuarto momento finito. Demuestre que para cualquier número real a ,

$$E(X^4) \leq E(X - a)^4.$$

183. Sea 1_A la función indicadora de un evento A . Demuestre que

- a) $E(1_A) = E(1_A^n) = P(A)$.
 b) $\text{Var}(1_A) = P(A)(1 - P(A)) \leq 1/4$.

184. Sea X con n -ésimo momento finito. Demuestre que

$$E|X|^n = n \int_0^\infty x^{n-1} (1 - F(x)) dx + n \int_{-\infty}^0 |x|^{n-1} F(x) dx.$$

185. Sea X discreta con valores en el conjunto $\{0, 1, 2, \dots\}$, y con segundo momento finito. Demuestre que

$$E(X^2) = \sum_{n=1}^{\infty} (2n - 1)P(X \geq n).$$

186. *Espacio L^1* . Demuestre que el espacio $L^1(\Omega, \mathcal{F}, P)$ consistente de todas las variables aleatorias X tales que $E|X| < \infty$, es un espacio vectorial. Para resolver este ejercicio suponga válida la igualdad $E(X + Y) = E(X) + E(Y)$, la cual se demuestra más adelante.

187. *Desigualdad de Cauchy-Schwarz*. Sean X y Y con segundo momento finito. Demuestre que

$$E^2(XY) \leq E(X^2)E(Y^2).$$

Sugerencia: Para cualquier valor real de t , la esperanza de $(tX + Y)^2$ es no negativa. Desarrolle el cuadrado y encuentre una ecuación cuadrática en t . ¿Qué puede decir de su discriminante?

188. *Espacio L^2* . Use la desigualdad de Cauchy-Schwarz para demostrar que el espacio $L^2(\Omega, \mathcal{F}, P)$ consistente de todas las variables aleatorias X tales que $E|X|^2 < \infty$, es un espacio vectorial.

189. *Desigualdad de Jensen.* Sea u una función convexa, y sea X una variable aleatoria con esperanza finita. Demuestre que

$$u(E(X)) \leq E(u(X)).$$

Sugerencia: La función u es convexa si para cada a existe un número m tal que $u(x) \geq u(a) + (x - a)m$, para todo x .

190. Sea X con esperanza finita. Use la desigualdad de Jensen para demostrar que

- a) $e^{E(X)} \leq E(e^X)$.
 b) $E^2(X) \leq E(X^2)$.
 c) $\frac{1}{E(X)} \leq E\left(\frac{1}{X}\right)$, suponiendo $X > 0$.

191. Demuestre que si X es una variable aleatoria acotada casi seguramente, es decir, existe $k > 0$ tal que $P(|X| \leq k) = 1$, entonces todos los momentos de X existen.

192. Sea X una variable aleatoria con función de densidad dada por

$$f(x) = \begin{cases} \frac{n}{x^{n+1}} & \text{si } x > 1, \\ 0 & \text{otro caso.} \end{cases}$$

Demuestre que esta función es de densidad para cualquier valor natural del parámetro n . Demuestre además que tal variable aleatoria tiene momentos finitos de orden: $1, 2, \dots, n - 1$, pero el n -ésimo momento y superiores no existen.

193. *Desigualdad C_r .* Sea $r > 0$. Demuestre que

$$E|X + Y|^r \leq C_r \cdot (E|X|^r + E|Y|^r),$$

en donde

$$C_r = \begin{cases} 1 & \text{si } 0 < r \leq 1, \\ 2^{r-1} & \text{si } r > 1. \end{cases}$$

Este resultado establece que si X y Y tienen r -ésimo momento absoluto finito, entonces $X + Y$ también. Sugerencia: Demuestre primero que para cualesquiera números reales x y y ,

$$|x + y|^r \leq C_r \cdot (|x|^r + |y|^r).$$

194. *Desigualdad de Hölder.* Sean r, s tales que $r > 1$ y $1/r + 1/s = 1$. Demuestre que

$$E|XY| \leq E^{1/r}|X|^r \cdot E^{1/s}|Y|^s.$$

Sugerencia: Use la desigualdad

$$|xy| \leq \frac{|x|^r}{r} + \frac{|y|^s}{s},$$

válida para cualesquiera números reales x y y , y para r y s con las condiciones mencionadas. El caso $r = s = 2$ corresponde a la desigualdad de Cauchy-Schwarz.

195. *Desigualdad de Minkowski.* Sea $r \geq 1$. Demuestre que

$$E^{1/r} |X + Y|^r \leq E^{1/r} |X|^r + E^{1/r} |Y|^r.$$

Sugerencia: $E |X + Y|^r \leq E (|X| \cdot |X + Y|^{r-1}) + E (|Y| \cdot |X + Y|^{r-1})$, ahora use la desigualdad de Hölder.

Mediana

196. *Minimización del error absoluto medio.* A la función $g(u) = E |X - u|$ se le conoce como error absoluto medio. Demuestre que si m una mediana de X , entonces para cualquier número real u ,

$$E |X - m| \leq E |X - u|.$$

Demuestre además que la igualdad se cumple si, y sólo si, u es cualquier otra mediana de X .

Distribución uniforme discreta

197. Sea X con distribución $\text{unif}\{1, \dots, n\}$. Demuestre que

- a) $E(X) = (n + 1)/2$.
- b) $E(X^2) = (n + 1)(2n + 1)/6$.
- c) $\text{Var}(X) = (n^2 - 1)/12$.

198. Se escogen al azar y de manera independiente dos números a y b dentro del conjunto $\{1, \dots, n\}$. Demuestre que la probabilidad de que el cociente a/b sea menor o igual a uno es $(n + 1)/2n$.

Distribución Bernoulli

199. Compruebe que la función de probabilidad de la distribución $\text{Ber}(p)$ efectivamente lo es. Obtenga además la correspondiente función de distribución. Grafique ambas funciones.

200. Sea X con distribución $\text{Ber}(p)$. Demuestre que

- a) $E(X) = p$.
- b) $E(X^n) = p$, para $n \geq 1$.
- c) $\text{Var}(X) = p(1 - p)$.

Distribución binomial

201. Use el teorema del binomio para comprobar que la función de probabilidad de la distribución $\text{bin}(n, p)$ efectivamente lo es.

202. Sea X con distribución $\text{bin}(n, p)$. Demuestre que

- a) $E(X) = np$.
 b) $E(X^2) = np(1 - p + np)$.
 c) $\text{Var}(X) = np(1 - p)$.
 d) $E(X - np)^3 = np(1 - p)(1 - 2p)$.
 e) $E(X - np)^4 = 3n^2p^2(1 - p)^2 + np(1 - p)(1 - 6(1 - p)p)$.
203. Sea X con distribución $\text{bin}(n, p)$. Demuestre que $Y = n - X$ tiene distribución $\text{bin}(n, 1 - p)$.
204. Sea X con distribución $\text{bin}(n, p)$. Demuestre que
- a) $P(X = x + 1) = \frac{p}{1 - p} \cdot \frac{n - x}{x + 1} \cdot P(X = x)$.
 b) $P(X = x - 1) \cdot P(X = x + 1) \leq P^2(X = x)$.
205. Sea X con distribución $\text{bin}(n, p)$. Demuestre que
- a) $P(X \in \{1, 3, 5, \dots\}) = \frac{1}{2}(1 - (1 - 2p)^n)$.
 b) $P(X \in \{0, 2, 4, \dots\}) = \frac{1}{2}(1 + (1 - 2p)^n)$.
206. Se lanza una moneda equilibrada 6 veces. Calcule la probabilidad de que cada cara caiga exactamente 3 veces.

Distribución geométrica

207. Compruebe que la función de probabilidad de la distribución $\text{geo}(p)$ efectivamente lo es. Demuestre que la correspondiente función de distribución es

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - (1 - p)^{\lfloor x \rfloor + 1} & \text{si } x > 0. \end{cases}$$

La expresión $\lfloor x \rfloor$ denota la parte entera de x .

208. Sea X con distribución $\text{geo}(p)$. Demuestre que
- a) $E(X) = (1 - p)/p$.
 b) $\text{Var}(X) = (1 - p)/p^2$.
209. Sea X con distribución $\text{geo}(p)$. Demuestre que $P(X \geq n) = (1 - p)^{n-1}$. Ahora use este resultado y la fórmula del Ejercicio 163 en la página 82 para demostrar que $E(X) = (1 - p)/p$.
210. *Pérdida de memoria de la distribución geométrica.* Sea X con distribución $\text{geo}(p)$. Demuestre que

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

211. Sea X una variable aleatoria discreta con valores en $\{0, 1, \dots\}$ y tal que cumple la igualdad

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

Demuestre que existe un número $p \in (0, 1)$ tal que X tiene distribución $\text{geo}(p)$.

Distribución Poisson

212. Compruebe que la función de probabilidad de la distribución Poisson(λ) efectivamente lo es.
213. Sea X con distribución Poisson(λ). Demuestre que
- a) $E(X) = \lambda$.
 - b) $E(X^2) = \lambda(\lambda + 1)$.
 - c) $\text{Var}(X) = \lambda$.
 - d) $E(X^3) = \lambda E(X + 1)^2$.
214. Sean X y Y independientes ambas con distribución Poisson con parámetros λ_1 y λ_2 respectivamente. Demuestre que $X + Y$ tiene distribución Poisson($\lambda_1 + \lambda_2$).
215. Sea X con distribución Poisson(λ). Demuestre que
- a) $P(X = x + 1) = \frac{\lambda}{x + 1} \cdot P(X = x)$.
 - b) $P(X = x - 1) \cdot P(X = x + 1) \leq P^2(X = x)$.
216. Sea X con distribución Poisson(λ). Demuestre que
- a) $P(X \in \{1, 3, 5, \dots\}) = \frac{1}{2}(1 - e^{-2\lambda})$.
 - b) $P(X \in \{0, 2, 4, \dots\}) = \frac{1}{2}(1 + e^{-2\lambda})$.
217. *Teorema de Poisson: convergencia de la distribución binomial a la distribución Poisson.* Para cada entero positivo n , sea X_n con distribución bin($n, \lambda/n$) con $\lambda > 0$. Demuestre que para cada $k = 0, 1, \dots$

$$\lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Distribución binomial negativa

218. Compruebe que la función de probabilidad de la distribución bin neg(r, p) efectivamente lo es.
219. Sea X con distribución bin neg(r, p). Demuestre que
- a) $E(X) = r(1 - p)/p$.
 - b) $\text{Var}(X) = r(1 - p)/p^2$.
220. Sea $\{X_n : n \in \mathbb{N}\}$ una sucesión de variables tal que cada X_n tiene distribución bin neg(n, p) con $p = n/(\lambda + n)$ para algún $\lambda > 0$. Demuestre que para cada $k = 0, 1, 2, \dots$

$$\lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Distribución hipergeométrica

221. Compruebe que la función de probabilidad de la distribución hipergeo(N, K, n) efectivamente lo es.
222. *Convergencia de la distribución hipergeométrica a la distribución binomial.* Sea X con distribución hipergeo(N, K, n). Demuestre que cuando N y K tienden a infinito de tal forma que $K/N \rightarrow p$, y $(N - K)/N \rightarrow (1 - p)$, entonces

$$\lim_{N, K \rightarrow \infty} P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

Distribución uniforme continua

223. Compruebe que la función de densidad de la distribución unif(a, b) efectivamente lo es. Calcule además la correspondiente función de distribución. Grafique ambas funciones.
224. Sea X con distribución unif(a, b). Demuestre que
- $E(X) = (a + b)/2$.
 - $E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n + 1)(b - a)}$.
 - $\text{Var}(X) = (b - a)^2/12$.
225. Sea X con distribución unif($0, 1$). Demuestre que $E(X^n) = 1/(n + 1)$.
226. Sea X con distribución unif($-1, 1$). Demuestre que

$$E(X^n) = \begin{cases} \frac{1}{n + 1} & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

227. Sea X con distribución unif($0, 1$). Obtenga la distribución de
- $Y = 10X - 5$.
 - $Y = 4X(1 - X)$.
228. Sea X con distribución unif($0, 1$) y sea $0 < p < 1$. Demuestre que la variable aleatoria $Y = \lfloor \ln X / \ln(1 - p) \rfloor$ tiene distribución geo(p). La expresión $\lfloor x \rfloor$ denota la parte entera de x .
229. Sea X con distribución unif($0, 1$). Defina a Y como el primer dígito decimal de X . Demuestre que Y tiene distribución uniforme en el conjunto $\{0, 1, \dots, 9\}$.

Distribución exponencial

230. Compruebe que la función de densidad de la distribución $\exp(\lambda)$ efectivamente lo es. Encuentre la correspondiente función de distribución.
231. Sea X con distribución $\exp(\lambda)$. Demuestre que
- $F(x) = 1 - e^{-\lambda x}$, para $x > 0$.
 - $F(x + y) - F(y) = F(x)[1 - F(y)]$, para $x, y > 0$.
232. Sea X con distribución $\exp(\lambda)$. Demuestre que $E(X) = 1/\lambda$, y $\text{Var}(X) = 1/\lambda^2$.
233. *Pérdida de memoria de la distribución exponencial.* Sea X con distribución $\exp(\lambda)$. Demuestre que

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

La distribución exponencial es la única distribución continua que satisface esta propiedad.

234. Sea X una variable aleatoria con función de distribución $F(x)$ continua, estrictamente creciente y tal que $0 < F(x) < 1$. Demuestre que la variable aleatoria $Y = -\ln F(X)$ tiene distribución exponencial con parámetro $\lambda = 1$.
235. Sea $a > 0$. Demuestre que si X se distribuye $\exp(\lambda)$, entonces aX se distribuye $\exp(\lambda/a)$.
236. Se dice que X tiene *distribución exponencial bilateral (o doble)* con parámetro $\lambda > 0$ si su función de densidad es, para x en \mathbb{R} ,

$$f(x) = \frac{1}{2} \lambda e^{-\lambda|x|}.$$

Demuestre que $E(X) = 0$ y $\text{Var}(X) = 2/\lambda^2$.

Distribución gama

237. Compruebe que la función de densidad de la distribución $\text{gama}(n, \lambda)$ efectivamente lo es. Verifique además que esta distribución se reduce a la distribución $\exp(\lambda)$ cuando $n = 1$.
238. Sea $a > 0$. Demuestre que si X se distribuye $\text{gama}(n, \lambda)$, entonces aX se distribuye $\text{gama}(n, \lambda/a)$.
239. Sea X con distribución $\text{gama}(n, \lambda)$. Demuestre que la función de distribución de X es, para $x > 0$,

$$F(x) = 1 - \sum_{k=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!}.$$

240. Sea X con distribución $\text{gama}(n, \lambda)$. Demuestre que
- $E(X) = n/\lambda$.

- b) $E(X^m) = \frac{\Gamma(m+n)}{\lambda^m \Gamma(n)}$ para $m = 1, 2, \dots$
 c) $\text{Var}(X) = n/\lambda^2$.

241. Demuestre las siguientes propiedades de la función gama.

- a) $\Gamma(n+1) = n\Gamma(n)$.
 b) $\Gamma(n+1) = n!$ para n entero.
 c) $\Gamma(2) = \Gamma(1) = 1$.
 d) $\Gamma(1/2) = \sqrt{\pi}$.
 e) $\Gamma(n+1/2) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi}$ para n entero.

Distribución beta

242. Compruebe que la función de densidad de la distribución beta(a, b) efectivamente lo es. Verifique además que esta distribución se reduce a la distribución unif($0, 1$) cuando $a = b = 1$.

243. Sea X con distribución beta(a, b). Demuestre que

- a) $E(X) = \frac{a}{a+b}$.
 b) $E(X^n) = \frac{B(a+n, b)}{B(a, b)}$.
 c) $\text{Var}(X) = \frac{ab}{(a+b+1)(a+b)^2}$.

244. Sea X con distribución beta(a, b). Demuestre que

- a) $a = E(X) \left[\frac{E(X)(1-E(X))}{\text{Var}(X)} - 1 \right]$.
 b) $b = (1-E(X)) \left[\frac{E(X)(1-E(X))}{\text{Var}(X)} - 1 \right]$.
 c) $a+b = \frac{E(X)(1-E(X))}{\text{Var}(X)} - 1$.

245. Demuestre las siguientes propiedades de la función beta.

- a) $B(a, b) = B(b, a)$.
 b) $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.
 c) $B(a, 1) = 1/a$.
 d) $B(1, b) = 1/b$.
 e) $B(a+1, b) = \frac{a}{b} B(a, b+1)$.
 f) $B(a+1, b) = \frac{a}{a+b} B(a, b)$.
 g) $B(a, b+1) = \frac{b}{a+b} B(a, b)$.

h) $B(1/2, 1/2) = \pi$.

246. Sea X con distribución beta(1/2, 1/2). En este caso se dice que X tiene una *distribución arcoseno*.

a) Calcule y grafique $f(x)$.

b) Demuestre directamente que $f(x)$ es una función de densidad.

c) Demuestre directamente que $E(X) = 1/2$, y $\text{Var}(X) = 1/8$.

247. Sea X con distribución beta(a, b). Demuestre que para $a > 0$ y $b = 1$,

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ x^a & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

248. Sea X con distribución beta(a, b). Demuestre que para $a = 1$ y $b > 0$,

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - (1 - x)^b & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

249. Demuestre que X tiene distribución beta(a, b) si, y sólo si, $1 - X$ tiene distribución beta(b, a).

Distribución normal

250. Demuestre que la función de densidad de la distribución $N(\mu, \sigma^2)$

a) es efectivamente una función de densidad.

b) es simétrica respecto de $x = \mu$.

c) alcanza su máximo en $x = \mu$.

d) tiene puntos de inflexión en $x = \mu \pm \sigma$.

251. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que $E(X) = \mu$ y $\text{Var}(X) = \sigma^2$.

252. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que

$$E|X - \mu|^n = \begin{cases} 0 & \text{si } n \text{ es impar,} \\ 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n & \text{si } n \text{ es par.} \end{cases}$$

253. Demuestre que X tiene distribución $N(\mu, \sigma^2)$ si, y sólo si, $Z = (X - \mu)/\sigma$ tiene distribución $N(0, 1)$.

254. Sea X con distribución normal estándar. Demuestre que

$$E(X^n) = \begin{cases} 0 & \text{si } n \text{ es impar,} \\ \frac{n!}{2^{n/2}(n/2)!} & \text{si } n \text{ es par.} \end{cases}$$

255. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que $Y = aX + b$, con $a \neq 0$, tiene una distribución normal. Encuentre los parámetros correspondientes.
256. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que la variable aleatoria $-X$ también tiene una distribución normal. Encuentre los parámetros correspondientes.
257. Sea X con distribución normal estándar. Demuestre que X^2 tiene una distribución $\chi^2(1)$. Recíprocamente, ¿será cierto que si Y tiene distribución $\chi^2(1)$ entonces \sqrt{Y} tiene distribución $N(0, 1)$?
258. Sea X con distribución normal estándar. Encuentre la función de densidad de la variable aleatoria $|X|$.
259. *El cociente de Mills*. Sea $\phi(x)$ la función de densidad de la distribución normal estándar, y sea $\Phi(x)$ la correspondiente función de distribución. Demuestre que

a) $\phi'(x) + x\phi(x) = 0$.

b) $\frac{1}{x} - \frac{1}{x^3} < \frac{1 - \Phi(x)}{\phi(x)} < \frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5}$, para $x > 0$.

Distribución log normal

260. Demuestre que la función de densidad de una distribución log normal (μ, σ^2) efectivamente lo es.
261. Sea X con distribución log normal (μ, σ^2) . Demuestre que
- a) $E(X) = \exp(\mu + \sigma^2/2)$.
- b) $\text{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$.
- c) $E(\ln X) = \mu$.
- d) $\text{Var}(\ln X) = \sigma^2$.

Capítulo 3

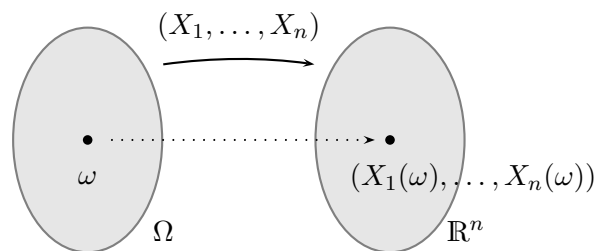
Vectores aleatorios

En este capítulo se extiende el concepto de variable aleatoria con valores reales a variables aleatorias con valores en \mathbb{R}^n . Se estudian además algunos conceptos importantes relacionados. Recuerde que en este capítulo, y a lo largo del texto, se tiene siempre como elemento base un espacio de probabilidad (Ω, \mathcal{F}, P) .

3.1. Vectores aleatorios

Definición (Vector aleatorio). Un vector aleatorio es una función $X : \Omega \rightarrow \mathbb{R}^n$ tal que para cualquier conjunto B en $\mathcal{B}(\mathbb{R}^n)$, se cumple que $X^{-1}B$ es un elemento de \mathcal{F} .

Todo vector aleatorio se puede representar en la forma $X = (X_1, \dots, X_n)$ en donde cada coordenada es una función de Ω en \mathbb{R} .



Un vector aleatorio es una función de Ω en \mathbb{R}^n .

Se demuestra a continuación que la condición que aparece en la definición anterior es equivalente a solicitar que cada coordenada del vector sea una variable aleatoria.

Proposición. Una función $X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ es un vector aleatorio si, y sólo si, cada coordenada es una variable aleatoria.

Demostración. Sea (X_1, \dots, X_n) un vector aleatorio. Entonces la imagen inversa de cualquier conjunto de Borel de \mathbb{R}^n es un elemento de la σ -álgebra del espacio de probabilidad. En particular, la imagen inversa del conjunto $B \times \Omega \times \dots \times \Omega$ pertenece a \mathcal{F} , para cualquier Boreliano B de \mathbb{R} . Pero esta imagen inversa es simplemente $X_1^{-1}B$. Esto demuestra que X_1 es variable aleatoria. De manera análoga se procede con las otras coordenadas del vector. Suponga ahora que cada coordenada de una función $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ es una variable aleatoria. Considere la colección $\mathcal{B} = \{B \in \mathcal{B}(\mathbb{R}^n) : (X_1, \dots, X_n)^{-1}B \in \mathcal{F}\}$. Como cada coordenada es una variable aleatoria, los conjuntos de Borel de \mathbb{R}^n de la forma $B_1 \times \dots \times B_n$, en donde cada factor de este producto es un Boreliano de \mathbb{R} , es un elemento de la colección \mathcal{B} . Entonces

$$\mathcal{B}(\mathbb{R}) \times \dots \times \mathcal{B}(\mathbb{R}) \subseteq \mathcal{B} \subseteq \mathcal{B}(\mathbb{R}^n).$$

Es fácil demostrar que la colección \mathcal{B} es una σ -álgebra. Así que

$$\sigma(\mathcal{B}(\mathbb{R}) \times \dots \times \mathcal{B}(\mathbb{R})) \subseteq \mathcal{B} \subseteq \mathcal{B}(\mathbb{R}^n).$$

Pero ambos extremos de esta ecuación coinciden. De modo que $\mathcal{B} = \mathcal{B}(\mathbb{R}^n)$, y por lo tanto la función (X_1, \dots, X_n) es un vector aleatorio. \square

En consecuencia, es correcto definir un vector aleatorio simplemente como un vector de variables aleatorias. Para simplificar la escritura donde sea posible se usan únicamente vectores aleatorios bidimensionales, esto es, de la forma (X, Y) . En la mayoría de los casos, las definiciones y resultados son fácilmente extendidos a dimensiones mayores. En general, sólo consideraremos vectores aleatorios como los que se definen a continuación.

Definición (Vector discreto y continuo). Se dice que el vector (X, Y) es discreto si cada coordenada es una variable aleatoria discreta, y se dice que es continuo en caso de que cada coordenada lo sea.

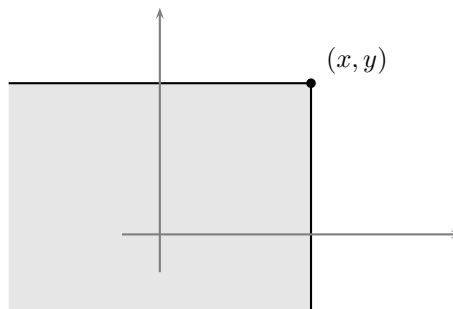
3.2. Distribución conjunta

A menudo es necesario considerar probabilidades de eventos que involucran a dos o más variables aleatorias a un mismo tiempo. El concepto fundamental en este caso es el de función de distribución conjunta.

Definición (Función de distribución conjunta). La función de distribución de un vector (X, Y) , denotada por $F(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$, se define como sigue

$$F(x, y) = P(X \leq x, Y \leq y).$$

El número $F(x, y)$ es entonces la probabilidad de que el vector aleatorio tome algún valor en la región $(-\infty, x] \times (-\infty, y]$, la cual se muestra a continuación:



El número $F(x, y) = P(X \leq x, Y \leq y)$ es la probabilidad de que el vector (X, Y) tome un valor en la región sombreada.

En palabras, la función $F(x, y)$ es la probabilidad de que X sea menor o igual a x , y al mismo tiempo Y sea menor o igual a y . Esto es simplemente la probabilidad del evento $(X \leq x) \cap (Y \leq y)$. A la función $F(x, y)$ se le conoce también como *función de distribución bivariada* de X y Y , y en general a la distribución conjunta de un vector aleatorio de cualquier dimensión finita se le llama *distribución multivariada*. Naturalmente, en el caso unidimensional, la distribución se llama *univariada*. Cuando sea necesario especificarlo se escribe $F_{X,Y}(x, y)$ en lugar de $F(x, y)$, y es evidente la forma de extender la definición para el caso de un vector aleatorio de más de dos coordenadas. Con el fin de mantener la notación simple, se mantiene la correspondencia de las letras, es decir, x es un valor asociado a X , y y esta asociada a Y .

Las funciones de distribución conjunta satisfacen propiedades semejantes al caso unidimensional, se estudian a continuación algunas de ellas.

Proposición. La distribución conjunta $F(x, y)$ satisface las siguientes propiedades.

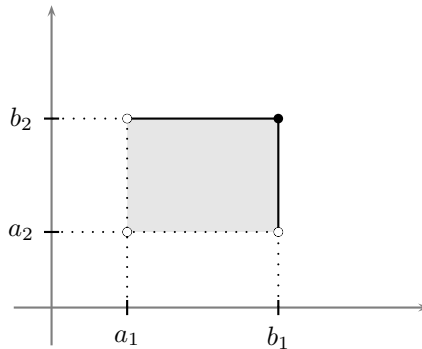
1. $\lim_{x, y \rightarrow \infty} F(x, y) = 1$, ambas variables.
2. $\lim_{x, y \rightarrow -\infty} F(x, y) = 0$, alguna de las variables.
3. $F(x, y)$ es no decreciente en cada variable.
4. $F(x, y)$ es continua por la derecha en cada variable.
5. Si $a_1 < b_1$ y $a_2 < b_2$, entonces

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0.$$

La demostración de las propiedades (1)-(4) es completamente análoga al caso unidimensional y por tanto la omitiremos. Respecto a la propiedad (5) observe que la expresión

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

corresponde a la probabilidad del evento $(a_1 < X \leq b_1, a_2 < Y \leq b_2)$. De modo que (5) se traduce simplemente en solicitar que la probabilidad de que (X, Y) tome valores en el rectángulo $(a_1, b_1] \times (a_2, b_2]$, sea no negativa. Este rectángulo se muestra en la siguiente figura.



La probabilidad asociada al rectángulo $(a_1, b_1] \times (a_2, b_2]$ es
 $P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$.

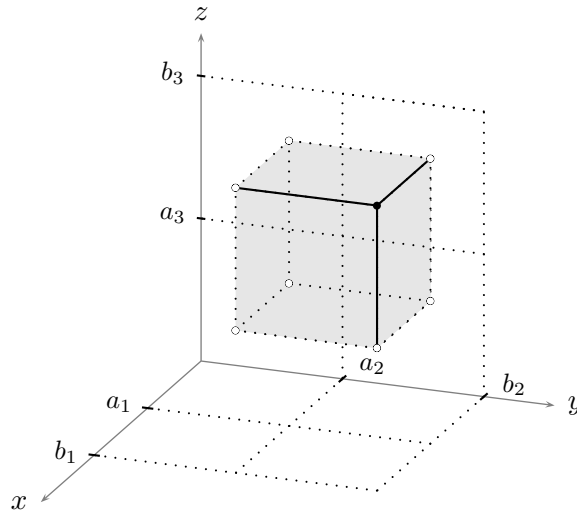
A diferencia del caso unidimensional, las propiedades (1) a (4) no son suficientes para asegurar que una función $F(x, y)$ asigna probabilidad no negativa a cualquier rectángulo. Por ejemplo, el Ejercicio 264 en la página 120 muestra una situación en donde esa condición falla. Por tanto en el caso de dimensión dos y superior, es necesario asegurarse de que tal propiedad se cumple.

Definición (Función de distribución conjunta). Una función cualquiera $F(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$, no necesariamente definida en términos de un vector aleatorio, es una función de distribución conjunta si cumple con las cinco propiedades enunciadas en la proposición anterior.

Para tres dimensiones se dice que $F(x_1, x_2, x_3) : \mathbb{R}^3 \rightarrow [0, 1]$ es una función de distribución si cumple las primeras cuatro propiedades anteriores y la quinta se reemplaza por la siguiente condición: Para cualesquiera números reales $a_1 < b_1$, $a_2 < b_2$, y $a_3 < b_3$,

$$\begin{aligned} &F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) \\ &+ F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) \\ &- F(a_1, a_2, a_3) \geq 0. \end{aligned}$$

Se puede demostrar que el lado izquierdo de esta desigualdad corresponde a la probabilidad del evento $(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3)$, y entonces el requisito naturalmente es que este número sea no negativo.



Región $(a_1, b_1] \times (a_2, b_2] \times (a_3, b_3]$.

Más generalmente, se tiene la siguiente definición.

Definición. Una función $F : \mathbb{R}^n \rightarrow [0, 1]$ es una función de distribución si cumple las primeras cuatro propiedades anteriores y, adicionalmente, para cualesquiera números reales $a_1 < b_1, a_2 < b_2, \dots, a_n < b_n$,

$$\sum_{x_i \in \{a_i, b_i\}} (-1)^{\#a} F(x_1, \dots, x_n) \geq 0,$$

en donde $\#a$ es el número de veces que alguna de las variables x_i toma el valor a_i en la evaluación de la función F .

Nuevamente la suma que aparece en esta definición corresponde a la probabilidad del evento $(a_1 < X_1 \leq b_1, \dots, a_n < X_n \leq b_n)$, y la condición requiere simplemente que este número sea no negativo. Finalmente enunciamos un resultado que establece la importancia de la función de distribución, y cuya demostración, que puede ser encontrada por ejemplo en [16], es análoga al caso unidimensional.

Proposición. Sea $F : \mathbb{R}^n \rightarrow [0, 1]$ una función de distribución. Entonces existe un espacio de probabilidad, y un vector aleatorio, cuya función de distribución es F .

3.3. Densidad conjunta

Como en el caso unidimensional, algunos vectores tienen asociada otra función llamada de probabilidad y la cual se define a continuación.

Definición (Función de probabilidad conjunta). La función de probabilidad de un vector discreto (X, Y) es la función $f(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$ dada por

$$f(x, y) = P(X = x, Y = y).$$

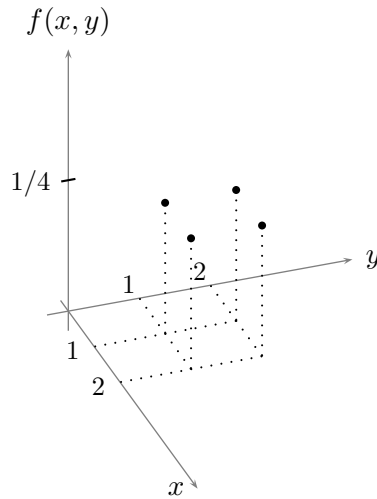
A esta función también se le llama función de probabilidad conjunta de X y Y .

Es evidente que la función de probabilidad de un vector discreto es una función no negativa y tal que

$$\sum_x \sum_y f(x, y) = 1.$$

Recíprocamente, toda función no negativa $f(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$ que sea estrictamente positiva únicamente en un subconjunto discreto de \mathbb{R}^2 y que sume uno, se llama *función de probabilidad conjunta*. La definición de la función de probabilidad para el caso discreto multidimensional es evidente.

Ejemplo. La función $f(x, y) = 1/4$, para $x, y = 1, 2$, es una función de probabilidad conjunta pues es no negativa y suma uno, corresponde a la distribución uniforme sobre el conjunto $\{1, 2\} \times \{1, 2\}$. La gráfica se muestra a continuación:

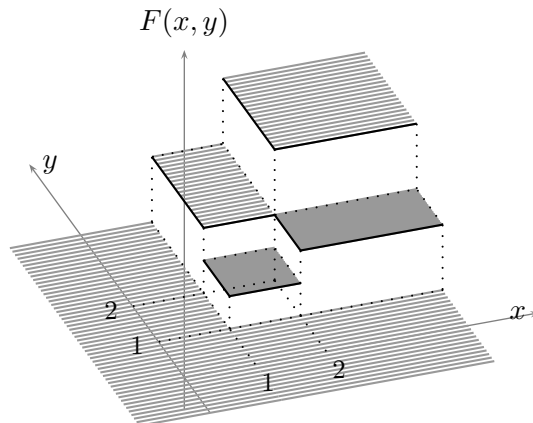


Función de probabilidad $f(x, y) = 1/4$, para $x, y = 1, 2$.

La correspondiente función de distribución es

$$F(x, y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v) = \begin{cases} 0 & \text{si } x < 1 \text{ ó } y < 1, \\ 1/4 & \text{si } 1 \leq x < 2, 1 \leq y < 2, \\ 2/4 & \text{si } 1 \leq x < 2, y \geq 2, \\ 2/4 & \text{si } x \geq 2, 1 \leq y < 2, \\ 1 & \text{si } x \geq 2 \text{ y } y \geq 2, \end{cases}$$

cuya gráfica es:



Función de distribución $F(x, y)$.

○

Ejemplo. La función definida por $f(x, y) = (1/2)^{x+y}$ para $x, y \in \mathbb{N}$, e idénticamente cero fuera de este conjunto discreto, es una función de probabilidad bivariada pues es no negativa y suma uno. En efecto,

$$\sum_{x,y=1}^{\infty} f(x, y) = \sum_{x,y=1}^{\infty} \frac{1}{2^{x+y}} = \left(\sum_{x=1}^{\infty} \frac{1}{2^x} \right)^2 = 1.$$

○

Para el caso de vectores continuos se tiene la siguiente definición.

Definición (Función de densidad conjunta). Sea (X, Y) un vector continuo con función de distribución $F(x, y)$. Se dice que (X, Y) es absolutamente continuo si existe una función no negativa e integrable $f(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$, tal que, para todo (x, y) en \mathbb{R}^2 , se cumple la igualdad

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du.$$

A la función $f(x, y)$ se le denota por $f_{X,Y}(x, y)$, y se le llama función de densidad conjunta de X y Y .

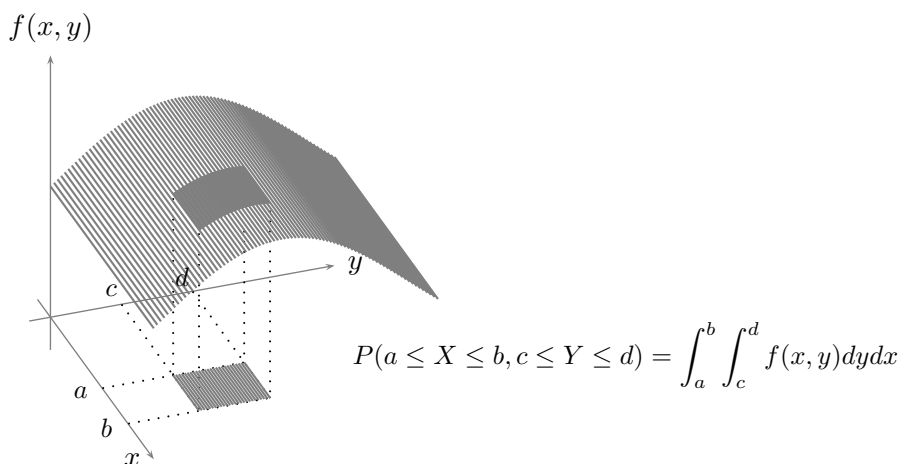
Así como en el caso unidimensional, no existe realmente unicidad para la función de densidad pues basta modificarla en algunos puntos para ser distinta pero seguir cumpliendo la igualdad anterior, la función de distribución y por tanto las probabilidades, permanecen sin cambio alguno. Es claro que la función de densidad conjunta $f(x, y)$ de un vector absolutamente continuo es no negativa y cumple la condición:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$$

Recíprocamente, toda función no negativa $f : \mathbb{R}^2 \rightarrow [0, \infty)$, que integre uno, se llama *función de densidad conjunta*. En particular, cuando $f(x, y)$ es continua:

$$f(x, y) = \frac{\partial^2}{\partial y \partial x} F(x, y).$$

Observe que, en el caso absolutamente continuo y conociendo la función de densidad conjunta, la probabilidad del evento $(a \leq X \leq b, c \leq Y \leq d)$ se calcula como la integral doble que se ilustra en la siguiente figura:

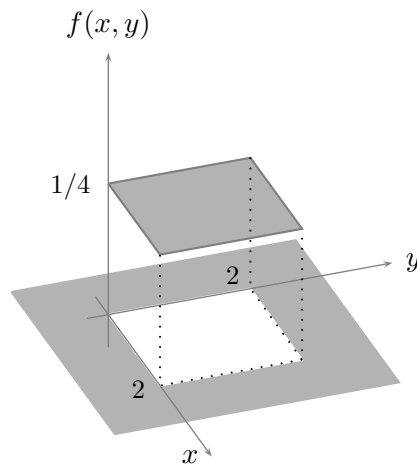


La probabilidad como el volumen bajo una superficie.

Ejemplo. La función $f : \mathbb{R}^2 \rightarrow [0, \infty)$ dada por

$$f(x, y) = \begin{cases} \frac{1}{4} & \text{si } x, y \in [0, 2], \\ 0 & \text{otro caso,} \end{cases}$$

es una función de densidad pues es no negativa e integra uno. Esta función de densidad conjunta corresponde a la distribución uniforme del vector (X, Y) en el cuadrado $[0, 2] \times [0, 2]$. La gráfica se muestra a continuación:

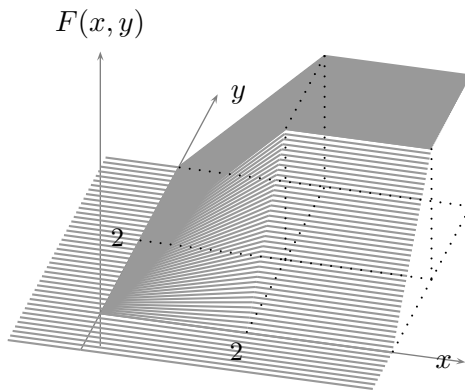


Función de densidad $f(x, y) = 1/4$, para $x, y \in [0, 2]$.

Calculando la doble integral para los distintos valores de x y y , se encuentra que la función de distribución conjunta es:

$$\begin{aligned}
 F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du \\
 &= \begin{cases} 0 & \text{si } x < 0 \text{ ó } y < 0, \\ xy/4 & \text{si } 0 \leq x, y \leq 2, \\ x/2 & \text{si } 0 \leq x \leq 2, y \geq 2, \\ y/2 & \text{si } 0 \leq y \leq 2, x \geq 2, \\ 1 & \text{si } x \geq 2 \text{ y } y \geq 2. \end{cases}
 \end{aligned}$$

Gráficamente:



Función de distribución $F(x, y)$.

o

3.4. Distribución marginal

Dada la función de distribución conjunta $F(x, y)$ de un vector aleatorio, es posible obtener la función de distribución de cada variable aleatoria por separado mediante el siguiente procedimiento.

Definición (Función de distribución marginal). Sea (X, Y) un vector con función de distribución $F(x, y)$. A la función

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

se le conoce como la función de distribución marginal de X . Análogamente se define la función de distribución marginal de Y como

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

No es difícil verificar que las funciones de distribución marginales son efectivamente funciones de distribución univariadas. En el caso de que se tenga una función de densidad conjunta, se pueden obtener las funciones de densidad individuales como indica la siguiente definición.

Definición (Función de densidad marginal). Sea (X, Y) un vector absolutamente continuo con función de densidad $f(x, y)$. A la función

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

se le conoce como la función de densidad marginal de X . Análogamente se define la función de densidad marginal de Y como

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

Si (X, Y) es un vector discreto la integral se reemplaza por una suma.

Tampoco es difícil comprobar que las funciones de densidad marginales son efectivamente funciones de densidad univariadas. Las dos definiciones anteriores pueden extenderse de manera evidente cuando se tenga un vector aleatorio de cualquier dimensión finita.

3.5. Distribución condicional

La siguiente definición es una extensión del concepto elemental de probabilidad condicional de eventos.

Definición (Función de densidad condicional). Sea (X, Y) un vector con función de densidad $f_{X,Y}(x, y)$, y sea y tal que $f_Y(y) \neq 0$. A la función

$$x \mapsto f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

se le conoce como la función de densidad condicional de X dado que Y toma el valor y .

No es difícil comprobar que la función $x \mapsto f_{X|Y}(x|y)$ es efectivamente una función de densidad, tanto en el caso discreto como en el continuo. Observe que el valor y permanece fijo y la función es vista como una función de la variable real x . Se pueden definir también funciones de distribución condicionales de la siguiente forma.

Definición (Función de distribución condicional). Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad $f_{X,Y}(x, y)$, y sea y tal que $f_Y(y) \neq 0$. A la función

$$x \mapsto F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(u|y) du$$

se le conoce como la función de distribución condicional de X dado que Y toma el valor y . Cuando el vector aleatorio (X, Y) es discreto la integral se substituye por la suma correspondiente.

Nuevamente resulta que la función $x \mapsto F_{X|Y}(x|y)$ es efectivamente una función de distribución. En el caso absolutamente continuo tenemos la relación

$$f_{X|Y}(x|y) = \frac{\partial}{\partial x} F_{X|Y}(x|y).$$

3.6. Independencia

Podemos ahora definir el importante concepto de independencia de variables aleatorias.

Definición (Independencia de dos variables aleatorias). Se dice que X y Y son independientes, y a menudo se escribe $X \perp Y$, si para cada par de conjuntos Borel medibles A, B , se cumple la igualdad

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B). \quad (3.1)$$

En términos de la siempre existente función de distribución, la independencia de dos variables aleatorias se puede expresar como indica el siguiente resultado.

Proposición (Independencia de dos variables aleatorias). Las variables aleatorias X y Y son independientes si, y sólo si, para cada (x, y) en \mathbb{R}^2 se cumple la igualdad

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y). \quad (3.2)$$

Demostración. Si X y Y son independientes, entonces tomando $A = (-\infty, x]$ y $B = (-\infty, y]$ en (3.1) se obtiene (3.2). Suponga ahora que se cumple (3.2) para cualesquiera x y y en \mathbb{R} . Defina la colección

$$\mathcal{A} = \{A \in \mathcal{B}(\mathbb{R}) : P(X \in A, Y \leq y) = P(X \in A) \cdot P(Y \leq y), \\ \text{para todo } y \in \mathbb{R}\}.$$

Usando la hipótesis es posible demostrar que $\mathcal{A} = \mathcal{B}(\mathbb{R})$. Sea ahora A un elemento cualquiera fijo de $\mathcal{B}(\mathbb{R})$. Defina la colección

$$\mathcal{B} = \{B \in \mathcal{B}(\mathbb{R}) : P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)\}.$$

Resulta nuevamente que $\mathcal{B} = \mathcal{B}(\mathbb{R})$, y de esta forma, para cualquier A y para cualquier B en $\mathcal{B}(\mathbb{R})$, se cumple (3.1). \square

El concepto de independencia de variables aleatorias es una extensión de la misma propiedad para eventos. Cuando la función de densidad conjunta $f_{X,Y}(x, y)$ existe, la condición de independencia de X y Y es equivalente a solicitar que para cualesquiera números reales x y y , se cumpla la identidad

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y). \quad (3.3)$$

En el caso discreto, la afirmación anterior es completamente correcta. Para el caso continuo hay una observación técnica que es necesario mencionar. Como en este caso las funciones de densidad pueden ser modificadas sin que cambie la función de distribución asociada, la igualdad (3.3) puede no cumplirse para cada $(x, y) \in \mathbb{R}^2$, entonces se permite que la igualdad no se cumpla en un conjunto de medida de Lebesgue cero, por ejemplo, un conjunto numerable de parejas (x, y) en \mathbb{R}^2 , y entonces habrá independencia en el caso continuo si se cumple (3.3), salvo conjuntos de medida de Lebesgue cero.

El concepto de independencia puede ser extendido claramente al caso de varias variables aleatorias de la forma siguiente:

Definición (Independencia de n variables aleatorias). Se dice que las variables X_1, X_2, \dots, X_n son independientes si para cualesquiera Borelianos A_1, A_2, \dots, A_n , se cumple

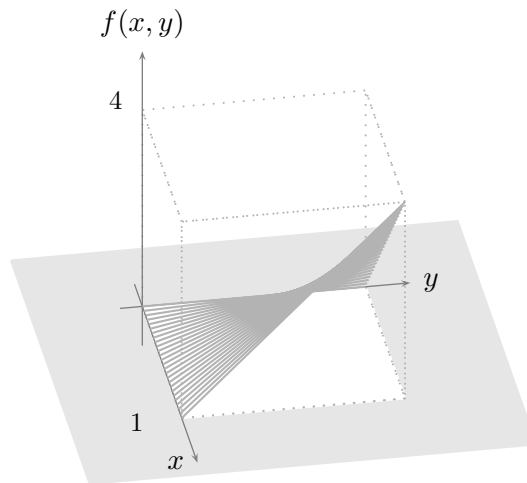
$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

Equivalentemente, si para cualquier (x_1, x_2, \dots, x_n) en \mathbb{R}^n se cumple

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdots F_{X_n}(x_n).$$

Más aún, una colección infinita de variables aleatorias es independiente si cualquier subconjunto finito de ella lo es.

Ejemplo. Sea (X, Y) un vector aleatorio con función de densidad $f(x, y) = 4xy$ para $0 \leq x, y \leq 1$. La gráfica de esta función aparece en la siguiente figura.



Función de densidad $f(x, y) = 4xy$ para $0 \leq x, y \leq 1$.

La función de densidad marginal de X se calcula de la siguiente forma. Para $0 \leq x \leq 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 4xy dy = 2x.$$

Por lo tanto, $f_X(x) = 2x$ para $0 \leq x \leq 1$. Análogamente $f_Y(y) = 2y$ para $0 \leq y \leq 1$. En consecuencia, X y Y son independientes pues para cada par (x, y) , se cumple $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$. \circ

Proposición. Sean X y Y independientes, y sean g y h dos funciones de \mathbb{R} en \mathbb{R} , Borel medibles. Entonces las variables $g(X)$ y $h(Y)$ también son independientes.

Demostración. Sean A y B cualesquiera dos conjuntos Borel medibles. Entonces

$$\begin{aligned} P(g(X) \in A, h(Y) \in B) &= P(X \in g^{-1}(A), Y \in h^{-1}(B)) \\ &= P(X \in g^{-1}(A)) \cdot P(Y \in h^{-1}(B)) \\ &= P(g(X) \in A) \cdot P(h(Y) \in B). \end{aligned}$$

□

Este resultado puede extenderse fácilmente al caso n -dimensional, y obtener que la composición de n funciones Borel medibles aplicadas, respectivamente, a n variables aleatorias independientes, produce nuevamente variables aleatorias independientes.

La definición de independencia de dos variables aleatorias puede extenderse al caso de dos vectores aleatorios de cualquier dimensión de la forma siguiente, y puede extenderse un poco más para incluir la independencia de un número finito de vectores aleatorios no necesariamente de la misma dimensión.

Definición (Independencia de dos vectores aleatorios). Se dice que los vectores aleatorios $X = (X_1, \dots, X_n)$ y $Y = (Y_1, \dots, Y_m)$ son independientes, si para cada A en $\mathcal{B}(\mathbb{R}^n)$, y cada B en $\mathcal{B}(\mathbb{R}^m)$, se cumple la igualdad

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B). \quad (3.4)$$

3.7. Esperanza de una función de un vector aleatorio

Si (X, Y) es un vector aleatorio y $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ es una función Borel medible, entonces $\varphi(X, Y)$ es una variable aleatoria y el problema nuevamente es encontrar su esperanza. Usando directamente la definición, la esperanza de $\varphi(X, Y)$ se calcula del siguiente modo:

$$E[\varphi(X, Y)] = \int_{-\infty}^{\infty} x dF_{\varphi(X, Y)}(x),$$

pero, así como en el caso unidimensional, ello requiere encontrar primero la distribución de $\varphi(X, Y)$, lo cual puede ser difícil. El siguiente resultado establece una forma alternativa de calcular la esperanza de $\varphi(X, Y)$, sin conocer su distribución, pero conociendo, por supuesto, la distribución del vector (X, Y) .

Teorema (Esperanza de una función de un vector aleatorio). Sea (X, Y) un vector aleatorio, y sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ una función Borel medible tal que la variable aleatoria $\varphi(X, Y)$ tiene esperanza finita. Entonces

$$E[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) dF_{X,Y}(x, y). \quad (3.5)$$

Nuevamente omitiremos la demostración de este resultado. Cuando el vector (X, Y) es discreto, la fórmula (3.5) se reduce a

$$E[\varphi(X, Y)] = \sum_{x,y} \varphi(x, y)P(X = x, Y = y),$$

en donde la suma se efectúa sobre todos los posibles valores (x, y) del vector. En este caso la demostración del teorema resulta no muy complicada, y se pide dar los detalles en el Ejercicio 328. En el caso cuando (X, Y) es absolutamente continuo, la expresión (3.5) se escribe

$$E[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) f_{X,Y}(x, y) dx dy.$$

Con ayuda de esta resultado podemos ahora demostrar que la esperanza separa sumas.

Proposición. Sean X y Y con esperanza finita. Entonces

$$E(X + Y) = E(X) + E(Y).$$

Demostración. Sean $\varphi(x, y) = x + y$, $\varphi_1(x, y) = x$, y $\varphi_2(x, y) = y$. Entonces

$$\begin{aligned} E(X + Y) &= E(\varphi(X, Y)) \\ &= \int_{\mathbb{R}^2} (x + y) dF_{X,Y}(x, y) \\ &= \int_{\mathbb{R}^2} x dF_{X,Y}(x, y) + \int_{\mathbb{R}^2} y dF_{X,Y}(x, y) \\ &= E(\varphi_1(X, Y)) + E(\varphi_2(X, Y)) \\ &= E(X) + E(Y). \end{aligned}$$

□

Proposición. Sean X y Y independientes, y sean g y h dos funciones Borel medibles tales que $g(X)$ y $h(Y)$ tienen esperanza finita. Entonces

$$E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)].$$

En particular, cuando X y Y son independientes,

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Demostración.

$$\begin{aligned} E[g(X) \cdot h(Y)] &= \int_{\mathbb{R}^2} g(x) \cdot h(y) dF_{X,Y}(x, y) \\ &= \int_{\mathbb{R}^2} g(x) \cdot h(y) dF_X(x) dF_Y(y) \\ &= E[g(X)] \cdot E[h(Y)]. \end{aligned}$$

□

Es interesante observar que el recíproco de la afirmación anterior es, en general, falso.

Ejemplo. Considere el vector aleatorio discreto (X, Y) con función de probabilidad

| | | | |
|------------------|-----|-----|-----|
| $x \backslash y$ | -1 | 0 | 1 |
| -1 | 1/5 | 0 | 1/5 |
| 0 | 0 | 1/5 | 0 |
| 1 | 1/5 | 0 | 1/5 |

Entonces es sencillo verificar que $E(XY) = E(X)E(Y) = 0$, sin embargo X y Y no son independientes pues $P(X = 0, Y = 0) = 1/5$, mientras que $P(X = 0)P(Y = 0) = 1/25$. Otros ejemplos de esta misma situación pueden encontrarse en el Ejercicio 329 en la página 130. ◻

3.8. Covarianza

En esta sección se define y estudia la covarianza entre dos variables aleatorias. Una interpretación de este número, ligeramente modificado, será dada en la siguiente sección.

Definición (Covarianza). La *covarianza* de X y Y , denotada por $\text{Cov}(X, Y)$, es el número

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

Para que la definición anterior tenga sentido es necesario suponer que las esperanzas $E(X)$, $E(Y)$ y $E(XY)$ son finitas. Se revisan a continuación algunas propiedades de la covarianza.

Proposición. La covarianza satisface las siguientes propiedades.

1. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
3. $\text{Cov}(X, X) = \text{Var}(X)$.
4. $\text{Cov}(a, Y) = 0$, a constante.
5. $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$, a constante.
6. $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.
7. X, Y independientes $\Rightarrow \text{Cov}(X, Y) = 0$.
8. En general, $\text{Cov}(X, Y) = 0 \not\Rightarrow X, Y$ independientes.

Demostración. Para probar (1) se usa la propiedad lineal de la esperanza,

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E[XY - YE(X) - XE(Y) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Las propiedades (2), (3) y (4) se siguen directamente de la definición, lo mismo que (5) y (6) al hacer uso de las propiedades de linealidad de la esperanza. La proposición (7) se obtiene fácilmente de (1) pues $E(XY) = E(X)E(Y)$ cuando X y Y son independientes. Finalmente damos un ejemplo para (8). Sea (X, Y) un vector aleatorio discreto con función de densidad

$$f_{X,Y}(x, y) = \begin{cases} 1/8 & \text{si } (x, y) \in \{(-1, -1), (-1, 1), (1, -1), (1, 1)\}, \\ 1/2 & \text{si } (x, y) = (0, 0), \\ 0 & \text{otro caso.} \end{cases}$$

Entonces X y Y tienen idénticas densidades marginales,

$$f_X(x) = \begin{cases} 1/4 & \text{si } x \in \{-1, 1\}, \\ 1/2 & \text{si } x = 0, \\ 0 & \text{otro caso.} \end{cases} \quad f_Y(y) = \begin{cases} 1/4 & \text{si } y \in \{-1, 1\}, \\ 1/2 & \text{si } y = 0, \\ 0 & \text{otro caso.} \end{cases}$$

Puede entonces comprobarse que $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$. Sin embargo X y Y no son independientes pues en particular $P(X = 0, Y = 0) = 1/2$, mientras que $P(X = 0)P(Y = 0) = 1/4$. \square

3.9. Coeficiente de correlación

El coeficiente de correlación de dos variables aleatorias es un número que mide el grado de *dependencia lineal* que existe entre ellas. Su definición es la siguiente.

Definición (Coeficiente de correlación). El coeficiente de correlación de las variables aleatorias X y Y , denotado por $\rho(X, Y)$, es el número

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Naturalmente en esta definición se necesita suponer que las varianzas son estrictamente positivas y finitas. La interpretación dada al coeficiente de correlación se justifica a partir de los siguientes resultados.

Proposición. El coeficiente de correlación satisface las siguientes propiedades.

1. Si X y Y son independientes, entonces $\rho(X, Y) = 0$.
2. $-1 \leq \rho(X, Y) \leq 1$.
3. $|\rho(X, Y)| = 1$ si, y sólo si, existen constantes a y b tales que, con probabilidad uno, $Y = aX + b$, con $a > 0$ si $\rho(X, Y) = 1$, y $a < 0$ si $\rho(X, Y) = -1$.

Demostración. (1) Si X y Y son independientes, entonces $\text{Cov}(X, Y) = 0$, y por lo tanto $\rho(X, Y) = 0$. (2) Suponga primero que X y Y son tales que $E(X) = E(Y) = 0$, y $\text{Var}(X) = \text{Var}(Y) = 1$. Para cualquier valor de λ ,

$$\begin{aligned} 0 &\leq \text{Var}(X + \lambda Y) \\ &= E[(X + \lambda Y)^2] - E^2[X + \lambda Y] \\ &= 1 + 2\lambda E(XY) + \lambda^2. \end{aligned}$$

El caso $\lambda = 1$ produce el resultado $E(XY) \geq -1$, mientras que para $\lambda = -1$ se obtiene $E(XY) \leq 1$. Es decir, $-1 \leq E(XY) \leq 1$. Ahora se aplica este resultado a

las variables aleatorias

$$\frac{X - \mu_X}{\sigma_X} \quad \text{y} \quad \frac{Y - \mu_Y}{\sigma_Y},$$

que evidentemente son centradas y con varianza unitaria. Entonces

$$-1 \leq E \left[\left(\frac{X - \mu_X}{\sigma_X} \right) \left(\frac{Y - \mu_Y}{\sigma_Y} \right) \right] \leq 1.$$

Esto es precisamente lo enunciado en (2) pues el término de enmedio es $\rho(X, Y)$. Ahora se demuestra (3). Si X y Y son tales que $Y = aX + b$ con $a \neq 0$ y b constantes, entonces

$$\rho(X, Y) = \frac{\text{Cov}(X, aX + b)}{\sqrt{\text{Var}(X)\text{Var}(aX + b)}} = \frac{a}{|a|}.$$

Por lo tanto $\rho(X, Y) = 1$ cuando $a > 0$, y $\rho(X, Y) = -1$ cuando $a < 0$. Inversamente, suponga que X y Y son tales que $|\rho(X, Y)| = 1$. Defina

$$U = \frac{X - \mu_X}{\sigma_X} \quad \text{y} \quad V = \frac{Y - \mu_Y}{\sigma_Y}.$$

Entonces claramente $E(U) = E(V) = 0$, y $\text{Var}(U) = \text{Var}(V) = 1$. Por lo tanto $\rho(U, V) = E(UV)$. Es fácil ver también que $|\rho(U, V)| = |\rho(X, Y)| = 1$. Si $\rho(U, V) = 1$, entonces

$$\begin{aligned} \text{Var}(U - V) &= E[(U - V)^2] - E^2(U - V) \\ &= E[(U - V)^2] \\ &= 2[1 - E(UV)] \\ &= 0. \end{aligned}$$

Esto significa que con probabilidad uno, la v.a. $U - V$ es constante. Esto es, para alguna constante c , con probabilidad uno, $U - V = c$. Pero esta constante c debe ser cero pues $E(U - V) = 0$. Por lo tanto,

$$\frac{X - \mu_X}{\sigma_X} = \frac{Y - \mu_Y}{\sigma_Y},$$

de donde se obtiene $Y = \mu_Y + (X - \mu_X)\sigma_Y/\sigma_X$. Esto establece una relación lineal directa entre X y Y . En cambio, si $\rho(U, V) = -1$ entonces

$$\begin{aligned} \text{Var}(U + V) &= E[(U + V)^2] - E^2(U + V) \\ &= E[(U + V)^2] \\ &= 2[1 + E(UV)] \\ &= 0. \end{aligned}$$

Esto significa nuevamente que con probabilidad uno, la v.a. $U + V$ es constante. Esto es, para alguna constante c , con probabilidad uno, $U + V = c$. Nuevamente la constante c es cero pues $E(U + V) = 0$. Por lo tanto,

$$\frac{X - \mu_X}{\sigma_X} = -\frac{Y - \mu_Y}{\sigma_Y},$$

de donde se obtiene $Y = \mu_Y - (X - \mu_X)\sigma_Y/\sigma_X$. Esto establece una relación lineal, ahora inversa, entre X y Y . Uniendo los últimos dos resultados se obtiene que, cuando $|\rho(X, Y)| = 1$, con probabilidad uno,

$$Y = \underbrace{(\rho(X, Y)\frac{\sigma_Y}{\sigma_X})}_a X + \underbrace{(\mu_Y - \rho(X, Y)\mu_X\frac{\sigma_Y}{\sigma_X})}_b.$$

□

Definición (Correlación positiva, negativa o nula). Cuando $\rho(X, Y) = 0$ se dice que X y Y son no correlacionadas. Cuando $|\rho(X, Y)| = 1$ se dice que X y Y están perfectamente correlacionadas positiva o negativamente, de acuerdo al signo de $\rho(X, Y)$.

En general, el recíproco de la primera afirmación de la proposición recién demostrada, es falso, es decir, la condición $\rho(X, Y) = 0$ no es suficiente para poder afirmar que X y Y son independientes. En el Ejercicio 358 se muestra una situación concreta de este resultado. Sin embargo, cuando la distribución de (X, Y) es normal y $\rho(X, Y) = 0$, entonces X y Y son independientes.

Proposición. Si (X, Y) es un vector con distribución normal bivariada tal que $\rho(X, Y) = 0$, entonces X y Y son independientes.

Demostración. Como veremos más adelante, la función de densidad normal bivariada está dada por la siguiente expresión:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right),$$

en donde $\mu_X = E(X)$, $\sigma_X^2 = \text{Var}(X)$, $\mu_Y = E(Y)$, $\sigma_Y^2 = \text{Var}(Y)$, y $\rho \in (-1, 1)$. Se pueden calcular directamente las funciones de densidad marginales y comprobar que

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp[-(x - \mu_X)^2/2\sigma_X^2]$$

y

$$f(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp[-(y - \mu_Y)^2/2\sigma_Y^2],$$

es decir, X tiene distribución $N(\mu_X, \sigma_X^2)$, y Y tiene distribución $N(\mu_Y, \sigma_Y^2)$. Después

de hacer algunos cálculos sencillos se puede demostrar que $\rho(X, Y) = \rho$, y comprobar finalmente que cuando $\rho = 0$, se verifica la igualdad $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

□

En resumen tenemos la siguiente tabla.

| Propiedades del coeficiente de correlación |
|--|
| $\rho(X, Y) \in [-1, 1]$. |
| $ \rho(X, Y) = 1 \Leftrightarrow Y = aX + b$. |
| $X \perp Y \Rightarrow \rho(X, Y) = 0$. |
| En general, $\rho(X, Y) = 0 \not\Rightarrow X \perp Y$. |
| Si (X, Y) tiene dist. normal y $\rho(X, Y) = 0$, entonces $X \perp Y$. |

3.10. Esperanza y varianza de un vector aleatorio

Definición (Esperanza y varianza de un vector). Sea X el vector aleatorio (X_1, \dots, X_n) . Cuando cada coordenada del vector tiene esperanza finita se define la esperanza de X como el vector numérico

$$E(X) = (E(X_1), \dots, E(X_n)).$$

Si cada coordenada tiene segundo momento finito, entonces la varianza de X se define como la matriz cuadrada

$$\text{Var}(X) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Var}(X_n) \end{pmatrix}_{n \times n}.$$

La varianza de un vector X puede expresarse como $E[(X - E(X))^t(X - E(X))]$, en donde X^t significa transpuesta del vector X . Observe que $(X - E(X))^t$ es un vector columna de dimensión $n \times 1$, mientras que $(X - E(X))$ es un vector renglón de dimensión $1 \times n$. De modo que el producto de estos dos vectores, en el orden

indicado, resulta en una matriz cuadrada de dimensión $n \times n$ cuya entrada (i, j) es

$$E[(X_i - E(X_i))(X_j - E(X_j))] = \text{Cov}(X_i, X_j).$$

Esta matriz también se llama *matriz de varianzas y covarianzas* y tiene las siguientes propiedades.

Proposición. La matriz $\text{Var}(X)$ es *simétrica* y *positiva definida*. Esto último significa que para cualquier vector $\theta = (\theta_1, \dots, \theta_n)$ de \mathbb{R}^n se cumple la desigualdad

$$\langle \text{Var}(X)\theta, \theta \rangle \geq 0,$$

en donde $\langle \cdot, \cdot \rangle$ denota el producto interior usual de \mathbb{R}^n .

Demostración. La simetría se sigue del hecho de que $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$. La propiedad de ser positiva definida se obtiene usando la bilinealidad de la covarianza,

$$\begin{aligned} \langle \text{Var}(X)\theta, \theta \rangle &= \sum_{i,j=1}^n \text{Cov}(X_i, X_j)\theta_i\theta_j \\ &= \sum_{i,j=1}^n \text{Cov}(\theta_i X_i, \theta_j X_j) \\ &= \text{Cov}\left(\sum_{i=1}^n \theta_i X_i, \sum_{j=1}^n \theta_j X_j\right) \\ &= \text{Var}\left(\sum_{i=1}^n \theta_i X_i\right) \geq 0. \end{aligned}$$

□

3.11. Distribuciones multivariadas discretas

En esta sección se estudian algunas distribuciones discretas de vectores aleatorios.

Distribución multinomial

Suponga que se tiene un experimento aleatorio con k posibles resultados distintos. Las probabilidades para cada uno de estos resultados son respectivamente p_1, \dots, p_k , en donde $p_1 + \dots + p_k = 1$. Ahora suponga que se tienen n ensayos sucesivos independientes del experimento anterior, y defina las variables aleatorias discretas

X_1, \dots, X_k , como aquellas que registran el número de veces que se obtienen cada uno de los k posibles resultados en los n ensayos. Entonces se dice que el vector $X = (X_1, \dots, X_k)$ tienen una distribución multinomial, y su función de densidad conjunta es

$$f(x_1, \dots, x_k) = \begin{cases} \binom{n}{x_1 \dots x_k} p_1^{x_1} \dots p_k^{x_k} & \text{si } x_1, \dots, x_k = 0, 1, \dots, n \\ & \text{con } x_1 + \dots + x_k = n, \\ 0 & \text{otro caso.} \end{cases}$$

Los parámetros de esta distribución son entonces el número de ensayos n , el número de resultados distintos k en cada ensayo, y las probabilidades p_1, \dots, p_k . El factor que aparece en paréntesis en la función de densidad conjunta se conoce como *coeficiente multinomial* y se define como sigue

$$\binom{n}{x_1 \dots x_k} = \frac{n!}{x_1! \dots x_k!}.$$

Se dice entonces que X tiene distribución multinomial(n, k, p_1, \dots, p_k). Observe que cuando únicamente hay dos posibles resultados en cada ensayo, es decir $k = 2$, la distribución multinomial se reduce a la distribución binomial. No es difícil probar que

$$E(X) = (np_1, \dots, np_k),$$

$$\text{y } [\text{Var}(X)]_{ij} = \begin{cases} np_i(1 - p_i) & \text{si } i = j, \\ -np_i p_j & \text{si } i \neq j. \end{cases}$$

Distribución hipergeométrica multivariada

Suponga que se tienen N objetos de los cuales N_1 son de un primer tipo, N_2 son de un segundo tipo y así sucesivamente con N_k objetos de tipo k . Entonces $N_1 + \dots + N_k = N$. Suponga que de la totalidad de objetos se obtiene una muestra sin reemplazo de tamaño n , y defina las variables X_1, \dots, X_k , como aquellas que representan el número de objetos seleccionados de cada tipo. Se dice entonces que X_1, \dots, X_k tienen una distribución hipergeométrica multivariada y su función de densidad conjunta es

$$f(x_1, \dots, x_k) = \frac{\binom{N_1}{x_1} \dots \binom{N_k}{x_k}}{\binom{N}{n}}$$

en donde cada x_i toma valores en el conjunto $\{0, 1, \dots, n\}$ pero sujeto a $x_i \leq N_i$ y además debe cumplirse que $x_1 + \dots + x_k = n$. Se dice entonces que (X_1, \dots, X_k) tiene distribución hipergeométrica multivariada (N, N_1, \dots, N_k, n) . Observe que cuando únicamente hay dos tipos de objetos, es decir $k = 2$, la distribución hipergeométrica multivariada se reduce a la distribución hipergeométrica univariada. Véase la sección de ejercicios para la esperanza y varianza de esta distribución.

3.12. Distribuciones multivariadas continuas

Ahora estudiamos algunas distribuciones continuas de vectores aleatorios.

Distribución uniforme bivariada

Se dice que las variables aleatorias continuas X y Y tienen una distribución conjunta uniforme en el rectángulo $(a, b) \times (c, d)$, si su función de densidad conjunta es

$$f(x, y) = \begin{cases} \frac{1}{(b-a)(d-c)} & \text{si } x \in (a, b), y \in (c, d), \\ 0 & \text{otro caso.} \end{cases}$$

Se escribe $(X, Y) \sim \text{unif}(a, b) \times (c, d)$. Observe que las distribuciones marginales son nuevamente uniformes, además X y Y siempre son independientes. Es inmediato comprobar que $E(X, Y) = ((a+b)/2, (c+d)/2)$, y que

$$\text{Var}(X, Y) = \begin{pmatrix} \frac{1}{12}(b-a)^2 & 0 \\ 0 & \frac{1}{12}(d-c)^2 \end{pmatrix}.$$

Distribución normal bivariada

Se dice que las variables aleatorias continuas X y Y tienen una distribución normal bivariada si su función de densidad conjunta es

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right),$$

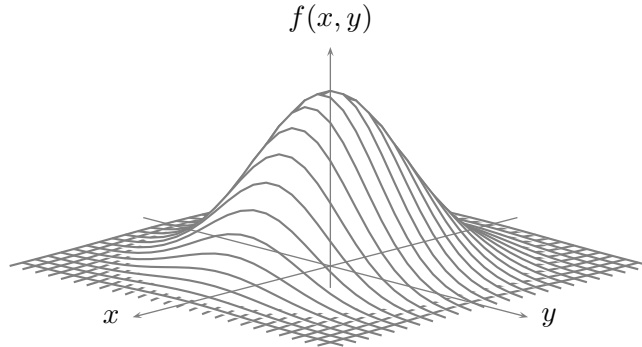
para cualesquiera valores reales de x y y , y en donde $-1 < \rho < 1$, $\sigma_X > 0$, $\sigma_Y > 0$, y μ_X, μ_Y dos constantes reales sin restricción. Se escribe $X \sim N(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Puede demostrarse que X tiene una distribución marginal $N(\mu_X, \sigma_X^2)$, y Y tiene distribución marginal $N(\mu_Y, \sigma_Y^2)$. El parámetro ρ resulta ser el coeficiente de correlación entre X y Y . Por lo tanto,

$$E(X, Y) = (\mu_X, \mu_Y),$$

y puede demostrarse que

$$\text{Var}(X, Y) = \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \cdot \sigma_Y \\ \rho \cdot \sigma_X \cdot \sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

Es interesante observar que existen distribuciones bivariadas con densidades marginales normales, pero cuya distribución conjunta no lo es. En el Ejercicio 376 en la página 136 se presenta un ejemplo al respecto. Cuando $\mu_X = \mu_Y = 0$, y $\sigma_X = \sigma_Y = 1$, la distribución se llama normal bivariada *estándar*, y su gráfica se muestra a continuación.



Función de densidad normal bivariada estándar.

3.13. Ejercicios

Distribución conjunta

262. Grafique y demuestre que las siguientes funciones son de distribución.

$$a) F(x, y) = (1 - e^{-x})\left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} y\right), \quad \text{para } x \geq 0.$$

$$b) F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y}, \quad \text{para } x, y \geq 0.$$

263. Investigue si las siguientes funciones son de distribución.

$$a) F(x, y) = 1 - e^{-xy}, \quad \text{para } x, y \geq 0.$$

$$b) F(x, y) = 1 - e^{-x-y}, \quad \text{para } x, y \geq 0.$$

264. Demuestre que la siguiente función no es de distribución.

$$F(x, y) = \begin{cases} 0 & \text{si } x + y < 0, \\ 1 & \text{si } x + y \geq 0. \end{cases}$$

Este es un ejemplo de una función que tiene el comportamiento límite adecuado en infinito, es continua por la derecha y no decreciente en cada variable, pero no es función de distribución pues asigna valores negativos a algunas regiones del plano. Por ejemplo calcule la probabilidad del cuadrado $(-1, 1] \times (-1, 1]$. De manera análoga demuestre que la siguiente función tampoco es de distribución.

$$F(x, y, z) = \begin{cases} 0 & \text{si } x + y + z < 0, \\ 1 & \text{si } x + y + z \geq 0. \end{cases}$$

Extienda este resultado al caso n -dimensional.

265. Demuestre que la siguiente función no es de distribución.

$$F(x, y) = \begin{cases} \min\{1, \max\{x, y\}\} & \text{si } x, y > 0, \\ 0 & \text{otro caso.} \end{cases}$$

266. Sean $F(x)$ y $G(x)$ dos funciones de distribución. Demuestre o proporcione un contraejemplo para las siguientes afirmaciones.

- a) $F(x)G(x)$ es una función de distribución univariada.
- b) $F(x)G(y)$ es una función de distribución bivariada.

267. Diga falso o verdadero. Justifique en cada caso.

- a) $P(X > x, Y > y) = 1 - P(X \leq x, Y \leq y)$.
- b) $P(X \leq x, Y \leq y) \leq P(X \leq x)$.
- c) $P(X \leq x) = P(X \leq x, Y \leq x) + P(X \leq x, Y > x)$.
- d) $P(X + Y \leq x) \leq P(X \leq x)$.
- e) $P(XY < 0) \leq P(X < 0)$.

268. Sean X y Y variables aleatorias con función de distribución conjunta $F(x, y)$. Demuestre que para cualesquiera números reales $a < b$ y $c < d$,

$$P(a < X \leq b, c < Y \leq d) = F(b, d) + F(a, c) - F(a, d) - F(b, c).$$

269. Sean X_1, X_2 y X_3 variables aleatorias con función de distribución conjunta $F(x_1, x_2, x_3)$. Demuestre que para cualesquiera números reales $a_1 < b_1, a_2 < b_2$ y $a_3 < b_3$, la probabilidad

$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3)$$

es igual a

$$\begin{aligned} &F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) \\ &+ F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) \\ &- F(a_1, a_2, a_3). \end{aligned}$$

270. Sea (X, Y) un vector con función de distribución conjunta $F_{X,Y}(x, y)$. Demuestre que para todo (x, y) en \mathbb{R}^2 ,

$$F_X(x) + F_Y(y) - 1 \leq F_{X,Y}(x, y) \leq \sqrt{F_X(x)F_Y(y)}.$$

271. Considere el espacio $\Omega = [0, 1] \times [0, 1]$ junto con $\sigma(\mathcal{B}[0, 1] \times \mathcal{B}[0, 1])$ y P la medida de probabilidad uniforme sobre Ω . Sea $X : \Omega \rightarrow \mathbb{R}^2$ el vector aleatorio dado por $X(\omega_1, \omega_2) = (\omega_1 \wedge \omega_2, \omega_1 \vee \omega_2)$. Demuestre que X es efectivamente un vector aleatorio y encuentre su función de distribución.

Densidad conjunta

272. Demuestre que la función de densidad de un vector (X, Y) absolutamente continuo puede ser encontrada, a partir de la función de distribución, de las siguientes formas alternativas:

$$a) f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} P(X > x, Y > y).$$

$$b) f_{X,Y}(x, y) = -\frac{\partial^2}{\partial x \partial y} P(X \leq x, Y > y).$$

$$c) f_{X,Y}(x, y) = -\frac{\partial^2}{\partial x \partial y} P(X > x, Y \leq y).$$

273. Grafique y demuestre que las siguientes funciones son de densidad.

$$a) f(x, y) = \frac{1}{ab}, \quad \text{para } 0 < x < a, \quad 0 < y < b.$$

$$b) f(x, y) = 4xy, \quad \text{para } 0 \leq x, y \leq 1.$$

$$c) f(x, y) = 6x^2y, \quad \text{para } 0 \leq x, y \leq 1.$$

$$d) f(x, y) = \frac{9}{4}x^2y^2, \quad \text{para } -1 \leq x, y \leq 1.$$

$$e) f(x, y) = e^{-x-y}, \quad \text{para } x, y > 0.$$

$$f) f(x, y) = e^{-x}, \quad \text{para } 0 < y < x.$$

274. Calcule la constante c que hace a f una función de densidad.

$$a) f(x) = cx, \quad \text{para } 0 \leq x \leq 1.$$

$$b) f(x, y) = cx, \quad \text{para } 0 < y < x < 1.$$

$$c) f(x, y) = c(x + y) \quad \text{para } 0 \leq x, y \leq 1.$$

$$d) f(x, y) = c(x^2 + \frac{1}{2}xy), \quad \text{para } 0 < x < 1, \quad 0 < y < 2.$$

$$e) f(x, y, z) = c(x + y + z), \quad \text{para } 0 \leq x, y, z \leq 1.$$

$$f) f(x_1, \dots, x_n) = c(x_1 + \dots + x_n), \quad \text{para } 0 \leq x_1, \dots, x_n \leq 1.$$

275. Encuentre la función de densidad del vector (X, Y) cuya función de distribución es

$$a) F(x, y) = (1 - e^{-x})\left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} y\right), \quad \text{para } x \geq 0.$$

$$b) F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y}, \quad \text{para } x, y \geq 0.$$

276. Encuentre la función de distribución del vector (X, Y) cuya función de densidad es

$$a) f(x, y) = \frac{1}{ab}, \quad \text{para } 0 < x < a, \quad 0 < y < b.$$

$$b) f(x, y) = e^{-x-y}, \quad \text{para } x, y > 0.$$

$$c) f(x, y) = e^{-y}, \quad \text{para } 0 < x < y.$$

$$d) f(x, y) = 2e^{-x-y}, \quad \text{para } 0 < x < y.$$

277. Sean $f(x)$ y $g(x)$ dos funciones de densidad. Demuestre o proporcione un contraejemplo para las siguientes afirmaciones:

- a) $f(x)g(x)$ es una función de densidad univariada.
- b) $f(x)g(y)$ es una función de densidad bivariada.

278. Sean X y Y independientes ambas con distribución $\exp(\lambda)$. Encuentre la función de densidad y de distribución de

- a) $W = \max\{X, Y\}$.
- b) $W = \min\{X, Y\}$.

Distribución marginal

279. Suponiendo el caso absolutamente continuo, demuestre que la función de densidad marginal,

$$x \mapsto f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

es efectivamente una función de densidad.

280. Demuestre que la función de distribución marginal

$$x \mapsto F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$$

es efectivamente una función de distribución.

281. Encuentre las funciones de distribución marginales del vector (X, Y) cuya función de distribución conjunta es

- a) $F(x, y) = (1 - e^{-x})(1 - e^{-y})$, para $x, y > 0$.
- b) $F(x, y) = (1 - e^{-x^2})(1 - e^{-y^2})$, para $x, y > 0$.

282. Encuentre las funciones de densidad marginales del vector (X, Y) cuya función de densidad conjunta es

- a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.
- b) $f(x, y) = 4xy$, para $0 < x, y < 1$.
- c) $f(x, y) = 24x(1 - x - y)$, para $x, y > 0$ y $x + y < 1$.
- d) $f(x, y) = (x + 2y)/4$, para $0 < x < 2$ y $0 < y < 1$.
- e) $f(x, y) = 2(4x + y)/5$, para $0 < x, y < 1$.
- f) $f(x, y) = 1/x$, para $0 < y < x < 1$.

283. Sea $0 < a < 1$ y defina la función $f(x, y) = a^x(1 - a)^y$ para $x, y \in \mathbb{N}$. Demuestre que $f(x, y)$ es una función de densidad y calcule las funciones de densidad marginales.

Distribución condicional

284. Demuestre que la función de distribución condicional

$$x \mapsto F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(u|y) du$$

es efectivamente una función de distribución.

285. Demuestre que la función de densidad condicional

$$x \mapsto f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

es efectivamente una función de densidad.

286. Sea (X, Y) un vector aleatorio absolutamente continuo. Demuestre la fórmula

$$f_{X|Y}(x|y) = \frac{\partial}{\partial x} F_{X|Y}(x|y).$$

287. *Pérdida de memoria en la distribución exponencial.* Sea X con distribución $\exp(\lambda)$ y sea $t > 0$ fijo. Demuestre que la distribución condicional de $X - t$, dado que $X \geq t$, sigue siendo $\exp(\lambda)$.

288. Calcule las funciones condicionales $f_{X|Y}(x|y)$ y $F_{X|Y}(x|y)$, para las siguientes funciones de densidad conjunta.

a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.

b) $f(x, y) = 4xy$, para $0 < x, y < 1$.

c) $f(x, y) = 24x(1 - x - y)$, para $x, y > 0$ y $x + y < 1$.

d) $f(x, y) = (x + 2y)/4$, para $0 < x < 2$ y $0 < y < 1$.

e) $f(x, y) = 2(4x + y)/5$, para $0 < x, y < 1$.

f) $f(x, y) = 1/x$, para $0 < y < x < 1$.

289. Calcule las funciones condicionales $F_{X|Y}(x|y)$ y $f_{X|Y}(x|y)$, para las siguientes funciones de distribución conjunta.

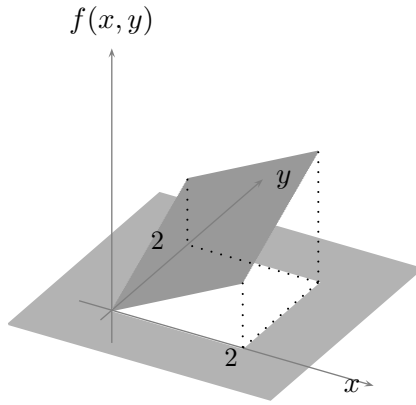
a) $F(x, y) = (1 - e^{-x})(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} y)$, para $x \geq 0$.

b) $F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y}$, para $x, y \geq 0$.

290. Se hacen tres lanzamientos de una moneda equilibrada cuyos resultados llamaremos cara y cruz. Sea X la v.a. que denota el número de caras que se obtienen en los dos primeros lanzamientos y sea Y la v.a. que denota el número de cruces en los dos últimos lanzamientos. Calcule $f_{X,Y}(x, y)$, $f_X(x)$, $f_Y(y)$ y $f_{Y|X}(y|x)$ para $x = 0, 1, 2$.

291. Sea (X, Y) un vector con función de densidad

$$f_{X,Y}(x, y) = \frac{1}{8}(x + y), \quad \text{para } 0 \leq x, y \leq 2.$$

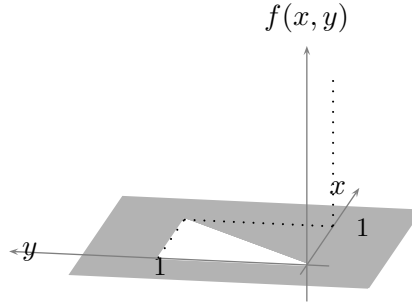


Función de densidad $f(x, y) = \frac{1}{8}(x + y)$, para $x, y \in [0, 2]$.

Compruebe que $f(x, y)$ es una función de densidad. Calcule $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$, $F_{X,Y}(x, y)$, $F_X(x)$, $F_Y(y)$, $F_{X|Y}(x|y)$, $F_{Y|X}(y|x)$, $P(Y > X)$ y $P(X > 1 | Y < 1)$.

292. Sea (X, Y) un vector con función de densidad

$$f_{X,Y}(x, y) = 8xy, \quad \text{para } 0 < x < y < 1.$$

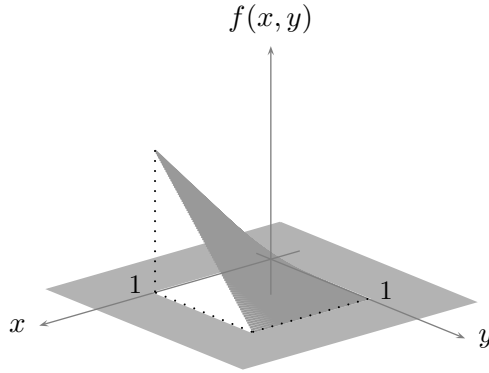


Función de densidad $f(x, y) = 8xy$, para $0 < x < y < 1$.

Compruebe que $f(x, y)$ es una función de densidad. Calcule $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$, $F_{X,Y}(x, y)$, $F_X(x)$, $F_Y(y)$, $F_{X|Y}(x|y)$, $F_{Y|X}(y|x)$, $P(X + Y < 1)$ y $P(Y < 1/2 | X < 1/2)$.

293. Sea (X, Y) un vector con función de densidad

$$f_{X,Y}(x, y) = 4x(1 - y), \quad \text{para } 0 < x, y < 1.$$



Función de densidad $f(x, y) = 4x(1 - y)$, para $0 < x, y < 1$.

Compruebe que $f(x, y)$ es efectivamente una función de densidad. Calcule $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$, $F_{X,Y}(x, y)$, $F_X(x)$, $F_Y(y)$, $F_{X|Y}(x|y)$, $F_{Y|X}(y|x)$, $P(X > 1/2)$ y $P(1/4 < Y < 3/4 \mid X < 1/2)$.

Independencia

294. Demuestre la variable aleatoria constante $X = c$ es independiente de cualquier otra variable aleatoria. Inversamente, suponga ahora que X es independiente de cualquier otra variable aleatoria, demuestre que X es constante.
295. Demuestre que los eventos A y B son independientes si, y sólo si, las variables aleatorias indicadoras 1_A y 1_B lo son.
296. Demuestre que si tres variables aleatorias son independientes, entonces cualesquiera dos de ellas lo son. Más generalmente, demuestre que cualquier subconjunto de un conjunto de variables aleatorias independientes también lo es.
297. Sean X_1, \dots, X_n independientes, y sean g_1, \dots, g_n funciones de \mathbb{R} en \mathbb{R} , Borel medibles. Demuestre que las variables aleatorias $g_1(X_1), \dots, g_n(X_n)$ son independientes.
298. Sean X_1, \dots, X_n independientes, y sea $1 \leq k < n$. Sean $g : \mathbb{R}^k \rightarrow \mathbb{R}$ y $h : \mathbb{R}^{n-k} \rightarrow \mathbb{R}$ funciones Borel medibles. Demuestre que las variables aleatorias $g(X_1, \dots, X_k)$ y $h(X_{k+1}, \dots, X_n)$ son independientes.
299. Determine si las siguientes son funciones de densidad de variables aleatorias independientes.
 - a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.
 - b) $f(x, y) = 2x$, para $0 < x, y < 1$.
 - c) $f(x, y) = 2e^{-x-y}$, para $0 < x < y$.
 - d) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

$$e) f(x, y) = \frac{3}{8}(x^2 + y^2), \text{ para } x, y \in [-1, 1].$$

300. Determine si las siguientes son funciones de distribución de variables aleatorias independientes.

$$a) F(x, y) = (1 - e^{-x})(1 - e^{-y}), \text{ para } x, y > 0.$$

$$b) F(x, y) = (1 - e^{-x^2})(1 - e^{-y^2}), \text{ para } x, y > 0.$$

301. Demuestre que X y Y son independientes si, y sólo si, para cualesquiera números reales x y y ,

$$P(X > x, Y > y) = P(X > x) \cdot P(Y > y).$$

302. Demuestre que X y Y son independientes si, y sólo si, para cualesquiera números reales $a < b$ y $c < d$,

$$P(a < X \leq b, c < Y \leq d) = P(a < X \leq b) \cdot P(c < Y \leq d).$$

303. Diga falso o verdadero. Demuestre en cada caso.

$$a) X, Y \text{ independientes} \Rightarrow X, Y^2 \text{ independientes.}$$

$$b) X, Y \text{ independientes} \Rightarrow X^2, Y^2 \text{ independientes.}$$

$$c) X, Y \text{ independientes} \Rightarrow X + Y, Y \text{ independientes.}$$

$$d) X, Y \text{ independientes} \Rightarrow X + Y, X - Y \text{ independientes.}$$

$$e) X, Y \text{ independientes} \Rightarrow XY, Y \text{ independientes.}$$

$$f) X^2, Y^2 \text{ independientes} \Rightarrow X, Y \text{ independientes.}$$

$$g) X, Y, Z \text{ independientes} \Rightarrow X + Y, Z \text{ independientes.}$$

304. Sean X y Y independientes ambas con distribución normal estándar. Demuestre que $Z = aX + bY + c$ tiene distribución normal cuando $ab \neq 0$. Encuentre $E(Z)$ y $\text{Var}(Z)$.

305. Sean X_1, \dots, X_n variables aleatorias independientes cada una con distribución $\text{Ber}(p)$. Calcule $P(X_1 + \dots + X_n = k)$ para $k = 0, 1, \dots, n$.

306. Sean X y Y independientes ambas con distribución $\text{unif}\{1, \dots, n\}$. Encuentre la distribución de $(U, V) = (X + Y, X - Y)$. Determine si U y V son independientes.

307. Sean $X \geq 0$ y $Y \geq 0$ independientes con valores enteros naturales y con esperanza finita. Demuestre que

$$E(\text{mín}\{X, Y\}) = \sum_{n=1}^{\infty} P(X \geq n)P(Y \geq n).$$

308. Sean X y Y independientes ambas con distribución $\text{unif}\{-1, 1\}$. Sea $Z = XY$. Demuestre que X, Y y Z son independientes dos a dos pero no lo son en su conjunto.

309. Sean X y Y independientes con distribución $\text{Poisson}(\lambda_1)$ y $\text{Poisson}(\lambda_2)$ respectivamente. Demuestre que la distribución condicional de X dado que $X+Y = n$ es $\text{bin}(n, \lambda_1/(\lambda_1 + \lambda_2))$.
310. Sean X y Y independientes con distribución $\text{unif}\{0, 1, \dots, n\}$ y $\text{unif}\{0, 1, \dots, m\}$ respectivamente. Encuentre la función de densidad de $X + Y$.
311. Sean X_1, \dots, X_n independientes con distribución $\text{geo}(p)$. Demuestre que $X_1 + \dots + X_n$ tiene distribución $\text{bin neg}(n, p)$.
312. Sean X y Y independientes. Encuentre la función de distribución de W en términos de $F_X(x)$ y $F_Y(y)$ cuando
- $W = \text{máx}\{X, Y\}$.
 - $W = \text{mín}\{X, Y\}$.
313. Sean X y Y independientes ambas con distribución $\text{exp}(\lambda)$, y sea a una constante. Calcule $P(X \vee Y \leq aX)$ y $P(X \wedge Y \leq aX)$.
314. Usando la siguiente tabla, construya la función de densidad $f(x, y)$ de un vector discreto (X, Y) con la condición de que X y Y sean independientes.

| | | |
|------------------|---------|---------|
| $x \backslash y$ | 0 | 1 |
| 0 | \cdot | \cdot |
| 1 | \cdot | \cdot |

315. Sea (X, Y) un vector discreto con distribución de probabilidad uniforme en el conjunto $\{1, \dots, n\} \times \{1, \dots, m\}$, con n y m enteros positivos. Demuestre que X y Y son independientes.
316. Sea (X, Y) un vector aleatorio con función de densidad

$$f_{X,Y}(x, y) = c(1 - x), \quad \text{para } 0 < x < y < 1.$$

- Encuentre el valor de c que hace a $f_{X,Y}(x, y)$ una función de densidad y grafique esta función.
 - Calcule $P(X + Y > 1)$ y $P(X \leq 1/2)$.
 - Encuentre las funciones de densidad marginales $f_X(x)$ y $f_Y(y)$.
 - Determine si X y Y son independientes.
317. Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = c \left(\frac{1}{2} \right)^{x+y}, \quad \text{para } x = 0, 1, 2 \text{ y } y = 1, 2.$$

Encuentre el valor de la constante c y determine si X y Y son independientes. Calcule además las probabilidades $P(X = 1)$, $P(X = 2 | Y = 2)$ y $P(XY = 2)$.

318. Sea (X, Y) un vector aleatorio con función de densidad

$$f_{X,Y}(x, y) = 2, \quad \text{para } 0 < x < y < 1.$$

- a) Grafique y demuestre que $f_{X,Y}(x, y)$ es una función de densidad.
- b) Encuentre las funciones de densidad marginales $f_X(x)$ y $f_Y(y)$.
- c) Determine si X y Y son independientes.
- d) Calcule $P(Y > X)$ y $P(Y > X^2)$.

319. Sea (X, Y) un vector aleatorio con función de densidad

$$f_{X,Y}(x, y) = c|x + y|, \quad \text{para } -1 < x, y < 1.$$

- a) Encuentre el valor de la constante c que hace a $f_{X,Y}(x, y)$ una función de densidad y grafique esta función.
- b) Calcule $P(X > 0)$, $P(XY > 0)$ y $P(0 < X + Y < 1)$.
- c) Encuentre las funciones de densidad marginales $f_X(x)$ y $f_Y(y)$.
- d) Determine si X y Y son independientes.

320. Sean X y Y independientes con distribución exponencial con parámetros λ_1 y λ_2 respectivamente. Demuestre que

$$P(Y > X) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

321. Sean X y Y independientes con distribución exponencial con parámetros λ_1 y λ_2 respectivamente. Demuestre que $X \wedge Y$ tiene distribución exponencial de parámetro $\lambda_1 + \lambda_2$.

322. Sean X y Y independientes con distribución $\text{bin}(n, p)$ y $\text{bin}(m, p)$ respectivamente. Demuestre que $X + Y$ tiene distribución $\text{bin}(n + m, p)$.

323. Sean X y Y independientes con distribución Poisson con parámetros λ_1 y λ_2 respectivamente. Demuestre que $X + Y$ tiene distribución $\text{Poisson}(\lambda_1 + \lambda_2)$.

324. Sea (X, Y, Z) un vector aleatorio con función de densidad

$$f_{X,Y,Z}(x, y, z) = 8xyz, \quad \text{para } 0 < x, y, z < 1.$$

- a) Compruebe que $f(x, y, z)$ es una función de densidad.
- b) Calcule $P(X < Y < Z)$ y $P(X + Y + Z < 1)$.
- c) Encuentre $f_{X,Y}(x, y)$, $f_{X,Z}(x, z)$ y $f_{Y,Z}(y, z)$.
- d) Determine si X , Y y Z son independientes.

325. Sea (X, Y, Z) un vector aleatorio con función de densidad

$$f_{X,Y,Z}(x, y, z) = 24xz, \quad \text{para } 0 < x < y < z < 1.$$

- a) Compruebe que $f(x, y, z)$ es una función de densidad.
- b) Calcule $P(X + Y < 1)$ y $P(Z - X > 1/2)$.
- c) Encuentre $f_{X,Y}(x, y)$, $f_{X,Z}(x, z)$ y $f_{Y,Z}(y, z)$.
- d) Determine si X , Y y Z son independientes.

326. Sea X_1, X_2, \dots una sucesión de v.a.s independientes cada una con distribución unif(0, 1). Demuestre que para cualquier $\lambda > 0$,

$$\lim_{n \rightarrow \infty} P(\text{máx}\{X_1, \dots, X_n\} \leq 1 - \frac{\lambda}{n}) = e^{-\lambda}.$$

327. Sean X y Y independientes con distribución Poisson de parámetros λ_1 y λ_2 respectivamente. Demuestre que

$$E(X | X + Y = n) = n \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Esperanza de una función de un vector aleatorio

328. *Esperanza de una función de un vector discreto.* Sea (X, Y) un vector aleatorio discreto con valores en el conjunto producto $\{x_1, x_2, \dots\} \times \{y_1, y_2, \dots\}$, y sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ una función Borel medible tal que la variable $\varphi(X, Y)$ tiene esperanza finita. Demuestre que

$$E[\varphi(X, Y)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varphi(x_i, y_j) P(X = x_i, Y = y_j).$$

329. Demuestre que la condición $E(XY) = E(X)E(Y)$ no implica necesariamente que X y Y son independientes. Para ello considere cualquiera de los siguientes ejemplos.

$$a) f(x, y) = \begin{cases} 1/8 & \text{si } (x, y) = (1, 1), (1, -1), (-1, 1), (-1, -1), \\ 1/2 & \text{si } (x, y) = (0, 0), \\ 0 & \text{otro caso.} \end{cases}$$

$$b) f(x, y) = 3(x^2 + y^2)/8, \text{ para } x, y \in [-1, 1].$$

$$c) X \text{ con distribución uniforme en } \{-1, 0, 1\} \text{ y } Y = 1_{(X \neq 0)}.$$

330. Demuestre que si X_1, X_2, \dots, X_n son independientes e integrables, entonces

$$E(X_1 X_2 \cdots X_n) = E(X_1) \cdot E(X_2) \cdots E(X_n).$$

331. Sean X y Y independientes. Diga falso o verdadero justificando en cada caso.

$$a) \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

$$b) \text{Var}(X - Y) = \text{Var}(X) - \text{Var}(Y).$$

$$c) \text{Var}(XY) = \text{Var}(X)\text{Var}(Y).$$

332. Sean X y Y independientes con varianza finita. Demuestre que

$$\text{Var}(XY) = \text{Var}(X) \cdot \text{Var}(Y) + E^2(X) \cdot \text{Var}(Y) + E^2(Y) \cdot \text{Var}(X).$$

333. Sean X_1, \dots, X_n independientes con la misma distribución, y sea $S_n = X_1 + \cdots + X_n$. Suponiendo que las esperanzas indicadas existen, demuestre que $E(X_1 / S_n) = 1/n$. Concluya que para $m \leq n$, $E(S_m / S_n) = m/n$.

334. Sea X_1, \dots, X_n variables aleatorias independientes con idéntica distribución y con esperanza finita. Demuestre que

$$E(X_1 | X_1 + \dots + X_n = k) = \frac{k}{n}.$$

335. Sea (X, Y) un vector aleatorio discreto con función de densidad dada por la siguiente tabla

| $x \backslash y$ | -1 | 0 | 1 |
|------------------|-----|-----|-----|
| 1 | .1 | .05 | .1 |
| 2 | .06 | .2 | .04 |
| 3 | .1 | .05 | .3 |

- Grafique $f(x, y)$ y compruebe que efectivamente se trata de una función de densidad conjunta.
 - Calcule y grafique las densidades marginales $f_X(x)$ y $f_Y(y)$. Verifique que ambas son funciones de densidad.
 - Demuestre que X y Y no son independientes.
 - Calcule $E(XY)$ y $f_{X+Y}(u)$.
336. Sea (X, Y) un vector discreto con función de densidad dada por la siguiente tabla

| $x \backslash y$ | 2 | 4 | 6 |
|------------------|------|------|------|
| 1 | 2/18 | 3/18 | 1/18 |
| 2 | 3/18 | 5/18 | 1/18 |
| 3 | 1/18 | 1/18 | 1/18 |

- Grafique $f(x, y)$ y compruebe que efectivamente es una función de densidad conjunta.
 - Calcule y grafique las densidades marginales $f_X(x)$ y $f_Y(y)$. Verifique que ambas son efectivamente funciones de densidad.
 - Demuestre que X y Y no son independientes.
 - Calcule $E(XY)$ y $f_{X+Y}(u)$.
337. Sea (X, Y) un vector aleatorio con función de densidad dada por

$$f(x, y) = \begin{cases} 8xy & \text{si } 0 < y < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

- Grafique $f(x, y)$ y compruebe que efectivamente es una función de densidad conjunta.
- Encuentre y grafique las densidades marginales $f_X(x)$ y $f_Y(y)$. Verifique que ambas son efectivamente funciones de densidad.
- Demuestre que X y Y no son independientes.
- Calcule $E(XY)$ y $f_{X+Y}(u)$.

Esperanza y varianza de un vector

338. Calcule la esperanza y varianza del vector aleatorio (X, Y) cuya función de densidad conjunta es

a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.

b) $f(x, y) = 4xy$, para $x, y \in [0, 1]$.

Covarianza

339. Sea a un número real cualquiera. Encuentre variables aleatorias X y Y tales que $\text{Cov}(X, Y) = a$.

340. Diga falso o verdadero. Demuestre en cada caso.

a) $X \geq 0, Y \geq 0 \Rightarrow \text{Cov}(X, Y) \geq 0$.

b) $\text{Cov}(X, Y) = 0, \text{Cov}(Y, Z) = 0 \Rightarrow \text{Cov}(X, Z) = 0$.

c) $\text{Cov}(X, Y) > 0, \text{Cov}(Y, Z) > 0 \Rightarrow \text{Cov}(X, Z) > 0$.

d) $\text{Cov}(X, Y) = a, \text{Cov}(Y, Z) = a \Rightarrow \text{Cov}(X, Z) = a$.

341. Diga falso o verdadero. Demuestre en cada caso.

a) $\text{Cov}(X, Y) \geq 0$.

b) $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$, con a, b constantes.

c) $\text{Cov}(X, aY + b) = a \text{Cov}(X, Y) + b$, con a, b constantes.

342. Demuestre que

a) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

b) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

c) $\text{Cov}(X, X) = \text{Var}(X)$.

d) $\text{Cov}(X, -X) = -\text{Var}(X)$.

e) $\text{Cov}(aX + b, Y) = a \text{Cov}(X, Y)$, con a, b constantes.

f) $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.

343. Demuestre que la condición $\text{Cov}(X, Y) = 0$ no es suficiente para concluir que X y Y son independientes. En el texto se proporciona un ejemplo para un vector discreto. Construya ahora un ejemplo para un vector continuo.

344. Demuestre que $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2 \text{Cov}(X, Y)$.

345. Demuestre que

a) $\text{Var}(X_1 + \cdots + X_n) = \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{j < k} \text{Cov}(X_j, X_k)$.

b) $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$.

346. Sea X_1, \dots, X_n independientes y con varianza finita. Demuestre que

$$\text{Var}(X_1 + \dots + X_n) = \sum_{k=1}^n \text{Var}(X_k).$$

347. Sean X_1, \dots, X_n independientes y con idéntica distribución. Defina $\bar{X} = (X_1 + \dots + X_n)/n$. Demuestre que para cada $k = 1, \dots, n$, se cumple

$$\text{Cov}(X_k - \bar{X}, \bar{X}) = 0.$$

348. Sea (X, Y) con distribución uniforme en el conjunto $\{1, \dots, n\} \times \{1, \dots, n\}$. Demuestre que $\text{Cov}(X, Y) = 0$.

349. Sea (X, Y) con distribución uniforme en el conjunto $(a, b) \times (c, d)$. Demuestre que $\text{Cov}(X, Y) = 0$.

350. Calcule la covarianza de X y Y cuya función de densidad conjunta está dada por la siguiente tabla.

| | | | |
|-----|------|------|------|
| x\y | -1 | 0 | 1 |
| -1 | 1/12 | 2/12 | 3/12 |
| 1 | 3/12 | 2/12 | 1/12 |

351. Calcule la covarianza de X y Y cuya función de densidad conjunta está dada por la siguiente tabla.

| | | | |
|-----|-----|-----|-----|
| x\y | 1 | 2 | 3 |
| 2 | .2 | .05 | .15 |
| 4 | .05 | .1 | .15 |
| 6 | .05 | .1 | .15 |

352. Calcule la covarianza de X y Y , cuya función de densidad conjunta es

- a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.
- b) $f(x, y) = 3x^2y$, para $-1 < x < 1$, $0 < y < 1$.
- c) $f(x, y) = \frac{1}{2}e^{-x}$, para $|y| < x$.
- d) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

353. Sea (X, Y) un vector con distribución normal bivariada $N(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Demuestre que $\text{Cov}(X, Y) = \rho \cdot \sigma_X \sigma_Y$.

Coefficiente de correlación

354. Sean a, b, c, d constantes tales que $ac > 0$. Demuestre que

$$\rho(aX + b, cY + d) = \rho(X, Y).$$

355. Diga falso o verdadero. Demuestre en cada caso.

- a) $\rho(X, Y) = 0, \rho(Y, Z) = 0 \Rightarrow \rho(X, Z) = 0.$
- b) $\rho(X, Y) > 0, \rho(Y, Z) > 0 \Rightarrow \rho(X, Z) > 0.$
- c) $\rho(X, Y) < 0, \rho(Y, Z) < 0 \Rightarrow \rho(X, Z) < 0.$
- d) $\rho(X, Y) = 1, \rho(Y, Z) = 1 \Rightarrow \rho(X, Z) = 1.$
- e) $\rho(X, Y) = -1, \rho(Y, Z) = -1 \Rightarrow \rho(X, Z) = -1.$
- f) $\rho(X, Y)\rho(Y, Z) = -1 \Rightarrow \rho(X, Z) = -1.$
- g) $\rho(X, Y) = a, \rho(Y, Z) = a \Rightarrow \rho(X, Z) = a.$

356. Diga falso verdadero. Demuestre en cada caso.

- a) $\rho(X, Y) = \rho(Y, X).$
- b) $\rho(aX, Y) = a\rho(X, Y), \quad a \text{ constante.}$
- c) $\rho(X + a, Y) = \rho(X, Y), \quad a \text{ constante.}$
- d) $\rho(aX + b, Y) = a\rho(X, Y) + b, \quad a, b \text{ constantes.}$
- e) $\rho(X_1 + X_2, Y) = \rho(X_1, Y) + \rho(X_2, Y).$

357. Sea a un número en $[-1, 1]$. Encuentre variables aleatorias X y Y tales que $\rho(X, Y) = a.$

358. Sean X y Y independientes con distribución $\text{Ber}(p)$ con $p = 1/2$. Demuestre que el coeficiente de correlación entre $X + Y$ y $|X - Y|$ es cero, y sin embargo estas variables aleatorias no son independientes.

359. Sean a y b constantes. Demuestre que

- a) $\rho(X, X) = 1.$
- b) $\rho(X, -X) = -1.$
- c) $\rho(X, aX + b) = 1, \quad \text{si } a > 0.$
- d) $\rho(X, aX + b) = -1, \quad \text{si } a < 0.$
- e) $\rho(X, aX + b) = 0, \quad \text{si } a = 0.$

360. Demuestre que $\rho(aX + b, cY + d) = \text{signo}(ac) \cdot \rho(X, Y)$, en donde $ac \neq 0$ y

$$\text{signo}(x) = \begin{cases} 1 & \text{si } x > 0, \\ 0 & \text{si } x < 0. \end{cases}$$

361. Calcule el coeficiente de correlación de X y Y cuya función de densidad conjunta está dada por la siguiente tabla.

| | | |
|------------------|-----|-----|
| $x \backslash y$ | 1 | 2 |
| 0 | 1/8 | 1/4 |
| 1 | 1/2 | 1/8 |

362. Calcule el coeficiente de correlación de X y Y cuya función de densidad conjunta está dada por la siguiente tabla.

| | | | |
|------------------|-----|-----|-----|
| $x \backslash y$ | 1 | 2 | 3 |
| 2 | 1/9 | 1/9 | 1/9 |
| 4 | 1/9 | 1/9 | 1/9 |
| 6 | 1/9 | 1/9 | 1/9 |

363. Calcule el coeficiente de correlación de X y Y con distribución conjunta uniforme en el conjunto
- $\{1, \dots, n\} \times \{1, \dots, n\}$.
 - $[-1, 1] \times [-1, 1]$.
364. Sea X con distribución $\text{bin}(n, p)$ y sea $Y = n - X$. Demuestre que $\text{Cov}(X, Y) = -np(1 - p)$ y por lo tanto $\rho(X, Y) = -1$.
365. Calcule el coeficiente de correlación de X y Y cuya función de densidad conjunta es
- $f(x, y) = \frac{1}{2} \sin(x + y)$, para $x, y \in [0, \pi/2]$.
 - $f(x, y) = \frac{1}{2} e^{-x}$, para $|y| < x$.
 - $f(x, y) = e^{-x-y}$, para $x, y > 0$.
366. Sean X y Y independientes e idénticamente distribuidas. Demuestre que $\rho(X + Y, X - Y) = 0$.
367. Sea (X, Y) un vector con distribución normal bivariada $N(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Demuestre que $\rho(X, Y) = \rho$.

Distribución multinomial

368. Demuestre que la función de densidad de la distribución multinomial efectivamente lo es.
369. Sea $X = (X_1, \dots, X_k)$ con distribución multinomial (n, k, p_1, \dots, p_k) . Demuestre que X_i tiene distribución marginal $\text{bin}(n, p_i)$, para $i = 1, \dots, k$.
370. Sea $X = (X_1, \dots, X_k)$ con distribución multinomial (n, k, p_1, \dots, p_k) . Demuestre que $E(X) = (np_1, \dots, np_k)$ y que

$$[\text{Var}(X)]_{ij} = \begin{cases} np_i(1 - p_i) & \text{si } i = j, \\ -np_i p_j & \text{si } i \neq j. \end{cases}$$

Distribución hipergeométrica multivariada

371. Demuestre que la función de densidad de la distribución hipergeométrica multivariada efectivamente lo es.
372. Sea $X = (X_1, \dots, X_k)$ con distribución hipergeométrica multivariada con parámetros (N, N_1, \dots, N_k, n) . Demuestre que X_i tiene distribución hipergeométrica univariada con parámetros (N, N_i, n) , para $i = 1, \dots, k$.
373. Sea $X = (X_1, \dots, X_k)$ con distribución hipergeométrica multivariada con parámetros (N, N_1, \dots, N_k, n) . Demuestre que $E(X) = (n \frac{N_1}{N}, \dots, n \frac{N_k}{N})$, y

que

$$[\text{Var}(X)]_{ij} = \begin{cases} n \cdot \frac{N_i}{N} \cdot \frac{N - N_i}{N} \cdot \frac{N - n}{N - 1} & \text{si } i = j, \\ n \cdot \frac{N_i}{N} \cdot \frac{N_j}{N} \cdot \frac{n - N}{N - 1} & \text{si } i \neq j. \end{cases}$$

Distribución normal bivariada

374. Demuestre que la función de densidad de la distribución normal bivariada efectivamente lo es.

375. Sea (X, Y) un vector con distribución normal bivariada $N(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Demuestre que X tiene distribución marginal $N(\mu_X, \sigma_X^2)$, y Y tiene distribución marginal $N(\mu_Y, \sigma_Y^2)$. Véase el siguiente ejercicio para verificar que el recíproco de este resultado es falso.

376. Sea $f(x, y)$ la función de densidad normal bivariada estándar con $\rho = 0$. Defina

$$g(x, y) = \begin{cases} 2f(x, y) & \text{si } xy < 0, \\ 0 & \text{si } xy \geq 0. \end{cases}$$

Demuestre que $g(x, y)$ es una función de densidad bivariada que no es normal pero cuyas densidades marginales son normales.

377. Sea (X, Y) un vector con distribución normal bivariada $(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Demuestre que $E(X) = (\mu_X, \mu_Y)$, y

$$\text{Var}(X, Y) = \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \cdot \sigma_Y \\ \rho \cdot \sigma_X \cdot \sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

378. Sea (X, Y) un vector con distribución normal bivariada $N(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Demuestre que la distribución condicional de X dado que $Y = y$ es normal con media $\mu_X + \rho \frac{\sigma_Y}{\sigma_X}(y - \mu_Y)$ y varianza $\sigma_X^2(1 - \rho^2)$, y que la distribución condicional de Y dado que $X = x$ es normal con media $\mu_Y + \rho \frac{\sigma_X}{\sigma_Y}(x - \mu_X)$ y varianza $\sigma_Y^2(1 - \rho^2)$.

Capítulo 4

Esperanza condicional

En este capítulo se define el importante concepto de esperanza condicional de una variable aleatoria respecto de una σ -álgebra, y se estudian algunas de sus propiedades elementales.

4.1. Esperanza condicional

Definición (Esperanza condicional). Sea X una variable aleatoria con esperanza finita, y sea \mathcal{G} una sub- σ -álgebra de \mathcal{F} . La esperanza condicional de X dado \mathcal{G} , es una variable aleatoria denotada por $E(X|\mathcal{G})$, que cumple las siguientes tres propiedades:

1. Es \mathcal{G} -medible.
2. Tiene esperanza finita.
3. Para cualquier evento G en \mathcal{G} ,

$$E[E(X|\mathcal{G}) \cdot 1_G] = E[X \cdot 1_G]. \quad (4.1)$$

Es importante enfatizar que la esperanza condicional, a pesar de su nombre, no es un número, aunque puede serlo, sino una variable aleatoria. Usando el teorema de Radon-Nikodym (véase por ejemplo [3]), puede demostrarse que esta variable aleatoria existe y es única casi seguramente, esto significa que si existe otra variable aleatoria con las tres propiedades de la definición anterior, entonces con probabilidad uno coincide con $E(X|\mathcal{G})$. Veamos algunas propiedades elementales de esta variable aleatoria.

Proposición. La esperanza condicional cumple las siguientes propiedades.

- a) $E(E(X | \mathcal{G})) = E(X)$.
- b) Si X es \mathcal{G} -medible, entonces $E(X | \mathcal{G}) = X$. En particular, si c es una constante, $E(c | \mathcal{G}) = c$.
- c) $E(aX + Y | \mathcal{G}) = aE(X | \mathcal{G}) + E(Y | \mathcal{G})$, a constante.

Demostración. La primera propiedad se obtiene tomando el caso particular $\mathcal{G} = \Omega$ en la igualdad (4.1), y establece que las variables aleatorias X y $E(X | \mathcal{G})$ tienen la misma esperanza. Para la segunda propiedad observe que si X es \mathcal{G} -medible, entonces X mismo cumple con las tres propiedades de la definición de $E(X | \mathcal{G})$, por la unicidad se obtiene la igualdad casi segura. La tercera propiedad es consecuencia de la linealidad de la esperanza, de (4.1) y de la unicidad. \square

Cuando la σ -álgebra \mathcal{G} es la mínima respecto de la cual una función $Y : \Omega \rightarrow \mathbb{R}$ es variable aleatoria, es decir $\mathcal{G} = \sigma(Y)$, entonces la esperanza condicional se escribe simplemente como $E(X | Y)$ en lugar de $E(X | \sigma(Y))$. Si ω es tal que $Y(\omega) = y$ entonces la variable aleatoria $E(X | Y)$ evaluada en ω es

$$E(X | Y)(\omega) = E(X | Y = y) = \int_{-\infty}^{\infty} x dF_{X|Y}(x|y).$$

Los siguientes casos particulares relacionan a la esperanza condicional con los conceptos elementales de esperanza y probabilidad condicional.

Proposición. Sea X con esperanza finita, y sean A y B eventos con $0 < P(B) < 1$. Entonces

- a) $E(X | \{\emptyset, \Omega\}) = E(X)$.
- b) $E(1_A | \{\emptyset, \Omega\}) = P(A)$.
- c) $E(1_A | \{\emptyset, B, B^c, \Omega\}) = P(A | B)1_B + P(A | B^c)1_{B^c}$.

Demostración. La primera igualdad se sigue del hecho que $E(X | \mathcal{G})$ es medible respecto de \mathcal{G} y de que cualquier función medible respecto de $\mathcal{G} = \{\emptyset, \Omega\}$ es constante. La tercera condición en la definición de esperanza condicional implica que esta constante debe ser $E(X)$. La segunda igualdad es evidentemente un caso particular de la primera. Para demostrar la tercera igualdad observe que toda función medible respecto de $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$ es constante tanto en B como en B^c . Además,

$$E[E(1_A | \mathcal{G}) \cdot 1_B] = E[1_A \cdot 1_B] = P(A \cap B).$$

Como la variable aleatoria $E(1_A | \mathcal{G})$ es constante en B , el lado izquierdo es igual a $E(1_A | \mathcal{G})(\omega) \cdot P(B)$, para cualquier ω en B . De donde se obtiene $E(1_A | \mathcal{G})(\omega) = P(A|B)$ para ω en B . El análisis es análogo al considerar el evento B^c y de esto se obtiene la fórmula de la tercera propiedad. \square

Una introducción a la esperanza condicional ligeramente más completa a la presentada en esta sección, aunque también sencilla y breve, puede encontrarse en [21]. Un tratamiento más completo y riguroso puede consultarse por ejemplo en [15] o [27].

4.2. Varianza condicional

Usando la esperanza condicional se puede obtener la varianza condicional respecto de una σ -álgebra de la siguiente forma:

Definición (Varianza condicional). Sea X con segundo momento finito, y sea \mathcal{G} una sub- σ -álgebra de \mathcal{F} . La varianza condicional de X dado \mathcal{G} , denotada por $\text{Var}(X | \mathcal{G})$, se define como la variable aleatoria dada por

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

Nuevamente cuando la sub- σ -álgebra \mathcal{G} es $\sigma(Y)$, para alguna variable aleatoria Y , entonces $\text{Var}(X | \mathcal{G})$ se escribe $\text{Var}(X | Y)$ y puede tomarse como definición la igualdad

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

Se demuestran a continuación dos propiedades sencillas de esta variable aleatoria. Otras propiedades de la varianza condicional se encuentran en la sección de ejercicios.

Proposición. La varianza condicional cumple las siguientes propiedades.

- a) $\text{Var}(X | \mathcal{G}) = E(X^2 | \mathcal{G}) - E^2(X | \mathcal{G})$.
- b) $\text{Var}(X) = E[\text{Var}(X | \mathcal{G})] + \text{Var}[E(X | \mathcal{G})]$.

Demostración. La primera fórmula se obtiene a partir de la definición al desarrollar el cuadrado y utilizar las propiedades de linealidad de la esperanza condicional. Para la segunda propiedad, tomando esperanza en (a) se obtiene

$$E[\text{Var}(X | \mathcal{G})] = E(X^2) - E[E^2(X | \mathcal{G})]. \quad (4.2)$$

Por otro lado

$$\begin{aligned}\text{Var}[E(X|\mathcal{G})] &= E[E^2(X|\mathcal{G})] - E^2[E(X|\mathcal{G})] \\ &= E[E^2(X|\mathcal{G})] - E^2(X).\end{aligned}\tag{4.3}$$

Sumando (4.2) y (4.3) se obtiene (b). □

4.3. Ejercicios

Esperanza condicional

379. Demuestre que si X es \mathcal{G} -medible, entonces $E(X|\mathcal{G}) = X$.
380. Demuestre que si c es una constante, entonces para cualquier sub- σ -álgebra \mathcal{G} , $E(c|\mathcal{G}) = c$.
381. Sea A un evento y sea $\mathcal{G} = \{\emptyset, \Omega\}$. Demuestre que $E(1_A|\mathcal{G}) = P(A)$.
382. Sea X una variable aleatoria con esperanza finita y sea $\mathcal{G} = \{\emptyset, \Omega\}$. Demuestre que $E(X|\mathcal{G}) = E(X)$.
383. Sean A y B dos eventos, y suponga que $0 < P(B) < 1$. Demuestre que

$$E(1_A|1_B) = P(A|B) \cdot 1_B + P(A|B^c) \cdot 1_{B^c}.$$

384. Sea (X, Y) un vector con función de densidad dada por $f_{X,Y}(x, y) = 3y$ para $0 < x < y < 1$. Compruebe que $f(x, y)$ es efectivamente una función de densidad y calcule
- $P(X + Y < 1/2)$.
 - $f_X(x)$ y $f_Y(y)$.
 - $E(Y)$ y $E(Y|X = x)$.
385. Sea (X, Y) un vector con distribución uniforme en el conjunto $\{1, \dots, 6\} \times \{1, \dots, 6\}$. Calcule
- $P(X = Y)$.
 - $P(X + Y \leq 6)$.
 - $f_X(x)$ y $f_Y(y)$.
 - $E(X|X + Y = 6)$.
386. Sea (X, Y) un vector con función de densidad dada por la siguiente tabla

| $x \backslash y$ | -1 | 0 | 1 |
|------------------|-----|-----|-----|
| 1 | .3 | .05 | .05 |
| 2 | .05 | .2 | .05 |
| 3 | .1 | .1 | .1 |

Calcule

- a) $P(X = 2)$, $P(X + Y = 1)$ y $P(Y \leq X)$.
- b) $f_X(x)$ y $f_Y(y)$.
- c) $f_{Y|X}(y|x)$ para $x = 1, 2, 3$.
- d) $E(Y|X = x)$ para $x = 1, 2, 3$.

Varianza condicional

387. Demuestre que

- a) $\text{Var}(X | \{\emptyset, \Omega\}) = \text{Var}(X)$.
- b) $\text{Var}(1_A | \{\emptyset, \Omega\}) = P(A)(1 - P(A))$.

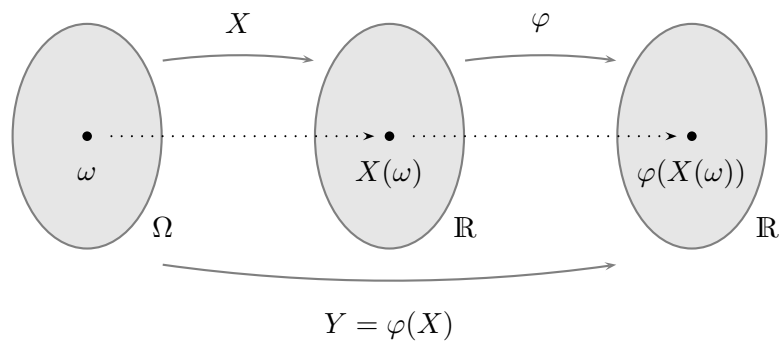
Capítulo 5

Transformaciones

Sea X una variable aleatoria con distribución conocida, y sea φ es una función tal que $Y = \varphi(X)$ es otra variable aleatoria. ¿Cuál es la distribución de Y ? En este capítulo se da respuesta a esta pregunta tanto en el caso unidimensional como en el caso de vectores aleatorios. En particular, se encuentran fórmulas explícitas para la función de densidad de la suma, diferencia, producto y cociente de dos variables aleatorias absolutamente continuas.

5.1. Transformación de una variable aleatoria

Suponga que X es una variable aleatoria y φ es una función tal que $Y = \varphi(X)$ es otra variable aleatoria. En esta sección se estudian un par de resultados que proveen de fórmulas para la función de densidad de Y , en términos de la función de densidad de X . Gráficamente:



Teorema de cambio de variable 1. Sea X una variable aleatoria continua con valores dentro de un intervalo $(a, b) \subseteq \mathbb{R}$, y con función de densidad $f_X(x)$. Sea $\varphi : (a, b) \rightarrow \mathbb{R}$ una función continua, estrictamente creciente o decreciente, y con inversa diferenciable. Entonces la variable aleatoria $Y = \varphi(X)$ toma valores dentro del intervalo $\varphi(a, b)$, y tiene función de densidad

$$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(y)) \left| \frac{d}{dy} \varphi^{-1}(y) \right| & \text{para } y \in \varphi(a, b), \\ 0 & \text{otro caso.} \end{cases}$$

Demostración. Suponga primero el caso φ estrictamente creciente. Entonces para $y \in \varphi(a, b)$,

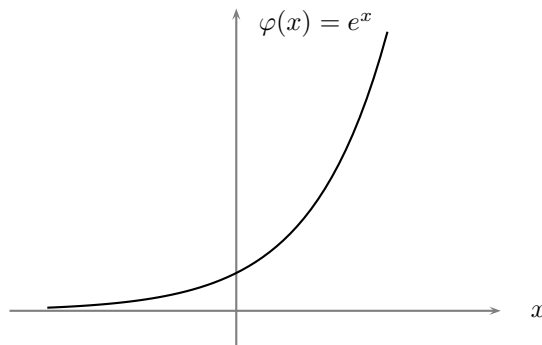
$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\varphi(X) \leq y) \\ &= P(X \leq \varphi^{-1}(y)) \\ &= F_X(\varphi^{-1}(y)). \end{aligned}$$

Derivando se obtiene $f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \frac{d}{dy} \varphi^{-1}(y)$. Para φ estrictamente decreciente

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\varphi(X) \leq y) \\ &= P(X \geq \varphi^{-1}(y)) \\ &= 1 - F_X(\varphi^{-1}(y)). \end{aligned}$$

Entonces $f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \left[-\frac{d}{dy} \varphi^{-1}(y) \right]$. En cualquiera caso se obtiene el resultado del teorema. \square

Por ejemplo, la función $\varphi(x) = e^x$, definida sobre toda la recta real cumple muy bien con las condiciones del teorema anterior. Usaremos esta función para mostrar con dos ejemplos la forma de aplicar este resultado.



Ejemplo (Distribución log normal). Sea X con distribución $N(\mu, \sigma^2)$, y sea φ la función estrictamente creciente $\varphi(x) = e^x$, con inversa diferenciable $\varphi^{-1}(y) = \ln y$. Entonces la variable aleatoria $Y = e^X$ toma valores en el intervalo $(0, \infty)$, y su distribución se conoce con el nombre de distribución log normal (μ, σ^2) . Su función de densidad es

$$f_Y(y) = \begin{cases} \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln y - \mu)^2}{2\sigma^2}\right] & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

◦

Ejemplo (Distribución log gama). Sea X con distribución gama (n, λ) , y sea nuevamente $\varphi(x) = e^x$, con inversa diferenciable $\varphi^{-1}(y) = \ln y$. Entonces la variable aleatoria $Y = e^X$ toma valores en el intervalo $(0, \infty)$, y su distribución se conoce como distribución log gama (n, λ) . Su función de densidad es

$$f_Y(y) = \begin{cases} \frac{(\lambda \ln y)^{n-1}}{\Gamma(n)} \lambda y^{-\lambda-1} & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

◦

El resultado anterior puede extenderse al caso en el que la transformación φ sea estrictamente monótona por pedazos. Se enuncia y demuestra a continuación este resultado cuando la transformación φ se descompone en dos partes monótonas, siendo fácil la extensión cuando se tiene un mayor número de secciones en donde se presente la monotonía estricta.

Teorema de cambio de variable 2. Sea X una variable aleatoria continua con valores dentro de un intervalo $(a, c) \subseteq \mathbb{R}$, y con función de densidad $f_X(x)$. Sea $\varphi : (a, c) \rightarrow \mathbb{R}$ una función tal que admite la descomposición

$$\varphi(x) = \begin{cases} \varphi_1(x) & \text{si } x \in (a, b), \\ \varphi_2(x) & \text{si } x \in (b, c), \end{cases}$$

en donde $a < b < c$, y cada una de las funciones $\varphi_1(x) : (a, b) \rightarrow \mathbb{R}$ y $\varphi_2(x) : (b, c) \rightarrow \mathbb{R}$ es continua, estrictamente creciente o decreciente, y con inversa diferenciable. Entonces la variable aleatoria $Y = \varphi(X)$ toma valores dentro del intervalo $\varphi(a, c)$, y tiene función de densidad

$$\begin{aligned} f_Y(y) &= f_X(\varphi_1^{-1}(y)) \left| \frac{d}{dy} \varphi_1^{-1}(y) \right| \cdot 1_{\varphi_1(a,b)}(y) \\ &+ f_X(\varphi_2^{-1}(y)) \left| \frac{d}{dy} \varphi_2^{-1}(y) \right| \cdot 1_{\varphi_2(b,c)}(y). \end{aligned}$$

Demostración. La prueba es análoga al caso anterior, únicamente hay que hacer el análisis sobre cada uno de los intervalos de monotonía estricta. Para cualquier y en \mathbb{R} ,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\varphi(X) \leq y) \\ &= P[(\varphi_1(X) \leq y) \cap (X \in (a, b))] + P[(\varphi_2(X) \leq y) \cap (X \in (b, c))]. \end{aligned}$$

Nos interesa el comportamiento de estas probabilidades como funciones de y , puesto que calcularemos la derivada de ellas para encontrar $f_Y(y)$. Por ejemplo, la primera probabilidad, vista como función de y , es

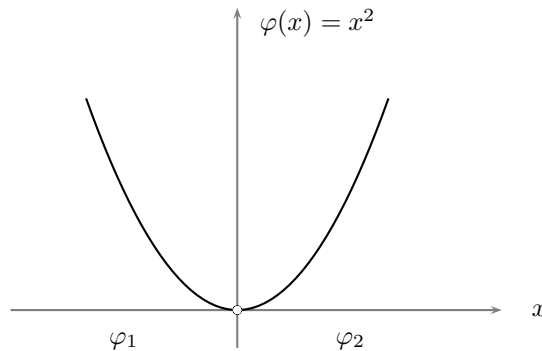
$$y \mapsto P[(\varphi_1(X) \leq y) \cap (X \in (a, b))],$$

que permanece constante para $y \notin \varphi_1(a, b)$, de modo que, suponiendo por ejemplo φ_1 creciente,

$$\begin{aligned} \frac{d}{dy} P[(\varphi_1(X) \leq y) \cap (X \in (a, b))] &= \frac{d}{dy} P[(\varphi_1(X) \leq y) \cap (X \in (a, b))] \cdot 1_{\varphi_1(a, b)}(y) \\ &= \frac{d}{dy} P[(X \leq \varphi_1^{-1}(y)) \cap (X \in (a, b))] \cdot 1_{\varphi_1(a, b)}(y) \\ &= \frac{d}{dy} P[a < X \leq \varphi_1^{-1}(y)] \cdot 1_{\varphi_1(a, b)}(y) \\ &= \frac{d}{dy} F_X(\varphi_1^{-1}(y)) \cdot 1_{\varphi_1(a, b)}(y) \\ &= f_X(\varphi_1^{-1}(y)) \frac{d}{dy} \varphi_1^{-1}(y) \cdot 1_{\varphi_1(a, b)}(y). \end{aligned}$$

De manera análoga se procede respecto del segundo sumando, considerando también el caso cuando se presenta la monotonía decreciente. De esta forma se obtiene la fórmula enunciada. \square

Ejemplo. Sea X continua con función de densidad $f_X(x)$. Considere la transformación $\varphi(x) = x^2$, la cual es estrictamente decreciente en $(-\infty, 0)$, y estrictamente creciente en $(0, \infty)$.



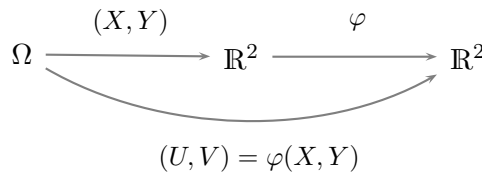
Defina entonces las funciones monótonas $\varphi_1(x) = x^2$ sobre $(-\infty, 0)$, y $\varphi_2(x) = x^2$ sobre $(0, \infty)$. Entonces sus inversas son $\varphi_1^{-1}(y) = -\sqrt{y}$, y $\varphi_2^{-1}(y) = \sqrt{y}$. La variable $Y = X^2$ tiene por lo tanto función de densidad

$$f_Y(y) = \begin{cases} f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} & \text{si } y > 0, \\ 0 & \text{otro caso.} \end{cases}$$

◦

5.2. Transformación de un vector aleatorio

Suponga ahora que (X, Y) es un vector con función de densidad conocida, y φ es una función tal que $(U, V) = \varphi(X, Y)$ es otro vector aleatorio. El problema es encontrar la función de densidad del nuevo vector (U, V) . Gráficamente



Teorema de cambio de variable 3. Sea (X, Y) un vector continuo con valores en $I \subseteq \mathbb{R}^2$, y con función de densidad $f_{X,Y}(x, y)$. Sea $\varphi(x, y) : I \rightarrow \mathbb{R}^2$ una función continua con inversa $\varphi^{-1}(u, v)$ diferenciable. Entonces el vector $(U, V) = \varphi(X, Y)$ toma valores en $\varphi(I)$ y tiene función de densidad

$$f_{U,V}(u, v) = \begin{cases} f_{X,Y}(\varphi^{-1}(u, v)) |J(u, v)| & \text{para } (u, v) \in \varphi(I), \\ 0 & \text{otro caso,} \end{cases} \quad (5.1)$$

en donde

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix}.$$

Una prueba rigurosa de este teorema resulta ser un tanto elaborada, y por simplicidad se omite. Sin embargo, puede usarse el siguiente argumento intuitivo para encontrar la fórmula enunciada. Sea

$$(U, V) = \varphi(X, Y) = (\varphi_1(X, Y), \varphi_2(X, Y))$$

con inversa

$$(X, Y) = \varphi^{-1}(U, V) = (\varphi_1^{-1}(U, V), \varphi_2^{-1}(U, V)).$$

Sea A el cuadrado de área infinitesimal de esquinas con coordenadas (x, y) , $(x + dx, y)$, $(x, y + dy)$ y $(x + dx, y + dy)$. Bajo la transformación φ las coordenadas de las esquinas del cuadrado A se transforman en las siguientes coordenadas

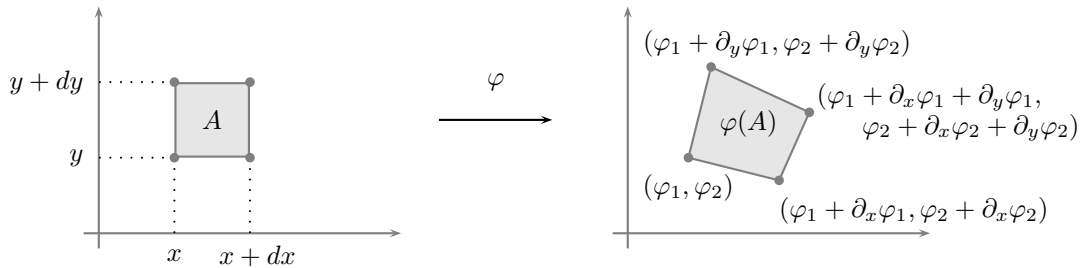
$$(x, y) \mapsto (\varphi_1(x, y), \varphi_2(x, y)).$$

$$\begin{aligned} (x + dx, y) &\mapsto (\varphi_1(x + dx, y), \varphi_2(x + dx, y)) \\ &\doteq (\varphi_1(x, y) + \partial_x \varphi_1(x, y)dx, \varphi_2(x, y) + \partial_x \varphi_2(x, y)dx). \end{aligned}$$

$$\begin{aligned} (x, y + dy) &\mapsto (\varphi_1(x, y + dy), \varphi_2(x, y + dy)) \\ &\doteq (\varphi_1(x, y) + \partial_y \varphi_1(x, y)dy, \varphi_2(x, y) + \partial_y \varphi_2(x, y)dy). \end{aligned}$$

$$\begin{aligned} (x + dx, y + dy) &\mapsto (\varphi_1(x + dx, y + dy), \varphi_2(x + dx, y + dy)) \\ &\doteq (\varphi_1(x, y) + \partial_x \varphi_1(x, y)dx + \partial_y \varphi_1(x, y)dy, \\ &\quad \varphi_2(x, y) + \partial_x \varphi_2(x, y)dx + \partial_y \varphi_2(x, y)dy). \end{aligned}$$

Gráficamente la transformación de estos puntos se muestra a continuación.



Entonces $P((X, Y) \in A) = P((U, V) \in \varphi(A))$. Por lo tanto

$$f_{X,Y}(x, y) dx dy = f_{U,V}(u, v) \times \text{“Área de } \varphi(A)\text{”}.$$

En donde

$$\begin{aligned} \text{“Área de } \varphi(A)\text{”} &= |\partial_x \varphi_1 \cdot \partial_y \varphi_2 - \partial_x \varphi_2 \cdot \partial_y \varphi_1| dx dy \\ &= \left\| \begin{vmatrix} \partial_x \varphi_1 & \partial_y \varphi_1 \\ \partial_x \varphi_2 & \partial_y \varphi_2 \end{vmatrix} \right\| dx dy \\ &= |J(x, y)| dx dy. \end{aligned}$$

Además $|J(x, y)| = \frac{1}{|J(u, v)|}$. Por lo tanto

$$f_{X,Y}(x, y) dx dy = f_{U,V}(u, v) \frac{dx dy}{|J(u, v)|}.$$

De esta ecuación se obtiene

$$f_{U,V}(u, v) = f_{X,Y}(\varphi_1^{-1}(u, v), \varphi_2^{-1}(u, v)) |J(u, v)|.$$

Como ejemplo de aplicación de la proposición anterior, en las secciones siguientes utilizaremos la fórmula (5.1) para encontrar expresiones para la función de densidad de la suma, diferencia, producto y cociente de dos variables aleatorias. Las fórmulas generales sobre transformaciones encontradas en estas dos primeras secciones se resumen en la siguiente tabla.

| Transformaciones | |
|--------------------------|--|
| $Y = \varphi(X)$ | $\Rightarrow f_Y(y) = f_X(\varphi^{-1}(y)) \left \frac{d}{dy} \varphi^{-1}(y) \right .$ |
| $(U, V) = \varphi(X, Y)$ | $\Rightarrow f_{U,V}(u, v) = f_{X,Y}(\varphi^{-1}(u, v)) J(u, v) $ |
| | en donde $J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix}.$ |

Distribución de la suma

El siguiente resultado proporciona una fórmula para la función de densidad de la suma de dos variables aleatorias absolutamente continuas.

Proposición. Sea (X, Y) un vector continuo con función de densidad conjunta $f_{X,Y}(x, y)$. Entonces $X + Y$ tiene función de densidad

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u - v, v) dv. \quad (5.2)$$

Demostración. Sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la transformación $\varphi(x, y) = (x + y, y)$, con inversa $\varphi^{-1}(u, v) = (u - v, v)$. El Jacobiano de la transformación inversa es

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1.$$

Por la fórmula (5.1), $f_{X+Y,Y}(u, v) = f_{X,Y}(u - v, v)$. Integrando respecto a v se obtiene (5.2). \square

Observe que haciendo el cambio de variable $z(v) = u - v$ en (5.2) se obtiene la expresión equivalente

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(z, u - z) dz. \quad (5.3)$$

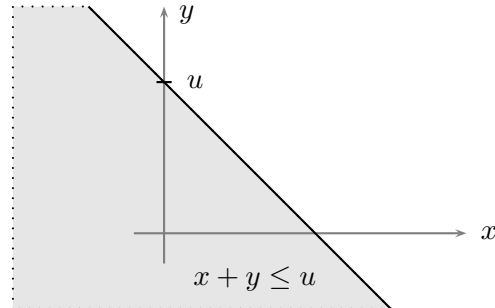
En particular, cuando X y Y son independientes, la fórmula (5.2) se reduce a

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(u - v)f_Y(v) dv. \quad (5.4)$$

Se puede demostrar la proposición anterior mediante el procedimiento usual de encontrar primero la función de distribución de $X+Y$ y después derivar para encontrar la función de densidad. Por definición,

$$\begin{aligned} F_{X+Y}(u) &= P(X + Y \leq u) \\ &= \iint_{\{(x,y) \mid x+y \leq u\}} f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{u-x} f_{X,Y}(x, y) dy dx. \end{aligned}$$

La región de integración es la siguiente:



Derivando respecto a u se obtiene

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, u - x) dx,$$

que corresponde a la expresión (5.3) equivalente a (5.2).

Convolución

Definición (Convolución). La convolución de dos funciones de densidad continuas f_1 y f_2 , es una función de densidad denotada por $f_1 * f_2$, y definida como sigue

$$(f_1 * f_2)(x) = \int_{-\infty}^{\infty} f_1(x-y)f_2(y) dy.$$

Más generalmente, la convolución de dos funciones de distribución F_1 y F_2 es la función de distribución

$$(F_1 * F_2)(x) = \int_{-\infty}^{\infty} F_1(x-y)dF_2(y).$$

En consecuencia, si X y Y son dos variables aleatorias continuas independientes con correspondientes funciones de densidad $f_1(x)$ y $f_2(x)$, entonces la función de densidad de $X + Y$ es la convolución $(f_1 * f_2)(x)$. Observe que hemos denotado la convolución por el mismo símbolo, primero cuando los argumentos son funciones de densidad y en el otro cuando son funciones de distribución. Para el caso de funciones de distribución absolutamente continuas, se tiene la relación

$$\frac{d}{dx}(F_1 * F_2)(x) = (f_1 * f_2)(x).$$

En el caso cuando X y Y son discretas, independientes y con valores enteros, entonces es sencillo verificar que la función de probabilidad de $X + Y$ es, en completa analogía con (5.4),

$$f_{X+Y}(u) = \sum_k f_X(u-k)f_Y(k).$$

Distribución de la diferencia

Se encontrará ahora una fórmula para la función de densidad de la diferencia de dos variables aleatorias.

Proposición. Sea (X, Y) un vector absolutamente continuo con función de densidad $f_{X,Y}(x, y)$. Entonces $X - Y$ tiene función de densidad

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u+v, v) dv. \quad (5.5)$$

Demostración. Procedemos como en la sección anterior. Sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la transformación $\varphi(x, y) = (x - y, y)$ con inversa $\varphi^{-1}(u, v) = (u + v, v)$. El Jacobiano de la transformación inversa es

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1.$$

Por la fórmula (5.1), $f_{X-Y, Y}(u, v) = f_{X, Y}(u + v, v)$. Integrando respecto a v se obtiene (5.5). \square

Con el cambio de variable $z(v) = u + v$ en (5.5) se obtiene la expresión equivalente

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_{X, Y}(z, z - u) dz. \quad (5.6)$$

Cuando X y Y son independientes la fórmula (5.5) se reduce a

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_X(u + v) f_Y(v) dv.$$

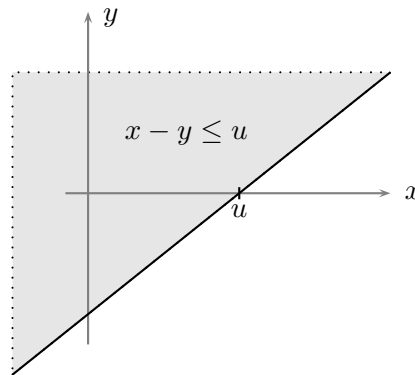
En el caso discreto cuando X y Y son independientes con valores enteros, la variable $X - Y$ también toma valores enteros, y tiene función de probabilidad

$$f_{X-Y}(u) = \sum_k f_X(u + k) f_Y(k).$$

Nuevamente se puede demostrar la proposición anterior mediante el procedimiento usual de encontrar primero la función de distribución y después derivar para encontrar la función de densidad. Por definición,

$$\begin{aligned} F_{X-Y}(u) &= P(X - Y \leq u) \\ &= \int \int_{\{(x, y) \mid x - y \leq u\}} f_{X, Y}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{x-u}^{\infty} f_{X, Y}(x, y) dy dx. \end{aligned}$$

La región de integración es la siguiente:



Derivando respecto a u se obtiene (5.6) equivalente a (5.5). A partir de la fórmula para la suma de dos variables aleatorias se puede construir una tercera demostración de (5.5). Por la fórmula para la suma,

$$f_{X-Y}(u) = f_{X+(-Y)}(u) = \int_{-\infty}^{\infty} f_{X,-Y}(u-v, v) dv.$$

Haciendo el cambio de variable $x = -v$, se obtiene

$$\begin{aligned} f_{X-Y}(u) &= \int_{-\infty}^{\infty} f_{X,-Y}(u+x, -x) dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(u+x, x) dx. \end{aligned}$$

Distribución del producto

Ahora se encontrará una fórmula para la función de densidad del producto de dos variables aleatorias absolutamente continuas.

Proposición. Sea (X, Y) un vector continuo con función de densidad conjunta $f_{X,Y}(x, y)$. Entonces XY tiene función de densidad

$$f_{XY}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u/v, v) \left| \frac{1}{v} \right| dv. \quad (5.7)$$

Demostración. Se usa nuevamente la fórmula (5.1). Sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la transformación $\varphi(x, y) = (xy, y)$ cuya inversa es, para $v \neq 0$, $\varphi^{-1}(u, v) = (u/v, v)$. El Jacobiano de la transformación inversa es

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix} = \begin{vmatrix} 1/v & u/v^2 \\ 0 & 1 \end{vmatrix} = \frac{1}{v}.$$

Por la fórmula (5.1), para $v \neq 0$, $f_{XY,Y}(u, v) = f_{X,Y}(u/v, v) |1/v|$. Integrando respecto a v se obtiene (5.7). \square

Haciendo $x(v) = u/v$ en (5.7) se obtiene la expresión equivalente

$$f_{XY}(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, u/x) \left| \frac{1}{x} \right| dx. \quad (5.8)$$

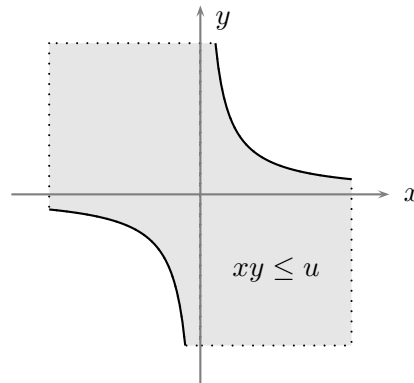
Cuando X y Y son independientes la fórmula (5.7) se reduce a

$$f_{XY}(u) = \int_{-\infty}^{\infty} f_X(u/v) f_Y(v) \left| \frac{1}{v} \right| dv.$$

Usaremos el procedimiento usual de encontrar primero la función de distribución de XY y después derivar para encontrar la función de densidad. Por definición,

$$\begin{aligned} F_{XY}(u) &= P(XY \leq u) \\ &= \int \int_{\{(x,y): xy \leq u\}} f_{X,Y}(x,y) dy dx \\ &= \int_{-\infty}^0 \int_{u/x}^{\infty} f_{X,Y}(x,y) dy dx + \int_0^{\infty} \int_{-\infty}^{u/x} f_{X,Y}(x,y) dy dx. \end{aligned}$$

La región de integración para $u > 0$ es la siguiente:



Derivando respecto a u ,

$$\begin{aligned} f_{XY}(u) &= \int_{-\infty}^0 f_{X,Y}(x, u/x)(-1/x) dy dx + \int_0^{\infty} f_{X,Y}(x, u/x)(1/x) dy dx. \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, u/x)|1/x| dx, \end{aligned}$$

que corresponde a (5.8), equivalente a (5.7).

Distribución del cociente

Finalmente se encontrará una fórmula para el cociente de dos variables aleatorias absolutamente continuas.

Proposición. Sea (X, Y) un vector continuo con función de densidad conjunta $f_{X,Y}(x, y)$ y tal que $Y \neq 0$. Entonces X/Y tiene función de densidad

$$f_{X/Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(uv, v) |v| dv. \quad (5.9)$$

Demostración. Procederemos como en las secciones anteriores. Sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la transformación $\varphi(x, y) = (x/y, y)$ para $y \neq 0$, y con inversa $\varphi^{-1}(u, v) = (uv, v)$. El Jacobiano de la transformación inversa es

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v.$$

Por la fórmula (5.1), $f_{X/Y}(u, v) = f_{X,Y}(uv, v) |v|$, de donde se obtiene (5.9). \square

Haciendo $x(v) = uv$ en (5.9) se obtiene la expresión equivalente

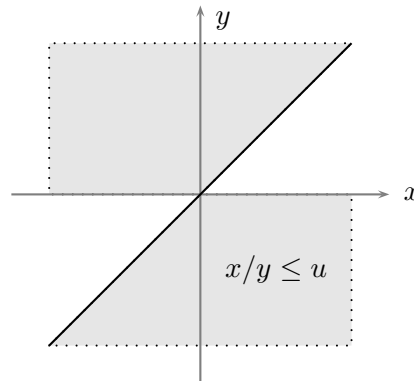
$$f_{X/Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, x/u) |x/u^2| dx. \quad (5.10)$$

Observe nuevamente que cuando X y Y son independientes el integrando en la fórmula (5.9) se escribe como el producto de las densidades marginales.

Ahora usaremos el procedimiento usual de encontrar primero la función de distribución y después derivar para encontrar la función de densidad.

$$\begin{aligned} F_{X/Y}(u) &= P(X/Y \leq u) \\ &= \int \int_{\{(x,y) : x/y \leq u\}} f_{X,Y}(x, y) dx dy \\ &= \int_{-\infty}^0 \int_{uy}^{\infty} f_{X,Y}(x, y) dx dy + \int_0^{\infty} \int_{-\infty}^{uy} f_{X,Y}(x, y) dx dy. \end{aligned}$$

La región de integración para $u > 0$ es la siguiente:



Derivando respecto a u ,

$$\begin{aligned} f_{X/Y}(u) &= - \int_{-\infty}^0 f_{X,Y}(uy, y) y dy + \int_0^{\infty} f_{X,Y}(uy, y) y dy \\ &= \int_{-\infty}^{\infty} f_{X,Y}(uy, y) |y| dy. \end{aligned}$$

A partir de la fórmula para el producto de dos variables aleatorias se puede construir una tercera demostración de (5.9) de la forma siguiente.

$$f_{X/Y}(u) = f_{X \cdot (1/Y)}(u) = \int_{-\infty}^{\infty} f_{X,1/Y}(u/v, v) \left| \frac{1}{v} \right| dv.$$

Haciendo el cambio de variable $x = 1/v$ se obtiene

$$\begin{aligned} f_{X/Y}(u) &= \int_{-\infty}^{\infty} f_{X,1/Y}(ux, 1/x) |x| dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(ux, x) |x| dx. \end{aligned}$$

Las fórmulas encontradas se resumen en la siguiente tabla.

Fórmulas para la suma, diferencia, producto y cociente de dos variables aleatorias absolutamente continuas

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u-v, v) dv$$

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u+v, v) dv$$

$$f_{XY}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u/v, v) \left| \frac{1}{v} \right| dv$$

$$f_{X/Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(uv, v) |v| dv$$

5.3. Ejercicios

Transformación de una v.a.

- 388. Sea X con distribución $\text{unif}(0,1)$ y sea $\lambda > 0$. Demuestre que la variable aleatoria $Y = -(\ln X)/\lambda$ tiene distribución $\exp(\lambda)$.
- 389. Sea X con distribución $\exp(\lambda)$. Encuentre la función de densidad y de distribución de $Y = 1 - \exp(-\lambda X)$.
- 390. Encuentre la distribución de $Y = 1/X$ cuando X tiene distribución
 - a) $\text{unif}(0,1)$.
 - b) $\exp(\lambda)$.

391. Sea X continua con función de densidad $f_X(x)$. Demuestre que

$$f_{|X|}(x) = \begin{cases} f_X(-x) + f_X(x) & \text{si } x > 0, \\ 0 & \text{otro caso.} \end{cases}$$

392. Sea X con distribución uniforme en el intervalo $(0, 2\pi)$. Encuentre la función de densidad de

- a) $Y = \text{sen}(X)$.
- b) $Y = \text{cos}(X)$.

393. Encuentre la distribución de $Y = X^n$ para cada n en \mathbb{N} , cuando X tiene distribución

- a) $\text{unif}(0, 1)$.
- b) $\text{unif}(-1, 1)$.
- c) $\exp(\lambda)$.

394. Sea X con distribución $\text{unif}(-1, 1)$. Encuentre la función de densidad de X^2 .

395. Sea X absolutamente continua con función de distribución $F(x)$. Demuestre que $Y = F(X)$ tiene distribución $\text{unif}[0, 1]$.

396. Encuentre la función de densidad de $Y = 1/X$ cuando X tiene función de densidad

$$f_X(x) = \begin{cases} 1/2 & \text{si } 0 < x \leq 1, \\ 1/(2x^2) & \text{si } x > 1, \\ 0 & \text{otro caso.} \end{cases}$$

397. Sea X con distribución $\text{unif}(a, b)$. Encuentre la distribución de la variable aleatoria $Y = X/(b - X)$.

Transformación de un vector aleatorio

398. Sean X y Y independientes ambas con distribución $\text{unif}(0, 1)$. Encuentre la función de densidad del vector

- a) $(X, X + Y)$.
- b) $(X + Y, X - Y)$.

399. Sean X y Y independientes ambas con distribución $\text{unif}(-1, 1)$. Encuentre la función de densidad del vector

- a) $(X + Y, X - Y)$.
- b) $|Y - X|$.
- c) $(X - Y, Y - X)$.

400. Sea (X, Y) un vector con distribución uniforme en el círculo unitario $\{(x, y) : x^2 + y^2 \leq 1\}$. Encuentre la función de densidad del vector

$$(R, \Theta) = (\sqrt{X^2 + Y^2}, \arctan(Y/X)).$$

Distribución de la suma

401. Encuentre la función de densidad de la suma de dos variables aleatorias cuya función de densidad conjunta es

a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.

b) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

c) $f(x, y) = e^{-y}$, para $0 < x < y$.

d) $f_{X,Y}(x, y) = 8xy$, para $0 < x < y < 1$.

e) $f_{X,Y}(x, y) = 4x(1 - y)$, para $0 < x, y < 1$.

402. Encuentre la función de densidad de la suma de dos variables aleatorias independientes cada una de ellas con distribución

a) $\text{unif}(0, 1)$.

b) $\exp(\lambda)$.

403. Encuentre la función de densidad de la suma de dos variables aleatorias independientes cada una de ellas con función de densidad

a) $f(x) = 2x$, para $0 < x < 1$.

b) $f(x) = 6x(1 - x)$, para $0 < x < 1$.

c) $f(x) = (1 + x)/2$, para $-1 < x < 1$.

404. Encuentre la función de densidad de la suma de dos variables aleatorias independientes X y Y , tales que

a) $X \sim \text{unif}(-1, 0)$ y $Y \sim \text{unif}(0, 1)$.

b) $X \sim \text{unif}(0, 1)$ y $Y \sim \exp(\lambda)$.

405. Sea (X, Y, Z) un vector absolutamente continuo. Demuestre que $X + Y + Z$ tiene función de densidad

$$f_{X+Y+Z}(u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(u - y - z, y, z) dy dz.$$

Aplique esta fórmula para encontrar la función de densidad de la suma de tres variables aleatorias independientes, en donde cada sumando tiene distribución $\text{unif}(0, 1)$.

406. Sea (X_1, \dots, X_n) un vector aleatorio absolutamente continuo. Demuestre que $X_1 + \dots + X_n$ tiene función de densidad

$$f(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(u - v_2 - \cdots - v_n, v_2, \dots, v_n) dv_2 \cdots dv_n.$$

Aplique esta fórmula para encontrar la función de densidad de la suma de n variables aleatorias independientes, en donde cada sumando tiene distribución $\text{unif}(0, 1)$.

407. Encuentre la función de densidad de la suma de dos variables aleatorias con distribución conjunta uniforme en el cuadrado $(-1, 1) \times (-1, 1)$.
408. Encuentre la función de densidad de la suma de tres variables aleatorias con distribución conjunta uniforme en el cubo $(-1, 1) \times (-1, 1) \times (-1, 1)$.
409. Encuentre la función de densidad de la suma de n variables aleatorias con distribución conjunta uniforme en el hipercubo

$$\underbrace{(-1, 1) \times \cdots \times (-1, 1)}_n.$$

410. Demuestre que la suma de dos variables aleatorias independientes, cada una de ellas con distribución normal, tiene nuevamente distribución normal, con media la suma de las medias, y varianza la suma de las varianzas.
411. Sean X_1, \dots, X_n independientes en donde X_k tiene distribución $N(\mu_k, \sigma_k^2)$ para $k = 1, \dots, n$. Sean c_1, \dots, c_n constantes dadas, no todas cero. Demuestre que

$$\sum_{k=1}^n c_k X_k \sim N\left(\sum_{k=1}^n c_k \mu_k, \sum_{k=1}^n c_k^2 \sigma_k^2\right).$$

412. Sean X_1, \dots, X_n independientes y con idéntica distribución $N(\mu, \sigma^2)$. Demuestre que el promedio $(X_1 + \cdots + X_n)/n$ tiene distribución $N(\mu, \sigma^2/n)$.
413. Demuestre que la suma de dos variables aleatorias independientes, cada una de ellas con distribución $\exp(\lambda)$, tiene distribución $\text{gama}(2, \lambda)$. Más generalmente, demuestre que la suma de n variables aleatorias independientes, cada una de ellas con distribución $\exp(\lambda)$, tiene distribución $\text{gama}(n, \lambda)$.
414. Demuestre que la suma de dos variables aleatorias independientes con distribución $\text{gama}(n, \lambda)$ y $\text{gama}(m, \lambda)$, tiene distribución $\text{gama}(n + m, \lambda)$.

Distribución de la diferencia

415. Sea (X, Y, Z) un vector absolutamente continuo con función de densidad $f_{X,Y,Z}(x, y, z)$. Demuestre que $X - Y - Z$ tiene función de densidad

$$f_{X-Y-Z}(u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(u + y + z, y, z) dy dz.$$

Aplique esta fórmula para encontrar la función de $X - Y - Z$, cuando estas variables son independientes y cada una de ellas tiene distribución $\text{unif}(0, 1)$.

416. Encuentre la función de densidad de $X - Y$, para (X, Y) un vector con función de densidad conjunta

- a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.
- b) $f(x, y) = e^{-x-y}$, para $x, y > 0$.
- c) $f(x, y) = e^{-y}$, para $0 < x < y$.

- d) $f_{X,Y}(x, y) = 8xy$, para $0 < x < y < 1$.
 e) $f_{X,Y}(x, y) = 4x(1 - y)$, para $0 < x, y < 1$.
417. Encuentre la función de densidad de $X - Y$, cuando X y Y son independientes y ambas con distribución
- a) $\text{unif}(0, 1)$.
 b) $\exp(\lambda)$.
418. Encuentre la función de densidad de $X - Y$, cuando X y Y son independientes y ambas con función de densidad
- a) $f(x) = 2x$, para $0 < x < 1$.
 b) $f(x) = 6x(1 - x)$, para $0 < x < 1$.
 c) $f(x) = (1 + x)/2$, para $-1 < x < 1$.
419. Encuentre la función de densidad de $X - Y$, cuando X y Y son independientes y tales que
- a) $X \sim \text{unif}(0, 1)$ y $Y \sim \text{unif}(1, 2)$.
 b) $X \sim \text{unif}(0, 1)$ y $Y \sim \exp(\lambda)$.
420. Demuestre que la diferencia entre dos variables aleatorias independientes ambas con distribución uniforme en el intervalo $(a - 1/2, a + 1/2)$ tiene función de densidad
- $$f(u) = \begin{cases} 1 - |u| & \text{si } -1 < u < 1, \\ 0 & \text{otro caso.} \end{cases}$$
421. Demuestre que la diferencia de dos variables aleatorias independientes, cada una de ellas con distribución normal, tiene nuevamente distribución normal, con media la diferencia de las medias, y varianza la suma de las varianzas.

Distribución del producto

422. Encuentre la función de densidad del producto de dos variables aleatorias independientes ambas con distribución
- a) $\text{unif}(0, 1)$.
 b) $\exp(\lambda)$.
 c) $N(0, 1)$.
423. Encuentre la función de densidad del producto de dos variables aleatorias cuya función de densidad conjunta es
- a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.
 b) $f(x, y) = e^{-x-y}$, para $x, y > 0$.
 c) $f(x, y) = e^{-y}$, para $0 < x < y$.
 d) $f_{X,Y}(x, y) = 8xy$, para $0 < x < y < 1$.

- e) $f_{X,Y}(x,y) = 4x(1-y)$, para $0 < x, y < 1$.
424. Encuentre la función de densidad del producto de dos variables aleatorias independientes cada una de ellas con función de densidad
- a) $f(x) = 2x$, para $0 < x < 1$.
 b) $f(x) = 6x(1-x)$, para $0 < x < 1$.
 c) $f(x) = (1+x)/2$, para $-1 < x < 1$.
425. Encuentre la función de densidad del producto de dos variables aleatorias independientes X y Y , tales que
- a) $X \sim \text{unif}(-1, 0)$ y $Y \sim \text{unif}(0, 1)$.
 b) $X \sim \text{unif}(0, 1)$ y $Y \sim \exp(\lambda)$.

Distribución del cociente

426. Encuentre la función de densidad de X/Y para (X, Y) un vector con función de densidad
- a) $f(x,y) = \frac{1}{ab}$ para $0 < x < a, 0 < y < b$.
 b) $f(x,y) = e^{-x-y}$, para $x, y > 0$.
 c) $f(x,y) = e^{-y}$, para $0 < x < y$.
 d) $f(x,y) = 8xy$, para $0 < x < y < 1$.
 e) $f(x,y) = 4x(1-y)$, para $0 < x, y < 1$.
 f) $f(x,y) = 2e^{-x-y}$, para $0 < x < y$.
427. Encuentre la función de densidad de X/Y cuando X y Y son independientes y ambas con distribución
- a) $\exp(\lambda)$.
 b) $\text{unif}(0, 1)$.
428. Encuentre la función de densidad de X/Y cuando X y Y son independientes y ambas con densidad
- a) $f(x) = 2x$, para $0 < x < 1$.
 b) $f(x) = 6x(1-x)$, para $0 < x < 1$.
 c) $f(x) = (1+x)/2$, para $-1 < x < 1$.
429. Encuentre la función de densidad de X/Y cuando X y Y son independientes y tales que
- a) $X \sim \text{unif}(-1, 1)$ y $Y \sim \text{unif}(0, 1)$.
 b) $X \sim \text{unif}(0, 1)$ y $Y \sim \exp(\lambda)$.
430. Sean X y Y independientes con distribución $\exp(\lambda)$. Encuentre la función de densidad de $X/(X+Y)$.
431. Sean X y Y independientes ambas con distribución normal estándar. Demuestre que la variable aleatoria X/Y tiene distribución Cauchy.

Capítulo 6

Distribuciones muestrales y estadísticas de orden

Se estudian ahora algunas distribuciones de probabilidad que surgen en la estadística y otras áreas de aplicación de la probabilidad.

Definición (Muestra aleatoria). Una muestra aleatoria es una colección de variables aleatorias X_1, \dots, X_n , que cumplen la condición de ser independientes y de tener cada una de ellas la misma distribución de probabilidad. Al número n se le llama tamaño de la muestra aleatoria.

A menudo se escribe m.a. para abreviar el término *muestra aleatoria*, y se usan las siglas v.a.i.i.d. para denotar el término *variables aleatorias independientes e idénticamente distribuidas*. Por lo tanto, una m.a. es una colección de v.a.i.i.d.

Definición (Estadística). Una estadística es una variable aleatoria de la forma: $g(X_1, \dots, X_n)$, en donde X_1, \dots, X_n es una muestra aleatoria, y $g : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función Borel medible.

Ejemplo (Media y varianza muestral). La *media muestral* es una estadística denotada por \bar{X} y definida como sigue

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Observe que \bar{X} es una combinación lineal de los elementos de la m.a. y por lo tanto es una v.a. Otro ejemplo importante de estadística es la *varianza muestral*, denotada

por S^2 y definida como sigue

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Observe que en el denominador aparece el número de sumandos menos uno. ◦

La media y la varianza muestrales tienen la característica de ser estimadores insesgados para la media y la varianza, respectivamente, de una distribución cualquiera. En particular, cuando la muestra aleatoria proviene de una distribución normal, resulta que la media y la varianza muestrales son variables aleatorias independientes. Utilizaremos este interesante e inesperado resultado más adelante, y cuya demostración puede encontrarse en [17].

Proposición. Sea X_1, \dots, X_n una m.a. de la distribución $N(\mu, \sigma^2)$. Entonces las estadísticas \bar{X} y S^2 son independientes.

Este resultado no es válido para cualquier distribución de probabilidad, por ejemplo, no es difícil verificar esta afirmación para una muestra aleatoria de la distribución Bernoulli. En la siguiente sección se estudian algunas distribuciones de probabilidad estrechamente relacionadas con la media y la varianza muestral.

6.1. Distribuciones muestrales

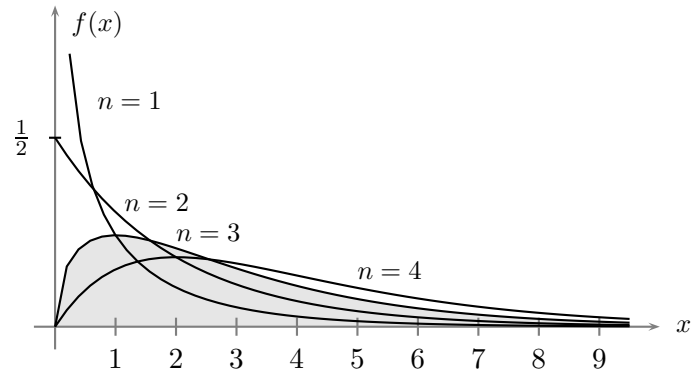
Se estudian a continuación algunas distribuciones de probabilidad que surgen en la estadística al considerar funciones de una muestra aleatoria.

Distribución ji-cuadrada

La variable aleatoria continua X tiene una distribución ji-cuadrada con $n > 0$ grados de libertad, si su función de densidad es

$$f(x) = \begin{cases} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

La gráfica de esta función es



Función de densidad $\chi^2(n)$.

En este caso se escribe $X \sim \chi^2(n)$, y puede demostrarse que $E(X) = n$, y $\text{Var}(X) = 2n$. Observe que la distribución $\chi^2(n)$ con $n = 2$ se reduce a la distribución $\exp(\lambda)$ con $\lambda = 1/2$. La distribución ji-cuadrada puede encontrarse como indican los siguientes resultados.

Proposición. Si $X \sim N(0, 1)$, entonces $X^2 \sim \chi^2(1)$.

Demostración. Para $x > 0$,

$$\begin{aligned}
 f_{X^2}(x) &= f_X(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_X(-\sqrt{x}) \frac{1}{2\sqrt{x}} \\
 &= f_X(\sqrt{x}) \frac{1}{\sqrt{x}} \\
 &= \frac{1}{\sqrt{2\pi}} e^{-x/2} \frac{1}{\sqrt{x}} \\
 &= \frac{1}{\Gamma(1/2)} \left(\frac{1}{2}\right)^{1/2} x^{1/2-1} e^{-x/2}.
 \end{aligned}$$

Esta expresión corresponde a la función de densidad de la distribución $\chi^2(1)$. \square

La suma de dos o más variables aleatorias independientes con distribución ji-cuadrada es nuevamente una variable aleatoria ji-cuadrada, y sus grados de libertad son la suma de los grados de libertad de cada uno de los sumandos. Este es el contenido de la siguiente proposición.

Proposición. Sean X_1, \dots, X_m independientes tales que cada X_i tiene distribución $\chi^2(n_i)$, para $i = 1, \dots, m$. Entonces

$$\sum_{i=1}^m X_i \sim \chi^2(n_1 + \dots + n_m).$$

Demostración. Es suficiente demostrar el resultado para el caso de dos variables aleatorias. Sean X y Y independientes con distribución ji-cuadrada con grados de libertad n y m , respectivamente. Este ligero cambio en la notación evitará el uso de subíndices. Por la fórmula (5.2), para $u > 0$,

$$\begin{aligned} f_{X+Y}(u) &= \int_0^u f_X(u-v)f_Y(v) dv \\ &= \int_0^u \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} (u-v)^{n/2-1} e^{-(u-v)/2} \\ &\quad \frac{1}{\Gamma(m/2)} \left(\frac{1}{2}\right)^{m/2} v^{m/2-1} e^{-v/2} dv \\ &= \frac{1}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{1}{2}\right)^{(n+m)/2} e^{-u/2} \\ &\quad \int_0^u (u-v)^{n/2-1} v^{m/2-1} dv. \end{aligned}$$

Haciendo el cambio de variable $w(v) = v/u$ en la integral se obtiene

$$\begin{aligned} f_{X+Y}(u) &= \frac{1}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{1}{2}\right)^{(n+m)/2} e^{-u/2} u^{(n+m)/2-1} \\ &\quad \int_0^1 (1-w)^{n/2-1} w^{m/2-1} dw. \end{aligned}$$

La integral resultante es $B(n/2, m/2)$. Entonces

$$\begin{aligned} f_{X+Y}(u) &= \frac{B(n/2, m/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{1}{2}\right)^{(n+m)/2} e^{-u/2} u^{(n+m)/2-1} \\ &= \frac{1}{\Gamma((n+m)/2)} \left(\frac{1}{2}\right)^{(n+m)/2} e^{-u/2} u^{(n+m)/2-1}. \end{aligned}$$

Esta última expresión es la función de densidad de la distribución $\chi^2(n+m)$. \square

El resultado anterior puede demostrarse de una manera más simple y elegante usando la función generadora de momentos o la función característica, presentadas en el siguiente capítulo.

Proposición. Sean X_1, \dots, X_n independientes con distribución $N(\mu, \sigma^2)$. Entonces

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n).$$

Demostración. Esto es una consecuencia sencilla de las dos proposiciones anteriores. Como cada X_i tiene distribución $N(\mu, \sigma^2)$, para $i = 1, \dots, n$, entonces $(X_i - \mu)/\sigma$ tiene distribución $N(0, 1)$. Por lo tanto,

$$\frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(1).$$

En consecuencia

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n).$$

□

Ahora se enuncia una proposición cuya demostración se pospone hasta que se cuente con la poderosa herramienta de las funciones generadoras de momentos. Este es el contenido del Ejercicio 538 en la página 217.

Proposición. Sean X y Y independientes tales que X tiene distribución $\chi^2(n)$, y $X + Y$ tiene distribución $\chi^2(m)$. Suponga $m > n$. Entonces Y tiene distribución $\chi^2(m - n)$.

Con ayuda de esta proposición se demuestra ahora el siguiente resultado de particular importancia en estadística.

Proposición. Sean X_1, \dots, X_n independientes con distribución $N(\mu, \sigma^2)$. Entonces

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1).$$

Demostración.

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2$$

$$= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.$$

Dividiendo entre σ^2 ,

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} = \frac{n-1}{\sigma^2} S^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2.$$

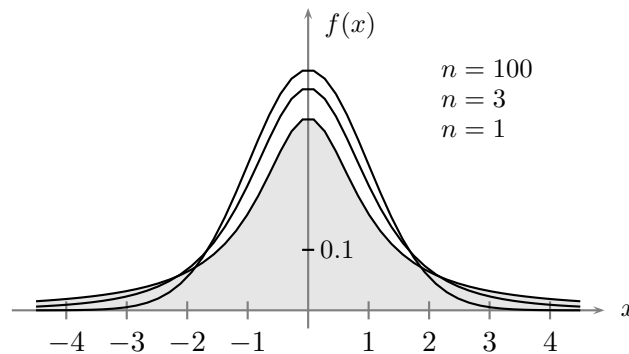
El término del lado izquierdo tiene distribución $\chi^2(n)$ mientras que el segundo sumando del lado derecho tiene distribución $\chi^2(1)$. Por la proposición anterior, y recordando que \bar{X} y S^2 son independientes, se concluye que el primer sumando del lado derecho tiene distribución $\chi^2(n-1)$. \square

Distribución t

La variable aleatoria continua X tiene una distribución t de Student con $n > 0$ grados de libertad si su función de densidad está dada por

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} (1 + x^2/n)^{-(n+1)/2}, \quad \text{para } -\infty < x < \infty,$$

cuya gráfica es:



Función de densidad $t(n)$.

En este caso se escribe $X \sim t(n)$. Esta distribución apareció por primera vez en 1908 en un trabajo publicado por William Gosset bajo el pseudónimo de *Student*. Se puede demostrar que $E(X) = 0$, y $\text{Var}(X) = n/(n-2)$, para $n > 2$. La primera igualdad establece que la distribución $t(n)$ se encuentra siempre centrada en cero para cualquier valor del parámetro n . Se muestran a continuación algunas formas en las que surge esta distribución.

Proposición. Sean $X \sim N(0, 1)$ y $Y \sim \chi^2(n)$ independientes. Entonces

$$\frac{X}{\sqrt{Y/n}} \sim t(n).$$

Demostración. Por independencia, la función de densidad conjunta de X y Y es, para $y > 0$,

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} y^{n/2-1} e^{-y/2}.$$

Se aplica la fórmula (5.1) para la transformación

$$\varphi(x, y) = (x, x/\sqrt{y/n}),$$

con inversa $\varphi^{-1}(s, t) = (s, ns^2/t^2)$. El Jacobiano de la transformación inversa es

$$J(s, t) = \begin{vmatrix} \partial x/\partial s & \partial x/\partial t \\ \partial y/\partial s & \partial y/\partial t \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 2sn/t^2 & -2ns^2/t^3 \end{vmatrix} = -2ns^2/t^3.$$

Por lo tanto

$$\begin{aligned} f_{S,T}(s, t) &= f_X(s) f_Y(ns^2/t^2) \cdot 2ns^2/t^3 \\ &= \frac{1}{\sqrt{2\pi}} e^{-s^2/2} \cdot \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} \frac{n^{n/2-1} s^{n-2}}{t^{n-2}} e^{-ns^2/2t^2} \cdot 2ns^2/t^3. \end{aligned}$$

Integrando respecto a s ,

$$f_T(t) = \frac{1}{\sqrt{2\pi}} \frac{n^{n/2}}{2^{n/2-1} \Gamma(n/2) t^{n+1}} \int_0^\infty s^n e^{-s^2(\frac{1}{2} + \frac{n}{2t^2})} ds.$$

Ahora efectuamos el cambio de variable $r(s) = s^2(\frac{1}{2} + \frac{n}{2t^2})$, de donde obtenemos $dr = 2s(\frac{1}{2} + \frac{n}{2t^2}) ds$, y entonces

$$\begin{aligned} f_T(t) &= \frac{1}{\sqrt{2\pi}} \frac{n^{n/2}}{2^{n/2-1} \Gamma(n/2) t^{n+1} 2 (\frac{1}{2} + \frac{n}{2t^2})^{(n+1)/2}} \int_0^\infty r^{(n-1)/2} e^{-r} dr \\ &= \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} \frac{1}{(1 + t^2/n)^{(n+1)/2}}, \end{aligned}$$

correspondiente a la función de densidad de la distribución $t(n)$. □

El siguiente resultado es de importancia en estadística para efectuar estimaciones del parámetro μ de una población normal cuando la varianza σ^2 es desconocida.

Proposición. Sea X_1, \dots, X_n una m.a. de una distribución $N(\mu, \sigma^2)$. Entonces

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Demostración. Use la proposición anterior aplicada a las variables aleatorias independientes

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1),$$

$$\text{y } \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1).$$

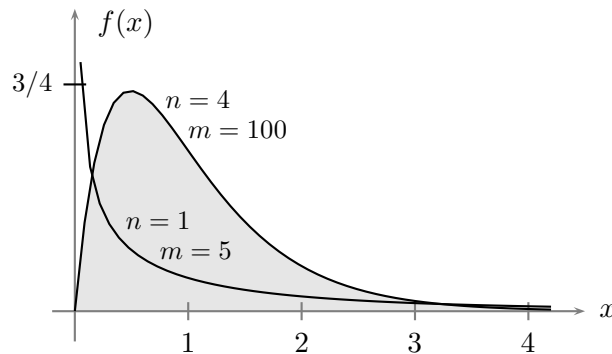
□

Distribución F

La variable aleatoria continua X tiene una distribución F de Snedecor con parámetros $n > 0$ y $m > 0$ si su función de densidad es

$$f(x) = \begin{cases} \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} x^{n/2-1} \left(1 + \frac{n}{m}x\right)^{-(n+m)/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Se escribe $X \sim F(n, m)$. En la siguiente figura se muestra el comportamiento de esta función de densidad.



Función de densidad $F(n, m)$.

Puede demostrarse que

$$E(X) = \frac{m}{m-2} \quad \text{para } m > 2,$$

$$\text{y } \text{Var}(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)} \quad \text{para } m > 4.$$

Los siguientes resultados indican la forma de obtener esta distribución.

Proposición. Sean $X \sim \chi^2(n)$ y $Y \sim \chi^2(m)$ independientes. Entonces

$$\frac{X/n}{Y/m} \sim F(n, m).$$

Este resultado se obtiene directamente de la aplicación de la fórmula para la función de densidad del cociente de dos variables aleatorias. Los detalles se dejan como ejercicio.

Proposición. Si $X \sim t(m)$, entonces $X^2 \sim F(1, m)$.

Para demostrar este resultado aplique la fórmula general para la densidad de X^2 , es decir, para $x > 0$,

$$f_{X^2}(x) = f_X(\sqrt{x})\frac{1}{2\sqrt{x}} + f_X(-\sqrt{x})\frac{1}{2\sqrt{x}}.$$

6.2. Estadísticas de orden

Dada una muestra aleatoria X_1, \dots, X_n , podemos evaluar cada una de estas variables en un punto muestral ω cualquiera y obtener una colección de números reales $X_1(\omega), \dots, X_n(\omega)$. Estos números pueden ser ordenados de menor a mayor incluyendo repeticiones. Si $X_{(i)}(\omega)$ denota el i -ésimo número ordenado, tenemos entonces la colección no decreciente de números reales

$$X_{(1)}(\omega) \leq \dots \leq X_{(n)}(\omega).$$

Ahora hacemos variar el argumento ω y lo que se obtiene son las así llamadas estadísticas de orden. Este proceso de ordenamiento resulta ser de importancia en algunas aplicaciones. Tenemos entonces la siguiente definición.

Definición (Estadísticas de orden). Sea X_1, \dots, X_n una muestra aleatoria. A las variables aleatorias ordenadas

$$\begin{aligned} X_{(1)} &= \text{mín} \{X_1, \dots, X_n\}, \\ X_{(2)} &= \text{mín} \{X_1, \dots, X_n\} \setminus \{X_{(1)}\}, \\ X_{(3)} &= \text{mín} \{X_1, \dots, X_n\} \setminus \{X_{(1)}, X_{(2)}\}, \\ &\vdots \\ X_{(n)} &= \text{máx} \{X_1, \dots, X_n\}, \end{aligned}$$

se les conoce con el nombre de estadísticas de orden. A $X_{(1)}$ se le llama primera estadística de orden, a $X_{(2)}$ se le llama segunda estadística de orden, etc. A $X_{(i)}$ se le llama i -ésima estadística de orden, $i = 1, \dots, n$.

Observe que, aunque los elementos de la muestra aleatoria son variables aleatorias independientes, las estadísticas de orden no lo son, pues deben mantener la relación $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Observe además que la i -ésima estadística de orden $X_{(i)}$ no necesariamente es igual a alguna variable de la muestra aleatoria en particular, sino que, en general, es una función de todas las variables de la muestra aleatoria. Nuestro objetivo en esta sección es encontrar algunas fórmulas relacionadas con las distribuciones de probabilidad de las estadísticas de orden, cuando se conoce la distribución de cada variable de la muestra aleatoria.

Distribuciones individuales

Comenzamos encontrando la distribución de la primera y de la última estadística de orden de manera individual.

Proposición. Sea X_1, \dots, X_n una m.a. de una distribución continua con función de densidad $f(x)$, y función de distribución $F(x)$. Entonces

1. $f_{X_{(1)}}(x) = nf(x)[1 - F(x)]^{n-1}$.
2. $f_{X_{(n)}}(x) = nf(x)[F(x)]^{n-1}$.

Demostración. Para verificar la primera fórmula se calcula primero la función de distribución

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) \\ &= P(\text{mín}\{X_1, \dots, X_n\} \leq x) \\ &= 1 - P(\text{mín}\{X_1, \dots, X_n\} > x) \end{aligned}$$

$$\begin{aligned}
&= 1 - P(X_1 > x, \dots, X_n > x) \\
&= 1 - [P(X_1 > x)]^n \\
&= 1 - [1 - F(x)]^n.
\end{aligned}$$

Entonces $f_{X_{(1)}}(x) = nf(x)[1 - F(x)]^{n-1}$. Para demostrar la segunda fórmula se procede de manera análoga,

$$\begin{aligned}
F_{X_{(n)}}(x) &= P(X_{(n)} \leq x) \\
&= P(\text{máx}\{X_1, \dots, X_n\} \leq x) \\
&= P(X_1 \leq x, \dots, X_n \leq x) \\
&= [P(X_1 \leq x)]^n \\
&= [F(x)]^n.
\end{aligned}$$

Por lo tanto $f_{X_{(n)}}(x) = nf(x)[F(x)]^{n-1}$. □

Ahora se presenta el resultado general de la función de densidad de la i -ésima estadística de orden.

Proposición. Sea X_1, \dots, X_n una m.a. de una distribución continua con función de densidad $f(x)$, y función de distribución $F(x)$. Entonces la función de densidad de la i -ésima estadística de orden es

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}.$$

Demostración. Sea Y_i la variable aleatoria dada por

$$Y_i = 1_{(-\infty, x]}(X_i) = \begin{cases} 1 & \text{si } X_i \leq x, \\ 0 & \text{si } X_i > x, \end{cases}$$

en donde X_i es el i -ésimo elemento de la muestra aleatoria. Las variables Y_1, \dots, Y_n son independientes y cada una de ellas puede considerarse un ensayo Bernoulli con probabilidad de éxito, es decir tomar el valor 1, igual a $P(X_i \leq x) = F(x)$. Entonces la suma $Y_1 + \dots + Y_n$ corresponde al número de v.a.s X_i que cumplen la condición $X_i \leq x$, y por lo tanto esta suma tiene distribución $\text{bin}(n, p)$, con $p = F(x)$. Entonces

$$\begin{aligned}
F_{X_{(i)}}(x) &= P(X_{(i)} \leq x) \\
&= P(Y_1 + \dots + Y_n \geq i) \\
&= \sum_{j=i}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}.
\end{aligned}$$

Derivando y después simplificando,

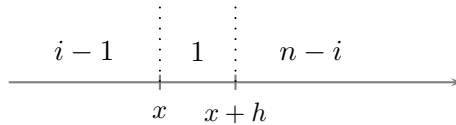
$$f_{X_{(i)}}(x) = \sum_{j=i}^n \binom{n}{j} f(x) [F(x)]^{j-1} [1 - F(x)]^{n-j-1} [j(1 - F(x)) - (n - j)F(x)]$$

$$\begin{aligned}
&= \sum_{j=i}^n \binom{n}{j} f(x)[F(x)]^{j-1}[1-F(x)]^{n-j-1}[j-nF(x)] \\
&= \binom{n}{i} i f(x)[F(x)]^{i-1}[1-F(x)]^{n-i}.
\end{aligned}$$

□

Observe que la fórmula recién demostrada se reduce a las encontradas antes, cuando el índice i toma los valores 1 y n .

A continuación se presenta un argumento corto e intuitivo que nos lleva al mismo resultado. Sea $h > 0$ arbitrario, y considere los siguientes tres intervalos ajenos $(-\infty, x]$, $(x, x+h]$ y $(x+h, \infty)$.



La probabilidad de que $i-1$ variables de la muestra tomen un valor en el intervalo $(-\infty, x]$, una de ellas en $(x, x+h]$, y el resto $n-i$ en $(x+h, \infty)$ es, de acuerdo a la distribución multinomial,

$$\frac{n!}{(i-1)!1!(n-i)!} [F(x)]^{i-1} [F(x+h) - F(x)] [1 - F(x+h)]^{n-i}.$$

Esta probabilidad es, aproximadamente, $f_{X_{(i)}}(x)h$. Dividiendo entre h , y después haciendo h tender a cero se obtiene nuevamente

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x)[F(x)]^{i-1}[1-F(x)]^{n-i}.$$

Definición (Rango). Sea X_1, \dots, X_n una muestra aleatoria. A la variable aleatoria $R = X_{(n)} - X_{(1)}$ se le conoce como el rango de la muestra.

El siguiente resultado provee de una fórmula para la función de densidad de esta variable.

Proposición. Sea X_1, \dots, X_n una m.a. de una distribución continua con función de densidad $f(x)$, y función de distribución $F(x)$. Entonces para $r > 0$,

$$f_R(r) = n(n-1) \int_{-\infty}^{\infty} f(v)f(r+v)[F(r+v) - F(v)]^{n-2} dv.$$

Demostración. Para $x < y$,

$$\begin{aligned} F_{X_{(1)}, X_{(n)}}(x, y) &= P(X_{(1)} \leq x, X_{(n)} \leq y) \\ &= P(X_{(n)} \leq y) - P(X_{(n)} \leq y, X_{(1)} > x) \\ &= [F(y)]^n - P(x < X_1 \leq y, \dots, x < X_n \leq y) \\ &= [F(y)]^n - [F(y) - F(x)]^n. \end{aligned}$$

Por lo tanto, $f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1)f(x)f(y)[F(y) - F(x)]^{n-2}$, para $n \geq 2$. Ahora se usa la fórmula

$$f_{Y-X}(u) = \int_{-\infty}^{\infty} f_{X,Y}(v, u+v) dv$$

equivalente a (5.5) para la diferencia de dos variables aleatorias. Entonces para $r > 0$,

$$f_{X_{(n)}-X_{(1)}}(r) = n(n-1) \int_{-\infty}^{\infty} f(v)f(r+v)[F(r+v) - F(v)]^{n-2} dv.$$

□

Distribuciones conjuntas

Se presentan a continuación dos resultados acerca de la distribución conjunta de las estadísticas de orden. El primer resultado trata acerca de la distribución conjunta de todas ellas, y después se considera la distribución conjunta de cualesquiera dos.

Proposición. Sea X_1, \dots, X_n una m.a. de una distribución continua con función de densidad $f(x)$. Entonces

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = \begin{cases} n!f(x_1) \cdots f(x_n) & \text{si } x_1 < \cdots < x_n, \\ 0 & \text{en otro caso.} \end{cases}$$

Demostración. Se considera la función de distribución conjunta de todas las estadísticas de orden, y después se deriva n veces para encontrar la función de densidad. Para

$$x_1 < x_2 < \dots < x_n,$$

$$F_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = P(X_{(1)} \leq x_1, X_{(2)} \leq x_2, \dots, X_{(n)} \leq x_n).$$

Como $(X_{(2)} \leq x_2) = (x_1 < X_{(2)} \leq x_2) \cup (X_{(2)} \leq x_1)$, se obtiene la expresión

$$F_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = P(X_{(1)} \leq x_1, x_1 < X_{(2)} \leq x_2, \dots, X_{(n)} \leq x_n) + P(X_{(1)} \leq x_1, X_{(2)} \leq x_1, \dots, X_{(n)} \leq x_n).$$

Observe que el segundo sumando no depende de x_2 , así es que al tomar la derivada respecto de esta variable, este término desaparece. De manera análoga procedemos con los eventos $(X_{(3)} \leq x_3)$ hasta $(X_{(n)} \leq x_n)$. Al final se obtiene

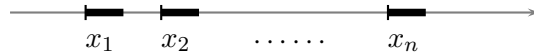
$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} P(X_{(1)} \leq x_1, x_1 < X_{(2)} \leq x_2, \dots, x_{n-1} < X_{(n)} \leq x_n).$$

Como ahora los intervalos involucrados son disjuntos, la distribución multinomial asegura que

$$\begin{aligned} P(X_{(1)} \leq x_1, x_1 < X_{(2)} \leq x_2, \dots, x_{n-1} < X_{(n)} \leq x_n) \\ &= n! P(X_1 \leq x_1, x_1 < X_2 \leq x_2, \dots, x_{n-1} < X_n \leq x_n) \\ &= n! F(x_1)[F(x_2) - F(x_1)] \dots [F(x_n) - F(x_{n-1})], \end{aligned}$$

en donde la última igualdad se sigue de la independencia e idéntica distribución de las variables de la muestra. Ahora solo resta derivar para encontrar el resultado buscado, siendo más sencillo encontrar las derivadas en el orden inverso. \square

La siguiente demostración es una prueba corta pero no formal del mismo resultado. Sea $x_1 < x_2 < \dots < x_n$, y $h > 0$ suficientemente pequeño tal que los intervalos $(x_1, x_1 + h], (x_2, x_2 + h], \dots, (x_n, x_n + h]$ son ajenos.



La probabilidad de que las variables aleatorias tomen valores, cada una de ellas, en uno y sólo uno de estos intervalos es, de acuerdo a la distribución multinomial,

$$\frac{n!}{1! \dots 1!} [F(x_1 + h) - F(x_1)] \dots [F(x_n + h) - F(x_n)].$$

Esta probabilidad es, aproximadamente, $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n)h \dots h$. Dividiendo entre h^n , y después haciendo h tender a cero se obtiene, una vez más,

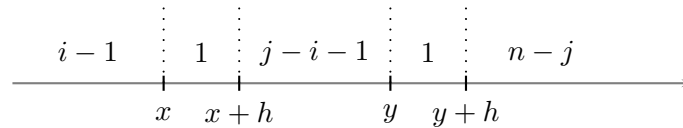
$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = n! f(x_1) \dots f(x_n).$$

Ahora nos interesa encontrar una fórmula para la densidad conjunta de cualesquiera dos estadísticas de orden.

Proposición. Sea X_1, \dots, X_n una m.a. de una distribución continua con función de distribución $F(x)$, y función de densidad $f(x)$. Sea $i < j$. Para $x < y$,

$$f_{X_{(i)}, X_{(j)}}(x, y) = \binom{n}{i, j-i, n-j} i(j-i) f(x) f(y) [F(x)]^{i-1} [F(y) - F(x)]^{j-i-1} [1 - F(y)]^{n-j}.$$

Para este resultado se presenta únicamente un argumento intuitivo. Sean $x < y$ y considere los intervalos ajenos $(-\infty, x]$, $(x, x+h]$, $(x+h, y]$, $(y, y+h]$, y $(y+h, \infty)$ para $h > 0$ suficientemente pequeña.



La probabilidad de que $i-1$ variables de la muestra tomen un valor en $(-\infty, x]$, una de ellas en $(x, x+h]$, $j-i+1$ variables en $(x+h, y]$, otra en $(y, y+h]$, y el resto, $n-j$ variables, tomen un valor en $(y+h, \infty)$ es, de acuerdo a la distribución multinomial,

$$\frac{n!}{(i-1)! 1! (j-i-1)! 1! (n-j)!} [F(x)]^{i-1} \cdot [F(x+h) - F(x)] [F(y) - F(x+h)]^{j-i-1} \cdot [F(y+h) - F(y)] \cdot [1 - F(y+h)]^{n-j}.$$

Esta probabilidad es aproximadamente igual a $f_{X_{(i)}, X_{(j)}}(x, y) \cdot h \cdot h$. Dividiendo entre h^2 , y después haciendo h tender a cero se obtiene la fórmula anunciada.

6.3. Ejercicios

432. Sea X_1, \dots, X_n una muestra aleatoria de una distribución con media μ y varianza σ^2 . Demuestre que $E(\bar{X}) = \mu$ y $E(S^2) = \sigma^2$. Estos resultados son de importancia en estadística y muestran que \bar{X} y S^2 son *estimadores insesgados* para la media y varianza de la distribución.
433. Sea X_1, \dots, X_n una m.a. de una distribución con media μ y varianza σ^2 . Demuestre que $\text{Var}(\bar{X}) = \sigma^2/n$. ¿Cuánto vale $\text{Var}(S^2)$?

434. Sea X_1, \dots, X_n una m.a. de una distribución $\text{Ber}(p)$. Demuestre que las estadísticas \bar{X} y S^2 no son independientes.

Distribución χ^2

435. Demuestre que la función de densidad de la distribución $\chi^2(n)$ efectivamente lo es.
436. Demuestre que la distribución $\chi^2(n)$, con $n = 2$, se reduce a la distribución $\exp(\lambda)$ con $\lambda = 1/2$.
437. Demuestre que la distribución $\text{gama}(n/2, \lambda)$, con $\lambda = 1/2$, se reduce a la distribución $\chi^2(n)$.
438. Sea X con distribución $\chi^2(n)$. Demuestre que

a) $E(X) = n$.

b) $E(X^m) = 2^m \frac{\Gamma(m + n/2)}{\Gamma(n/2)}$, para $m = 1, 2, \dots$

c) $\text{Var}(X) = 2n$.

439. Sean X_1, \dots, X_n independientes con distribución $N(\mu, \sigma^2)$. Demuestre que

$$\frac{(\bar{X} - \mu)^2}{\sigma^2/n} \sim \chi^2(1).$$

440. Sean X_1, \dots, X_n independientes con distribución $N(0, 1)$. Demuestre que

$$\sum_{i=1}^n X_i^2 \sim \chi^2(n).$$

441. Sean X_1, \dots, X_n independientes tales que cada X_i tiene distribución $N(\mu_i, \sigma_i^2)$ para $i = 1, \dots, n$. Demuestre que

$$\sum_{i=1}^n \frac{(X_i - \mu_i)^2}{\sigma_i^2} \sim \chi^2(n).$$

442. Sean X y Y independientes ambas con distribución normal estándar. Sean $R = \sqrt{X^2 + Y^2}$ y $\theta = \tan^{-1}(Y/X)$. Demuestre que

a) R^2 tiene distribución $\chi^2(n)$ con $n = 2$ grados de libertad.

b) $\tan \theta$ tiene distribución Cauchy.

c) R y θ son independientes.

Distribución t

443. Demuestre que la función de densidad de una distribución $t(n)$ efectivamente lo es.
444. Sea X con distribución $t(n)$. Demuestre que
- a) $E(X) = 0$.
 - b) $\text{Var}(X) = \frac{n}{n-2}$, para $n > 2$.
445. Demuestre que la distribución $t(n+1)$ tiene momentos finitos de orden menor o igual a n , pero ningún otro momento de orden superior.

Distribución F

446. Demuestre que la función de densidad de la distribución $F(n, m)$ efectivamente lo es.
447. Sea X con distribución $F(n, m)$. Demuestre que
- a) $E(X) = \frac{m}{m-2}$, para $m > 2$.
 - b) $\text{Var}(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}$, para $m > 4$.
448. Sea X con distribución $F(n, m)$. Demuestre que $Y = 1/X$ tiene distribución $F(m, n)$. Observe el cambio en el orden de los parámetros. Este resultado es útil para obtener valores de F que no aparecen en tablas de esta distribución que son comunes en textos de estadística.
449. Sea X con distribución $F(n, m)$. Demuestre que cuando m tiende a infinito la función de densidad de nX converge puntualmente a la función de densidad de la distribución $\chi^2(n)$.
450. Sean $X \sim \chi^2(n)$ y $Y \sim \chi^2(m)$ independientes. Demuestre que

$$\frac{X/n}{Y/m} \sim F(n, m).$$

451. Demuestre que si $X \sim t(n)$, entonces $X^2 \sim F(1, n)$.

Estadísticas de orden: distribuciones individuales

452. Sea X_1, \dots, X_n una m.a. de una distribución continua $F(x)$ con función de densidad $f(x)$. Las funciones de densidad de la primera y última estadísticas de orden son
- a) $f_{X_{(1)}}(x) = nf(x)[1 - F(x)]^{n-1}$.
 - b) $f_{X_{(n)}}(x) = nf(x)[F(x)]^{n-1}$.

Demuestre que estas funciones son efectivamente funciones de densidad.

453. Sea X_1, \dots, X_n una m.a. de una distribución continua $F(x)$ con función de densidad $f(x)$. La función de densidad de la i -ésima estadística de orden es

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}.$$

Compruebe que ésta es efectivamente una función de densidad, y que esta expresión se reduce a la fórmulas encontradas antes, cuando $i = 1$ e $i = n$, respectivamente.

454. Sea X_1, \dots, X_n una m.a. de una distribución $\text{unif}(0, 1)$. Demuestre que la i -ésima estadística de orden tiene distribución $\text{beta}(i, n + 1 - i)$. Encuentre por lo tanto su esperanza y varianza.
455. Sea X_1, \dots, X_n una m.a. de una distribución $\exp(\lambda)$. Encuentre la función de densidad de la i -ésima estadística de orden.
456. Sean $X_{(1)}, X_{(2)}$ las estadísticas de orden de una m.a. de tamaño dos de una distribución $N(\mu, \sigma^2)$. Demuestre que $E[X_{(1)}] = \mu - \sigma/\sqrt{\pi}$.
457. Sean $X_{(1)}, X_{(2)}$ las estadísticas de orden de una m.a. de tamaño dos de una distribución $N(\mu, \sigma^2)$. Calcule $E[X_{(2)}]$.
458. Sea X_1, \dots, X_n una m.a. de una distribución $F(x)$. Sea x un número real cualquiera, y para cada $i = 1, \dots, n$ defina

$$Y_i = 1_{(-\infty, x]}(X_i).$$

Demuestre que las variables Y_1, \dots, Y_n son independientes, y cada una de ellas tiene distribución $\text{Ber}(n, p)$, con $p = F(x)$. Este hecho fue utilizado en el procedimiento para encontrar la función de densidad de la i -ésima estadística de orden.

459. Sean X_1 y X_2 absolutamente continuas e independientes, y defina $Y = \text{máx}\{X_1, X_2\}$. Demuestre que
- $F_Y(y) = F_{X_1}(y)F_{X_2}(y)$.
 - $f_Y(y) = F_{X_1}(y)f_{X_2}(y) + f_{X_1}(y)F_{X_2}(y)$.
 - $f_Y(y) = 2F(y)f(y)$, cuando X_1 y X_2 tienen la misma distribución.
460. Use el ejercicio anterior para encontrar la función de densidad de $Y = \text{máx}\{X_1, X_2\}$ cuando X_1 y X_2 son independientes cada una con distribución
- $\text{unif}(0, 1)$.
 - $\exp(\lambda)$.
461. Sean X_1 y X_2 absolutamente continuas e independientes. Defina $Y = \text{mín}\{X_1, X_2\}$. Demuestre que
- $F_Y(y) = 1 - [1 - F_{X_1}(y)][1 - F_{X_2}(y)]$.

b) $f_Y(y) = [1 - F_{X_1}(y)]f_{X_2}(y) + f_{X_1}(y)[1 - F_{X_2}(y)]$.

c) $f_Y(y) = 2[1 - F(y)]f(y)$, cuando X_1 y X_2 tienen la misma distribución.

462. Use el ejercicio anterior para encontrar la función de densidad del mínimo de dos variables aleatorias independientes cada una con distribución uniforme en el intervalo $(0, 1)$.

463. Demuestre que el mínimo de n variables aleatorias independientes con distribución exponencial es nuevamente exponencial con parámetro la suma de los parámetros.

464. Sea X_1, \dots, X_n una m.a. de una distribución continua $F(x)$ con función de densidad $f(x)$. Sea $R = X_{(n)} - X_{(1)}$ el rango de la muestra. Demuestre que para $r > 0$ y $n \geq 2$,

$$f_R(r) = n(n-1) \int_{-\infty}^{\infty} f(y)f(y-r)[F(y) - F(y-r)]^{n-2} dy.$$

465. Se escogen n puntos al azar con distribución uniforme en el intervalo unitario $(0, 1)$. Demuestre que la función de densidad de la distancia máxima entre cualesquiera dos puntos es

$$f(r) = \begin{cases} n(n-1)r^{n-2}(1-r) & \text{si } 0 < r < 1, \\ 0 & \text{otro caso.} \end{cases}$$

Estadísticas de orden: distribuciones conjuntas

466. Sea X_1, \dots, X_n una m.a. de una distribución continua con función de densidad $f(x)$. Se ha demostrado que para $x_1 < x_2 < \dots < x_n$,

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = n!f(x_1) \cdots f(x_n).$$

Compruebe que ésta es efectivamente una función de densidad.

467. A partir de la fórmula para $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n)$, calcule la función de densidad marginal de $X_{(1)}$, encontrando nuevamente que

$$f_{X_{(1)}}(x) = nf(x)[1 - F(x)]^{n-1}.$$

468. A partir de la fórmula para $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n)$, calcule la función de densidad marginal de $X_{(n)}$, encontrando nuevamente que

$$f_{X_{(n)}}(x) = nf(x)[F(x)]^{n-1}.$$

469. A partir de la fórmula para $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n)$, calcule la función de densidad marginal de $X_{(i)}$, para $i = 1, \dots, n$, encontrando nuevamente que

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x)[F(x)]^{i-1}[1 - F(x)]^{n-i}.$$

470. Sea X_1, \dots, X_n una m.a. de una distribución continua con función de distribución $F(x)$, y función de densidad $f(x)$. Sea $i < j$. Se ha demostrado que para $x < y$,

$$f_{X_{(i)}, X_{(j)}}(x, y) = \binom{n}{i, j-i, n-j} i(j-i) f(x) f(y) [F(x)]^{i-1} [F(y) - F(x)]^{j-i-1} [1 - F(y)]^{n-j}.$$

Compruebe que ésta es efectivamente una función de densidad bivariada.

471. A partir de la fórmula para $f_{X_{(i)}, X_{(j)}}(x, y)$, calcule la función de densidad marginal de $X_{(i)}$, encontrando nuevamente que

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}.$$

472. Sea X_1, \dots, X_n una m.a. de una distribución $\text{unif}(0, 1)$. Encuentre la función de densidad de

- a) $X_{(1)}$ y $X_{(2)}$ conjuntamente.
 b) $R = X_{(n)} - X_{(1)}$.

473. *Mediana muestral.* La mediana de una muestra aleatoria X_1, \dots, X_n , denotada por $\text{Med}(X_1, \dots, X_n)$, se define del siguiente modo: Considere las estadísticas de orden $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Entonces

$$\text{Med}(X_1, \dots, X_n) = \begin{cases} X_{(\frac{n+1}{2})} & \text{si } n \text{ es impar,} \\ \frac{1}{2} [X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}] & \text{si } n \text{ es par.} \end{cases}$$

Encuentre la función de densidad de la mediana de una muestra aleatoria de la distribución $\text{unif}(0, 1)$, primero suponiendo que el tamaño de la muestra n es impar, y después para n par.

474. Sea X_1, \dots, X_n una m.a. de una distribución $\text{unif}(0, 1)$. Encuentre la función de densidad del vector $(X_{(1)}, \dots, X_{(n)})$.

475. Sea X_1, \dots, X_n una m.a. de una distribución $\text{unif}(0, 1)$. Calcule el coeficiente de correlación entre $X_{(i)}$ y $X_{(j)}$.

476. Sea X_1, \dots, X_n una m.a. de una distribución continua $F(x)$ con función de densidad $f(x)$. Demuestre directamente que para $x < y$,

$$f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1) f(x) f(y) [F(y) - F(x)]^{n-2}.$$

477. Encuentre la función de densidad conjunta de $X_{(1)}$ y $X_{(n)}$ para una m.a. de tamaño n de una distribución

- a) $\text{unif}(0, 1)$.
 b) $\exp(\lambda)$.

478. Calcule la covarianza entre $X_{(1)}$ y $X_{(n)}$ para una m.a. de tamaño n de una distribución

a) $\text{unif}(0, 1)$.

b) $\text{exp}(\lambda)$.

Capítulo 7

Convergencia

En este capítulo se presenta una introducción al tema de convergencia de variables aleatorias. Se estudian distintas formas en que una sucesión de variables aleatorias puede converger.

7.1. Convergencia puntual

Sea X_1, X_2, \dots una sucesión infinita de variables aleatorias. Al evaluar cada una de estas variables en un elemento ω de Ω se obtiene la sucesión numérica $X_1(\omega), X_2(\omega), \dots$. Suponga que esta sucesión converge a un cierto número real denotado por $X(\omega)$. Si lo anterior se cumple para todos y cada uno de los elementos de Ω , entonces se dice que la sucesión de variables aleatorias *converge puntualmente*, y su límite es la función $X : \Omega \rightarrow \mathbb{R}$ definida naturalmente por

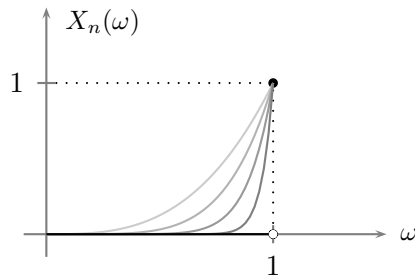
$$X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega).$$

Se ha demostrado antes que en esta situación la función límite X es efectivamente una variable aleatoria. Formalmente se tiene entonces la siguiente definición.

Definición (Convergencia puntual). La sucesión X_1, X_2, \dots converge puntualmente a X si para cada ω en Ω ,

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Ejemplo Considere el espacio medible $([0, 1], \mathcal{B}[0, 1])$, y defina la sucesión de variables aleatorias $X_n(\omega) = \omega^n$. Como en este caso el espacio muestral es un subconjunto de números reales, podemos graficar las variables aleatorias de la siguiente forma:



Gráfica de la variable aleatoria X_n .

Entonces para $\omega \in [0, 1)$, $X_n(\omega) \rightarrow 0$. Mientras que para $\omega = 1$, $X_n(\omega) = 1$. De esta manera la sucesión converge puntualmente a la variable aleatoria

$$X(\omega) = \begin{cases} 0 & \text{si } \omega \in [0, 1), \\ 1 & \text{si } \omega = 1. \end{cases}$$

◦

En algunas situaciones la convergencia puntual resulta ser muy fuerte pues se pide la convergencia de la sucesión evaluada en todos y cada uno de los elementos de Ω . Se puede ser menos estricto y pedir, por ejemplo, que la convergencia se efectúe en todo el espacio Ω excepto en un subconjunto de probabilidad cero. Este tipo de convergencia menos restrictiva se llama *convergencia casi segura*, y se estudia en las siguientes secciones junto con otros tipos de convergencia.

7.2. Convergencia casi segura

Definición (Convergencia casi segura). La sucesión X_1, X_2, \dots converge casi seguramente a X , si

$$P\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = 1.$$

Por lo tanto, en la convergencia casi segura se permite que para algunos valores de ω , la sucesión numérica $X_1(\omega), X_2(\omega), \dots$ pueda no converger, sin embargo el subconjunto de Ω en donde esto suceda debe tener probabilidad cero. Para indicar la convergencia casi segura se escribe $X_n \xrightarrow{c.s.} X$, o bien $\lim_{n \rightarrow \infty} X_n = X$ c.s. A menudo se utiliza el término *convergencia casi dondequiera*, o bien *convergencia casi siempre*

para denotar este tipo de convergencia. Observe que omitiendo el argumento ω , la condición para la convergencia casi segura se escribe en la forma más corta:

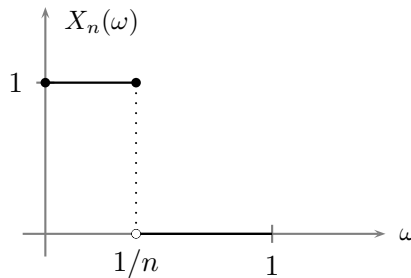
$$P(\lim_{n \rightarrow \infty} X_n = X) = 1,$$

o simplemente $P(X_n \rightarrow X) = 1$. Observe que el conjunto $(X_n \rightarrow X)$ debe ser medible para que tenga sentido aplicar la probabilidad. Puede demostrarse que bajo este tipo de convergencia, el límite es único casi seguramente.

Ejemplo. Considere el espacio de probabilidad $([0, 1], \mathcal{B}[0, 1], P)$ con P la medida uniforme, es decir, $P(a, b) = b - a$. Defina la sucesión de variables aleatorias

$$X_n(\omega) = \begin{cases} 1 & \text{si } 0 \leq \omega \leq 1/n, \\ 0 & \text{otro caso.} \end{cases}$$

Cuyas gráficas son:



Gráfica de la variable aleatoria X_n .

Observe que X_n tiene distribución Bernoulli con parámetro $p = 1/n$. La sucesión X_n converge casi seguramente a la variable aleatoria constante cero. Para demostrar esto se necesita verificar que $P(X_n \rightarrow 0) = 1$. Pero esta igualdad es evidente a partir del hecho de que

$$\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = 0\} = (0, 1],$$

cuya probabilidad es uno. El punto $\omega = 0$ es el único punto muestral para el cual $X_n(\omega)$ no converge a cero. Esto demuestra que $X_n \xrightarrow{c.s.} 0$. ◦

7.3. Convergencia en probabilidad

Definición (Convergencia en probabilidad). La sucesión X_1, X_2, \dots converge en probabilidad a X , si para cada $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\} = 0.$$

Para denotar la convergencia en probabilidad se escribe $X_n \xrightarrow{p} X$, y omitiendo el argumento ω la condición se escribe

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0.$$

Más adelante se demostrará que la convergencia en probabilidad es un tipo de convergencia aún menos restrictiva que la convergencia casi segura. Veamos ahora otras maneras en que una sucesión de variables aleatorias puede converger.

7.4. Convergencia en media

Definición (Convergencia en media). La sucesión X_1, X_2, \dots converge en media a X , si

$$\lim_{n \rightarrow \infty} E|X_n - X| = 0.$$

Observe que para este tipo de convergencia tanto los elementos de la sucesión como el límite mismo deben ser variables aleatorias con esperanza finita. A este tipo de convergencia también se le llama *convergencia en L^1* y se le denota por $X_n \xrightarrow{m} X$, o $X_n \xrightarrow{L^1} X$.

7.5. Convergencia en media cuadrática

Definición (Convergencia en media cuadrática). La sucesión X_1, X_2, \dots converge en media cuadrática a X , si

$$\lim_{n \rightarrow \infty} E|X_n - X|^2 = 0.$$

En la convergencia en media cuadrática se presupone que tanto los elementos de la sucesión como el límite mismo son variables aleatorias con segundo momento finito. A este tipo de convergencia también se le llama *convergencia en L^2* , y se le denota por $X_n \xrightarrow{m.c.} X$, o $X_n \xrightarrow{L^2} X$.

7.6. Convergencia en distribución

Definición (Convergencia en distribución). La sucesión X_1, X_2, \dots converge en distribución a X , si para todo punto x en donde $F_X(x)$ es continua, se cumple que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

En este caso se escribe $X_n \xrightarrow{d} X$. A este tipo de convergencia se le conoce también con el nombre de *convergencia débil*, y ello se debe a que esta forma de convergencia es la menos restrictiva de todas las mencionadas anteriormente.

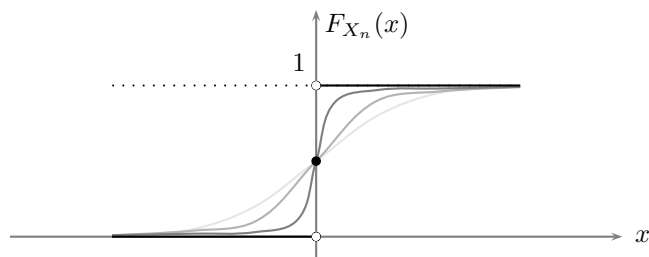
Ejemplo. Considere la sucesión X_1, X_2, \dots , en donde cada X_n tiene distribución $N(0, \sigma^2/n)$. Demostraremos que $X_n \xrightarrow{d} 0$. Como

$$F_{X_n}(x) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \int_{-\infty}^x e^{-u^2/2(\sigma^2/n)} du,$$

entonces

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1/2 & \text{si } x = 0, \\ 1 & \text{si } x > 0. \end{cases}$$

Gráficamente,



Sucesión y límite de las funciones de distribución $F_{X_n}(x)$.

Observe que la variable aleatoria constante $X = 0$ tiene función de distribución

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 & \text{si } x \geq 0. \end{cases}$$

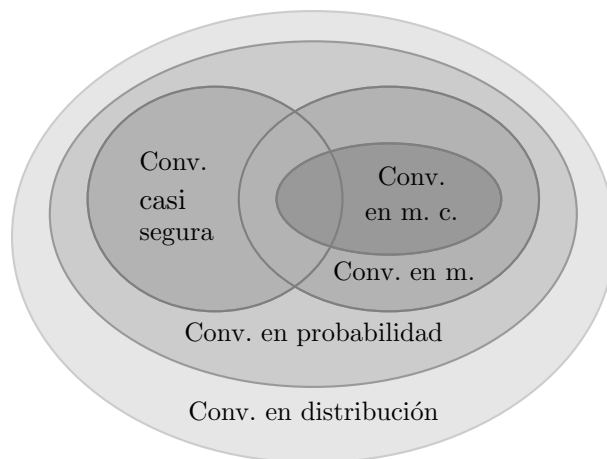
Tenemos entonces que $X_n \xrightarrow{d} 0$ pues $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ para todo punto x donde $F_X(x)$ es continua, esto es, para todo x en el conjunto $\mathbb{R} \setminus \{0\}$. Observe que las funciones $F_{X_n}(x)$ no convergen a $F(x)$ cuando $x = 0$. \circ

A manera de resumen se presenta en la siguiente tabla las definiciones de los distintos tipos de convergencia mencionados. En la siguiente sección se estudian las relaciones entre estos tipos de convergencia.

| Convergencia | Definición |
|---------------------|---|
| puntual | $X_n(\omega) \rightarrow X(\omega)$ para cada ω en Ω . |
| casi segura | $P(X_n \rightarrow X) = 1$. |
| en media | $E X_n - X \rightarrow 0$. |
| en media cuadrática | $E X_n - X ^2 \rightarrow 0$. |
| en probabilidad | $P(X_n - X > \epsilon) \rightarrow 0$. |
| en distribución | $F_{X_n}(x) \rightarrow F_X(x)$ en puntos de continuidad x de F_X . |

7.7. Relaciones generales entre los tipos de convergencia

En esta sección se establecen algunas relaciones generales entre los tipos de convergencia de variables aleatorias mencionados en la sección anterior. En la siguiente figura se ilustran de manera gráfica estas relaciones:



Relación entre los tipos de convergencia.

En este diagrama la contención se interpreta como implicación, por ejemplo, la convergencia casi segura implica la convergencia en probabilidad, y ésta a su vez implica la convergencia en distribución. Éstos y otros resultados se demuestran a continuación.

Proposición. Convergencia c.s. \Rightarrow convergencia en prob.

Demostración. Suponga $X_n \xrightarrow{c.s.} X$. Sea $\epsilon > 0$ y defina los eventos

$$A_n = \bigcup_{k=n}^{\infty} (|X_k - X| > \epsilon).$$

Esta sucesión es decreciente y su límite es entonces la intersección de todos los eventos. Como $(|X_n - X| > \epsilon) \subseteq A_n$, entonces $P(|X_n - X| > \epsilon) \leq P(A_n)$. Por lo tanto,

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) &\leq \lim_{n \rightarrow \infty} P(A_n) \\ &= P\left(\lim_{n \rightarrow \infty} A_n\right) \\ &= P\left(\bigcap_{n=1}^{\infty} A_n\right) \\ &= P(|X_n - X| > \epsilon, \text{ para cada } n \geq 1) \\ &= P\left(\lim_{n \rightarrow \infty} X_n \neq X\right) \\ &= 0. \end{aligned}$$

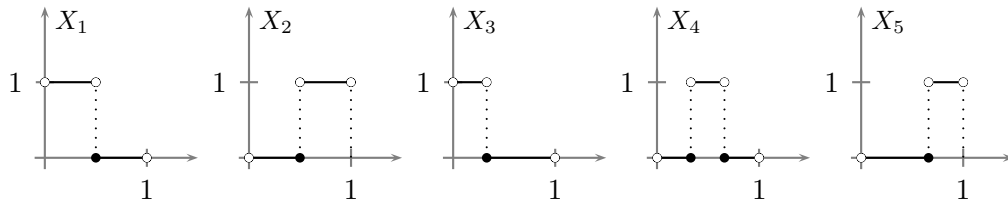
□

El recíproco de la proposición anterior es, en general, falso, es decir, la convergencia en probabilidad no implica necesariamente la convergencia casi siempre. Para ilustrar esta afirmación se proporciona a continuación un ejemplo.

Ejemplo (En general, convergencia en prob. $\not\Rightarrow$ convergencia c.s.). Considere el espacio de probabilidad $((0, 1), \mathcal{B}(0, 1), P)$, con P la medida uniforme. Defina los eventos

$$\begin{aligned} A_1 &= (0, 1/2), & A_2 &= (1/2, 1), \\ A_3 &= (0, 1/3), & A_4 &= (1/3, 2/3), & A_5 &= (2/3, 1), \\ A_6 &= (0, 1/4), & A_7 &= (1/4, 2/4), & A_8 &= (2/4, 3/4), & A_9 &= (3/4, 1), \\ & \dots \end{aligned}$$

Sea $X_n = 1_{A_n}$. Las gráficas de estas primeras variables aleatorias son las siguientes:



Gráficas de las primeras variables aleatorias X_n .

Entonces $X_n \xrightarrow{p} 0$ pues para cualquier $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - 0| > \epsilon) = \lim_{n \rightarrow \infty} P(A_n) = 0.$$

Sin embargo la sucesión no converge casi seguramente pues

$$\{w \in \Omega : \lim_{n \rightarrow \infty} X_n(w) \text{ existe} \} = \emptyset.$$

o

Ejemplo (En general, convergencia en media $\not\Rightarrow$ convergencia c.s.). Considere la sucesión de variables X_n del ejemplo anterior. Entonces $X_n \xrightarrow{m} 0$ pues $E|X_n - 0| = P(A_n) \rightarrow 0$. Sin embargo esta sucesión no converge c.s. pues $P(\lim X_n = 0) = P(\emptyset) = 0$.

o

El ejemplo anterior sirve también para mostrar que, en general, la convergencia en media cuadrática no implica la convergencia casi segura. Veamos otra no implicación.

Ejemplo (En general, convergencia c.s. $\not\Rightarrow$ convergencia en media). Considere el espacio $((0, 1), \mathcal{B}(0, 1), P)$, con P la medida de probabilidad uniforme. Defina la sucesión $X_n = n \cdot 1_{(0, 1/n)}$. Entonces X_n converge a cero casi seguramente pues $P(\lim X_n = 0) = P(\Omega) = 1$. Sin embargo no hay convergencia en media pues

$$E|X_n - 0| = E(X_n) = 1 \not\rightarrow 0.$$

◦

Estos ejemplos pueden ser usados también para demostrar que la convergencia casi segura no implica necesariamente la convergencia en media cuadrática.

Proposición. Convergencia en m.c. \Rightarrow convergencia en media.

Demostración. La desigualdad de Jensen establece que para u convexa,

$$u(E(X)) \leq E(u(X)).$$

Tomando $u(x) = x^2$ se obtiene $E^2|X_n - X| \leq E|X_n - X|^2$, de donde se sigue el resultado. Alternativamente la última desigualdad es consecuencia de la desigualdad de Cauchy-Schwarz. \square

Ejemplo (En general, convergencia en media $\not\Rightarrow$ convergencia en m.c.). Sea $X_n = n \cdot 1_{(0,1/n^2)}$ sobre el espacio $((0,1), \mathcal{B}(0,1), P)$, con P la medida uniforme. Entonces X_n converge a cero en media pues

$$E|X_n - 0| = E(X_n) = n \cdot 1/n^2 \rightarrow 0.$$

Sin embargo, no hay convergencia en media cuadrática pues

$$E|X_n - 0|^2 = E(X_n^2) = n^2 \cdot 1/n^2 = 1 \not\rightarrow 0.$$

◦

Proposición. Convergencia en media \Rightarrow convergencia en prob.

Demostración. Para cada $\epsilon > 0$ defina el evento $A_n = (|X_n - X| > \epsilon)$. Entonces

$$\begin{aligned} E|X_n - X| &= E(|X_n - X| \cdot 1_{A_n}) + E(|X_n - X| \cdot 1_{A_n^c}) \\ &\geq E(|X_n - X| \cdot 1_{A_n}) \\ &\geq \epsilon P(|X_n - X| > \epsilon). \end{aligned}$$

Por hipótesis, el lado izquierdo tiende a cero cuando n tiende a infinito. Por lo tanto $P(|X_n - X| > \epsilon) \rightarrow 0$. \square

El recíproco del resultado anterior es, en general, falso.

Ejemplo (En general, convergencia en prob. $\not\Rightarrow$ convergencia en media). Considere nuevamente el espacio $((0,1), \mathcal{B}(0,1), P)$, con P la medida uniforme, y

defina las variables $X_n = n \cdot 1_{(0,1/n)}$. Entonces X_n converge en probabilidad a cero pues para cualquier $\epsilon > 0$, $P(|X_n - 0| > \epsilon) = P(X_n > \epsilon) = 1/n \rightarrow 0$. Sin embargo, la sucesión no converge en media pues $E|X_n - 0| = E(X_n) = 1 \not\rightarrow 0$. \circ

Proposición. Convergencia en prob. \Rightarrow convergencia en dist.

Demostración. Suponga que $X_n \xrightarrow{p} X$, y sea x un punto de continuidad de $F_X(x)$. Para cualquier $\epsilon > 0$,

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(X_n \leq x, |X_n - X| \leq \epsilon) + P(X_n \leq x, |X_n - X| > \epsilon) \\ &\leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon). \end{aligned}$$

Por hipótesis el segundo sumando del lado derecho tiende a cero cuando n tiende a infinito. Entonces para cualquier $\epsilon > 0$,

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \epsilon).$$

Por la continuidad lateral,

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x).$$

Ahora se demuestra la desigualdad inversa. Para cualquier $\epsilon > 0$

$$\begin{aligned} F_X(x - \epsilon) &= P(X \leq x - \epsilon) \\ &= P(X \leq x - \epsilon, |X_n - X| \leq \epsilon) + P(X \leq x - \epsilon, |X_n - X| > \epsilon) \\ &\leq P(X_n \leq x) + P(|X_n - X| > \epsilon). \end{aligned}$$

Nuevamente el segundo sumando tiende a cero cuando n tiende a infinito. Entonces

$$F_X(x - \epsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x).$$

Por la continuidad en x ,

$$F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x).$$

En resumen,

$$F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x).$$

□

El recíproco de la proposición anterior es falso, es decir, la convergencia en distribución no siempre implica la convergencia en probabilidad.

Ejemplo (En general, convergencia en dist. $\not\Rightarrow$ convergencia en prob.).
 Sea X con distribución normal estándar, y sea

$$X_n = \begin{cases} X & \text{si } n \text{ es par,} \\ -X & \text{si } n \text{ es impar.} \end{cases}$$

Entonces claramente cada X_n también tiene distribución normal estándar y por lo tanto para cualquier número real x , $F_{X_n}(x) \rightarrow F_X(x)$, es decir, $X_n \xrightarrow{d} X$. Sin embargo la sucesión no converge en probabilidad a X , pues para valores impares de n y para valores pequeños de $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) = P(2|X| > \epsilon) > 1/2.$$

Lo anterior demuestra que $\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) \neq 0$. ◦

7.8. Dos resultados importantes de convergencia

Sea X_1, X_2, \dots una sucesión de variables aleatorias con esperanza finita. Suponga que X_n converge puntualmente a X . Es natural preguntarse si la sucesión de números $E(X_n)$ converge a $E(X)$. Tal convergencia numérica equivaldría a poder intercambiar las operaciones de límite y esperanza, es decir,

$$\lim_{n \rightarrow \infty} E(X_n) = E(\lim_{n \rightarrow \infty} X_n) = E(X).$$

En esta sección se estudian dos resultados que establecen condiciones bajo las cuales es válido este intercambio.

Teorema de convergencia monótona. Sea $0 \leq X_1 \leq X_2 \leq \dots$ una sucesión de variables aleatorias convergente puntualmente a una variable X . Entonces

$$\lim_{n \rightarrow \infty} E(X_n) = E(X).$$

Demostración. Como $0 \leq X_n \leq X$, entonces $0 \leq E(X_n) \leq E(X)$. Por lo tanto $\lim_{n \rightarrow \infty} E(X_n) \leq E(X)$. Ahora resta demostrar la desigualdad contraria. Primero se aproxima a X de la siguiente forma: Sea $\epsilon > 0$ arbitrario, y para cada entero $k \geq 0$ defina el evento

$$A_k = (k\epsilon \leq X < (k+1)\epsilon).$$

Esta es una colección disjunta de eventos, cuya unión es Ω . Defina ahora la variable aleatoria discreta aproximante

$$Y(\omega) = k\epsilon \quad \text{si} \quad k\epsilon \leq X(\omega) < (k+1)\epsilon.$$

Observe que Y aproxima a X de la siguiente forma: $Y \leq X < Y + \epsilon$. O bien $X - \epsilon < Y \leq X$. Por lo tanto,

$$E(X) - \epsilon \leq E(Y) \leq E(X).$$

Para cada número natural n defina el evento $B_n = (X_n \geq Y)$. No es difícil comprobar que $B_n \nearrow \Omega$. Por lo tanto, para k fijo, $A_k \cap B_n \nearrow A_k$ cuando $n \rightarrow \infty$, y entonces $P(A_k \cap B_n) \nearrow P(A_k)$. Ahora considere la variable aleatoria discreta $Y \cdot 1_{B_n}$ dada por

$$Y \cdot 1_{B_n}(\omega) = \begin{cases} Y(\omega) & \text{si } \omega \in B_n, \\ 0 & \text{si } \omega \notin B_n. \end{cases}$$

Entonces $0 \leq Y \cdot 1_{B_n} \leq X_n$, y por lo tanto $0 \leq E(Y \cdot 1_{B_n}) \leq E(X_n)$. Entonces

$$\begin{aligned} \lim_{n \rightarrow \infty} E(X_n) &\geq \lim_{n \rightarrow \infty} E(Y \cdot 1_{B_n}) \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} E(Y \cdot 1_{B_n \cap A_k}) \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} k\epsilon \cdot P(B_n \cap A_k) \\ &\geq \lim_{n \rightarrow \infty} \sum_{k=0}^m k\epsilon \cdot P(B_n \cap A_k) \\ &= \sum_{k=0}^m k\epsilon \cdot P(A_k) \end{aligned}$$

Como esta desigualdad es válida para cualquier $m \geq 0$, entonces

$$\lim_{n \rightarrow \infty} E(X_n) \geq \sum_{k=0}^{\infty} k\epsilon \cdot P(A_k) = E(Y) \geq E(X) - \epsilon.$$

Dado que $\epsilon > 0$ es arbitrario, se tiene que

$$\lim_{n \rightarrow \infty} E(X_n) \geq E(X).$$

□

El siguiente resultado establece otro tipo de condición suficiente para obtener la misma conclusión.

Teorema de convergencia dominada. Sea X_1, X_2, \dots una sucesión de variables aleatorias para la cual existe otra variable Y integrable tal que $|X_n| \leq Y$, para $n \geq 1$. Si $\lim_{n \rightarrow \infty} X_n = X$ puntualmente, entonces X y X_n son integrables y

$$\lim_{n \rightarrow \infty} E(X_n) = E(X).$$

Demostración. Sea $Y_n = \inf\{X_n, X_{n+1}, \dots\}$. Entonces $Y_n \nearrow X$ cuando $n \rightarrow \infty$. Por lo tanto $(Y_n + Y) \nearrow (X + Y)$, en donde $Y_n + Y \geq 0$, pues como $-X_n \leq Y$, entonces $X_n \geq -Y$ para toda n , y por lo tanto $Y_n \geq -Y$. Por el teorema de convergencia monótona, $E(Y_n + Y) \nearrow E(X + Y)$. De donde se obtiene

$$E(Y_n) \nearrow E(X).$$

Sea ahora $Z_n = \sup\{X_n, X_{n+1}, \dots\}$. Entonces $Z_n \searrow X$ cuando $n \rightarrow \infty$. Por lo tanto $(Y - Z_n) \nearrow (Y - X)$, en donde $Y - Z_n \geq 0$, pues como $X_n \leq Y$ para toda n , entonces $Z_n \leq Y$. Por el teorema de convergencia monótona, $E(Y - Z_n) \nearrow E(Y - X)$. De donde se obtiene

$$E(Z_n) \searrow E(X).$$

Ahora observe que $Y_n \leq X_n \leq Z_n$. Por lo tanto $E(Y_n) \leq E(X_n) \leq E(Z_n)$. Al hacer n tender a infinito se obtiene el resultado. \square

Estos dos teoremas son herramientas fuertes en la teoría de la probabilidad. En particular, se usarán en la última parte del curso para formalizar algunas demostraciones.

7.9. Ejercicios

Convergencia casi segura

479. Demuestre que en la convergencia casi segura, el límite es único casi seguramente, es decir, si $X_n \xrightarrow{c.s.} X$, y $X_n \xrightarrow{c.s.} Y$, entonces $X = Y$ c.s.
480. Sean a y b constantes. Demuestre que si $X_n \xrightarrow{c.s.} X$, entonces

$$aX_n + b \xrightarrow{c.s.} aX + b.$$

481. Suponga que $X_n \xrightarrow{c.s.} X$ y $Y_n \xrightarrow{c.s.} Y$. Demuestre que

a) $X_n + Y_n \xrightarrow{c.s.} X + Y$.

b) $X_n Y_n \xrightarrow{c.s.} XY$.

482. Considere el espacio de probabilidad $([0, 1], \mathcal{B}[0, 1], P)$, con P la medida de probabilidad uniforme. Demuestre que

$$n1_{[0, 1/n]} \xrightarrow{c.s.} 0.$$

Convergencia en probabilidad

483. Demuestre que en la convergencia en probabilidad, el límite es único casi seguramente, es decir, si $X_n \xrightarrow{P} X$, y $X_n \xrightarrow{P} Y$, entonces $X = Y$ c.s.
484. Sean a y b constantes. Demuestre que si $X_n \xrightarrow{P} X$, entonces

$$aX_n + b \xrightarrow{P} aX + b.$$

485. Suponga que $X_n \xrightarrow{p} x$ y $Y_n \xrightarrow{p} y$, en donde x y y son dos números reales fijos. Demuestre que

a) $X_n + Y_n \xrightarrow{p} x + y$.

b) $X_n Y_n \xrightarrow{p} xy$.

c) Si g es continua en x , entonces $g(X_n) \xrightarrow{p} g(x)$.

486. Demuestre que si $X_n \xrightarrow{p} X$ y $Y_n \xrightarrow{p} Y$ entonces

$$X_n + Y_n \xrightarrow{p} X + Y.$$

487. Sean X_1, X_2, \dots variables aleatorias independientes cada una con distribución unif(a, b). Demuestre que cuando n tiende a infinito

a) $\min\{X_1, \dots, X_n\} \xrightarrow{p} a$.

b) $\max\{X_1, \dots, X_n\} \xrightarrow{p} b$.

Convergencia en media

488. Demuestre que en la convergencia en media, el límite es único casi seguramente, es decir, si $X_n \xrightarrow{m} X$, y $X_n \xrightarrow{m} Y$, entonces $X = Y$ c.s.

489. Sean a y b constantes. Demuestre que si $X_n \xrightarrow{m} X$, entonces

$$aX_n + b \xrightarrow{m} aX + b.$$

490. Suponga que $X_n \xrightarrow{m} X$ y $Y_n \xrightarrow{m} Y$. Demuestre que

$$X_n + Y_n \xrightarrow{m} X + Y.$$

Proporcione un contraejemplo para la siguiente afirmación:

$$X_n Y_n \xrightarrow{m} XY.$$

491. Demuestre que si $X_n \xrightarrow{m} X$, entonces $E(X_n) \rightarrow E(X)$.

Convergencia en media cuadrática

492. Demuestre que en la convergencia en media cuadrática, el límite es único casi seguramente, es decir, si $X_n \xrightarrow{m.c.} X$, y $X_n \xrightarrow{m.c.} Y$, entonces $X = Y$ c.s.

493. Sean a y b constantes. Demuestre que si $X_n \xrightarrow{m.c.} X$, entonces

$$aX_n + b \xrightarrow{m.c.} aX + b.$$

494. Use la desigualdad de Cauchy-Schwarz para demostrar que si $X_n \xrightarrow{m.c.} X$ y $Y_n \xrightarrow{m.c.} Y$, entonces

$$X_n + Y_n \xrightarrow{m.c.} X + Y.$$

495. Demuestre que si $X_n \xrightarrow{m.c.} X$, entonces $E(X_n^2) \rightarrow E(X^2)$.

Convergencia en distribución

496. Demuestre que en la convergencia en distribución, el límite es único en distribución, es decir, si $X_n \xrightarrow{d} X$, y $X_n \xrightarrow{d} Y$, entonces X y Y tienen la misma distribución.
497. Considere el espacio de probabilidad $([0, 1], \mathcal{B}[0, 1], P)$ en donde P es la medida de probabilidad uniforme. Sea $X_n = 1_{[0, 1/2+1/n]}$ y $X = 1_{[0, 1/2]}$. Demuestre que $X_n \xrightarrow{d} X$.
498. Sea X_n con distribución $\text{unif}[a - 1/n, a + 1/n]$, en donde a es una constante. Demuestre que $X_n \xrightarrow{d} a$.
499. Sea X_n con distribución uniforme en el conjunto $\{0, 1, \dots, n\}$. Demuestre que

$$\frac{1}{n}X_n \xrightarrow{d} \text{unif}[0, 1].$$

500. Sea c una constante. Demuestre que $X_n \xrightarrow{p} c$ si, y sólo si, $X_n \xrightarrow{d} c$.

Relaciones generales entre los tipos de convergencia

501. *Otro ejemplo de que la convergencia casi segura no implica la convergencia en media.* Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que para cada número natural n ,

$$\begin{aligned} P(X_n = 0) &= 1/4, \\ P(X_n = 1) &= 1/2, \\ \text{y } P(X_n = 2) &= 1/4. \end{aligned}$$

Defina $Y_n = X_1 \cdot X_2 \cdots X_n$. Demuestre que Y_n converge a cero, casi seguramente, pero no así en media, ni en media cuadrática.

502. Sea A_1, A_2, \dots una sucesión de eventos tal que $\lim_{n \rightarrow \infty} A_n = A$. ¿En qué sentido la sucesión de variables aleatorias 1_{A_n} converge a 1_A ?
503. Demuestre que si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} Y$ entonces no necesariamente
- $cX_n \xrightarrow{d} cX$, c constante.
 - $X_n + Y_n \xrightarrow{d} X + Y$.
504. Sea X_n con distribución $N(\mu_n, \sigma_n^2)$ y X con distribución $N(\mu, \sigma^2)$. Suponga $\mu_n \rightarrow \mu$ y $\sigma_n^2 \rightarrow \sigma^2$, con $\sigma_n^2, \sigma^2 > 0$. ¿En qué sentido $X_n \rightarrow X$?
505. Suponga $X_n \xrightarrow{d} X$ en donde X_n y X son variables aleatorias absolutamente continuas. ¿Bajo qué condiciones $f_{X_n}(x) \rightarrow f_X(x)$?

Capítulo 8

Funciones generadoras

En este capítulo se estudia la función generadora de probabilidad, la función generadora de momentos y la función característica. Estas funciones son transformaciones de las distribuciones de probabilidad, y constituyen una herramienta muy útil en la teoría moderna de la probabilidad.

8.1. Función generadora de probabilidad

Definición (Función generadora de probabilidad). La función generadora de probabilidad de X es la función

$$G(t) = E(t^X),$$

definida para valores reales de t tal que la esperanza sea convergente absolutamente.

Cuando sea necesario especificarlo se escribe $G_X(t)$ en lugar de $G(t)$, y se usan las letras f.g.p. en lugar de *función generadora de probabilidad*. Esta función se utiliza principalmente, aunque no únicamente, en el caso de variables aleatorias con valores enteros. Por comodidad supondremos que éstas toman valores en el conjunto $\{0, 1, \dots\}$, que corresponde al caso de las variables aleatorias discretas estudiadas en este curso. En tal situación,

$$G(t) = \sum_{k=0}^{\infty} t^k \cdot P(X = k).$$

Por lo tanto la f.g.p. es una serie de potencias en t , con coeficientes dados por la distribución de probabilidad, por ende el nombre. Es importante observar que el

radio de convergencia de esta serie es por lo menos uno, pues para $|t| < 1$,

$$|G(t)| = \left| \sum_{k=0}^{\infty} t^k \cdot P(X = k) \right| \leq \sum_{k=0}^{\infty} |t|^k \cdot P(X = k) \leq \sum_{k=0}^{\infty} P(X = k) = 1.$$

Ejemplo. Sea X con distribución Poisson(λ). Entonces la f.g.p. de X puede calcularse de la siguiente forma:

$$\begin{aligned} G(t) &= \sum_{k=0}^{\infty} t^k \cdot e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \\ &= e^{-\lambda} e^{\lambda t} \\ &= e^{-\lambda(1-t)}. \end{aligned}$$

Observe que en este caso la f.g.p. se encuentra definida para todo valor real de t . \circ

En la siguiente tabla se muestran ejemplos de funciones generadoras de probabilidad para algunas distribuciones discretas.

| Distribución | Función generadora de probabilidad |
|----------------------------------|---|
| $\text{unif}\{x_1, \dots, x_n\}$ | $G(t) = \frac{1}{n}(t^{x_1} + \dots + t^{x_n})$ |
| $\text{Ber}(p)$ | $G(t) = 1 - p + pt$ |
| $\text{bin}(n, p)$ | $G(t) = (1 - p + pt)^n$ |
| $\text{geo}(p)$ | $G(t) = p/[1 - t(1 - p)]$ |
| $\text{Poisson}(\lambda)$ | $G(t) = e^{-\lambda(1-t)}$ |
| $\text{bin neg}(r, p)$ | $G(t) = (p/[1 - t(1 - p)])^r$ |

La función generadora de probabilidad determina de manera única a la distribución en el siguiente sentido: si X y Y tienen la misma distribución de probabilidad, entonces naturalmente $G_X(t) = G_Y(t)$, para valores de t donde esta esperanza exista. Inversamente, sean X y Y tales que $G_X(t)$ y $G_Y(t)$ existen y coinciden en algún intervalo alrededor del cero, entonces X y Y tienen la misma distribución de probabilidad. Éstas y otras propiedades generales de la f.g.p. se estudian a continuación, y más adelante se ilustran estos resultados con algunos ejemplos.

Proposición (Propiedades de la f.g.p.).

1. Sean X y Y variables aleatorias con valores en $\{0, 1, \dots\}$ tales que $G_X(t)$ y $G_Y(t)$ existen y coinciden en algún intervalo alrededor de $t = 0$. Entonces X y Y tienen la misma distribución de probabilidad.

2. Si el n -ésimo momento factorial de X existe, entonces

$$\lim_{t \nearrow 1} \frac{d^n}{dt^n} G_X(t) = E[X(X-1)\cdots(X-n+1)].$$

3. Si X y Y son independientes y cuyas f.g.p. existen, entonces

$$G_{X+Y}(t) = G_X(t) \cdot G_Y(t).$$

Demostración. (1) Sean $a_k = P(X = k)$ y $b_k = P(Y = k)$, para cada $k \geq 0$. La igualdad $G_X(t) = G_Y(t)$ se escribe entonces de la forma siguiente:

$$\sum_{k=0}^{\infty} t^k a_k = \sum_{k=0}^{\infty} t^k b_k.$$

Para que estas dos series de potencias en t coincidan en algún intervalo no trivial alrededor del cero, sus coeficientes deben forzosamente coincidir, es decir, $a_k = b_k$ para cada $k \geq 0$. Esto significa que las distribuciones de probabilidad coinciden. (2) Como las series de potencia se pueden derivar término a término conservándose el mismo radio de convergencia, se tiene que

$$\begin{aligned} G'(t) &= \frac{d}{dt} \sum_{k=0}^{\infty} t^k P(X = k) \\ &= \sum_{k=0}^{\infty} \frac{d}{dt} t^k P(X = k) \\ &= \sum_{k=1}^{\infty} k t^{k-1} P(X = k). \end{aligned}$$

Por lo tanto, suponiendo que la esperanza existe, por el lema de Abel,

$$\lim_{t \nearrow 1} G'(t) = \sum_{k=1}^{\infty} k P(X = k) = E(X).$$

Para la segunda derivada se tiene

$$G''(t) = \sum_{k=2}^{\infty} k(k-1)t^{k-2} P(X = k),$$

de modo que cuando el segundo momento existe,

$$\lim_{t \nearrow 1} G''(t) = \sum_{k=2}^{\infty} k(k-1)P(X = k) = E(X(X-1)).$$

De manera análoga se demuestra para las derivadas superiores. (3) Cuando X y Y son independientes,

$$\begin{aligned} G_{X+Y}(t) &= E(t^{X+Y}) \\ &= E(t^X \cdot t^Y) \\ &= E(t^X) \cdot E(t^Y) \\ &= G_X(t) \cdot G_Y(t). \end{aligned}$$

□

Debido a la segunda propiedad, a la f.g.p. también se le conoce como *función generadora de momentos factoriales*.

Ejemplo. Se había encontrado que la f.g.p. de una variable aleatoria X con distribución Poisson(λ) es

$$G(t) = e^{-\lambda(1-t)}.$$

Usando esta función encontraremos la esperanza y varianza de X . Al derivar una vez se obtiene $G'(t) = \lambda e^{-\lambda(1-t)}$, y al evaluar en $t = 1$,

$$E(X) = G'(1) = \lambda.$$

Derivando por segunda vez, $G''(t) = \lambda^2 e^{-\lambda(1-t)}$, y en $t = 1$ se obtiene $E(X(X-1)) = G''(1) = \lambda^2$. Por lo tanto

$$\text{Var}(X) = E(X^2) - E^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

○

Ahora se muestra el uso de la f.g.p. para determinar la distribución de una variable aleatoria.

Ejemplo. Suponga que X y Y son independientes con distribución Poisson(λ_1) y Poisson(λ_2), respectivamente. Entonces

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= e^{-\lambda_1(1-t)} \cdot e^{-\lambda_2(1-t)} \\ &= e^{-(\lambda_1+\lambda_2)(1-t)}. \end{aligned}$$

Esta expresión corresponde a la f.g.p. de la distribución Poisson con parámetro $\lambda_1 + \lambda_2$. Debido a la unicidad, se concluye entonces que $X + Y$ tiene distribución Poisson($\lambda_1 + \lambda_2$). ○

8.2. Función generadora de momentos

La función generadora de momentos es otra función que se puede asociar a algunas distribuciones de probabilidad, su existencia no está garantizada en todos los

casos. Cuando existe, determina de manera única a la distribución de probabilidad asociada, y tiene propiedades semejantes a las de la f.g.p. estudiada en la sección anterior. La función generadora de momentos se utiliza tanto para variables aleatorias discretas como continuas.

Definición (Función generadora de momentos). La función generadora de momentos de X es la función

$$M(t) = E(e^{tX}),$$

definida para valores reales de t tales que esta esperanza sea absolutamente convergente.

Nuevamente, cuando sea necesario especificarlo se escribe $M_X(t)$ en lugar de $M(t)$, y se usan las letras f.g.m. en lugar de *función generadora de momentos*. La parte importante de esta función es su existencia en una vecindad no trivial alrededor del cero. Observe que la f.g.m. y la f.g.p. están relacionadas, cuando existen, por la igualdad $M(t) = G(e^t)$.

Ejemplo. Sea X con distribución $\text{gama}(n, \lambda)$. Entonces la f.g.m. de X puede calcularse de la siguiente forma:

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx} \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} dx \\ &= \lambda^n (\lambda - t)^{-n} \int_0^\infty \frac{[(\lambda - t)x]^{n-1}}{\Gamma(n)} (\lambda - t) e^{-(\lambda - t)x} dx \\ &= \lambda^n (\lambda - t)^{-n}. \end{aligned}$$

Observe que $M(t)$ esta definida para $t < \lambda$. ◦

La siguiente tabla muestra algunos otros ejemplos de funciones generadoras de momentos para ciertas distribuciones.

| Distribución | Función generadora de momentos |
|---------------------------|---|
| $\text{unif}(a, b)$ | $M(t) = (e^{bt} - e^{at}) / (bt - at)$ |
| $\text{exp}(\lambda)$ | $M(t) = \lambda / (\lambda - t)$ |
| $\text{gama}(n, \lambda)$ | $M(t) = [\lambda / (\lambda - t)]^n$ |
| $N(\mu, \sigma^2)$ | $M(t) = \exp(\mu t + \sigma^2 t^2 / 2)$ |
| $\chi^2(n)$ | $M(t) = (1 - 2t)^{-n/2}$ |
| $t(n)$ | $M(t)$ no existe para $t \neq 0$ |

Se demuestran a continuación algunas propiedades básicas de la f.g.m., y después se muestra su utilidad mediante algunos ejemplos.

Proposición (Propiedades de la f.g.m.). Sea X tal que su f.g.m. $M(t)$ es finita para cada $t \in (-s, s)$, con $s > 0$. Entonces

1. Todos los momentos de X son finitos.

$$2. M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n).$$

3. $M(t)$ tiene derivadas continuas de cualquier orden en $(-s, s)$, y se cumple

$$\left. \frac{d^n}{dt^n} M(t) \right|_{t=0} = E(X^n).$$

Demostración.

(1) La prueba se basa en las identidades:

$$E|X|^n = n \int_0^{\infty} (1 - F(x)) x^{n-1} dx + n \int_{-\infty}^0 F(x) |x|^{n-1} dx,$$

$$\text{y } M(t) = 1 + t \int_0^{\infty} (1 - F(x)) e^{tx} dx - t \int_{-\infty}^0 F(x) e^{tx} dx,$$

en donde, por hipótesis, las dos integrales de $M(t)$ son finitas para cualquier $t \in (-s, s)$. Para $x > 0$ se toma cualquier $t \in (0, s)$, y entonces

$$\frac{(tx)^n}{n!} \leq e^{tx}.$$

Es decir, $x^n \leq (n!/t^n)e^{tx}$. De modo que, salvo constantes, la primera integral de $E|X|^n$ es menor o igual a la primera integral de $M(t)$, siendo ésta última finita, la otra también.

Para $x < 0$ conviene tomar $t \in (-s, 0)$, pues en tal caso $tx > 0$ y entonces

$$\frac{|tx|^n}{n!} \leq e^{|tx|} = e^{tx}.$$

Es decir, $|x|^n \leq (n!/|t|^n)e^{tx}$. Ahora la segunda integral de $E|X|^n$ es menor o igual a la segunda integral de $M(t)$, siendo ésta última finita, la otra también. De esta forma todos los momentos de X existen cuando $M(t)$ es finita en algún intervalo no trivial alrededor del cero.

(2) Se usa la fórmula

$$E(X^n) = n \int_0^{\infty} (1 - F(x)) x^{n-1} dx - n \int_{-\infty}^0 F(x) x^{n-1} dx.$$

Entonces para cualquier $t \in (-s, s)$, y $m \geq 1$,

$$\begin{aligned} \sum_{n=0}^m \frac{t^n}{n!} E(X^n) &= 1 + \sum_{n=1}^m \frac{t^n}{n!} \cdot n \int_0^\infty (1 - F(x)) \cdot x^{n-1} dx \\ &\quad - \sum_{n=1}^m \frac{t^n}{n!} \cdot n \int_{-\infty}^0 F(x) \cdot x^{n-1} dx \\ &= 1 + t \int_0^\infty (1 - F(x)) \cdot \sum_{n=0}^{m-1} \frac{t^n}{n!} \cdot x^n dx \\ &\quad - t \int_{-\infty}^0 F(x) \cdot \sum_{n=0}^{m-1} \frac{t^n}{n!} \cdot x^n dx. \end{aligned}$$

Usando el teorema de convergencia monótona, o el de convergencia dominada, dependiendo de los valores de t y x , cada una de estas integrales es convergente, para cualquier $t \in (-s, s)$, cuando $m \rightarrow \infty$. De modo que

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n) &= 1 + t \int_0^\infty (1 - F(x)) \cdot e^{tx} dx - t \int_{-\infty}^0 F(x) \cdot e^{tx} dx \\ &= M(t). \end{aligned}$$

(3) Dado que $M(t)$ se puede expresar como una serie de potencias en t , diferenciando y evaluando en cero se obtienen los coeficientes $E(X^n)$. \square

Si el n -ésimo momento de una variable aleatoria existe, no implica que éste puede ser hallado a través de la n -ésima derivada de la f.g.m. evaluada en cero. Es decir, es necesario conocer la existencia de la f.g.m. para que pueda ser utilizada para obtener los momentos. Por ejemplo, una variable aleatoria con distribución $t(n)$ tiene esperanza cero pero su f.g.m. $M(t)$ no existe para t distinto de cero.

Ejemplo. Sea X con distribución $\text{gama}(n, \lambda)$. Entonces hemos encontrado antes que para $t < \lambda$,

$$M(t) = \lambda^n (\lambda - t)^{-n}.$$

Calcularemos ahora la esperanza y varianza de X con ayuda de la f.g.m. Derivando una vez,

$$M'(t) = \lambda^n n (\lambda - t)^{-n-1}.$$

Al evaluar en $t = 0$ se obtiene $E(X) = n/\lambda$. Derivando nuevamente,

$$M''(t) = \lambda^n n(n+1) (\lambda - t)^{-n-2}.$$

Por lo tanto $E(X^2) = M''(0) = n(n+1)/\lambda^2$. Entonces

$$\text{Var}(X) = n(n+1)/\lambda^2 - n^2/\lambda^2 = n/\lambda^2.$$

o

Ejemplo. Suponga ahora que X y Y son independientes cada una con distribución gama(n, λ) y gama(m, λ), respectivamente. Entonces la f.g.m. de $X + Y$ es

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= \lambda^n (\lambda - t)^{-n} \cdot \lambda^m (\lambda - t)^{-m} \\ &= \lambda^{n+m} (\lambda - t)^{-n-m}. \end{aligned}$$

Ésta es nuevamente la expresión de la f.g.m. de la distribución gama, ahora con parámetros $n+m$ y λ . Se concluye entonces $X+Y$ tiene distribución gama($n+m, \lambda$). ◦

Nuevamente, en el caso de independencia la función generadora de la suma es el producto las de funciones generadoras individuales.

Proposición. Sean X y Y son independientes, y cuyas f.g.m. existen en una vecindad no trivial alrededor del cero. Entonces para cualquier $t \in (-s, s)$ para algún $s > 0$,

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$$

Demostración. Bajo la hipótesis de independencia,

$$\begin{aligned} M_{X+Y}(t) &= E(e^{t(X+Y)}) \\ &= E(e^{tX} \cdot e^{tY}) \\ &= E(e^{tX}) \cdot E(e^{tY}) \\ &= M_X(t) \cdot M_Y(t). \end{aligned}$$

□

Es interesante observar que la condición $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ no es suficiente para concluir que X y Y son independientes. En el Ejercicio 542 se pide dar los detalles de tal afirmación.

Como hemos mencionado antes, no todas las distribuciones de probabilidad permiten calcular la función generadora de momentos dentro de un intervalo no trivial alrededor del cero, ni todos los cálculos son tan sencillos como en el ejemplo mostrado. Por ejemplo, la f.g.m. de la distribución Cauchy estándar no existe para valores de t distintos de cero como se pide comprobar en el Ejercicio 543.

Cuando se tienen dos variables X y Y con la misma distribución, entonces sus funciones generadoras de momentos coinciden pues éstas se obtienen a través de la función de distribución común. Por el contrario, si $M_X(t) = M_Y(t)$ en una vecindad no trivial alrededor del cero, entonces puede demostrarse que sus distribuciones coinciden, éste resultado y otro relativo a convergencia es el contenido de la siguiente proposición, cuya demostración omitiremos.

Proposición (Otras propiedades de la f.g.m.).

1. Las variables X y Y tienen la misma distribución si, y sólo si, $M_X(t) = M_Y(t)$ para valores de t en una vecindad no trivial alrededor del cero.
2. Sea X_1, X_2, \dots una sucesión de variables aleatorias cuyas funciones generadoras de momentos existen todas ellas en algún intervalo no trivial alrededor del cero. Entonces $X_n \xrightarrow{d} X$ si, y sólo si, $M_{X_n}(t) \rightarrow M_X(t)$.

En la sección de ejercicios se pueden encontrar las funciones generadoras de momentos de algunas otras distribuciones de probabilidad, tanto discretas como continuas, así como en el primer apéndice al final del libro.

8.3. Función característica

En esta última sección se estudia la función característica y se enuncian algunas de sus propiedades. Ésta es una función definida para cada distribución de probabilidad, y a diferencia de las funciones generadoras de probabilidad y de momentos estudiadas antes, siempre existe. Su definición es la siguiente.

Definición (Función característica). La función característica de X es la función

$$\phi(t) = E(e^{itX}),$$

definida para cualquier número real t . El número i es la unidad de los números imaginarios.

Observe que la función característica es una función de los números reales en los números complejos, y puede escribirse en la forma siguiente:

$$\phi(t) = E(\cos tX) + iE(\sin tX).$$

Nuevamente se escribe $\phi_X(t)$ cuando sea necesario especificar que se trata de la función característica de X . Se escribe simplemente f.c. en lugar de *función característica*. Observe que la f.c., la f.g.m. y la f.g.p. están relacionadas, cuando existen las dos últimas, por las igualdades $\phi(t) = M(it) = G(e^{it})$. Se muestran a continuación algunos ejemplos de la forma de encontrar la función característica a partir de una distribución de probabilidad.

Ejemplo. Sea X con distribución $\text{bin}(n, p)$. Entonces

$$\phi(t) = E(e^{itX})$$

$$\begin{aligned}
&= \sum_{x=0}^n e^{itx} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^n \binom{n}{x} (pe^{it})^x (1-p)^{n-x} \\
&= (1-p + pe^{it})^n.
\end{aligned}$$

◦

Ejemplo. Sea X con distribución Poisson(λ). Entonces $\phi(t) = e^{-\lambda(1-e^{it})}$. En efecto,

$$\begin{aligned}
\phi(t) &= E(e^{itX}) \\
&= \sum_{x=0}^{\infty} e^{itx} \left[e^{-\lambda} \frac{\lambda^x}{x!} \right] \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{it})^x}{x!} \\
&= e^{-\lambda(1-e^{it})}.
\end{aligned}$$

◦

Otros ejemplos de funciones características de distribuciones discretas se muestra en la siguiente tabla:

| Distribución discreta | Función característica |
|-----------------------|-------------------------------------|
| Ber(p) | $\phi(t) = 1 - p + pe^{it}$ |
| bin(n, p) | $\phi(t) = (1 - p + pe^{it})^n$ |
| Poisson(λ) | $\phi(t) = e^{-\lambda(1-e^{it})}$ |
| geo(p) | $\phi(t) = p/(1 - (1-p)e^{it})$ |
| bin neg(r, p) | $\phi(t) = [p/(1 - (1-p)e^{it})]^r$ |

Ahora se mostrará la forma de encontrar la función característica para dos distribuciones continuas: la distribución normal y la distribución gama.

Ejemplo. Sea X con distribución $N(\mu, \sigma^2)$. Entonces

$$\begin{aligned}
\phi(t) &= E(e^{itX}) \\
&= \int_{-\infty}^{\infty} e^{itx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx
\end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2-2x(\mu-it\sigma^2)+\mu^2)/2\sigma^2} dx \\
&= e^{(-\mu^2+(\mu-it\sigma^2)^2)/2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[x-(\mu-it\sigma^2)]^2/2\sigma^2} dx \\
&= e^{it\mu-t^2\sigma^2/2}.
\end{aligned}$$

Observe que el último integrando es la función de densidad normal con media el número complejo $\mu - it\sigma^2$, y varianza σ^2 . El hecho de que esta integral también vale uno es consecuencia del *principio de continuación analítica* de la teoría de variable compleja. \circ

Ejemplo. Sea X con distribución Gama(z, λ). Entonces $\phi(t) = (\frac{\lambda}{\lambda-it})^z$. En efecto,

$$\begin{aligned}
\phi(t) &= E(e^{itX}) \\
&= \int_0^{\infty} e^{itx} \left[\frac{(\lambda x)^{z-1}}{\Gamma(z)} \lambda e^{-\lambda x} \right] dx \\
&= \int_0^{\infty} \frac{\lambda}{\Gamma(z)} (\lambda x)^{z-1} e^{-(\lambda-it)x} dx \\
&= \frac{\lambda^z}{(\lambda-it)^z} \int_0^{\infty} \frac{[(\lambda-it)x]^{z-1}}{\Gamma(z)} (\lambda-it) e^{-(\lambda-it)x} dx \\
&= \left(\frac{\lambda}{\lambda-it} \right)^z.
\end{aligned}$$

El último integrando es la función de densidad de la distribución $\text{gama}(z, \lambda - it)$. Usando la teoría de variable compleja puede demostrarse que esta integral vale uno. \circ

La siguiente tabla muestra algunos otros ejemplos de funciones características para variables aleatorias continuas:

| Distribución continua | Función característica |
|---------------------------|---|
| $\text{unif}(a, b)$ | $\phi(t) = (e^{ibt} - e^{iat}) / (ibt - iat)$ |
| $\text{exp}(\lambda)$ | $\phi(t) = \lambda / (\lambda - it)$ |
| $\text{gama}(n, \lambda)$ | $\phi(t) = [\lambda / (\lambda - it)]^n$ |
| $N(\mu, \sigma^2)$ | $\phi(t) = \exp(i\mu t - \sigma^2 t^2 / 2)$ |
| $\chi^2(n)$ | $\phi(t) = (1 - 2it)^{-n/2}$ |
| $t(n)$ | $\phi(t) = e^{- t }$, cuando $n = 1$. |

La existencia de la función característica se sigue del siguiente resultado.

Proposición (Existencia de la f.c.). Para cualquier número real t ,

$$|\phi(t)| \leq 1.$$

En particular, $\phi(0) = 1$.

Demostración. Para cualquier número real t ,

$$\begin{aligned} |\phi(t)| &= \left| \int_{-\infty}^{\infty} e^{itx} dF(x) \right| \\ &\leq \int_{-\infty}^{\infty} |e^{itx}| dF(x) \\ &= \int_{-\infty}^{\infty} dF(x) \\ &= 1. \end{aligned}$$

□

De modo que $\phi(t)$ es un número complejo de módulo menor o igual a uno, para cualquier valor de t . Veremos a continuación algunas otras propiedades de esta importante función. En particular, demostraremos que los momentos de una variable aleatoria X pueden ser generados, cuando existen, con la f.c. a través de la fórmula $\phi^{(n)}(0) = i^n \cdot E(X^n)$, y como en el caso de las funciones generadoras anteriores, cuando X y Y son independientes se cumple que $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$.

Proposición. Si X tiene n -ésimo momento finito, entonces

$$\frac{d^n}{dt^n} \phi(t) = E[(iX)^n \cdot e^{itX}].$$

En consecuencia,

$$\left. \frac{d^n}{dt^n} \phi(t) \right|_{t=0} = i^n \cdot E(X^n).$$

Demostración. Para cualquier h distinto de cero,

$$\begin{aligned} \frac{\phi(t+h) - \phi(t)}{h} &= \int_{-\infty}^{\infty} \frac{e^{i(t+h)x} - e^{itx}}{h} dF(x) \\ &= \int_{-\infty}^{\infty} e^{itx} \cdot \frac{e^{ihx} - 1}{h} dF(x) \\ &= E\left[e^{itX} \cdot \frac{e^{ihX} - 1}{h} \right]. \end{aligned}$$

Como $\lim_{h \rightarrow 0} \frac{e^{ihx} - 1}{h} = ix$, entonces, puntualmente,

$$\lim_{h \rightarrow 0} e^{itX} \cdot \frac{e^{ihX} - 1}{h} = iX \cdot e^{itX}.$$

Se comprueba que las variables aleatorias de esta sucesión, parametrizada por h , son uniformemente dominadas por una variable aleatoria integrable, en efecto,

$$\begin{aligned} \left| e^{itX} \cdot \frac{e^{ihX} - 1}{h} \right| &= \left| \frac{e^{ihX} - 1}{h} \right| \\ &= \left| \frac{1}{h} \int_0^h iX \cdot e^{isX} ds \right| \\ &\leq |X| \cdot \frac{1}{h} \int_0^h |e^{isX}| ds \\ &= |X|. \end{aligned}$$

Por hipótesis, $E|X| < \infty$, entonces por el teorema de convergencia dominada,

$$\frac{d}{dt} \phi(t) = \int_{-\infty}^{\infty} ix \cdot e^{itx} dF(x) = E[iX \cdot e^{itX}].$$

Usando el mismo procedimiento, se encuentra que

$$\frac{d^n}{dt^n} \phi(t) = \int_{-\infty}^{\infty} (ix)^n \cdot e^{itx} dF(x) = E[(iX)^n \cdot e^{itX}].$$

Tomando el límite cuando $t \rightarrow 0$, y usando nuevamente el teorema de convergencia dominada,

$$\left. \frac{d^n}{dt^n} \phi(t) \right|_{t=0} = i^n \cdot E(X^n).$$

□

Proposición. Si X y Y son independientes, entonces

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t).$$

Demostración. Por independencia,

$$\begin{aligned} \phi_{X+Y}(t) &= E(e^{it(X+Y)}) \\ &= E(e^{itX} \cdot e^{itY}) \\ &= E(e^{itX}) \cdot E(e^{itY}) \\ &= \phi_X(t) \cdot \phi_Y(t). \end{aligned}$$

□

En particular, este resultado establece que el producto de dos funciones características es nuevamente una función característica. Por otro lado, es necesario señalar que, en general, el recíproco de la última propiedad es falso. En el Ejercicio 565 se pide demostrar que la condición $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$ no es suficiente para concluir que las variables aleatorias X y Y son independientes. Otra de las propiedades fundamentales de la función característica es su capacidad de determinar de manera única a las distintas distribuciones de probabilidad. A este respecto se tienen los siguientes resultados.

Proposición (Fórmula de inversión de Lèvy). Sea X con función de distribución $F(x)$, y función característica $\phi(t)$. Si $x < y$ son puntos de continuidad de F , entonces

$$F(y) - F(x) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \cdot \phi(t) dt.$$

Cuando x y y no necesariamente son puntos de continuidad de F , el lado izquierdo es $\frac{1}{2}(F(y) + F(y-)) - \frac{1}{2}(F(x) + F(x-))$.

Demostración. Sea

$$\begin{aligned} I(T) &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \cdot \phi(t) dt \\ &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \cdot \left[\int_{-\infty}^{\infty} e^{itz} dF(z) \right] dt \\ &= \frac{1}{2\pi} \int_{-T}^T \int_{-\infty}^{\infty} \frac{e^{it(z-x)} - e^{it(z-y)}}{it} dF(z) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^T \frac{e^{it(z-x)} - e^{it(z-y)}}{it} dt dF(z). \end{aligned}$$

El cambio en el orden de integración es permitido pues el integrando es una función continua y acotada en $t \in [-T, T]$ y $z \in \mathbb{R}$, incluyendo cuando $t = 0$, pues puede definirse esta función de acuerdo a su comportamiento límite en ese punto, es decir,

$$\lim_{t \rightarrow 0} \frac{e^{it(z-x)} - e^{it(z-y)}}{it} = y - x.$$

Desarrollando las exponenciales en términos de senos y cosenos se obtiene

$$\begin{aligned} I(T) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^T \frac{1}{it} [\cos t(z-x) + i \operatorname{sen} t(z-x) \\ &\quad - \cos t(z-y) - i \operatorname{sen} t(z-y)] dt dF(z), \end{aligned}$$

en donde para cualquier número real a , por ser coseno una función par, y seno una función impar,

$$\int_{-T}^T \frac{\cos(at)}{t} dt = 0,$$

$$y \quad \int_{-T}^T \frac{\text{sen}(at)}{t} dt = 2 \int_0^T \frac{\text{sen}(at)}{t} dt.$$

Por lo tanto

$$I(T) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[2 \int_0^T \frac{\text{sen } t(z-x)}{t} dt - 2 \int_0^T \frac{\text{sen } t(z-y)}{t} dt \right] dF(z).$$

El siguiente paso consiste en aplicar el teorema de convergencia dominada cuando $T \rightarrow \infty$. La integral $I(T)$ es la esperanza de la variable aleatoria

$$X_T = \frac{1}{2\pi} \left[2 \int_0^T \frac{\text{sen } t(X-x)}{t} dt - 2 \int_0^T \frac{\text{sen } t(X-y)}{t} dt \right].$$

Nos interesa encontrar el límite de esta variable cuando $T \rightarrow \infty$. Para ello se hace uso del siguiente resultado:

$$\begin{aligned} \lim_{T \rightarrow \infty} 2 \int_0^T \frac{\text{sen } at}{t} dt &= \begin{cases} \pi & \text{si } a > 0, \\ -\pi & \text{si } a < 0, \\ 0 & \text{si } a = 0, \end{cases} \\ &= \pi \cdot \text{signo}(a), \end{aligned}$$

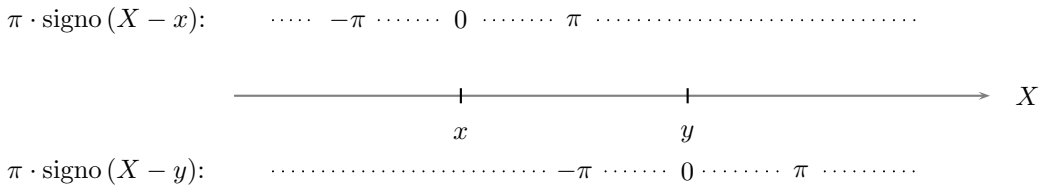
en donde la función signo es

$$\text{signo}(a) = \begin{cases} +1 & \text{si } a > 0, \\ -1 & \text{si } a < 0, \\ 0 & \text{si } a = 0. \end{cases}$$

Entonces, puntualmente,

$$\begin{aligned} \lim_{T \rightarrow \infty} X_T &= \frac{1}{2\pi} \left[\pi \cdot \text{signo}(X-x) - \pi \cdot \text{signo}(X-y) \right] \\ &= \begin{cases} 0 & \text{si } X < x, \\ 1/2 & \text{si } X = x, \\ 1 & \text{si } x < X < y, \\ 1/2 & \text{si } X = y, \\ 0 & \text{si } X > y. \end{cases} \\ &= \frac{1}{2} \cdot 1_{\{x,y\}}(X) + 1_{(x,y)}(X). \end{aligned}$$

Estas evaluaciones pueden encontrarse fácilmente usando el siguiente diagrama:



Además las variables X_T están acotadas en valor absoluto por una constante pues para cualquier número real a ,

$$\left| \int_0^T \frac{\operatorname{sen} at}{t} dt \right| \leq \sup_{T>0} \left| \int_0^T \frac{\operatorname{sen} t}{t} dt \right| < \infty.$$

Por lo tanto

$$\begin{aligned} \lim_{T \rightarrow \infty} I(T) &= \int_{-\infty}^{\infty} \left[\frac{1}{2} \cdot 1_{\{x,y\}}(z) + 1_{(x,y)}(z) \right] dF(z) \\ &= \frac{1}{2}P(X = x) + \frac{1}{2}P(X = y) + P(x < X < y) \\ &= P(x < X \leq y) + \frac{1}{2}P(X = x) - \frac{1}{2}P(X = y) \\ &= F(y) - F(x) + \frac{1}{2}P(X = x) - \frac{1}{2}P(X = y) \\ &= \frac{1}{2} (F(y) + F(y-)) - \frac{1}{2} (F(x) + F(x-)). \end{aligned}$$

En particular, si x y y son puntos de continuidad de F , entonces la integral vale $F(y) - F(x)$. \square

Como corolario del teorema de inversión probaremos que las funciones de distribución determinan de manera única a las distribuciones de probabilidad.

Teorema de unicidad. Si X y Y son tales que $\phi_X(t) = \phi_Y(t)$ para todo valor real de t , entonces X y Y tienen la misma distribución.

Demostración. Sea $\phi(t)$ la función característica común, y sea z cualquier número real. Escóganse x y y tales que $x < z < y$. Haciendo x tender a $-\infty$, y $y \searrow z$, en la fórmula de inversión de Lèvy, se obtiene una única función de distribución dada por

$$F(z) = \lim_{y \searrow z} \lim_{x \searrow -\infty} \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \cdot \phi(t) dt.$$

\square

En el caso absolutamente continuo se tiene la siguiente fórmula.

Proposición (Fórmula de inversión en el caso abs. continuo). Sea X absolutamente continua con función de densidad $f(x)$, y función característica $\phi(t)$. Entonces

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \cdot \phi(t) dt.$$

Demostración. Para $x < y$, dos puntos de continuidad de F , por el teorema de inversión de Lèvy, y después usando el teorema de Fubini,

$$\begin{aligned}
 F(y) - F(x) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \cdot \phi(t) dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \cdot \phi(t) dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_x^y e^{-itx} dx \right] \cdot \phi(t) dt. \\
 &= \int_x^y \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \cdot \phi(t) dt \right] dx.
 \end{aligned}$$

Por lo tanto el integrando debe ser la función de densidad de X . □

Es necesario señalar que el uso de esta fórmula requiere conocer de antemano que la función característica proviene de una variable aleatoria absolutamente continua. Surge entonces el problema de encontrar condiciones sobre $\phi(t)$ que garanticen que la correspondiente variable aleatoria sea absolutamente continua. Ahora se demuestra un resultado que será de utilidad en la última parte del curso y que establece que la convergencia en distribución es equivalente a la convergencia puntual de las correspondientes funciones características.

Teorema de Continuidad. Sean X, X_1, X_2, \dots variables aleatorias. Entonces $X_n \xrightarrow{d} X$ si, y sólo si, $\phi_{X_n}(t) \rightarrow \phi_X(t)$.

Demostración. Suponga que $\phi_{X_n}(t) \rightarrow \phi_X(t)$. Entonces para dos puntos de continuidad $x < y$ de F_X , el teorema de inversión de Lèvy establece que

$$\begin{aligned}
 F_X(y) - F_X(x) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \cdot \phi(t) dt. \\
 &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \cdot \left[\lim_{n \rightarrow \infty} \phi_{X_n}(t) \right] dt. \\
 &= \lim_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \cdot [\phi_{X_n}(t)] dt. \\
 &= \lim_{n \rightarrow \infty} F_{X_n}(y) - F_{X_n}(x).
 \end{aligned}$$

Haciendo x tender a $-\infty$ se obtiene $F_X(y) = \lim_{n \rightarrow \infty} F_{X_n}(y)$. □

8.4. Ejercicios

Función generadora de probabilidad

506. Sea X con varianza finita y con f.g.p. $G(t)$. Demuestre que

- a) $E(X) = G'(1)$.
- b) $E(X^2) = G''(1) + G'(1)$.
- c) $\text{Var}(X) = G''(1) + G'(1) - [G'(1)]^2$.

507. Sea X con f.g.p. $G_X(t)$, y sean a y b dos constantes. Demuestre que

$$G_{aX+b}(t) = t^b G_X(t^a).$$

508. Sea X con distribución $\text{Ber}(p)$. Demuestre que

- a) $G(t) = 1 - p + pt$.
- b) $E(X) = p$, usando $G(t)$.
- c) $\text{Var}(X) = p(1 - p)$, usando $G(t)$.

509. Sea X con distribución $\text{bin}(n, p)$. Demuestre que

- a) $G(t) = (1 - p + pt)^n$.
- b) $E(X) = np$, usando $G(t)$.
- c) $\text{Var}(X) = np(1 - p)$, usando $G(t)$.

510. Sean X_1, \dots, X_n independientes, cada una con distribución $\text{Ber}(p)$. Use la f.g.p. para demostrar que $X_1 + \dots + X_n$ tiene distribución $\text{bin}(n, p)$.

511. Sean X y Y independientes con distribución $\text{bin}(n, p)$ y $\text{bin}(m, p)$, respectivamente. Use la f.g.p. para demostrar que $X + Y$ tiene distribución $\text{bin}(n + m, p)$.

512. Sea X con distribución $\text{bin}(N, p)$, en donde N es una v.a. con distribución $\text{bin}(m, r)$. Use la f.g.p. para demostrar que X tiene distribución $\text{bin}(m, rp)$.

513. Sea X con distribución $\text{geo}(p)$. Demuestre que

- a) $G(t) = p/[1 - t(1 - p)]$.
- b) $E(X) = (1 - p)/p$, usando $G(t)$.
- c) $\text{Var}(X) = (1 - p)/p^2$, usando $G(t)$.

514. Sea X con distribución $\text{Poisson}(\lambda)$. Demuestre que

- a) $G(t) = e^{-\lambda(1-t)}$.
- b) $E(X) = \lambda$, usando $G(t)$.
- c) $\text{Var}(X) = \lambda$, usando $G(t)$.

515. Sean X y Y independientes con distribución Poisson con parámetros λ_1 y λ_2 respectivamente. Use la f.g.p. para demostrar nuevamente que $X + Y$ tiene distribución $\text{Poisson}(\lambda_1 + \lambda_2)$.

516. Sea X con distribución bin neg(r, p). Demuestre que

- a) $G(t) = [p/(1 - t(1 - p))]^r$.
- b) $E(X) = r(1 - p)/p$, usando $G(t)$.
- c) $\text{Var}(X) = r(1 - p)/p^2$, usando $G(t)$.

517. Sean X_1, \dots, X_n independientes tales que X_k tiene f.g.p. $G_k(t)$, para $k = 1, \dots, n$. Demuestre que

$$G_{X_1 + \dots + X_n}(t) = \prod_{k=1}^n G_k(t).$$

518. Investigue si la condición $G_{X+Y}(t) = G_X(t)G_Y(t)$, válida para valores de t en algún intervalo no trivial alrededor del cero, es suficiente para concluir que X y Y son independientes.

519. Sea X_1, X_2, \dots una sucesión de v.a.i.i.d. con f.g.p. $G_X(t)$. Sea N otra variable aleatoria con valores en \mathbb{N} , independiente de la sucesión y con f.g.p. $G_N(t)$. Sea $Y = X_1 + \dots + X_N$. Demuestre que

- a) $G_Y(t) = G_N(G_X(t))$.
- b) $E(Y) = E(N)E(X)$, usando $G_Y(t)$.
- c) $\text{Var}(Y) = E^2(X)\text{Var}(N) + E(N)\text{Var}(X)$, usando $G_Y(t)$.

Función generadora de momentos

520. Sea X con varianza finita y con f.g.m. $M(t)$. Demuestre que

- a) $E(X) = M'(0)$.
- b) $E(X^2) = M''(0)$.
- c) $\text{Var}(X) = M''(0) - (M'(0))^2$.

521. Sean X y Y independientes e idénticamente distribuidas con f.g.m. $M(t)$. Demuestre que $M_{X-Y}(t) = M(t) \cdot M(-t)$.

522. Sea X con f.g.m. $M_X(t)$, y sean a y b dos constantes. Demuestre que

$$M_{aX+b}(t) = e^{tb}M_X(at).$$

523. Sea X con f.g.m. $M_X(t)$. Diga falso o verdadero. Demuestre en cada caso.

- a) $M_X(t) \geq 0$.
- b) $M_{2X}(t) = M_X(2t)$.

524. Sea X con distribución Ber(p). Demuestre que

- a) $M(t) = 1 - p + pe^t$.
- b) $E(X) = p$, usando $M(t)$.
- c) $E(X^n) = p$, usando $M(t)$.

- d) $\text{Var}(X) = p(1 - p)$, usando $M(t)$.
525. Sea X con distribución $\text{bin}(n, p)$. Demuestre que
- $M(t) = (1 - p + pe^t)^n$.
 - $E(X) = np$, usando $M(t)$.
 - $\text{Var}(X) = np(1 - p)$, usando $M(t)$.
526. Sean X_1, \dots, X_n independientes cada una con distribución $\text{Ber}(p)$. Use la f.g.m. para demostrar que $X_1 + \dots + X_n$ tiene distribución $\text{bin}(n, p)$.
527. Sean X y Y independientes con distribución $\text{bin}(n, p)$ y $\text{bin}(m, p)$ respectivamente. Use la f.g.m. para demostrar que $X + Y$ tiene distribución $\text{bin}(n + m, p)$.
528. Sea X con distribución $\text{geo}(p)$. Demuestre que
- $M(t) = p/[1 - (1 - p)e^t]$.
 - $E(X) = (1 - p)/p$, usando $M(t)$.
 - $\text{Var}(X) = (1 - p)/p^2$, usando $M(t)$.
529. Sea X con distribución $\text{Poisson}(\lambda)$. Demuestre que
- $M(t) = \exp[\lambda(e^t - 1)]$.
 - $M''(t) = M'(t) + \lambda e^t M'(t)$.
 - $E(X) = \lambda$, usando $M(t)$.
 - $\text{Var}(X) = \lambda$, usando $M(t)$.
 - $E[(X - \lambda)^3] = \lambda$, usando $M(t)$.
530. Sea X con distribución $\text{unif}(a, b)$. Demuestre que
- $M(t) = \frac{e^{bt} - e^{at}}{(b - a)t}$.
 - $E(X) = (a + b)/2$, usando $M(t)$.
 - $\text{Var}(X) = (b - a)^2/12$, usando $M(t)$.
531. Sea X con distribución $\text{exp}(\lambda)$. Demuestre que
- $M(t) = \lambda/(\lambda - t)$, para $t < \lambda$.
 - $E(X) = 1/\lambda$, usando $M(t)$.
 - $\text{Var}(X) = 1/\lambda^2$, usando $M(t)$.
532. Sea X con distribución $\text{N}(\mu, \sigma^2)$. Demuestre que
- $M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.
 - $E(X) = \mu$, usando $M(t)$.
 - $\text{Var}(X) = \sigma^2$, usando $M(t)$.
533. Sean X y Y variables aleatorias independientes con distribución $\text{N}(\mu_1, \sigma_1^2)$ y $\text{N}(\mu_2, \sigma_2^2)$ respectivamente. Use la f.g.m. para demostrar que $X + Y$ tiene distribución $\text{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

534. Sea X con distribución gama(n, λ). Demuestre que
- $M(t) = [\lambda/(\lambda - t)]^n$, para $t < \lambda$.
 - $E(X) = n/\lambda$, usando $M(t)$.
 - $\text{Var}(X) = n/\lambda^2$, usando $M(t)$.
535. Sean X y Y independientes ambas con distribución $\exp(\lambda)$. Use la f.g.m. para demostrar que $X + Y$ tiene distribución gama($2, \lambda$).
536. Sean X y Y independientes con distribución gama(n, λ) y gama(m, λ) respectivamente. Use la f.g.m. para demostrar que $X + Y$ tiene distribución gama($n + m, \lambda$).
537. Sea X con distribución $\chi^2(n)$. Demuestre que
- $M(t) = [1/(1 - 2t)]^{n/2}$, para $t < 1/2$.
 - $E(X) = n$, usando $M(t)$.
 - $\text{Var}(X) = 2n$, usando $M(t)$.
538. Use la f.g.m. para demostrar que si X y Y son independientes tales que X tiene distribución $\chi^2(n)$ y $X + Y$ tiene distribución $\chi^2(m)$ con $m > n$, entonces Y tiene distribución $\chi^2(m - n)$.
539. Sean X y Y independientes con distribución $\chi^2(n)$ y $\chi^2(m)$ respectivamente. Use la f.g.m. para demostrar que $X + Y$ tiene distribución $\chi^2(n + m)$.
540. Sea X con distribución $N(\mu, \sigma^2)$. Use la f.g.m. para demostrar que
- $-X$ tiene distribución $N(-\mu, \sigma^2)$.
 - $aX + b$ tiene distribución $N(a\mu + b, a^2\sigma^2)$, con $a \neq 0$.
 - X^2 tiene distribución $\chi^2(1)$.
541. Sean X_1, \dots, X_n independientes tales que X_k tiene f.g.m. $M_k(t)$ para $k = 1, \dots, n$. Demuestre que

$$M_{X_1 + \dots + X_n}(t) = \prod_{k=1}^n M_k(t).$$

542. Demuestre que la condición $M_{X+Y}(t) = M_X(t)M_Y(t)$ no implica, necesariamente, que X y Y son independientes. Considere la distribución conjunta

$$f(x, y) = \frac{1}{4}[1 + xy(x^2 - y^2)] \quad \text{para} \quad -1 < x, y < 1.$$

543. Sea X con distribución Cauchy estándar. Demuestre que

$$M_X(t) = \begin{cases} 1 & \text{si } t = 0, \\ \infty & \text{si } t \neq 0. \end{cases}$$

544. Sea X con distribución $t(n)$. Demuestre que

$$M_X(t) = \begin{cases} 1 & \text{si } t = 0, \\ \infty & \text{si } t \neq 0. \end{cases}$$

545. Demuestre que la f.g.m. de la siguiente función de densidad, en donde n es un número natural, no existe.

$$f(x) = \begin{cases} \frac{n}{x^{n+1}} & \text{si } x > 1, \\ 0 & \text{otro caso.} \end{cases}$$

Función característica

546. Encuentre la f.c. de una v.a. con función de densidad

a) $f(x) = \frac{1}{x!(e-1)}$, para $x = 1, 2, \dots$

b) $f(x) = \frac{1}{2}e^{-|x|}$, para $-\infty < x < \infty$.

547. Sea X con función característica $\phi_X(t)$, y sean a y b dos constantes. Demuestre que

$$\phi_{aX+b}(t) = e^{itb} \phi_X(at).$$

548. Sea F una función de distribución con función característica ϕ . Demuestre que $F(x)$ es simétrica si, y sólo si, $\phi(t)$ es real.

549. Demuestre que la función característica es una función uniformemente continua.

550. Demuestre que la f.c. satisface $\phi(-t) = \overline{\phi(t)}$, en donde \bar{z} denota el complejo conjugado de z .

551. Sean $\phi_1(t)$ y $\phi_2(t)$ dos funciones características, y sea $\alpha \in [0, 1]$. Demuestre que $\phi(t) = \alpha\phi_1(t) + (1 - \alpha)\phi_2(t)$ es también una función característica.

552. Sean X y Y independientes y con idéntica distribución. Demuestre que $\phi_{X-Y}(t) = |\phi_X(t)|^2$.

553. Sea X con distribución $\text{Ber}(p)$. Demuestre que

a) $\phi(t) = 1 - p + pe^{it}$.

b) $E(X) = p$, usando $\phi(t)$.

c) $\text{Var}(X) = p(1 - p)$, usando $\phi(t)$.

d) $E(X^n) = p$, usando $\phi(t)$, con $n \geq 1$ entero.

554. Sea X con distribución $\text{bin}(n, p)$. Hemos demostrado que la función característica de esta distribución es $\phi(t) = (1 - p + pe^{it})^n$. Usando $\phi(t)$ demuestre ahora que

a) $E(X) = np$.

- b) $\text{Var}(X) = np(1 - p)$.
555. Sea X con distribución Poisson(λ). Hemos demostrado que la función característica de esta distribución es $\phi(t) = \exp[-\lambda(1 - e^{it})]$. Usando $\phi(t)$ compruebe que
- a) $E(X) = \lambda$.
- b) $\text{Var}(X) = \lambda$.
556. Sea X con distribución geo(p). Demuestre que
- a) $\phi(t) = p/(1 - (1 - p)e^{it})$.
- b) $E(X) = (1 - p)/p$, usando $\phi(t)$.
- c) $\text{Var}(X) = (1 - p)/p^2$, usando $\phi(t)$.
557. Sea X tiene distribución bin neg(r, p). Demuestre que
- a) $\phi(t) = [p/(1 - (1 - p)e^{it})]^r$.
- b) $E(X) = r(1 - p)/p$, usando $\phi(t)$.
- c) $\text{Var}(X) = r(1 - p)/p^2$, usando $\phi(t)$.
558. Sea X con distribución unif($-a, a$). Demuestre que $\phi(t) = \frac{\text{sen } at}{at}$.
559. Sea X con distribución unif(a, b). Demuestre que
- a) $\phi(t) = [e^{ibt} - e^{iat}]/[it(b - a)]$.
- b) $E(X) = (a + b)/2$, usando $\phi(t)$.
- c) $\text{Var}(X) = (b - a)^2/12$, usando $\phi(t)$.
560. Sea X con distribución $N(\mu, \sigma^2)$. Hemos demostrado que la función característica de esta distribución es $\phi(t) = \exp(i\mu t - \sigma^2 t^2/2)$. Usando $\phi(t)$ demuestre ahora que
- a) $E(X) = \mu$.
- b) $\text{Var}(X) = \sigma^2$.
561. Sea X con distribución exp(λ). Demuestre que
- a) $\phi(t) = \lambda/(\lambda - it)$.
- b) $E(X) = 1/\lambda$, usando $\phi(t)$.
- c) $\text{Var}(X) = 1/\lambda^2$, usando $\phi(t)$.
562. Sea X con distribución gama(n, λ). Hemos encontrado que la función característica de esta distribución es $\phi(t) = [\lambda/(\lambda - it)]^n$. Usando $\phi(t)$ compruebe nuevamente que
- a) $E(X) = n/\lambda$.
- b) $\text{Var}(X) = n/\lambda^2$.

563. Sean X y Y independientes ambas con distribución $\exp(\lambda)$. Use la f.c. para demostrar que $X + Y$ tiene distribución $\text{gama}(2, \lambda)$.
564. Sean X y Y independientes con distribución $\text{gama}(n, \lambda)$ y $\text{gama}(m, \lambda)$ respectivamente. Use la f.c. para demostrar que $X + Y$ tiene distribución $\text{gama}(n + m, \lambda)$.
565. Se sabe que si X y Y son independientes, entonces $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$. Sin embargo la condición $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$ no es suficiente para concluir que X y Y son independientes. Demuestre esta afirmación considerando la siguiente distribución conjunta:

$$f(x, y) = \frac{1}{4}[1 + xy(x^2 - y^2)] \quad \text{para} \quad -1 < x, y < 1.$$

566. Sea X con función de distribución

$$F(x) = \frac{e^x}{1 + e^x}.$$

Demuestre que $F(x)$ es efectivamente una función de distribución, y calcule $\phi(t)$. Con ayuda de ésta encuentre $E(X)$ y $\text{Var}(X)$.

567. Sean X y Y independientes. Demuestre que

$$\phi_{XY}(t) = \int_{-\infty}^{\infty} \phi_Y(tx) dF_X(x) = \int_{-\infty}^{\infty} \phi_X(ty) dF_Y(y).$$

568. Mediante el *cálculo de residuos* de la teoría de variable compleja puede demostrarse que la distribución Cauchy estándar tiene función característica

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\pi(1+x^2)} dx = e^{-|t|}.$$

Suponiendo este resultado, encuentre el error en el siguiente argumento para encontrar la f.g.m. de la distribución Cauchy: “Como $\phi(t) = e^{-|t|}$ y $M(t) = \phi(-it)$, entonces $M(t) = e^{-|-it|} = e^{-|t|}$.”

569. Sean X_1, \dots, X_n v.a.i.i.d. con distribución Cauchy estándar, es decir, la función característica de cada una de estas variables es

$$\phi(t) = e^{-|t|}.$$

Use este resultado para demostrar que la v.a. $S_n = (X_1 + \dots + X_n)/n$ tiene distribución Cauchy estándar para cualquier valor de n .

Capítulo 9

Teoremas límite

En este último capítulo se estudian dos de los teoremas más importantes en probabilidad: la ley de los grandes números y el teorema central del límite. Antes de ello se revisan dos desigualdades de interés general.

9.1. Desigualdad de Markov

Proposición (Desigualdad de Markov). Sea $X \geq 0$ con esperanza finita. Para cualquier $\epsilon > 0$,

$$P(X > \epsilon) \leq \frac{E(X)}{\epsilon}.$$

Demostración.

$$\begin{aligned} E(X) &= E(X \cdot 1_{(X>\epsilon)} + X \cdot 1_{(X\leq\epsilon)}) \\ &\geq E(X \cdot 1_{(X>\epsilon)}) \\ &\geq E(\epsilon \cdot 1_{(X>\epsilon)}) \\ &= \epsilon P(X > \epsilon). \end{aligned}$$

□

En palabras, este resultado establece que la probabilidad de que X exceda un valor ϵ positivo está acotada superiormente por la media entre ϵ . Existen otras versiones equivalentes de esta desigualdad. Por ejemplo, la desigualdad de Markov aplicada a la variable aleatoria no negativa $|X|$ establece que $P(|X| > \epsilon) \leq E|X|/\epsilon$, y para $|X|^n$ con n cualquier número natural, $P(|X| > \epsilon) \leq E|X|^n/\epsilon^n$.

9.2. Desigualdad de Chebyshev

La desigualdad de Chebyshev es un resultado de bastante utilidad en algunas situaciones, en particular se usará en la siguiente sección para demostrar la ley débil de los grandes números.

Proposición (Desigualdad de Chebyshev). Sea X con media μ y varianza $\sigma^2 < \infty$. Para cualquier $\epsilon > 0$,

$$P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}. \quad (9.1)$$

Demostración.

$$\begin{aligned} \sigma^2 &= E[(X - \mu)^2] \\ &= E[(X - \mu)^2 \cdot 1_{(|X - \mu| > \epsilon)} + (X - \mu)^2 \cdot 1_{(|X - \mu| \leq \epsilon)}] \\ &\geq E[(X - \mu)^2 \cdot 1_{(|X - \mu| > \epsilon)}] \\ &\geq E[\epsilon^2 \cdot 1_{(|X - \mu| > \epsilon)}] \\ &= \epsilon^2 P(|X - \mu| > \epsilon). \end{aligned}$$

□

En palabras, la desigualdad dice que la probabilidad de que X difiera de su media en más de ϵ está acotada superiormente por la varianza entre ϵ^2 . A este resultado se le conoce también con el nombre de *desigualdad de Chebyshev-Bienaymé*. Existen otras versiones de esta desigualdad equivalentes a la demostrada, por ejemplo,

- a) $P(|X - \mu| > \epsilon\sigma) \leq 1/\epsilon^2$.
- b) $P(|X - \mu| \leq \epsilon\sigma) \geq 1 - 1/\epsilon^2$.
- c) $P(|X - \mu| \leq \epsilon) \geq 1 - \sigma^2/\epsilon^2$.

Proposición (Desigualdad de Chebyshev extendida) Sea X una variable aleatoria, y sea $g \geq 0$ una función no decreciente tal que $g(X)$ es una variable aleatoria con esperanza finita. Para cualquier $\epsilon > 0$,

$$P(X > \epsilon) \leq \frac{E[g(X)]}{g(\epsilon)}. \quad (9.2)$$

Demostración.

$$\begin{aligned} E[g(X)] &= E[g(X) \cdot 1_{(X>\epsilon)} + g(X) \cdot 1_{(X\leq\epsilon)}] \\ &\geq E[g(X) \cdot 1_{(X>\epsilon)}] \\ &\geq E[g(\epsilon) \cdot 1_{(X>\epsilon)}] \\ &= g(\epsilon)P(X > \epsilon). \end{aligned}$$

□



Pafnuty Lvovich Chebyshev
(Rusia, 1821–1894)



Andrei Andreyevich Markov
(Rusia, 1856–1922)

Profesor y alumno.

Fuente: Archivo MacTutor, Universidad de St. Andrews.

A partir de la desigualdad de Chebyshev extendida y con una función g adecuada se pueden obtener tanto la desigualdad de Chebyshev como la desigualdad de Markov. En resumen se tiene la siguiente tabla:

| Versiones de las desigualdades de Markov y de Chebyshev | |
|---|---|
| Markov: | Para $\epsilon > 0$, a) $X \geq 0 \Rightarrow P(X > \epsilon) \leq E(X)/\epsilon$. b) $P(X > \epsilon) \leq E X /\epsilon$. c) $P(X > \epsilon) \leq E X ^n/\epsilon^n$. |
| Chebyshev: | Para $\epsilon > 0$, a) $P(X - \mu > \epsilon) \leq \text{Var}(X)/\epsilon^2$. b) $P(X > \epsilon) \leq E[g(X)]/g(\epsilon)$, con $g \geq 0$ no decreciente. |

9.3. Ley de los grandes números

En esta sección se estudia uno de los teoremas más importantes de la teoría clásica de la probabilidad. Este interesante resultado se conoce como la ley de los grandes números y establece que, bajo ciertas condiciones, el promedio de variables aleatorias converge a una constante cuando el número de sumandos crece a infinito. Más precisamente el resultado es el siguiente.

Teorema (Ley débil de los grandes números). Sean X_1, X_2, \dots independientes tales que $E(X_i) = \mu$. Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

Demostración. (Suponiendo segundo momento finito.) Sea $S_n = (X_1 + \dots + X_n)/n$. Entonces $E(S_n) = \mu$ y $\text{Var}(S_n) \leq \sigma^2/n$, suponiendo $\text{Var}(X_i) \leq \sigma^2 < \infty$. La desigualdad de Chebyshev aplicada a S_n asegura que para cualquier $\epsilon > 0$ se cumple $P(|S_n - \mu| \geq \epsilon) \leq \sigma^2/n\epsilon^2$. Basta ahora tomar el límite cuando n tiende a infinito para obtener el resultado requerido. \square

La ley débil de los grandes números establece entonces que la variable aleatoria $S_n = (X_1 + \dots + X_n)/n$ converge en probabilidad a la media común μ . Observe que para la demostración de este resultado no hemos supuesto idéntica distribución para las variables aleatorias involucradas, únicamente que tengan la misma media, que sean independientes y aunque las varianzas pueden ser diferentes, se ha necesitado que sean uniformemente acotadas. Damos a continuación un ejemplo sencillo de aplicación de este resultado y más adelante demostraremos una versión más fuerte.

Ejemplo (Definición de probabilidad frecuentista). Considere un experimento aleatorio cualquiera y sea A un evento. Se efectúan realizaciones independientes del experimento, y se observa en cada ensayo la ocurrencia o no ocurrencia del evento A . Sea X_k la variable que toma el valor uno si en el k -ésimo ensayo se observa A , y cero en caso contrario. Entonces X_1, X_2, \dots son variables aleatorias independientes con distribución $\text{Ber}(p)$, en donde p es la probabilidad desconocida del evento A . Por lo tanto $E(X_k) = p$ y $\text{Var}(X_k) = p(1 - p)$. La ley débil de los grandes números asegura que la fracción de ensayos en los que se observa el evento A converge, en probabilidad, a la constante desconocida p cuando el número de ensayos crece a infinito. Esta es la definición frecuentista de la probabilidad, y hemos entonces corroborado su validez con ayuda de la ley de los grandes números. \circ

La siguiente versión de la ley de los grandes números asegura que bajo ciertas condiciones la convergencia de la variable aleatoria $(X_1 + \dots + X_n)/n$ a la media común

μ es más fuerte, es casi segura.

Teorema (Ley fuerte de los grandes números). Sean X_1, X_2, \dots independientes e idénticamente distribuidas, tales que $E(X_i) = \mu$. Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{c.s.} \mu.$$

Demostración. (Suponiendo cuarto momento finito.) Dada la idéntica distribución de los elementos de la sucesión, cualquier elemento de ésta se denota simplemente por X . Suponga que $E|X - \mu|^2 = \sigma^2$ y observe que $E(X - \mu) = 0$. Entonces por independencia,

$$E \left| \sum_{i=1}^n (X_i - \mu) \right|^4 = nE|X - \mu|^4 + 3n(n-1)\sigma^4.$$

Por la desigualdad de Chebyshev (9.2) aplicada a la variable $|\sum_{i=1}^n (X_i - \mu)|$ y la función $g(x) = x^4$ se obtiene, para $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\sum_{i=1}^n (X_i - \mu)\right| > n\epsilon\right) &\leq \frac{E\left|\sum_{i=1}^n (X_i - \mu)\right|^4}{(n\epsilon)^4} \\ &= \frac{nE|X - \mu|^4 + 3n(n-1)\sigma^4}{(n\epsilon)^4}. \end{aligned}$$

Sea el evento $A_n = (|\frac{1}{n} \sum_{i=1}^n X_i - \mu| > \epsilon)$. Entonces $\sum_{n=1}^{\infty} P(A_n) < \infty$. Por el lema de Borel-Cantelli la probabilidad de que ocurra una infinidad de eventos A_n es cero, es decir, con probabilidad uno, sólo un número finito de estos eventos ocurre. Por lo tanto con probabilidad uno, existe un número natural n a partir del cual ningún evento A_n se verifica. Es decir,

$$P\left(\lim_{n \rightarrow \infty} \left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \leq \epsilon\right) = 1.$$

Como esta afirmación vale para cualquier $\epsilon > 0$, se cumple que

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mu\right) = 1.$$

□

Ejemplo (El problema del mono, nuevamente). Usaremos la ley fuerte de los grandes números para dar otra solución al problema del mono. Considere entonces un mono que escribe caracteres al azar. Nos interesa encontrar la probabilidad de que el mono eventualmente escriba las obras completas de Shakespeare, las cuales,

supondremos, tienen una longitud total de N caracteres. Nuevamente se consideran bloques de longitud N de la siguiente forma

$$\underbrace{x_1, \dots, x_N}_{\text{bloqueo 1}}, \underbrace{x_{N+1}, \dots, x_{2N}}_{\text{bloqueo 2}}, \dots$$

Sea A_k el evento correspondiente a que en el k -ésimo bloque el mono ha tenido éxito, y sea X_k la variable aleatoria indicadora del evento A_k , es decir,

$$X_k = \begin{cases} 1 & \text{si } A_k \text{ ocurre,} \\ 0 & \text{si } A_k \text{ no ocurre.} \end{cases}$$

Se tiene entonces una sucesión de variables aleatorias X_1, X_2, \dots independientes e idénticamente distribuidas Bernoulli(p), con $p = P(A_k) = (1/m)^N$, suponiendo que el total de caracteres disponibles es m . En particular, la media de cada una de estas variables es $E(X_k) = p$. Considere ahora la suma $X_1 + X_2 + \dots + X_n$. Si para algún valor de n esta suma es positiva, significa que alguno de los sumandos es distinto de cero, y por lo tanto que el mono ha tenido éxito. Pero esto es justamente lo que garantiza la ley fuerte de los grandes números pues

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = p\right) = 1.$$

Es decir, con probabilidad uno la suma de esta ecuación es positiva. Esto implica que debe existir un valor de k tal que $X_k = 1$, y esto a su vez significa que en el k -ésimo bloque ¡el mono ha tenido éxito! Más aún, para que el promedio que aparece en esta ecuación sea positivo necesariamente la suma debe ser infinita, y por lo tanto, deben existir una infinidad de valores de k tal que $X_k = 1$. Esto quiere decir que con probabilidad uno ¡el mono escribirá una infinidad de veces las obras completas de Shakespeare! \circ

9.4. Teorema central del límite

Concluimos el curso con el célebre y famoso teorema central del límite. Este resultado es de amplio uso en estadística y otras ramas de aplicación de la probabilidad, y una de sus primeras versiones tiene el nombre de teorema de De Moivre-Laplace, que se enuncia a continuación.

Teorema de De Moivre-Laplace. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes tal que cada una de ellas tiene distribución Bernoulli con parámetro $p \in (0, 1)$. Entonces para cualesquiera números reales $a < b$,

$$\lim_{n \rightarrow \infty} P\left(a < \frac{(X_1 + \dots + X_n) - np}{\sqrt{np(1-p)}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

En palabras este resultado establece que la variable aleatoria $((X_1 + \dots + X_n) - np)/\sqrt{np(1-p)}$ converge en distribución a una variable aleatoria normal estándar. Una demostración directa del teorema de De Moivre-Laplace puede ser encontrada en [5]. Este teorema fue descubierto por De Moivre alrededor de 1733 en el caso cuando las variables aleatorias tienen distribución Bernoulli con $p = 1/2$. Años después Laplace demostró el mismo resultado para valores arbitrarios de p . El teorema de De Moivre-Laplace es un caso particular del siguiente resultado fundamental.

Teorema central del límite. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que para cada natural n , $E(X_n) = \mu$ y $\text{Var}(X_n) = \sigma^2 < \infty$. Entonces

$$\frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0, 1).$$

Este resultado fue demostrado rigurosamente por Lyapunov alrededor de 1901, aunque Laplace había dado antes argumentos intuitivos que mostraban su validez. Observe que no hay ninguna hipótesis adicional sobre la distribución de las variables de la sucesión, es decir, éstas pueden tener cualquier distribución, sólo requiriendo la existencia de la media y la varianza. El resultado es válido como está enunciado, sin embargo, a fin de dar una demostración simple, supondremos adicionalmente que los elementos de la sucesión tienen momentos finitos de cualquier orden. La demostración que se presenta hace uso de esta hipótesis y de la función característica.

Demostración. (Suponiendo todos los momentos finitos.) Observe que

$$\frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n}\sigma} = \frac{(X_1 - \mu)/\sigma + \dots + (X_n - \mu)/\sigma}{\sqrt{n}},$$

en donde cada sumando del numerador en el lado derecho es una variable con media cero y varianza uno. Así pues, sin pérdida de generalidad, supondremos que cada variable de la sucesión tiene media cero y varianza uno, y consideraremos la suma

$$Z_n = \frac{X_1 + \dots + X_n}{\sqrt{n}}.$$

Se desea probar que $Z_n \xrightarrow{d} N(0, 1)$. Para ello es suficiente demostrar que $\phi_{Z_n}(t) \rightarrow e^{-t^2/2}$. Por independencia e idéntica distribución,

$$\phi_{Z_n}(t) = E \left[e^{it(X_1 + \dots + X_n)/\sqrt{n}} \right] = \left[\phi_X(t/\sqrt{n}) \right]^n.$$

Por lo tanto,

$$\ln \phi_{Z_n}(t) = n \ln \left(1 + \frac{it}{\sqrt{n}} E(X) + \frac{i^2 t^2}{2!n} E(X^2) + \frac{i^3 t^3}{3!n^{3/2}} E(X^3) + \dots \right).$$

Usando la fórmula: $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots$, y factorizando potencias de it se obtiene

$$\begin{aligned} \ln \phi_{Z_n}(t) &= [E(X^2) - E^2(X)] i^2 t^2 / 2 \\ &+ \left[\frac{E(X^3)}{3! \sqrt{n}} - \frac{E(X)E(X^2)}{2\sqrt{n}} + \frac{E^3(X)}{3\sqrt{n}} \right] \frac{i^3 t^3}{\sqrt{n}} + \dots \end{aligned}$$

El primer sumando es $-t^2/2$, y todos los términos a partir del segundo sumando se anulan cuando n tiende a infinito. Por lo tanto,

$$\lim_{n \rightarrow \infty} \ln \phi_{Z_n}(t) = -t^2/2.$$

Como la función logaritmo es una función continua se tiene que

$$\ln \left(\lim_{n \rightarrow \infty} \phi_{Z_n}(t) \right) = -t^2/2.$$

De donde se obtiene

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(t) = e^{-t^2/2}.$$

□

El teorema central del límite establece entonces que para cualquier x en \mathbb{R} ,

$$\lim_{n \rightarrow \infty} P \left[\frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n}\sigma} \leq x \right] = P(Z \leq x),$$

en donde Z tiene distribución normal estándar. Observe que la suma $X_1 + \dots + X_n$ tiene media $n\mu$ y varianza $n\sigma^2$, de modo que la expresión de arriba es una especie de estandarización de esta variable. Equivalentemente este resultado puede enunciarse del siguiente modo:

$$\frac{\frac{1}{n}(X_1 + \dots + X_n) - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1).$$

9.5. Ejercicios

Desigualdad de Markov

570. Demuestre la desigualdad de Markov siguiendo los siguientes pasos: Suponga $X \geq 0$, y para $\epsilon > 0$ defina

$$X_\epsilon = \begin{cases} \epsilon & \text{si } X > \epsilon, \\ 0 & \text{si } X \leq \epsilon. \end{cases}$$

Compruebe que $X_\epsilon \leq X$. Ahora tome esperanza de ambos lados y calcule $E(X_\epsilon)$.

571. Demuestre directamente las siguientes versiones de la desigualdad de Markov. Para cualquier $\epsilon > 0$,

- a) $P(|X| > \epsilon) \leq \frac{E|X|}{\epsilon}$.
- b) $P(|X| > \epsilon) \leq \frac{E|X|^n}{\epsilon^n}$, con n cualquier número natural.

572. Use la desigualdad de Markov para demostrar que si X es una variable aleatoria no negativa con esperanza cero, entonces $X = 0$ casi seguramente.
573. Demuestre que la convergencia en media implica la convergencia en probabilidad usando la desigualdad de Markov aplicada a la variable aleatoria no negativa $|X_n - X|$.

Desigualdad de Chebyshev

574. Use la desigualdad de Chebyshev (9.2) para demostrar directamente que la convergencia en media cuadrática implica la convergencia en probabilidad.
575. Demuestre la desigualdad de Chebyshev (9.1) usando la desigualdad de Markov aplicada a la variable aleatoria no negativa $|X - \mu|$.
576. Use la desigualdad de Chebyshev para demostrar que si X es una variable aleatoria tal que $E(X) = a$ y $\text{Var}(X) = 0$, entonces X es constante casi seguramente, es decir, $P(X = a) = 1$.
577. Sea X con media μ y varianza σ^2 . Use la desigualdad de Chebyshev para estimar la probabilidad de que X tome valores entre $\mu - \epsilon\sigma$ y $\mu + \epsilon\sigma$ para cualquier $\epsilon > 0$ constante.
578. A partir de la desigualdad de Chebyshev extendida (9.2) demuestre la desigualdad de Chebyshev (9.1) y la desigualdad de Markov.
579. Demuestre que $P(|X| > \epsilon) \leq \frac{E|X|}{\epsilon}$, para $\epsilon > 0$,
- a) usando la desigualdad de Chebyshev extendida.
- b) de manera directa.
580. Demuestre que $P(|X| > \epsilon) \leq \frac{E|X|^n}{\epsilon^n}$, para $\epsilon > 0$ y $n \in \mathbb{N}$,
- a) usando la desigualdad de Chebyshev extendida.
- b) de manera directa.
581. Demuestre que $P(X > \epsilon) \leq \frac{E(e^{tX})}{e^{t\epsilon}}$ para $\epsilon > 0$ y $t > 0$,
- a) usando la desigualdad de Chebyshev extendida.
- b) de manera directa.
582. Sea X con función de densidad

$$f(x) = \begin{cases} 1/18 & \text{si } x = -1, 1, \\ 16/18 & \text{si } x = 0, \\ 0 & \text{otro caso.} \end{cases}$$

Demuestre que $P(|X - \mu| > 3\sigma)$ y la estimación dada por la desigualdad de Chebyshev para esta probabilidad coinciden. Este ejercicio demuestra que en general la cota superior dada por la desigualdad de Chebyshev es óptima, es decir, no puede establecerse una cota superior más pequeña.

583. Considere la siguiente versión de la desigualdad de Chebyshev

$$P(|X - \mu| \leq \epsilon\sigma) \geq 1 - 1/\epsilon^2.$$

Encuentre el mínimo valor de $\epsilon > 0$ de tal modo que la probabilidad de que una variable aleatoria tome valores entre $\mu - \epsilon\sigma$ y $\mu + \epsilon\sigma$ sea al menos 0.90.

584. *Desigualdad de Cantelli.* Demuestre que si $\text{Var}(X) < \infty$, entonces para cualquier $\epsilon > 0$,

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

Ley de los grandes números

585. Use la ley débil de los grandes números para demostrar que si X_n tiene distribución $\text{bin}(n, p)$, entonces cuando $n \rightarrow \infty$,

$$\frac{1}{n}X_n \xrightarrow{p} p.$$

586. *Ley de los grandes números en media cuadrática.* Demuestre que si X_1, X_2, \dots es una sucesión de v.a.s independientes con media μ y varianza σ^2 , entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{m.c.} \mu.$$

Observe que no se pide la hipótesis de idéntica distribución para las variables aleatorias y que este resultado no es consecuencia de la ley fuerte.

587. Sean X_1, \dots, X_n independientes con distribución $N(\mu, \sigma^2)$. Para cualquier valor de n el promedio $(X_1 + \dots + X_n)/n$ tiene distribución $N(\mu, \sigma^2/n)$. ¿Contradice esto la ley de los grandes números?
588. En el Ejercicio 569 se pide usar la f.c. para demostrar que si X_1, \dots, X_n son v.a.i.i.d. con distribución Cauchy estándar, entonces el promedio $S_n = (X_1 + \dots + X_n)/n$ tiene distribución Cauchy estándar, independientemente del valor de n . ¿Contradice esto la ley de los grandes números?
589. Se lanza una moneda equilibrada $2n$ veces. Calcule la probabilidad de que ambas caras caigan el mismo número de veces. ¿Qué le sucede a esta probabilidad cuando n tiende a infinito?. ¿Contradice esto la ley de los grandes números?.

Teorema central del límite

590. Use el teorema central del límite para estimar la probabilidad de obtener mas de 520 águilas en 1000 lanzamientos de una moneda honesta.
591. Sea X_1, X_2, \dots una sucesión de v.a.i.i.d. con distribución Poisson(λ) con $\lambda = 1$. Use el teorema central del límite para demostrar que

$$\lim_{n \rightarrow \infty} \frac{1}{e^n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}.$$

592. La probabilidad de ocurrencia de un evento en un ensayo es de 0.3. ¿Cuál es la probabilidad de que la frecuencia relativa de este evento en 100 ensayos se encuentre entre 0.2 y 0.5?

Apéndice A

Distribuciones de probabilidad

Se presenta a continuación una lista con algunas distribuciones de probabilidad univariadas de uso común. Como es costumbre, la función de probabilidad o de densidad se denota por $f(x)$, y la función de distribución por $F(x)$. Además, se recuerda la notación para las siguientes funciones asociadas a una distribución.

$G(t)$: Función generadora de probabilidad.

$M(t)$: Función generadora de momentos.

$\phi(t)$: Función característica.

DISTRIBUCIONES UNIVARIADAS DISCRETAS

Distribución uniforme

$X \sim \text{unif}\{x_1, \dots, x_n\}$ con $n \in \mathbb{N}$.

$f(x) = 1/n$ para $x = x_1, \dots, x_n$.

$E(X) = \frac{1}{n}(x_1 + \dots + x_n)$.

$\text{Var}(X) = \frac{1}{n}[(x_1 - \mu)^2 + \dots + (x_n - \mu)^2]$.

$G(t) = \frac{1}{n}(t^{x_1} + \dots + t^{x_n})$.

$M(t) = \frac{1}{n}(e^{x_1 t} + \dots + e^{x_n t})$.

Distribución Bernoulli

$X \sim \text{Ber}(p)$ con $p \in (0, 1)$.

$f(x) = p^x(1-p)^{1-x}$ para $x = 0, 1$.

$E(X) = p$.

$$\begin{aligned}\text{Var}(X) &= p(1-p). \\ G(t) &= 1-p+pt. \\ M(t) &= 1-p+pe^t.\end{aligned}$$

Distribución binomial

$$\begin{aligned}X &\sim \text{bin}(n, p) \text{ con } n \in \mathbb{N} \text{ y } p \in (0, 1). \\ f(x) &= \binom{n}{x} p^x (1-p)^{n-x} \text{ para } x = 0, 1, \dots, n. \\ E(X) &= np. \\ \text{Var}(X) &= np(1-p). \\ G(t) &= (1-p+pt)^n. \\ M(t) &= [1-p+pe^t]^n.\end{aligned}$$

Distribución geométrica

$$\begin{aligned}X &\sim \text{geo}(p), \text{ con } p \in (0, 1) \text{ y } q = 1-p. \\ f(x) &= p(1-p)^x \text{ para } x = 0, 1, \dots \\ E(X) &= q/p. \\ \text{Var}(X) &= q/p^2. \\ G(t) &= p/[1-t(1-p)]. \\ M(t) &= p/[1-(1-p)e^t].\end{aligned}$$

Distribución Poisson

$$\begin{aligned}X &\sim \text{Poisson}(\lambda), \text{ con } \lambda > 0. \\ f(x) &= e^{-\lambda} \frac{\lambda^x}{x!} \text{ para } x = 0, 1, \dots \\ E(X) &= \lambda. \\ \text{Var}(X) &= \lambda. \\ G(t) &= e^{-\lambda(1-t)}. \\ M(t) &= \exp[\lambda(e^t - 1)].\end{aligned}$$

Distribución binomial negativa

$X \sim \text{bin neg}(r, p)$ con $r \in \mathbb{N}$ y $p \in (0, 1)$.

$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x \quad \text{para } x = 0, 1, \dots$$

$$E(X) = r(1-p)/p.$$

$$\text{Var}(X) = r(1-p)/p^2.$$

$$G(t) = [p/(1-t(1-p))]^r.$$

$$M(t) = [p/(1-qt)]^r.$$

Distribución hipergeométrica

$X \sim \text{hipergeo}(N, K, n)$ con $N, K, n \in \mathbb{N}$ y $n \leq K \leq N$.

$$f(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} \quad \text{para } x = 0, 1, \dots, n.$$

$$E(X) = nK/N.$$

$$\text{Var}(X) = n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}.$$

DISTRIBUCIONES UNIVARIADAS CONTINUAS

Distribución uniforme continua

$X \sim \text{unif}(a, b)$ con $a < b$.

$$f(x) = 1/(b-a) \quad \text{para } x \in (a, b).$$

$$F(x) = (x-a)/(b-a) \quad \text{para } x \in (a, b).$$

$$E(X) = (a+b)/2.$$

$$\text{Var}(X) = (b-a)^2/12.$$

$$M(t) = (e^{bt} - e^{at})/(bt - at).$$

Distribución exponencial

$X \sim \text{exp}(\lambda)$ con $\lambda > 0$.

$$f(x) = \lambda e^{-\lambda x} \quad \text{para } x > 0.$$

$$F(x) = 1 - e^{-\lambda x} \quad \text{para } x > 0.$$

$$E(X) = 1/\lambda.$$

$$\text{Var}(X) = 1/\lambda^2.$$

$$M(t) = \lambda/(\lambda - t) \quad \text{para } t < \lambda.$$

Distribución gama

$$X \sim \text{gama}(n, \lambda) \quad \text{con } n > 0 \text{ y } \lambda > 0.$$

$$f(x) = \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} \quad \text{para } x > 0.$$

$$F(x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-1} (\lambda x)^k / k!, \quad \text{para } x > 0 \text{ y } n \text{ entero.}$$

$$E(X) = n/\lambda.$$

$$\text{Var}(X) = n/\lambda^2.$$

$$M(t) = [\lambda/(\lambda - t)]^n \quad \text{para } t < \lambda.$$

Distribución beta

$$X \sim \text{beta}(a, b) \quad \text{con } a > 0, b > 0.$$

$$f(x) = x^{a-1}(1-x)^{b-1}/B(a, b), \quad \text{para } x \in (0, 1).$$

$$E(X) = a/(a+b).$$

$$\text{Var}(X) = ab/[(a+b+1)(a+b)^2].$$

Distribución normal

$$X \sim N(\mu, \sigma^2) \quad \text{con } \mu \in \mathbb{R} \text{ y } \sigma^2 > 0.$$

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

$$E(X) = \mu.$$

$$\text{Var}(X) = \sigma^2.$$

$$M(t) = \exp(\mu t + \sigma^2 t^2/2).$$

$$\phi(t) = \exp(i\mu t - \sigma^2 t^2/2).$$

Cuando $\mu = 0$ y $\sigma^2 = 1$ se obtiene la distribución normal estándar.

Distribución ji-cuadrada

$$X \sim \chi^2(n) \quad \text{con } n > 0.$$

$$f(x) = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} \quad \text{para } x > 0.$$

$$E(X) = n.$$

$$\text{Var}(X) = 2n.$$

$$M(t) = (1 - 2t)^{-n/2} \text{ para } t < 1/2.$$

Distribución t

$$X \sim t(n) \text{ con } n > 0.$$

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}.$$

$$E(X) = 0.$$

$$\text{Var}(X) = n/(n-2) \text{ para } n > 2.$$

$$M(t) \text{ no existe para } t \neq 0.$$

$$\phi(t) = \exp(-|t|), \text{ cuando } n = 1.$$

La expresión de $\phi(t)$ resulta complicada para valores $n \geq 2$.

Distribución log normal

$$X \sim \text{log normal}(\mu, \sigma^2) \text{ con } \mu \in \mathbb{R} \text{ y } \sigma^2 > 0.$$

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp[-(\ln x - \mu)^2/2\sigma^2] \text{ para } x > 0.$$

$$E(X) = \exp(\mu + \sigma^2/2).$$

$$E(X^n) = \exp(n\mu + n^2\sigma^2/2).$$

$$\text{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2).$$

Distribución Pareto

$$X \sim \text{Pareto}(a, b) \text{ con } a, b > 0.$$

$$f(x) = \frac{ab^a}{(b+x)^{a+1}} \text{ para } x > 0.$$

$$F(x) = 1 - [b/(b+x)]^a \text{ para } x > 0.$$

$$E(X) = b/(a-1) \text{ para } a > 1.$$

$$\text{Var}(X) = ab^2/[(a-1)^2(a-2)] \text{ para } a > 2.$$

Distribución Weibull

$$X \sim \text{Weibull}(r, \lambda) \text{ con } r, \lambda > 0.$$

$$f(x) = e^{-(\lambda x)^r} r\lambda^r x^{r-1} \text{ para } x > 0.$$

$$F(x) = 1 - e^{-(\lambda x)^r} \text{ para } x > 0.$$

$$E(X) = \Gamma(1 + 1/r)/\lambda.$$
$$\text{Var}(X) = [\Gamma(1 + 2/r) - \Gamma^2(1 + 1/r)]/\lambda^2.$$

Distribución Cauchy

$X \sim \text{Cauchy}(a, b)$ con $a > 0$ y $b > 0$.

$$f(x) = \frac{1}{b\pi[1 + ((x - a)/b)^2]}.$$

La esperanza y varianza no existen.

$M(t)$ no existe para $t \neq 0$.

$$\phi(t) = \exp(-iat - b|t|).$$

Cuando $a = 0$ y $b = 1$ se obtiene la distribución Cauchy estándar, y coincide con la distribución $t(n)$ con $n = 1$. En este caso,

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

$$F(x) = 1/2 + (\arctan x)/\pi.$$

Apéndice B

Formulario

B.1. El alfabeto griego

| | | | | | |
|----------------------------|---------|-------------------|---------|----------------------------|----------|
| A α | alfa | I ι | iota | P ρ, ϱ | rho |
| B β | beta | K κ | kapa | $\Sigma \sigma, \varsigma$ | sigma |
| $\Gamma \gamma$ | gama | $\Lambda \lambda$ | lambda | T τ | tau |
| $\Delta \delta$ | delta | M μ | mu | $\Upsilon \upsilon$ | upsilon |
| E ϵ, ε | épsilon | N ν | nu | $\Phi \phi, \varphi$ | phi |
| Z ζ | zeta | $\Xi \xi$ | xi | X χ | ji ó chi |
| H η | eta | O o | omikron | $\Psi \psi$ | psi |
| $\Theta \theta, \vartheta$ | theta | $\Pi \pi$ | pi | $\Omega \omega$ | omega |

B.2. Notación

$\mathcal{B}(\mathbb{R})$: Conjuntos de Borel de \mathbb{R} .

$a \vee b$: $\max\{a, b\}$.

$a \wedge b$: $\min\{a, b\}$.

$A \perp B$: El evento A es independiente del evento B .

$[x]$: Parte entera de x .

$F(x+)$: Límite por la derecha de la función F en el punto x .

$F(x-)$: Límite por la izquierda de la función F en el punto x .

B.3. Límite superior e inferior

Sea $\{a_n : n \geq 1\}$ una sucesión de números reales. Para cada m natural defina

$$b_m = \inf \{a_m, a_{m+1}, \dots\},$$

$$y \quad c_m = \sup \{a_m, a_{m+1}, \dots\}.$$

Entonces claramente $b_m \leq b_{m+1}$, y $c_m \geq c_{m+1}$. Es decir, ambas sucesiones son monótonas, una no decreciente y la otra no creciente. Por lo tanto estas sucesiones son convergentes, no excluyendo con ello el valor infinito. Entonces sean

$$b = \lim_{m \rightarrow \infty} b_m,$$

$$y \quad c = \lim_{m \rightarrow \infty} c_m.$$

Al valor b se le llama *límite inferior*, y a c se le llama *límite superior* de la sucesión, y se denotan de la forma siguiente:

$$b = \liminf_{n \rightarrow \infty} a_n,$$

$$y \quad c = \limsup_{n \rightarrow \infty} a_n.$$

No es difícil demostrar que

$$\liminf_{n \rightarrow \infty} a_n \leq \limsup_{n \rightarrow \infty} a_n,$$

y que la sucesión es convergente si, y sólo si, $\liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n$.

B.4. Imagen inversa

Sean \mathbb{A} y \mathbb{B} dos conjuntos. Considere una función $X : \mathbb{A} \rightarrow \mathbb{B}$. La imagen inversa de $B \subseteq \mathbb{B}$ es un subconjunto de \mathbb{A} , denotado por $X^{-1}B$, y definido como sigue:

$$X^{-1}B = \{a \in \mathbb{A} : X(a) \in B\}.$$

Observe que X es una función puntual, es decir, lleva puntos de \mathbb{A} en puntos de \mathbb{B} , mientras que X^{-1} es una función conjuntista, es decir, lleva subconjuntos de \mathbb{B} en subconjuntos de \mathbb{A} . No es difícil verificar que la imagen inversa cumple las siguientes propiedades.

1. $X^{-1}\mathbb{B} = \mathbb{A}$.
2. $X^{-1}(B^c) = (X^{-1}B)^c$.
3. Si $B_1 \subseteq B_2$, entonces $X^{-1}B_1 \subseteq X^{-1}B_2$.
4. $X^{-1}(B_2 - B_1) = X^{-1}B_2 - X^{-1}B_1$.
5. $X^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right) = \bigcup_{k=1}^{\infty} X^{-1}B_k$.
6. $X^{-1}\left(\bigcap_{k=1}^n B_k\right) = \bigcap_{k=1}^n X^{-1}B_k$.

7. $X(X^{-1}B) \subseteq B$ [igualdad si, y sólo si, X es sobre].
8. $X^{-1}(XA) \subseteq A$ [igualdad si, y sólo si, X es inyectiva].

Si se tienen dos funciones $X : \mathbb{A} \rightarrow \mathbb{B}$ y $Y : \mathbb{B} \rightarrow \mathbb{C}$ entonces para cualquier C en \mathbb{C} , se cumple $(X \circ Y)^{-1}C = X^{-1}(Y^{-1}C)$.

B.5. Función indicadora

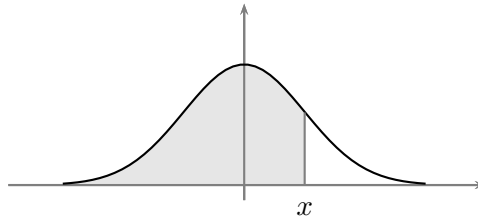
La *función indicadora* de un conjunto $A \subseteq \Omega$ es la función $1_A : \Omega \rightarrow \{0, 1\}$ dada por

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

De este modo la función 1_A toma el valor uno dentro del conjunto A , y cero fuera de él. Es sencillo verificar que esta función resulta ser una variable aleatoria cuando el conjunto A es un evento. La función indicadora cumple las siguientes propiedades.

- a) $1_{A \cup B} = \max\{1_A, 1_B\} = 1_A + 1_B - 1_A \cdot 1_B$.
- b) $1_{A \cap B} = \min\{1_A, 1_B\} = 1_A \cdot 1_B$.
- c) $1_{A^c} = 1 - 1_A$.
- d) $1_{A-B} = 1_A - 1_A \cdot 1_B$.
- e) $1_{A \Delta B} = |1_A - 1_B| = |1_A - 1_B|^2 = 1_A + 1_B - 2 \cdot 1_A \cdot 1_B$.
- f) $A \subseteq B \Rightarrow 1_A \leq 1_B$.

B.6. Tabla de la distribución normal estándar



$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

| x | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8399 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |

Bibliografía

- [1] Barry R. J. (1981) *Probabilidade: Um Curso em Nível Intermediário*. IMPA-CNPq.
- [2] Blake I.F. (1979) *An Introduction to Applied Probability*. Wiley.
- [3] Cohn D.L. (1980) *Measure Theory*. Birkhäuser.
- [4] Feller W. (1978) *Introducción a la Teoría de las Probabilidades y sus Aplicaciones*. Vol. I y II. Limusa.
- [5] García Álvarez M. A. (2005) *Introducción a la Teoría de la Probabilidad. Primer Curso*. Fondo de Cultura Económica.
- [6] García Álvarez M. A. (2005) *Introducción a la Teoría de la Probabilidad. Segundo Curso*. Fondo de Cultura Económica.
- [7] Grimmett G.R., Stirzaker D.R. (1982) *Probability and Random Processes*. Clarendon Press.
- [8] Grimmett G.R., Stirzaker D.R. (1986) *Probability: an Introduction*. Oxford University Press.
- [9] Grimmett G.R., Stirzaker D.R. (2001) *One Thousand Exercises in Probability*. Oxford University Press.
- [10] Gut A. (2005) *Probability: a Graduate Course*. Springer.
- [11] Halmos P.R. (1960) *Measure Theory*. Van Nostrand.
- [12] Harris B. (1966) *Theory of Probability*. Addison–Wesley.
- [13] Hoel P., Port S., Stone C. (1971) *Probability Theory*. Houghton Mifflin Co.
- [14] Jacod D., Protter P. (2000) *Probability Essentials*. Springer–Verlag.
- [15] Karr A.F. (1993) *Probability*. Springer–Verlag.
- [16] Laha R. G., Rohatgi V. K. (1979) *Probability Theory*. John Wiley & Sons.
- [17] Miller I., Miller M. (1999) *John E. Freund's Mathematical Statistics - 6th ed.* Prentice–Hall.

- [18] Mood A.M., Graybill F.A., Boes D.C. (1974) *Introduction to the Theory of Statistics*. McGraw Hill.
- [19] Parzen E. (1960) *Modern Probability Theory and its Applications*. Wiley.
- [20] Resnick S.I. (1999) *A Probability Path*. Birkhäuser.
- [21] Rincón L. (2004) ¿Qué es la esperanza condicional? *Miscelánea Matemática*. No. **37**, Agosto 2004, SMM.
- [22] Romano J.P. , Siegel A.F. (1986) *Counterexamples in Probability and Statistics*. Chapman & Hall.
- [23] Rosenthal J.S. (2000) *A First Look at Rigorous Probability Theory*. World Scientific.
- [24] Ross S. (1976) *A First Course in Probability*. Prentice Hall.
- [25] Villa Morales J. (2005) *Introducción a la Medida e Integración*. Textos Universitarios, Ciencias Básicas. Universidad Autónoma de Aguascalientes.
- [26] Von Plato J. (1994) *Creating Modern Probability*. Cambridge University Press.
- [27] Williams D. (1991) *Probability with Martingales*. Cambridge University Press.
- [28] Williams D. (2001) *Weighing the Odds: A Course in Probability and Statistics*. Cambridge University Press.

Índice

- σ -álgebra, 6
 - de Borel de \mathbb{R} , 12
 - de Borel de \mathbb{R}^n , 14
 - generada, 9
 - producto, 33
- Álgebra, 11
- Aditividad finita, 21
- Borel-Cantelli, 29
- Coefficiente
 - de correlación, 113
 - multinomial, 118
- Conjunto
 - Borel medible, 12
 - Boreliano, 12
 - de Borel, 12
 - medible, 6
- Continuidad de la prob, 23–25
- Convergencia
 - casi dondequiera, 183
 - casi segura, 183
 - casi siempre, 183
 - débil, 186
 - de eventos, 16
 - en distribución, 186
 - en media, 185
 - en media cuadrática, 186
 - en probabilidad, 185
 - puntual, 182
- Convolución, 150
- Covarianza, 112
- Desigualdad
 - C_r , 86
 - de Bonferroni, 38
 - de Boole, 21
 - de Cantelli, 230
 - de Cauchy-Schwarz, 85
 - de Chebyshev, 222
 - de Hölder, 86
 - de Jensen, 86
 - de Kounias, 38
 - de Markov, 221
 - de Minkowski, 87
- Desviación estándar, 60
- Distribución
 - arcoseno, 93
 - Bernoulli, 63, 232
 - beta, 70, 235
 - binomial, 64, 233
 - binomial negativa, 66, 234
 - bivariada, 97
 - Cauchy, 237
 - exponencial, 68, 234
 - exponencial doble, 91
 - F de Snedecor, 168
 - gama, 69, 235
 - geométrica, 65, 233
 - hipergeométrica, 67, 234
 - multivariada, 118
 - ji-cuadrada, 162, 235
 - log gama, 144
 - log normal, 72, 144, 236
 - multinomial, 117
 - multivariada, 97
 - normal, 71, 235
 - bivariada, 119
 - estándar, 71
 - Pareto, 236
 - Poisson, 66, 233
 - singular, 52
 - t de Student, 166, 236
 - uniforme
 - bivariada, 119
 - continua, 68, 234
 - discreta, 63, 232
 - univariada, 97
 - Weibull, 236

Error
 absoluto medio, 87
 cuadrático medio, 84
 Espacio
 de probabilidad, 5, 6
 medible, 6
 muestral, 5
 Esperanza
 condicional, 137
 de un vector, 116
 de una función de un vector, 110
 de una función de una v.a., 59
 de una v.a., 58
 Estadística, 161
 Estadísticas de orden, 170
 Evento, 5
 casi seguro, 55
 compuesto, 5
 simple, 5
 Función
 Borel medible, 43
 indicadora, 240
 medible, 76
 signo, 75
 Función característica, 205
 fórmula de inversión, 210, 212
 teorema de continuidad, 213
 teorema de unicidad, 212
 Función de densidad, 52
 condicional, 106
 marginal, 105
 Función de distribución, 47
 condicional, 106
 conjunta, 97
 marginal, 105
 Función de probabilidad
 conjunta, 100
 Función generadora
 de momentos, 201
 de momentos factoriales, 200
 de probabilidad, 197
 Igualdad
 casi segura, 55
 en distribución, 55
 Imagen inversa, 239
 Independencia
 de σ -álgebras, 28
 de eventos, 27
 de v.a.s, 106
 de vectores, 109
 Integral de Riemann-Stieltjes, 55
 Límite inferior
 de eventos, 15
 de números, 238
 Límite superior
 de eventos, 15
 de números, 238
 Ley de los grandes números, 224
 débil, 224
 en media cuadrática, 230
 fuerte, 225
 Matriz de covarianzas, 117
 Media, 58
 muestral, 161
 Mediana
 de una v.a., 62
 muestral, 180
 Medida de probabilidad, 6, 18
 inducida por una v.a., 76
 Momentos, 61
 absolutos, 61
 centrales, 61
 centrales absolutos, 61
 factoriales, 61
 Muestra aleatoria, 161
 Probabilidad
 clásica, 19
 geométrica, 19
 Semiálgebra, 11
 Teorema
 central del límite, 227
 de cambio de variable, 143, 144, 146
 de convergencia dominada, 193
 de convergencia monótona, 192
 de De Moivre-Laplace, 226
 de Poisson, 89
 Valor
 esperado, 58
 medio, 58

- promedio, 58
- Variable aleatoria, 40
 - continua, 52
 - continua singular, 52
 - discreta, 52
 - mixta, 53
- Varianza
 - condicional, 139
 - de un vector, 116
 - de una v.a., 60
 - muestral, 161
- Vector aleatorio, 95
 - continuo, 96
 - discreto, 96