Variables Aleatorias y Simulación Estocástica

ISBN: 978-956-306-068-3
Registro de Propiedad Intelectual: 200.524
Colección: Herramientas para la formación de profesores de matemáticas.
Diseño: Jessica Jure de la Cerda
Diseño de Ilustraciones: Catalina Frávega Thomas, Cristina Felmer Plominsky
Diagramación: Pedro Montealegre Barba, Francisco Santibáñez Palma
Financiamiento: Proyecto Fondef D05I-10211

Datos de contacto para la adquisición de los libros:
Para Chile:
1. En librerías para clientes directos.
2. Instituciones privadas directamente con:
   Juan Carlos Sáez C.
   Director Gerente
   Comunicaciones Noreste Ltda.
   J.C. Sáez Editor
   jcsaezc@vtr.net
   www.jcsaezeditor.blogspot.com
   Oficina: (56 2) 3260104 - (56 2) 3253148
3. Instituciones públicas o fiscales: www.chilecompra.cl

Desde el extranjero:
1. Liberalia Ediciones: www.liberalia.cl
2. Librería Antártica: www.antartica.cl
3. Argentina: Ediciones Manantial: www.emanantial.com.ar
4. Colombia: Editorial Siglo del Hombre
   Fono: (571) 3377700
5. España: Tarahumara, tarahumara@tarahumaralibros.com
   Fono: (34 91) 3656221
6. México: Alejandría Distribución Bibliográfica, alejandria@alejandrialibros.com.mx
   Fono: (52 5) 556161319 - (52 5) 6167509
7. Perú: Librería La Familia, Avenida República de Chile # 661
8. Uruguay: Dolmen Ediciones del Uruguay
   Fono: 00-598-2-7124857

Variables Aleatorias y Simulación Estocástica — Manuel Lladser
University of Colorado
manuel.lladser@colorado.edu

ESTA PRIMERA EDICIÓN DE 2.000 EJEMPLARES
Se terminó de imprimir en febrero de 2011 en WORLDCOLOR CHILE S.A.

VARIABLES ALEATORIAS Y SIMULACIÓN ESTOCÁSTICA

Manuel Lladser

University of Colorado

Dirección del Proyecto Fondef D05I-10211
Herramientas para la Formación de Profesores de Matemática

Patricio Felmer, Director del Proyecto
Universidad de Chile.

Leonor Varas, Directora Adjunta del Proyecto
Universidad de Chile.

Salomé Martínez, Subdirectora de Monografías
Universidad de Chile.

Cristián Reyes, Subdirector de Estudio de Casos
Universidad de Chile.

Presentación de la Colección

La colección de monografías que presentamos es el resultado del generoso esfuerzo de los autores, quienes han dedicado su tiempo y conocimiento a la tarea de escribir un texto de matemática. Pero este esfuerzo y generosidad no se encuentra plenamente representado en esta labor, sino que también en la enorme capacidad de aprendizaje que debieron mostrar, para entender y comprender las motivaciones y necesidades de los lectores: Futuros profesores de matemática.

Los autores, encantados una y otra vez por la matemática, sus abstracciones y aplicaciones, enfrentaron la tarea de buscar la mejor manera de traspasar ese encanto a un futuro profesor de matemática. Éste también se encanta y vibra con la matemática, pero además se apasiona con la posibilidad de explicarla, enseñarla y entregarla a los jóvenes estudiantes secundarios. Si la tarea parecía fácil en un comienzo, esta segunda dimensión puso al autor, matemático de profesión, un tremendo desafío. Tuvo que salir de su oficina a escuchar a los estudiantes de pedagogía, a los profesores, a los formadores de profesores y a sus pares. Tuvo que recibir críticas, someterse a la opinión de otros y reescribir una y otra vez su texto. Capítulos enteros resultaban inadecuados, el orden de los contenidos y de los ejemplos era inapropiado, se hacía necesario escribir una nueva versión y otra más. Conversaron con otros autores, escucharon sus opiniones, sostuvieron reuniones con los editores. Escuchar a los estudiantes de pedagogía significó, en muchos casos, realizar eventos de acercamiento, desarrollar cursos en base a la monografía, o formar parte de cursos ya establecidos. Es así que estas monografías recogen la experiencia de los autores y del equipo del proyecto, y también de formadores de profesores y estudiantes de pedagogía. Ellas son el fruto de un esfuerzo consciente y deliberado de acercamiento, de apertura de caminos, de despliegue de puentes entre mundos, muchas veces, separados por falta de comunicación y cuya unión es vital para el progreso de nuestra educación.

La colección de monografías que presentamos comprende una porción importante de los temas que usualmente encontramos en los currículos de formación de profesores de matemática de enseñanza media, pero en ningún caso pretende ser exhaustiva. Del mismo modo, se incorporan temas que sugieren nuevas formas de abordar los contenidos, con énfasis en una matemática más pertinente para el futuro profesor, la que difiere en su enfoque de la matemática para un ingeniero o para un licenciado en matemática, por ejemplo. El formato de monografía, que aborda temas específicos con extensión moderada, les da flexibilidad para que sean usadas de muy diversas maneras, ya sea como texto de un curso, material complementario, documento básico de un seminario, tema de memoria y también como lectura personal. Su utilidad ciertamente va más allá de las aulas universitarias, pues esta colección puede convertirse en la base de una biblioteca personal del futuro profesor o profesora, puede ser usada como material de consulta por profesores en ejercicio y como texto en cursos de especialización y post-títulos. Esta colección de monografías puede ser usada en concepciones curriculares muy distintas. Es, en suma, una herramienta nueva y valiosa, que a partir de ahora estará a disposición de estudiantes de pedagogía en matemática, formadores de profesores y profesores en ejercicio.

El momento en que esta colección de monografías fue concebida, hace cuatro años, no es casual. Nuestro interés por la creación de herramientas que contribuyan a la formación de profesores de matemática coincide con un acercamiento entre matemáticos y formadores de profesores que ha estado ocurriendo en Chile y en otros lugares del mundo. Nuestra motivación nace a partir de una creciente preocupación en todos los niveles de la sociedad, que ha ido abriendo paso a una demanda social y a un interés nacional por la calidad de la educación, expresada de muy diversas formas. Esta preocupación y nuestro interés encontró eco inmediato en un grupo de matemáticos, inicialmente de la Universidad de Chile, pero que muy rápidamente fue involucrando a matemáticos de la Pontificia Universidad Católica de Chile, de la Universidad de Concepción, de la Universidad Andrés Bello, de la Universidad Federico Santa María, de la Universidad Adolfo Ibáñez, de la Universidad de La Serena y también de la Universidad de la República de Uruguay y de la Universidad de Colorado de Estados Unidos.

La matemática ha adquirido un rol central en la sociedad actual, siendo un pilar fundamental que sustenta el desarrollo en sus diversas expresiones. Constituye el cimiento creciente de todas las disciplinas científicas, de sus aplicaciones en la tecnología y es clave en las habilidades básicas para la vida. Es así que la matemática actualmente se encuentra en el corazón del currículo escolar en el mundo y en particular en Chile. No es posible que un país que pretenda lograr un desarrollo que involucre a toda la sociedad, descuide el cultivo de la matemática o la formación de quienes tienen la misión de traspasar de generación en generación los conocimientos que la sociedad ha acumulado a lo largo de su historia.

Nuestro país vive cambios importantes en educación. Se ha llegado a la convicción que la formación de profesores es la base que nos permitirá generar los cambios cualitativos en calidad que nuestra sociedad ha impuesto. Conscientes de que la tarea formativa de los profesores de matemática y de las futuras generaciones de jóvenes es extremadamente compleja, debido a que confluyen un sinnúmero de factores y disciplinas, a través de esta colección de monografías, sus editores, autores y todos los que han participado del proyecto en cada una de sus etapas, contribuyen a esta tarea, poniendo a disposición una herramienta adicional que ahora debe tomar vida propia en los formadores, estudiantes, futuros profesores y jóvenes de nuestro país.

Patricio Felmer y Salomé Martínez
Editores

Agradecimientos

Agradecemos a todos quienes han hecho posible la realización de este proyecto Fondef: "Herramientas para la formación de Profesores de Matemáticas". A Cristián Cox, quien apoyó con decisión la idea original y contribuyó de manera crucial para obtener la participación del Ministerio de Educación como institución asociada. Agradecemos a Carlos Eugenio Beca por su apoyo durante toda la realización del proyecto. A Rafael Correa, Edgar Kausel y Juan Carlos Sáez, miembros del Comité Directivo. Agradecemos a Rafael Benguria, Servet Martínez y Fidel Oteiza, miembros del Comité Editorial de la colección, quienes realizaron valiosos aportes a los textos. A José Sánchez, entonces Decano de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Concepción y a Guillermo Marshall, quién fuera Decano de la Facultad de Matemáticas de la Pontificia Universidad Católica de Chile. A ambos agradecemos por su decisiva contribución para lograr la integridad de la colección de 15 monografías. Agradecemos a Víctor Campos, Ejecutivo de Proyectos de Fondef, por su colaboración y ayuda en las distintas etapas del proyecto.

En este volumen manisfestamos nuestro especial agradecimiento a Jaime San Martín, director del Centro de Modelamiento Matemático de la Universidad de Chile, por su constante apoyo durante toda la realización de este proyecto. Más aun, su apoyo decidido y generoso que permitió que esta monografía sea parte de la colección. También queremos reconocer su valioso aporte a la educación manifiestado desde la dirección del Centro de Modelamiento Matemático, el cual ha permitido un fuerte impulso al involucramiento de matemáticos activos en esta importante tarea.

Agradecemos también a Bárbara Ossandón de la Universidad de Santiago, a Jorge Ávila de la Universidad Católica Silva Henríquez, a Víctor Díaz de la Universidad de Magallanes, a Patricio Canelo de la Universidad de Playa Ancha en San Felipe y a Osvaldo Venegas y Silvia Vidal de la Universidad Católica de Temuco, quienes hicieron posible las visitas que realizamos a las carreras de pedagogía en matemática. Agradecemos a todos los evaluadores, alumnos, académicos y profesores -cuyos nombres no incluimos por ser más de una centena- quienes entregaron sugerencias, críticas y comentarios a los autores, que ayudaron a enriquecer cada uno de los textos.

Agradecemos a Marcela Lizana por su impecable aporte en todas las labores administrativas del proyecto, a Aldo Muzio por su colaboración en la etapa de evaluación, y también a Anyel Alfaro por sus contribuciones en la etapa final del proyecto y en la difusión de los logros alcanzados.

Dirección del Proyecto

Índice General

Prefacio 19

Capítulo 1: Axiomas de Probabilidades 23

1.1 Relaciones y operaciones entre sucesos 23
1.2 Sigma aditividad 26
1.3 Probabilidades condicionales e independencia 31
1.4 Independencia de varios sucesos 35
1.5 Secuencias monótonas de sucesos 38
1.6 Ejercicios propuestos 45

Capítulo 2: Variables aleatorias 53

2.1 ¿Qué es una variable aleatoria? 53
2.2 Función de distribución 56
2.3 Variables aleatorias discretas 61
2.4 Variables aleatorias continuas 62
2.5 Igualdad en distribución 73
2.6 Independencia de variables aleatorias 78
2.7 Ejercicios propuestos 88

Capítulo 3: La Ley de los Grandes Números 97

3.1 Esperanza de una variable aleatoria 97
3.2 Esperanza de variables aleatorias discretas 102
3.3 Esperanza de variables aleatorias continuas 105
3.4 La Ley de los Grandes Números 107
3.5 Propiedades de la esperanza 110
3.6 Varianza de variables aleatorias 116
3.7 Ejercicios propuestos 122

Capítulo 4: Distribuciones Destacadas y el Teorema Central Límite 133

4.1 Distribución de Bernoulli 133
4.2 Distribución Geométrica 135
4.3 Distribución Binominal 139
4.4 Distribución de Poisson 144

4.5 Distribución Exponencial 151
4.6 Distribución Normal 156
4.7 Teorema Central Límite 164
4.8 Ejercicios propuestos 173

Capítulo 5: Simulación Estocástica 181

5.1 Método de la Transformada Inversa 181
5.2 Método de Rechazo 194
5.8 Ejercicios propuestos 199

Bibliografía 203
Índice de Términos 205

A mis padres y a mi hermana Paulette

Prefacio

El objetivo general de esta monografía es desarrolllar una discusión rigurosa de los aspectos más destacados de la teoría de probabilidades, pero sin el formalismo propio de la teoría de la medida. Junto a esto se busca desarrollar la intuición probabilística del lector y ayudarle a apreciar de forma más holística algunos de los conceptos más importantes de las Probabilidades.

La monografía es autónoma y puede utilizarse, tanto como material de estudio y profundización acerca del tema, como texto en un curso avanzado del mismo. Está en su mayor parte dirigida a estudiantes de Pedagogía en Matemáticas, que hayan tenido un curso básico de probabilidades. También será de utilidad para aquellos profesores de Enseñanza Básica o Media, interesados en un estudio más avanzado pero ad hoc de las probabilidades. La monografía es un punto de transición entre dos niveles de aprendizaje de probabilidades: un nivel básico e intuitivo abordado de forma muy original y didáctica en *Probabilidades Doctas con Discos, Árboles, Bolitas y Urnas* [**4**], y una teoría mucho más general y abstracta abordada en textos como [**5, 6**]. Un curso de estadística básica al nivel de *Introducción a la Estadística* [**3**] es recomendable aunque no esencial para un uso provechoso.

A lo largo de la monografía se han incluido una variedad de ejemplos y ejercicios relacionados con el quehacer diario de un profesor de matemáticas. También han sido incluidos ejemplos relacionados con otras áreas de la ciencia o incluso con la vida cotidiana. Una vez finalizado el estudio de la monografía el lector podrá entre otras cosas:

(a) determinar la probabilidad que no hayan suspensiones de clases en un periodo de una semana, cuando generalmente ocurren e.g. cuatro suspensiones en ese lapso de tiempo;
(b) explicar por qué el promedio de notas de un alumno, se usa como una medida objetiva de su comprensión en la materia de un curso y decidir si es o no razonable aprobar a un estudiante cuyo promedio final de notas es un 3,91;
(c) determinar la cantidad mínima de agua necesaria para un paseo de curso, a modo de estar 99 % seguro que habrá suficiente líquido para beber durante toda la jornada.

La monografía se concentra principalmente en los siguientes temas: (i) los axiomas y el cálculo de probabilidades, (ii) la igualdad en distribución e independencia de variables aleatorias, (iii) la interpretación frecuentista de probabilidades y su conexión con el cálculo de esperanzas y varianzas, (iv) las distribuciones probabilísticas más destacadas y el Teorema Central Límite y (v) métodos básicos de simulación. Estos contenidos se desarrollan en un orden no convencional, que a la vez permite una discusión autónoma y concisa de los temas seleccionados. Se da énfasis a los conceptos de igualdad en distribución e independencia de variables aleatorias, que son luego interpretados en el contexto de variables aleatorias discretas y continuas. La definición de esperanza es motivada a partir de la interpretación frecuentista de probabilidades. Esto permite definir la esperanza y estudiar sus propiedades generales, sin la necesidad de distinguir entre variables aleatorias discretas y continuas. Por otro lado, también son proporcionadas; cotas para el error en la aproximación de Poisson de la distribución binomial y para el Teorema Central Límite. Los tópicos de simulación permitirán al lector desarrollar sus propios algoritmos para simular el lanzamiento de dados, monedas, o incluso la selección de un punto al azar en una cierta región del plano. Enfatizamos que para un estudio exitoso de este tema no es imprescindible el uso de un computador. Sin embargo, aquellos lectores con un entrenamiento básico en computación, podrán simular los experimentos de esencialmente cualquier ejemplo o ejercicio propuesto en esta monografía. Con relación a la lectura y estudio de la misma, el autor ha puesto entre paréntesis la pregunta "¿por qué?", donde ciertos detalles ya discutidos, y que debieran ser más familiares para el lector, han sido omitidos. Para un buen uso de la monografía, se invita al lector a reflexionar sobre estos detalles y, en lo posible, completarlos.

Aunque muchos otros temas podrían haber sido considerados, los cinco tópicos abordados en esta monografía, uno por cada capítulo, son los fundamentales, para cualquier estudiante que aspire a una comprensión genuina de esta materia. Por otro lado, debido a limitaciones de espacio, algunas demostraciones han sido dejadas como ejercicio y las demostraciones de ciertos teoremas técnicos han sido completamente omitidas. Sin embargo, en un primer estudio riguroso de cualquier tema, es mucho más importante entender las implicaciones de los resultados más técnicos, que sus demostraciones.

El autor está muy agradecido a su ex-profesor Dr. Patricio Felmer (U. de Chile) y a su amiga y colega Dra. Salomé Martínez (U. de Chile) por haberlo considerado en este interesante proyecto, y por su constante dedicación y numerosas sugerencias a lo largo de los casi tres años que tomó completar esta monografía. El autor también agradece a la Dra. Javiera Barrera (U. Técnica Federico Santa María), Dra. Gladys Bobadilla (U. de Santiago), Dr. Raúl Gouet (U. de Chile) y el Dr. Ernesto San Martín (Pontificia U. Católica de Chile) por sus valiosos comentarios y sugerencias como evaluadores de la monografía. Un especial agradecimiento al comité editorial, y en especial al Dr. Servet Martinez (U. de Chile), por la revisión cuidadosa y comentarios finales sobre la monografía. También quisiera agradecer al Dr. Juan Restrepo (U.

de Colorado), Dr. Pierre-Paul Romagnoli (U. Andrés Bello), Amrik Sen (U. de Colorado), y la Dra. María Soledad Torres (U. de Valparaíso) por sus comentarios sobre versiones preliminares de la monografía, y al Proyecto Núcleo Milenio Información y Aleatoriedad que facilitó una visita a Chile.

Para finalizar el autor sólo puede esperar que las sorpresas y desafíos que el lector encontrará a lo largo de la monografía sean gratificantes y marquen un hito en su exploración del fascinante mundo de las probabilidades.

Manuel E. Lladser B.
Boulder - Colorado
Octubre 2010

Capítulo 1: Axiomas de Probabilidades

En este capítulo revisaremos las nociones de suceso y de espacio muestral, y enfatizaremos la importancia de la sigma aditividad y del concepto de independencia entre sucesos, para modelar probabilísticamente una variedad de experimentos. Las secciones 1.1-1.4 revisarán material ya discutido en la monografía [**4**], aunque con un énfasis diferente. La Sección 1.5 introduce material no cubierto en [**4**]. El resultado principal en esta sección caracterizará la probabilidad de una unión contable y creciente de sucesos como un valor límite de probabilidades. Aplicando esta herramienta podremos e.g. determinar la probabilidad que, al tirar indefinidamente una moneda, observemos solamente caras a partir de un cierto momento, o también la probabilidad que el apellido González desaparezca algún día.

1.1 Relaciones y operaciones entre sucesos

Imagine un experimento cuyo resultado es incierto. Un *suceso* asociado con dicho experimento es una afirmación o propiedad acerca de su resultado, que puede cumplirse o no dependiendo de lo que suceda una vez llevado a cabo el experimento. Representaremos sucesos usualmente como conjuntos de resultados. Diremos que un cierto suceso *ocurrió* si el resultado del experimento pertenece a dicho suceso. En caso contrario, diremos que el suceso *no ocurrió*. Usaremos generalmente las letras A, B, C, etc o a veces A_1, A_2, A_3, etc para denotar sucesos.

Para establecer ideas, considere el experimento que consiste en registrar la nota final de un alumno en un cierto curso. Los resultados de este experimento son números decimales entre 1 y 7 con un único dígito significativo. Ejemplos de sucesos asociados con este experimento son:

$$\begin{aligned} A &= \text{“el alumno aprobará el curso”}, \\ B &= \text{“el alumno aprobará el curso con una nota menor o igual a 5.5”}, \\ C &= \text{“el alumno reprobará el curso con la peor nota posible”}. \end{aligned}$$

Si el alumno en cuestión obtiene un 6.4 al final del curso, entonces solamente el suceso A habrá ocurrido, pero no así los sucesos B y C. Por otro lado, si el alumno obtuviera un 4.9 entonces los sucesos A y B habrán occurrido pero no el suceso C.

Una manera conveniente de representar un suceso, es como el conjunto de resultados del experimento que garantizan su ocurrencia. Reservaremos la letra griega Ω para denotar el conjunto de todos los resultados posibles en un cierto experimento. Nos referiremos a este conjunto como el *espacio muestral*. Este suceso siempre ocurre y, por lo mismo, rara vez lo mencionaremos explícitamente. Otro suceso especial es el *suceso vacío* que denotaremos como $\emptyset$. ¡Este suceso nunca ocurre sin importar el resultado del experimento!

Por ejemplo, si representamos los sucesos anteriores como subconjuntos del espacio muestral $\Omega = \{1, 1.1, \ldots, 6.9, 7\}$ obtenemos las siguientes identidades

$$\begin{aligned} A &= \{4, 4.1, \ldots, 6.9, 7\}, \\ B &= \{4, 4.1, \ldots, 5.4, 5.5\}, \\ C &= \{1\}. \end{aligned}$$

Al representar sucesos como conjuntos de resultados, podemos comparar dos de ellos usando las relaciones de inclusión ($\subset$) o de igualdad ($=$). De acuerdo a la teoría de conjuntos, $A \subset B$ si y solo si $\omega \in A$ implica $\omega \in B$. Por otro lado, $A = B$ si y solo si $\omega \in A$ es equivalente a $\omega \in B$. Al pensar en la variable ω como el resultado del experimento, y en términos de la ocurrencia de sucesos, podemos definir estas dos relaciones como sigue:

Definición 1.1. Dados dos sucesos A y B, diremos que A *está incluido (o contenido) en* B si la ocurrencia de A siempre asegura la ocurrencia de B. Si además de esto, la ocurrencia de B siempre asegura la ocurrencia de A entonces diremos que A *es igual (o idéntico) a* B.

Una consecuencia directa de la definición es la siguiente equivalencia lógica:

$$A = B \iff (A \subset B \text{ y } B \subset A). \tag{1.1}$$

En otras palabras, $A = B$ solamente cuando cada vez que A ocurre B también lo hace, y cada vez que B ocurre A también lo hace.

Retomando el ejemplo anterior, vemos que A no está incluido en B. Esto se debe a que si la nota final del alumno es e.g. un 5.7, entonces A habrá ocurrido pero no así B. Por otro lado, $B \subset A$ ya que la ocurrencia de B siempre garantiza la de A. Claramente, C no está incluido ni en A ni tampoco en B. En particular, C no puede ser igual a ninguno de estos sucesos.

Una ventaja al identificar sucesos como conjuntos de resultados es que podemos usar las operaciones usuales sobre conjuntos i.e. complementación (^{c}), unión ($\cup$) e intersección ($\cap$) para definir sucesos elaborados a partir de sucesos simples. Recordemos que de acuerdo a la teoría de conjuntos, se definen

$$\begin{aligned} A^c &= \{\omega \in \Omega \mid \omega \notin A\}, \\ A \cup B &= \{\omega \in \Omega \mid \omega \in A \text{ o } \omega \in B\}, \\ A \cap B &= \{\omega \in \Omega \mid \omega \in A \text{ y } \omega \in B\}. \end{aligned}$$

En términos de la ocurrencia de sucesos, podemos reescribir las definiciones anteriores como sigue.

Definición 1.2. Considere dos sucesos A y B. El *complemento* de A es aquel suceso que ocurre solamente cuando A no ocurre. La *unión* A con B es aquel suceso que ocurre sólo cuando al menos uno de los sucesos ocurre. Finalmente, la *intersección* de A con B es aquel suceso que ocurre solamente cuando ambos sucesos ocurren simultáneamente.

Retomando el experimento, que consiste en registrar la nota final de un alumno, tenemos lo siguiente. El complemento de A es el suceso "el alumno reprobará el curso". En particular, $A^c = \{1, \ldots, 3.9\}$. Por otro lado, el complemento de B es $\{1, \ldots, 3.9, 5.6, \ldots, 7\}$ i.e. el suceso "el alumno reprobará el curso o lo pasará con al menos un 5.6". Por otro lado, usando que $A \subset B$ tenemos que $(A \cup B) = B$. Sin embargo, note que $(A \cup C) = \{1, 4, 4.1, \ldots, 6.9, 7\}$. Finalmente, como $B \subset A$ entonces $(A \cap B) = B$.

Diremos que dos sucesos son *disjuntos* cuando no pueden ocurrir simultáneamente. Esto es equivalente a decir que su intersección es el suceso vacío. Por ejemplo, los sucesos A y C del párrafo anterior son disjuntos. En particular, $(A \cap C) = \emptyset$.

Ejemplo 1.1. Para fijar ideas considere el siguiente experimento: se escogen dos niños al azar y se les pregunta su número favorito entre 1 y 5. Un espacio muestral posible para este experimento es el conjunto $\Omega = \{1, 2, 3, 4, 5\} \times \{1, 2, 3, 4, 5\}$ i.e. el conjunto de todos los pares ordenados de la forma (x, y), con $1 \leq x, y \leq 5$ números enteros. Aquí, x denota el número favorito del primer niño e y el del segundo.

Ejemplos de sucesos asociados con este experimento y su representación como subconjuntos de Ω vienen dados por

$$\begin{aligned}
A &= \text{“el número favorito de ambos niños es par”}, \\
&= \{(x, y) \in \Omega \mid x, y \in \{2, 4\}\}, \\
&= \{(2, 2), (2, 4), (4, 2), (4, 4)\}, \\
B &= \text{“los niños comparten el mismo número favorito”}, \\
&= \{(x, y) \in \Omega \mid x = y\}, \\
&= \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\}, \\
C &= \text{“el número favorito del primer niño es menor que el del segundo”}, \\
&= \{(x, y) \in \Omega \mid x < y\}, \\
&= \{(1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5)\}.
\end{aligned}$$

En particular, obtenemos que

$$\begin{aligned}
A^c &= \text{“el número favorito de al menos uno de los niños es impar”}, \\
&= \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (2, 1), (2, 3), (2, 5), (3, 1), (3, 2), \\
&\quad (3, 3), (3, 4), (3, 5), (4, 1), (4, 3), (4, 5), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5)\}, \\
(A \cup B) &= \text{“el número favorito de ambos niños es el mismo o es par”}, \\
&= \{(1, 1), (2, 2), (2, 4), (3, 3), (4, 2), (4, 4), (5, 5)\}, \\
(A \cap B) &= \text{“el número favorito de ambos niños es el mismo y es par”}, \\
&= \{(2, 2), (4, 4)\}.
\end{aligned}$$

Finalmente, observe que B y C no pueden ocurrir simultáneamente; en particular, B y C son sucesos disjuntos.

1.2 Sigma aditividad

La probabilidad de un suceso en un cierto experimento es un número entre cero y uno. Este número representa la certeza que se tiene que el suceso ocurra y depende de la información disponible sobre el experimento. La probabilidad asignada a un cierto suceso A será usualmente denotada como $\mathbb{P}(A)$.

De acuerdo a la *interpretación frecuentista de probabilidades*, la probabilidad de un cierto suceso representa su fracción de ocurrencia asintótica al repetir indefinidamente y bajo circunstancias similares el mismo experimento. Desde este punto de vista, cuando decimos que un cierto suceso A tiene probabilidad $\mathbb{P}(A)$, de alguna manera estamos diciendo que si repitiéramos el mismo experimento un número grande n de veces, entonces el suceso A ocurrirá aproximadamente en $n \cdot \mathbb{P}(A)$ de los experimentos. Por ejemplo, si $\mathbb{P}(A) = 1/2$, entonces esperamos que A ocurra aproximadamente en la mitad de los experimentos. Por otro lado, si $\mathbb{P}(A) = 1/10$, entonces esperamos que A ocurra aproximadamente en uno de cada diez experimentos.

Tan pronto se asuma la interpretación anterior, la función $\mathbb{P}$ definida sobre sucesos debe satisfacer tres propiedades fundamentales. Específicamente, diremos que una función $\mathbb{P}$ que asigna a cada suceso A un número real, es una *medida de probabilidad* cuando las siguientes propiedades son satisfechas:

(1.2) $\mathbb{P}(\Omega) = 1$,

(1.3) para todo suceso A, $\mathbb{P}(A) \geq 0$,

(1.4) si $A_1, A_2, \ldots$ es una secuencia contable de sucesos disjuntos entonces

$$\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i).$$

En (1.4), $\cup_i A_i$ denota la *unión de los sucesos* $A_1, A_2, \ldots$ Como es de esperarse de la Definición 1.2, este suceso ocurre solamente cuando al menos uno de los sucesos que participa de la unión ocurre. En términos matemáticos, esto significa que:

$$\omega \in \bigcup_i A_i \Longleftrightarrow \text{existe } i \text{ tal que } \omega \in A_i.$$

Cuando el número de términos que participan en la unión es finito, $\sum_i \mathbb{P}(A_i)$ es una suma común y corriente. Sin embargo, cuando el número de términos es infinito, esta suma corresponde al límite de una serie.

La primera propiedad en la definición de medida de probabilidad se basa en el hecho que el suceso Ω ocurrirá en cada una las realizaciones del experimento. En particular, su frecuencia de ocurrencia será siempre igual a uno. La segunda propiedad incorpora el hecho que la frecuencia con que un suceso será observado es una cantidad no negativa.

La última propiedad en la definición entregada es lo que se llama *sigma aditividad* y es una de las propiedades más importantes de la definición, aunque para desarrollar la teoría las tres propiedades son necesarias. Esta propiedad incorpora la siguiente idea: cuando los sucesos $A_1, A_2, \ldots$ son disjuntos la frecuencia con que su unión será observada corresponde a la suma de las frecuencias con que cada uno de estos sucesos será observado. (Si los sucesos en cuestión no son disjuntos, entonces la suma de sus frecuencias usualmente excederá la frecuencia con que su unión ocurre (*¿por qué?*).)

Cuando decimos que $A_1, A_2, \ldots$ es una *secuencia contable* queremos decir que el conjunto de índices asociado a estos sucesos es contable. Un conjunto I se dice *contable* si es finito o tiene la misma cardinalidad de los números naturales. En este último caso el término *enumerable* es comúnmente utilizado. Esto es equivalente a la existencia de una transformación $h : I \to \mathbb{N}$ que es inyectiva. Ejemplos de conjuntos contables no finitos incluyen $\mathbb{N}$, $\mathbb{Z}$ y $\mathbb{Q}$. Sin embargo, $\mathbb{R}$ y, más generalmente, cualquier subconjunto de los números reales que contiene un intervalo de largo estrictamente positivo, no es contable. En particular, por ejemplo, el intervalo $[0,1]$ no es contable.

Ejemplo 1.2. Considere el experimento donde se escogen dos niños al azar y se les pregunta su número favorito entre 1 y 5 (Ejemplo 1.1). Asociaremos a este experimento el espacio muestral $\Omega = \{1,2,3,4,5\} \times \{1,2,3,4,5\}$, el cual consta de 25 elementos. *¿Cuál es la probabilidad de cada uno de los siguientes sucesos?*

$$\begin{aligned}
A &= \text{“el número favorito de ambos niños es par”},\\
&= \{(2,2),(2,4),(4,2),(4,4)\},\\
B &= \text{“los niños comparten el mismo número favorito”},\\
&= \{(1,1),(2,2),(3,3),(4,4),(5,5)\},\\
C &= \text{“el número favorito del primer niño es menor que el del segundo”},\\
&= \{(1,2),(1,3),(1,4),(1,5),(2,3),(2,4),(2,5),(3,4),(3,5),(4,5)\}.
\end{aligned}$$

Para determinar la probabilidad del primer suceso observe que

$$A = \{(2,2)\} \cup \{(2,4)\} \cup \{(4,2)\} \cup \{(4,4)\},$$

y que los sucesos que participan de esta unión finita son disjuntos. En particular, debido a la sigma aditividad, tenemos que

$$\mathbb{P}(A) = \mathbb{P}(\{(2,2)\}) + \mathbb{P}(\{(2,4)\}) + \mathbb{P}(\{(4,2)\}) + \mathbb{P}(\{(4,4)\}).$$

Para computar el lado derecho de lo anterior basta con determinar la probabilidad de cada uno de los sucesos $\{(x,y)\}$, con $(x,y) \in \Omega$. Pero observe lo siguiente: *como los niños han sido escogidos al azar es razonable asumir que cada uno de los 25 sucesos $\{(x,y)\}$ es igualmente probable que cualquier otro i.e. existe un número p tal que*

$$\mathbb{P}(\{(x,y)\}) = p,$$

para cada $(x,y) \in \Omega$. Más aún, $p = 1/25$. Esto se debe a que $\mathbb{P}(\Omega) = 1$ y, por otro lado, $\mathbb{P}(\Omega) = 25p$, debido a que Ω corresponde a la unión finita y disjunta de 25 sucesos de la forma $\{(x,y)\}$. Por lo tanto, concluimos que

$$\mathbb{P}(A) = 4p = 4/25.$$

Similarmente, tenemos que $\mathbb{P}(B) = 5p = 1/5$ y $\mathbb{P}(C) = 10p = 2/5$.

El ejemplo recién discutido es un caso particular del siguiente.

Ejemplo 1.3. Considere un cierto experimento que tiene asociado un espacio muestral contable. Por ejemplo, si el experimento consiste en lanzar un dado entonces $\Omega = \{1,2,3,4,5,6\}$ es finito. Por otro lado, si el experimento consiste en lanzar un dado hasta observar la cara 6 por primera vez, entonces una posibilidad de espacio muestral es $\Omega = \{(6),(1,6),(2,6),(3,6),(4,6),(5,6),(1,1,6),(1,2,6),\ldots\}$. En este caso e.g. el elemento $(1,1,6)$ representa el resultado donde un 6 fue observado por primera vez en el tercer lanzamiento, habiendo visto un 1 en el primer y segundo lanzamiento.

Para cada $\omega \in \Omega$ denotaremos como p_ω la fracción asintótica de veces con que ω será observado al repetir el experimento indefinidamente. Una vez que la fracción asintótica de ocurrencia de cada elemento en el espacio muestral ha sido especificada, es natural definir la probabilidad de cada suceso $A \subset \Omega$ como

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega.$$

Si $p_\omega \geq 0$ y $\sum_{\omega\in\Omega} p_\omega = 1$ entonces $\mathbb{P}$ es una medida de probabilidad. En efecto, el lector podrá verificar fácilmente que las condiciones dadas en (1.2) y (1.3) aplican en este caso. Para demostrar (1.4) observe que si $A_1, A_2, \ldots$ es una secuencia contable y disjunta de sucesos entonces

$$\mathbb{P}(\cup_i A_i) = \sum_{\omega\in\cup_i A_i} p_\omega = \sum_i \sum_{\omega\in A_i} p_\omega = \sum_i \mathbb{P}(A_i).$$

Esto demuestra la sigma aditividad y, por lo tanto, $\mathbb{P}$ es efectivamente una medida de probabilidad.

Si Ω es finito y si cada uno de los resultados del experimento es igualmente probable a otro (e.g. el caso discutido en el Ejemplo 1.2) entonces

$$p_\omega = \frac{1}{|\Omega|}, \tag{1.5}$$

donde $|\Omega|$ es la *cardinalidad de* Ω i.e. el número de elementos diferentes del espacio muestral. En este caso, es obvio que $p_\omega \geq 0$ y $\sum_{\omega\in\Omega} p_\omega = 1$. De hecho, uno encuentra que

$$\mathbb{P}(A) = \sum_{\omega\in A} \frac{1}{|\Omega|} = \frac{1}{|\Omega|} \sum_{\omega\in A} 1 = \frac{|A|}{|\Omega|}.$$

En palabras: *cuando el espacio muestral es finito y cada resultado es tan probable como otro, la probabilidad de un suceso se calcula como la razón entre el número de casos favorables para el suceso y el número total de casos posibles.* Por ejemplo, si $\Omega = \{1,2,3,4,5,6\}$ es el espacio muestral asociado al lanzamiento de un dado equilibrado entonces $p_\omega = 1/6$. Bajo este modelo, la probabilidad del suceso "la cara lanzada es mayor o igual a cuatro" es $|\{4,5,6\}|/6 = 1/2$. Por otro lado, la probabilidad del suceso "la cara lanzada es un número primo" viene dada por $|\{1,2,3,5\}|/6 = 2/3$.

Si Ω es enumerable la discusión en el párrafo anterior no tiene sentido ya que $1/|\Omega| = 0$. Otras consideraciones son necesarias para definir los coeficientes p_ω en este caso. Por ejemplo, si $\Omega = \{(6),(1,6),(2,6),(3,6),\ldots\}$ es el espacio muestral asociado al lanzamiento de un dado equilibrado hasta observar por primera vez un 6, justificaremos en el Ejemplo 1.15 que

$$p_\omega = \left(\frac{1}{6}\right)^{n(\omega)}, \tag{1.6}$$

donde $n(\omega)$ denota el número de coordenadas de ω. Por ejemplo, $n(1,1,6) = 3$ y $p_{(1,1,6)} = (1/6)^3 = 1/216$. Claramente, $p_\omega \geq 0$. Por otro lado, debido al principio de multiplicación (ver Capítulo 4 en [**4**]), hay 5^{k-1} elementos en Ω tal que $n(\omega) = k$. Como $p_\omega = (1/6)^k$ en este caso, obtenemos que

$$\sum_{\omega\in\Omega} p_\omega = \sum_{k=1}^{\infty} 5^{k-1}\left(\frac{1}{6}\right)^k = \frac{1}{6}\sum_{k=1}^{\infty}\left(\frac{5}{6}\right)^{k-1} = \frac{1}{6}\cdot\frac{1}{1-5/6} = 1.$$

La penúltima identidad anterior se justifica debido a la *serie geométrica*:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x},\ \text{para } |x| \leq 1. \tag{1.7}$$

Los coeficientes en (1.6) inducen la medida de probabilidad dada por la fórmula

$$\mathbb{P}(A) = \sum_{\omega\in A}\left(\frac{1}{6}\right)^{n(\omega)},$$

para cada $A \subset \Omega$. Así, por ejemplo, la probabilidad del suceso

$$A = \text{"un 6 sale por primera vez en el cuarto lanzamiento del dado"}$$

viene dada por

$$\mathbb{P}(A) = \sum_{\omega\in A}\left(\frac{1}{6}\right)^{n(\omega)} = \sum_{\omega\in A}\left(\frac{1}{6}\right)^4 = \left(\frac{1}{6}\right)^4\cdot|A| = \left(\frac{1}{6}\right)^4\cdot 5\cdot 5\cdot 5\cdot 1 = \frac{125}{1296}.$$

Por otro lado, el suceso

$$B = \text{"un 6 sale por primera vez después del tercer lanzamiento"}$$

corresponde a la unión de los sucesos B_k="un 6 sale por primera vez en el k-ésimo lanzamiento del dado" con $k \geq 4$. Un argumento similar al recién utilizado muestra que $\mathbb{P}(B_k) = (5/6)^{k-1} \cdot (1/6)$. Como estos sucesos son disjuntos, la sigma aditividad implica que

$$\begin{aligned}
\mathbb{P}(B) = \sum_{k=4}^{\infty} \mathbb{P}(B_k) &= \frac{1}{6} \sum_{k=4}^{\infty} \left(\frac{5}{6}\right)^{k-1}, \\
&= \frac{1}{6} \left(\frac{5}{6}\right)^{3} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^{k} = \frac{1}{6} \left(\frac{5}{6}\right)^{3} \cdot \frac{1}{1-5/6} = \frac{125}{216}.
\end{aligned}$$

Es importante que el lector no olvide que la sigma aditividad requiere uniones contables. El siguiente ejemplo aclarará este punto.

Ejemplo 1.4. Considere un experimento que consiste en escoger un número al azar en el intervalo $\Omega = [0, 1]$. Veremos diversas situaciones como ésta en la monografía y, por lo mismo, no explicaremos el mecanismo como el punto es escogido. Para cada $x \in [0, 1]$, considere el suceso A_x definido como "el número escogido es x". Observe que si $i \neq j$ entonces $(A_i \cap A_j) = \emptyset$.

Claramente el suceso B definido como "el número escogido es racional" corresponde a la unión de los sucesos A_i, con $i \in \mathbb{Q} \cap [0, 1]$, i.e. $B = \cup_{i \in \mathbb{Q} \cap [0,1]} A_i$. Como $\mathbb{Q}$ es contable, también lo es el conjunto de índices $\mathbb{Q} \cap [0, 1]$. En particular, sea cual sea la medida de probabilidad $\mathbb{P}$ que usamos para modelar este experimento, y como los sucesos que participan de la unión son disjuntos, la sigma aditividad implica que

$$\mathbb{P}(B) = \sum_{i \in \mathbb{Q} \cap [0,1]} \mathbb{P}(A_i).$$

Por otro lado, si consideramos el suceso C="el número escogido es mayor o igual a cero y menor o igual a uno" entonces $C = \cup_{i \in [0,1]} A_i$. Como el conjunto $[0, 1]$ no es contable, no podemos usar la sigma aditividad para concluir que $\mathbb{P}(C)$ corresponde a $\sum_{i \in [0,1]} \mathbb{P}(A_i)$. ¡De hecho esta sumatoria no tiene sentido! Sin embargo, observe que $\mathbb{P}(C) = 1$ debido a que $C = \Omega$ en este caso.

Las propiedades (1.2)-(1.4) que definen una medida de probabilidad implican una serie de propiedades satisfecha por cualquier medida de probabilidad. Entre las que más usaremos se encuentran las siguientes:

(1.8) $\mathbb{P}(\emptyset) = 0,$

(1.9) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A),$

(1.10) $0 \leq \mathbb{P}(A) \leq 1,$

(1.11) Si $A \subset B$ entonces $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A),$

(1.12) Si $A \subset B$ entonces $\mathbb{P}(A) \leq \mathbb{P}(B),$

(1.13) Si $\mathbb{P}(A) = 1$ entonces $\mathbb{P}(A \cap B) = \mathbb{P}(B).$

En las propiedades, el suceso $B \setminus A$ denota lo que se llama la *diferencia de B con A*, o a veces *B menos A*. Este suceso ocurre solamente cuando B ocurre y A no lo hace.

Como A^c ocurre precisamente cuando A no lo hace, la identidad $B \setminus A = (B \cap A^c)$ aplica.

Ejemplo 1.5. La demostración de las propiedades en (1.8)-(1.12) aparecen como ejercicio al final del Capítulo. En este ejemplo, veremos cómo aplicar las propiedades en (1.9) y (1.12) para demostrar la propiedad en (1.13). En efecto, observe que (*¿por qué?*):

$$B = (A \cap B) \cup (A^c \cap B).$$

Como $(A \cap B)$ y $(A^c \cap B)$ son sucesos disjuntos, la sigma aditividad implica que $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$. Pero observe que $(A^c \cap B) \subset A^c$ y, por lo tanto, $0 \leq \mathbb{P}(A^c \cap B) \leq \mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 0$ i.e. $\mathbb{P}(A^c \cap B) = 0$. Concluimos de este modo que $\mathbb{P}(B) = \mathbb{P}(A \cap B)$, lo que demuestra la propiedad en (1.13).

Otra identidad importante que usaremos con frecuencia es la siguiente.

Teorema 1.3. *(Fórmula de Probabilidades Totales.) Si $A_1, A_2, \ldots$ es una secuencia contable de sucesos disjuntos y $\Omega = \cup_i A_i$ entonces para todo suceso A aplica que*

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap A_i).$$

Demostración. Observe que $A = A \cap \Omega = A \cap (\cup_i A_i) = \cup_i (A \cap A_i)$. Como los sucesos $A_1, A_2, \ldots$ son disjuntos, entonces también lo son $(A \cap A_1), (A \cap A_2), \ldots$ El teorema es ahora una consecuencia directa de la sigma aditividad. □

1.3 Probabilidades condicionales e independencia

Dados dos sucesos A y B, la *probabilidad condicional de A dado B* es la cantidad definida como

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \tag{1.14}$$

siempre y cuando $\mathbb{P}(B) > 0$. Esta cantidad representa la posibilidad del suceso A cuando se tiene la certeza que el suceso B ha ocurrido en el experimento. La definición resulta natural de la interpretación frecuentista de probabilidades: *si el mismo experimento se repitiera un número grande n de veces, entonces el suceso B debiera ocurrir alrededor de $n \cdot \mathbb{P}(B)$ veces, mientras que el suceso $(A \cap B)$ debiera ocurrir alrededor de $n \cdot \mathbb{P}(A \cap B)$ veces. En particular, la fracción de veces que esperamos que A ocurra cada vez que B ocurre está dada por la razón*

$$\frac{n \cdot \mathbb{P}(A \cap B)}{n \cdot \mathbb{P}(B)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

La probabilidad condicional de A dado B representa, por lo tanto, la fracción de experimentos en los que esperamos que el suceso A ocurra, cuando sólo consideramos aquellos experimentos donde B ocurrirá.

Ejemplo 1.6. Si en un cierto experimento los sucesos A y B ocurren simultáneamente 1 de cada 13 veces, mientras que B ocurre 4 de cada 13 veces, entonces la posibilidad de A cuando B ocurre está dada por la probabilidad condicional

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/13}{4/13} = \frac{1}{4}.$$

Una situación específica del caso anterior, es el experimento donde se escoge una carta al azar de una baraja inglesa y los sucesos A="la pinta de la carta es trébol" y B="el número de la carta es A, J, Q o K". Es intuitivamente claro que la pinta de la carta escogida será trébol sólo 1 entre 4 veces que ésta tenga marcado el número A, J, Q o K. ¡El cálculo anterior corrobora esta intuición!

Ejemplo 1.7. Para fijar ideas considere el experimento que consiste en lanzar un dado equilibrado y los sucesos

$$\begin{aligned} A &= \text{“la cara lanzada es un número par”}, \\ B &= \text{“la cara lanzada es menor o igual a cinco”}. \end{aligned}$$

Claramente $\mathbb{P}(A) = 1/2$, debido a que sólo tres de las seis caras del dado son números pares. Veremos, sin embargo, que $\mathbb{P}(A \mid B) = 2/5$; en particular, A es menos probable cuando B ha sucedido. Una manera intuitiva de ver que éste es el caso, es notar que sólo dos de las caras menores o iguales a cinco son pares. Como ninguna de estas cinco caras es más probable que otra, la posibilidad de A cuando sabemos que la cara lanzada es menor o igual a cinco debiera ser $2/5$. Esta intuición es consistente con la definición dada. De hecho observe que $(A \cap B) = \{2, 4\}$ y $B = \{1, 2, 3, 4, 5\}$. En particular, tenemos que

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{2/6}{5/6} = \frac{2}{5}.$$

Dados dos sucesos A y B tales que $\mathbb{P}(B) > 0$, observe que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B). \tag{1.15}$$

Usaremos esta identidad a menudo para determinar la probabilidad que A y B ocurran simultáneamente.

Ejemplo 1.8. Si en un cierto experimento el suceso A tiene una chance de un 60 % y, entre las veces que A ocurre, el suceso B también ocurre el 10 % de las veces, *¿cuál es la probabilidad que A y B ocurran simultáneamente?* De acuerdo a la información dada $\mathbb{P}(A) = 0{,}6$ y $\mathbb{P}(B \mid A) = 0{,}1$. En particular, la respuesta a la pregunta viene dada por

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A) = 0{,}1 \cdot 0{,}6 = 0{,}06.$$

La identidad en (1.15) resulta particularmente útil en combinación con la Fórmula de Probabilidades Totales (Teorema 1.3): *si $A_1, \ldots, A_n$ son sucesos disjuntos, cada* *uno con una probabilidad estrictamente positiva, y* $\Omega = \cup_{i=1}^n A_i$ *entonces para todo suceso* A *aplica que*

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid A_i) \cdot \mathbb{P}(A_i). \tag{1.16}$$

Ejemplo 1.9. Considere tres monedas que se ven idénticas, sin embargo, sólo una de ellas es equilibrada, mientras que las otras dos tienen probabilidad $47/100$ y $51/100$ de salir cara, respectivamente. Al tirar una de estas monedas, *¿cuál es la probabilidad que la moneda salga cara?*

En lo que sigue hablaremos de la moneda 1, 2 y 3, donde la moneda 1 es la que tiene la menor probabilidad de cara y la 3, la mayor. Este experimento consiste de dos partes: escoger una moneda y luego tirarla. Denotaremos como A_i el suceso "la i-ésima moneda fue escogida". Como las monedas son visualmente idénticas, es razonable definir $\mathbb{P}(A_i) = 1/3$.

Por otro lado, si A denota el suceso "la moneda tirada salió cara" entonces la probabilidad de este suceso usando la moneda 1 es $47/100$ i.e. $\mathbb{P}(A \mid A_1) = 47/100$. Similarmente tenemos que $\mathbb{P}(A \mid A_2) = 1/2$ y $\mathbb{P}(A \mid A_3) = 51/100$.

Sea cual sea el espacio muestral escogido para este experimento, observe que $\Omega = (A_1 \cup A_2 \cup A_3)$. Como estos tres sucesos son disjuntos, la Fórmula de Probabilidades Totales dada en (1.16) implica que

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \mid A_1) \cdot \mathbb{P}(A_1) + \mathbb{P}(A \mid A_2) \cdot \mathbb{P}(A_2) + \mathbb{P}(A \mid A_3) \cdot \mathbb{P}(A_3), \\ &= 47/100 \cdot 1/3 + 1/2 \cdot 1/3 + 51/100 \cdot 1/3 = 37/75. \end{aligned}$$

Imagine que al tirar la moneda ésta salió cara, *¿cuál es la probabilidad que la moneda 1 haya sido la tirada?* La respuesta no es $1/3$ ya que si la moneda sale cara, la moneda 1 es la menos probable entre las tres. La respuesta es de hecho la probabilidad condicional $\mathbb{P}(A_1 \mid A)$, la cual viene dada por

$$\mathbb{P}(A_1 \mid A) = \frac{\mathbb{P}(A_1 \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid A_1) \cdot \mathbb{P}(A_1)}{\mathbb{P}(A)} = \frac{47/100 \cdot 1/3}{37/75} = \frac{47}{148} \approx 31{,}7\,\%,$$

donde para la segunda identidad hemos usado (1.15).

Para motivar la siguiente definición será instructivo considerar un ejemplo simple. Considere el experimento donde se lanza un dado equilibrado y los sucesos

$$\begin{aligned} A &= \text{“la cara lanzada es un número impar”}, \\ B &= \text{“la cara lanzada es mayor o igual a tres”}. \end{aligned}$$

Observe que

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{2/6}{4/6} = \frac{1}{2}.$$

Como $\mathbb{P}(A) = 1/2$, vemos que $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ i.e. la posibilidad de A no aumenta ni disminuye cuando el suceso B ocurre. Debido a esto diremos que A y B son sucesos independientes. Usando la identidad en (1.15), esto equivale a decir que $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Este tipo de identidad motiva la siguiente definición.

Definición 1.4. Dados dos sucesos A y B, diremos que son *independientes* cuando $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Es casi directo ver que lo anterior es equivalente a tener que $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, cuando $\mathbb{P}(B) > 0$, o también que $\mathbb{P}(B \mid A) = \mathbb{P}(B)$, cuando $\mathbb{P}(A) > 0$. En palabras: *dos sucesos son independientes cuando la ocurrencia de uno no cambia (i.e. no aumenta ni disminuye) la posibilidad del otro.* Para fijar ideas consideraremos algunos ejemplos.

Ejemplo 1.10. Considere el experimento donde se escoge una carta al azar de una baraja inglesa y los sucesos

$$\begin{aligned} A &= \text{“el número de la carta es 10, J, Q, K o A”}, \\ B &= \text{“la pinta de la carta es corazón”}. \end{aligned}$$

Claramente $\mathbb{P}(A) = 5/13$, $\mathbb{P}(B) = 1/4$ y $\mathbb{P}(A \cap B) = 5/52$; en particular, $\mathbb{P}(A) \cdot \mathbb{P}(B)$ es igual a $\mathbb{P}(A \cap B)$ i.e. A y B son independientes.

Ejemplo 1.11. Imagine que usted lanza una moneda equilibrada tres veces. *¿Son independientes los siguientes sucesos?*

$$\begin{aligned} A &= \text{“la moneda sale cara en el primer lanzamiento”}, \\ B &= \text{“los resultados de los tres lanzamientos son idénticos”}. \end{aligned}$$

Observe que los ocho resultados posibles para este experimento son igualmente probables. En particular, como el suceso $(A \cap B)$ consiste de un único resultado, $\mathbb{P}(A \cap B) = 1/8$. Similarmente, se tiene que $\mathbb{P}(A) = 1/2$ y $\mathbb{P}(B) = 1/4$. Por lo tanto, $\mathbb{P}(A \cap B) = 1/8 = 1/2 \cdot 1/4 = \mathbb{P}(A) \cdot \mathbb{P}(B)$, y concluimos que A y B son sucesos independientes.

Ejemplo 1.12. Imagine un juego basado en tres lanzamientos de una misma moneda equilibrada y que usted jugará con un amigo. Para comenzar, y antes que usted tire las monedas, su amigo tiene que apostar a un cierto suceso que tiene una probabilidad exacta de 1/2 (e.g. que el primer lanzamiento saldrá cara). Siguiendo esto, usted tira a escondidas las monedas y está obligado a revelar algo no obvio pero cierto, que observó en los lanzamientos. La gracia del juego es que, basado en lo que usted revela, su amigo tendrá la opción de mantener su apuesta o cambiarla por lo completamente opuesto. Usted gana el round si la opción final de su amigo es incorrecta. *¿Cuál es una buena estrategia para usted en este juego?*

Para fijar ideas, suponga que su amigo apuesta al suceso

$$A = \text{“la moneda saldrá cara en el primer lanzamiento”}.$$

Intuitivamente, cualquier afirmación acerca del segundo o tercer lanzamiento debiera ser independiente de A. Así, por ejemplo, si al tirar las monedas usted obtiene $(cara, cara, cara)$ entonces usted podría revelar a su amigo la ocurrencia del suceso

$$B = \text{“el segundo y tercer lanzamiento resultaron idénticos”}.$$

¡Usted no estará entonces dando ninguna información útil a su amigo! De hecho, si $\Omega = \{(x, y, z) : x, y, z \in \{cara, sello\}\}$ entonces el suceso $(A \cap B)$ consta de sólo dos resultados, mientras que B consta de cuatro. Como los ocho elementos de Ω debieran ser igualmente probables, obtenemos que:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{2/8}{4/8} = \frac{1}{2} = \mathbb{P}(A).$$

1.4 Independencia de varios sucesos

En esta sección extenderemos el concepto de independencia para incluir más de dos sucesos. Para motivar la definición considere el experimento donde se tiran simultáneamente tres monedas. Considere los sucesos

$$\begin{aligned} A_1 &= \text{“la primera moneda sale cara”}, \\ A_2 &= \text{“la segunda moneda sale cara”}, \\ A_3 &= \text{“la tercera moneda sale cara”}. \end{aligned}$$

Claramente A_1 y A_2 son sucesos independientes, al igual que lo son A_1 y A_3, y también A_2 y A_3. En particular, las siguientes identidades aplican

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2), \tag{1.17}$$

$$\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_3), \tag{1.18}$$

$$\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2) \cdot \mathbb{P}(A_3). \tag{1.19}$$

Por otro lado, como el resultado de la primera y segunda moneda no debiera afectar el resultado de la tercera, el suceso A_3 debiera ser independiente de $(A_1 \cap A_2)$ ya que este último hace referencia sólo a la primera y segunda moneda. En otras palabras, incluso si sabemos que A_1 y A_2 ocurrieron, la probabilidad del suceso A_3 seguirá siendo $1/2$. Un cálculo simple corrobora esta intuición. En efecto, como cada uno de los resultados de este experimento tiene una posibilidad de $1/8$, obtenemos que

$$\mathbb{P}(A_3 \mid A_1 \cap A_2) = \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} = \frac{1/8}{2/8} = \frac{1}{2} = \mathbb{P}(A_3).$$

Como A_1 y A_2 son independientes, la identidad de arriba implica que (*¿por qué?*)

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3). \tag{1.20}$$

Las identidades en (1.17)-(1.20) implican cada una de las siguientes identidades:

$$\begin{array}{lll} \mathbb{P}(A_1 \mid A_2) = \mathbb{P}(A_1), & \mathbb{P}(A_1 \mid A_3) = \mathbb{P}(A_1), & \mathbb{P}(A_1 \mid A_2 \cap A_3) = \mathbb{P}(A_1), \\ \mathbb{P}(A_2 \mid A_1) = \mathbb{P}(A_2), & \mathbb{P}(A_2 \mid A_3) = \mathbb{P}(A_2), & \mathbb{P}(A_2 \mid A_1 \cap A_3) = \mathbb{P}(A_2), \\ \mathbb{P}(A_3 \mid A_1) = \mathbb{P}(A_3), & \mathbb{P}(A_3 \mid A_2) = \mathbb{P}(A_3), & \mathbb{P}(A_3 \mid A_1 \cap A_2) = \mathbb{P}(A_3). \end{array}$$

En palabras, podemos describir estas identidades diciendo que la posibilidad de cada uno de los sucesos A_1, A_2, y A_3 no se ve afectada por la ocurrencia de los otros. Sorprendentemente, en un experimento arbitrario, las identidades en (1.17)-(1.19) no garantizan la identidad en (1.20) ni viceversa (Ejemplo 1.13). Debido a esto, para definir la independencia de varios sucesos requeriremos los siguiente.

Definición 1.5. Los sucesos $A_1, \ldots, A_n$ se dirán *independientes* cuando, para todo conjunto no vacío de índices $I \subset \{1, \ldots, n\}$, la probabilidad del suceso $\cap_{i \in I} A_i$ corresponde al producto de las probabilidades $\mathbb{P}(A_i)$, con $i \in I$, i.e.

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Una consecuencia inmediata de lo anterior es que

$$\mathbb{P}\left(\bigcap_{i \in I} A_i \,\middle|\, \bigcap_{j \in J} A_j\right) = \frac{\prod_{k \in (I \cup J)} \mathbb{P}(A_k)}{\prod_{j \in J} \mathbb{P}(A_j)} = \prod_{i \in I} \mathbb{P}(A_i) = \mathbb{P}\left(\bigcap_{i \in I} A_i\right),$$

para todo par de conjuntos disjuntos de índices $I, J \subset \{1, \ldots, n\}$ tales que el suceso $\bigcap_{j \in J} A_j$ tiene probabilidad estrictamente positiva. Esto significa que la ocurrencia de cada uno de los sucesos A_j, con $j \in J$, no hace más favorable ni menos favorable la ocurrencia simultánea de los sucesos A_i, con $i \in I$ (cuando I y J no tienen índices en común).

Ejemplo 1.13. Considere el experimento donde una persona escoge al azar y equiprobablemente uno de los siguientes números: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18. *¿Son independientes los siguientes sucesos?*

$$\begin{aligned} A_1 &= \{1, 2, 3, 4, 5, 6\}, \\ A_2 &= \{4, 5, 6, 7, 8, 9, 10, 11, 12\}, \\ A_3 &= \{1, 2, 10, 11, 12, 13\}. \end{aligned}$$

Observe que $(A_1 \cap A_2 \cap A_3) = \emptyset$; en particular, $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 0$. Como $\mathbb{P}(A_1) > 0$, $\mathbb{P}(A_2) > 0$ y $\mathbb{P}(A_3) > 0$ concluimos que

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) \neq \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3),$$

i.e. A_1, A_2 y A_3 no son sucesos independientes.

Note que $\mathbb{P}(A_1 \cap A_2) = 1/6 = 1/3 \cdot 1/2 = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$; en particular, A_1 y A_2 son independientes. Similarmente, $\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_3)$ y también $\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2) \cdot \mathbb{P}(A_3)$. Este ejemplo muestra que las identidades en (1.17)-(1.19) no son en general suficientes para concluir la identidad en (1.20).

En los ejemplos presentados, hemos abordado la cuestión de independencia calculando probabilidades directamente. Desafortunadamente, para comprobar la independencia de n sucesos hay que verificar $(2^n - n - 1)$ identidades. Esto hace casi imposible de verificar por definición la independencia de más que unos pocos sucesos. Sin embargo, en muchas situaciones prácticas, la independencia de dos o más sucesos es postulada basada en la intuición que se tiene sobre el experimento. Por ejemplo, considere el experimento donde se registra la información del próximo paciente que visitará un cierto dentista y los sucesos

$$\begin{aligned} A &= \text{“el paciente es una mujer”}, \\ B &= \text{“el paciente padece de dolor en una muela del juicio”}, \\ C &= \text{“el paciente acarrea varios anillos en sus manos”}. \end{aligned}$$

Es natural postular que A y B son sucesos independientes: la posibilidad de A no debiera realmente cambiar si se sabe que el paciente tiene dolor muelas, en símbolos $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. La intuición detrás de esto es que la anatomía de la boca de un hombre no es diferente a la de una mujer. Sin embargo, es menos claro que A es independiente de C, y de hecho uno esperaría que $\mathbb{P}(A \mid C) > 1/2 = \mathbb{P}(A)$.

Ejemplo 1.14. Jorge es un colegial muy olvidadizo: el 1 % de las veces se olvida de su corbata, el 5 % de su estuche de lápices, y el 10 % de su cuaderno de matemáticas, *¿qué fracción de días Jorge se olvida de traer su estuche, cuaderno o corbata?* La respuesta viene dada por la probabilidad del suceso $(A \cup B \cup C)$, donde

$$\begin{aligned} A &= \text{“Jorge se olvidó de su corbata”}, \\ B &= \text{“Jorge se olvidó de su estuche de lápices”}, \\ C &= \text{“Jorge se olvidó de su cuaderno de matemáticas”}. \end{aligned}$$

Para calcular esta probabilidad usaremos la llamada *Fórmula de Inclusión-Exclusión* (Ejercicio 22) que establece lo siguiente:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Bajo el supuesto que los tres sucesos de arriba son independientes obtenemos que

$$\begin{aligned} \mathbb{P}(A \cup B \cup C) &= 0{,}01 + 0{,}05 + 0{,}1 - 0{,}01 \cdot 0{,}05 - 0{,}01 \cdot 0{,}1 - 0{,}05 \cdot 0{,}1 + 0{,}01 \cdot 0{,}05 \cdot 0{,}1, \\ &= 0{,}15355. \end{aligned}$$

Es importante destacar que sin el supuesto de independendencia, no es posible determinar las probabilidades de los sucesos $(A \cap B)$, $(A \cap C)$, $(B \cap C)$ ni $(A \cap B \cap C)$ a partir de las probabilidades de los sucesos A, B y C solamente. En este caso, necesitamos mayor información acerca del comportamiento olvidadizo de Jorge para determinar la probabilidad del suceso $(A \cup B \cup C)$ e.g. necesitamos la probabilidad condicional que éste se olvide de su corbata dado que se olvidó de su estuche y de su cuaderno. Note, sin embargo, que si conocemos las probabilidades condicionales $\mathbb{P}(A \mid B)$, $\mathbb{P}(A \mid C)$, $\mathbb{P}(A \mid B \cap C)$ y $\mathbb{P}(B \mid C)$ entonces la probabilidad de $(A \cup B \cup C)$ puede calcularse como (*¿por qué?*)

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

Ejemplo 1.15. Considere el experimento donde se tira un dado equilibrado hasta observar la cara 6 por primera vez (Ejemplo 1.3). Un espacio muestral natural para este experimento es $\Omega = \{(6), (1,6), (2,6), (3,6), (4,6), (5,6), (1,1,6), \ldots\}$. Para cada $\omega \in \Omega$, denotaremos como p_ω la probabilidad del suceso $\{w\}$.

Bajo el supuesto intuitivo que, cualquier suceso asociado a un cierto lanzamiento del dado es independiente de todos los lanzamientos previos y futuros del mismo, podemos fácilmente determinar p_ω como sigue.

Por ejemplo, si definimos los sucesos

$$\begin{aligned} A_1 &= \text{“un 1 sale en el primer lanzamiento”}, \\ A_2 &= \text{“un 1 sale en el segundo lanzamiento”}, \\ A_3 &= \text{“un 6 sale en el tercer lanzamiento”}, \end{aligned}$$

entonces bajo el postulado que son independientes obtenemos que

$$p_{(1,1,6)} = \mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) = (1/6)^3 .$$

Similarmente, si definimos los sucesos

$$\begin{aligned} B_1 &= \text{“un 5 sale en el primer lanzamiento”}, \\ B_2 &= \text{“un 4 sale en el segundo lanzamiento”}, \\ B_3 &= \text{“un 4 sale en el tercer lanzamiento”}, \\ B_4 &= \text{“un 6 sale en el cuarto lanzamiento”}, \end{aligned}$$

entonces al postular su independencia obtenemos que

$$p_{(5,4,4,6)} = \mathbb{P}(B_1 \cap B_2 \cap B_3 \cap B_4) = \mathbb{P}(B_1) \cdot \mathbb{P}(B_2) \cdot \mathbb{P}(B_3) \cdot \mathbb{P}(B_4) = (1/6)^4 .$$

Análogamente, para cada $\omega \in \Omega$, se puede demostrar la fórmula dada sin justificación en el Ejemplo 1.3: $p_\omega = (1/6)^{n(\omega)}$, donde $n(\omega)$ denota el número de coordenadas de ω.

1.5 Secuencias monótonas de sucesos

En lo que sigue $A_1, A_2, \ldots$ denotará una secuencia enumerable de sucesos asociada con un cierto experimento. En esta sección discutiremos dos resultados que, en ciertas situaciones, nos permitirán determinar la probabilidad que al menos uno o también cada uno de estos sucesos ocurra una vez realizado el experimento. El resultado principal se basa en las siguientes definiciones.

Definición 1.6. La secuencia de sucesos $A_1, A_2, \ldots$ se dice *creciente* si para todo $n \geq 1$, $A_n \subset A_{n+1}$. Por otro lado, $A_1, A_2, \ldots$ se dirá *decreciente* si para todo $n \geq 1$, $A_{n+1} \subset A_n$.

Para que la secuencia $A_1, A_2, , \ldots$ sea creciente la propiedad que define cada suceso A_n tiene que ser más exigente que la que define al suceso A_{n+1}. Por otro lado, para que $A_1, A_2, \ldots$ sea decreciente, la propiedad que define a A_n tiene que ser menos exigente que la que define a A_{n+1}. Para fijar ideas consideremos el siguiente ejemplo.

Ejemplo 1.16. Fernando es un estudiante de colegio y un apasionado jugador de pichanga durante sus recreos. Para cada $n \geq 1$ considere el suceso A_n definido como "el equipo de Fernando ha perdido al menos una de las primeras n pichangas". Claramente la ocurrencia de A_n implica la ocurrencia de A_{n+1}: si el equipo del joven ha perdido al menos uno de los primeros n encuentros entonces su equipo también ha perdido al menos uno de los primeros $(n+1)$ encuentros. Esto muestra que $A_n \subset A_{n+1}$ y, por lo tanto, la secuencia de sucesos $A_1, A_2, \ldots$ es creciente.

Por otro lado, considere el suceso B_n definido como "Fernando ha llegado tarde a clases después de los primeros n recreos". Claramente, si Fernando ha llegado atrasado después de los primeros $(n+1)$ recreos, entonces él también ha llegado atrasado después de los primeros n recreos i.e. $B_{n+1} \subset B_n$. Por lo tanto $B_1, B_2, \ldots$ es una secuencia decreciente de sucesos.

Finalmente, considere el suceso C_n definido como "el equipo de Fernando ganó la n-ésima pichanga". Observe que la ocurrencia de C_n no asegura la ocurrencia de C_{n+1} ni viceversa. Por lo tanto, la secuencia de sucesos $C_1, C_2, \ldots$ no es creciente ni tampoco decreciente.

El ejemplo recién propuesto muestra que es posible definir muchas secuencias crecientes o decrecientes de sucesos asociadas a un mismo experimento, aunque hay secuencias de sucesos que no son crecientes ni decrecientes.

El siguiente resultado nos permitirá calcular la probabilidad de una unión creciente de sucesos.

Teorema 1.7. *Si $A_1, A_2, \ldots$ es una secuencia creciente de sucesos entonces*

$$\mathbb{P}\left(\begin{array}{c} \text{al menos uno de los} \\ \text{sucesos } A_n \text{ ocurre} \end{array}\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Demostración. La secuencia de números reales $\mathbb{P}(A_1), \mathbb{P}(A_2), \ldots$ es creciente y acotada entre cero y uno. Esto se debe a las propiedades (1.10) y (1.12): como $A_n \subset A_{n+1}$ entonces $0 \leq \mathbb{P}(A_n) \leq \mathbb{P}(A_{n+1}) \leq 1$. En particular, el límite al lado derecho superior existe y es finito. Para demostrar el teorema reescribiremos la unión de los sucesos $A_1, A_2, \ldots$ de la siguiente manera (ver Figura 1.1):

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup \bigcup_{n=2}^{\infty} (A_n \setminus A_{n-1}).$$

Observe que $\mathbb{P}(A_n \setminus A_{n-1}) = \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})$ debido a la propiedad (1.11). Como los sucesos A_1, $(A_2 \setminus A_3)$, $(A_3 \setminus A_2), \ldots$ son disjuntos, la sigma aditividad implica que

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbb{P}(A_1) + \sum_{n=2}^{\infty} \mathbb{P}(A_n \setminus A_{n-1}), \\
&= \mathbb{P}(A_1) + \sum_{n=2}^{\infty} \{\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})\}, \\
&= \mathbb{P}(A_1) + \lim_{k\to\infty} \sum_{n=2}^{k} \{\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})\}, \\
&= \mathbb{P}(A_1) + \lim_{k\to\infty} \{\mathbb{P}(A_k) - \mathbb{P}(A_1)\} = \lim_{k\to\infty} \mathbb{P}(A_k),
\end{aligned}$$

donde hemos usado que la última sumatoria es telescópica. Esto completa la demostración del teorema. □

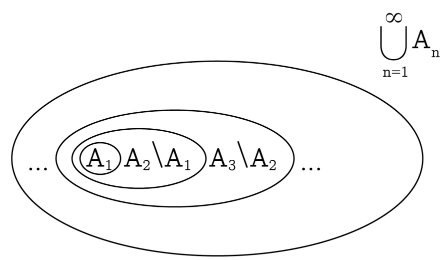


FIGURA 1.1. Representación de una unión creciente como una unión contable y disjunta de sucesos.

Ejemplo 1.17. *Muestre que si $A_1, A_2, \ldots$ es una secuencia creciente de sucesos tales que $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$ entonces $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A \cap A_n)$, para todo suceso A.*

Primero note que $(A \cap A_n) \subset (A \cap A_{n+1})$ debido a que $A_n \subset A_{n+1}$. Por lo tanto, la secuencia de sucesos $(A \cap A_1), (A \cap A_2), \ldots$ es creciente y el Teorema 1.7 implica que

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} (A \cap A_n)\right) = \lim_{n\to\infty} \mathbb{P}(A \cap A_n). \tag{1.21}$$

Ahora demostraremos que

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

usando la equivalencia lógica en (1.1). En efecto, si $\omega \in \bigcup_{n=1}^{\infty}(A \cap A_n)$ entonces existe n tal que $\omega \in (A \cap A_n)$; en particular, $\omega \in A$ y $\omega \in \bigcup_{n=1}^{\infty} A_n$, por lo tanto, $\omega \in A \cap (\bigcup_{n=1}^{\infty} A_n)$. Recíprocamente, si $\omega \in A \cap (\bigcup_{n=1}^{\infty} A_n)$ entonces $\omega \in A$ y existe n tal que $\omega \in A_n$; en particular, existe n tal que $\omega \in (A \cap A_n)$ y, por lo tanto, $\omega \in \bigcup_{n=1}^{\infty}(A \cap A_n)$. Hemos así demostrado que $\omega \in \bigcup_{n=1}^{\infty}(A \cap A_n)$ si y solo si $\omega \in A \cap (\bigcup_{n=1}^{\infty} A_n)$, lo que demuestra la igualdad de sucesos anteriores. En particular, usando la identidad en (1.21) obtenemos que

$$\mathbb{P}\left(A \cap \left(\bigcup_{n=1}^{\infty} A_n\right)\right) = \lim_{n\to\infty} \mathbb{P}(A \cap A_n). \tag{1.22}$$

Para finalizar note que $A_k \subset \bigcup_{n=1}^{\infty} A_n$, para cada $k \geq 1$, y por lo tanto $\mathbb{P}(A_k) \leq \mathbb{P}(\bigcup_{n=1}^{\infty} A_n)$, debido a la identidad en (1.12). Como el lado izquierdo de la última desigualdad converge a 1 a medida que n tiende a infinito, concluimos que $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = 1$. En particular, debido a la identidad en (1.13), tenemos que

$$\mathbb{P}\left(A \cap \left(\bigcup_{n=1}^{\infty} A_n\right)\right) = \mathbb{P}(A). \tag{1.23}$$

El resultado del ejemplo es ahora una consecuencia inmediata de las identidades en (1.22) y (1.23).

Toda secuencia creciente de sucesos puede transformarse en una secuencia decreciente y viceversa al tomar complementos. Esto se debe a que al tomar complementos invertimos la relación de inclusión i.e. si $A \subset B$ entonces $A^c \supset B^c$ (*¿por qué?*). Por otro lado, el complemento de una unión corresponde siempre a la intersección de los complementos. Esta última propiedad es parte de lo que se conoce en la teoría de conjuntos como las *Leyes de Morgan*. Matemáticamente tenemos que

$$\left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty}(A_n^c). \tag{1.24}$$

Al lado izquierdo de arriba, $\cap_{n=1}^{\infty} A_n$ denota la *intersección de los sucesos* $A_1, A_2, \ldots$ Al igual que en el caso de un número finito de intersecciones, este suceso ocurre solamente cuando cada uno de los sucesos que participa de la intersección se produce una vez llevado a cabo el experimento. Matemáticamente, esto significa que

$$\omega \in \bigcap_{n=1}^{\infty} A_n \iff \omega \in A_n, \text{ para todo } n.$$

Usando las observaciones de arriba podemos demostrar lo siguiente.

Teorema 1.8. *Si $A_1, A_2, \ldots$ es una secuencia decreciente de sucesos entonces*

$$\mathbb{P}\begin{pmatrix} \textit{cada uno de los} \\ \textit{sucesos } A_n \textit{ ocurre} \end{pmatrix} = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Demostración. Debido a la Ley de Morgan en (1.24), el suceso $\cup_{n=1}^{\infty} A_n^c$ es el complemento del suceso $\cap_{n=1}^{\infty} A_n$. En particular, y usando la propiedad en (1.9) obtenemos que

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

Note, sin embargo, que $A_1^c, A_2^c, \ldots$ etc es una secuencia creciente de sucesos (*¿por qué?*). Por lo tanto, y de acuerdo al Teorema 1.7, tenemos que

$$\begin{aligned}\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1 - \lim_{n\to\infty} \mathbb{P}(A_n^c), \\ &= 1 - \lim_{n\to\infty} \{1 - \mathbb{P}(A_n)\} = \lim_{n\to\infty} \mathbb{P}(A_n).\end{aligned}$$

Esto demuestra el teorema. □

Ejemplo 1.18. Considere el experimento que consiste en lanzar una moneda equilibrada un número infinito de veces, *¿cuál es la probabilidad de que todos los lanzamientos sean cara?* ¡Intuitivamente la respuesta debiera ser cero! Matemáticamente esto puede justificarse como sigue. Considere los sucesos

$$\begin{aligned} A &= \text{“la moneda sale cara en todos los lanzamientos”}, \\ A_n &= \text{“la moneda sale cara en los primeros } n \text{ lanzamientos”}. \end{aligned}$$

Demostraremos usando la equivalencia lógica en (1.1) que

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

En efecto, observe que $A \subset \cap_{n=1}^{\infty} A_n$ debido a que si A ocurre entonces cada uno de los sucesos A_n tiene que haber ocurrido también. Por otro lado, si para cada $n \geq 1$, la moneda salió cara en los primeros n lanzamientos, entonces ésta tiene que haber salido cara en todos los lanzamientos. Esto significa que $\cap_{n=1}^{\infty} A_n \subset A$, lo que demuestra la identidad anterior.

Note ahora que si la moneda sale cara en los primeros $(n+1)$ lanzamientos, entonces ésta tiene que haber salido cara en los primeros n lanzamientos. En particular, $A_{n+1} \subset A_n$ y, por lo tanto, $A_1, A_2, \ldots$ es una secuencia decreciente. Usando el Teorema 1.8, la identidad de arriba implica que

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

Pero observe que para cada n, $A_n = (C_1 \cap \cdots \cap C_n)$, donde C_i es el suceso “la moneda sale cara en el i-ésimo lanzamiento”. Como los lanzamientos de la moneda son independientes, estos sucesos también debieran ser independientes. En particular,

$\mathbb{P}(A_n) = \mathbb{P}(C_1) \cdots \mathbb{P}(C_n) = (1/2)^n$. Concluimos finalmente que

$$\mathbb{P}(A) = \lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0,$$

i.e. la probabilidad que todos los lanzamientos salgan cara es cero.

Para finalizar esta sección consideraremos un ejemplo más elaborado, pero que ilustra aplicaciones insospechadas de las ideas que hemos estado discutiendo.

Ejemplo 1.19. Mateo es un joven soltero muy interesado en la genealogía de su familia. En sus estudios ha determinado que sus antepasados han tenido exactamente cero, uno o dos descendientes con frecuencias aproximadas de 1/4, 1/12 y 2/3, respectivamente. Un día Mateo se pregunta: *¿cuál es la probabilidad que mi descendencia desaparezca algún día?*

Para responder la pregunta debemos determinar la probabilidad del suceso

$$A = \text{“la descendencia de Mateo desaparecerá algún día”}.$$

Considere para cada $n \geq 0$ el suceso

$$A_n = \begin{array}{c}\text{“la descendencia de Mateo se extinguió}\\ \text{dentro de las primeras } n \text{ generaciones”}\end{array}.$$

Observe que $A_0 = \emptyset$ debido a que Mateo constituye la generación cero. Claramente, $A_n \subset A_{n+1}$; en particular, la secuencia $A_1, A_2, \ldots$ es creciente. Como $A = \cup_{n=1}^{\infty} A_n$ (*¿por qué?*), el Teorema 1.7 implica que $\mathbb{P}(A) = \lim_{n\to\infty} p_n$, donde $p_n = \mathbb{P}(A_n)$.

La secuencia de números $(p_n)_{n\geq 0}$ es creciente, $p_0 = 0$, y

$$\mathbb{P}(A) = \lim_{n\to\infty} p_n. \tag{1.25}$$

Para determinar este valor límite debemos entender mejor la secuencia p_n. Para esto, y de acuerdo a la Fórmula de Probabilidades Totales en la identidad (1.16), observe que la siguiente igualdad aplica (*¿por qué?*):

$$\begin{aligned} p_{n+1} = {} & \mathbb{P}[A_{n+1} \mid \text{Mateo no tuvo descendientes}] \cdot 1/4 \\ & + \mathbb{P}[A_{n+1} \mid \text{Mateo tuvo sólo un descendiente}] \cdot 1/12 \\ & + \mathbb{P}[A_{n+1} \mid \text{Mateo tuvo sólo dos descendientes}] \cdot 2/3. \end{aligned} \tag{1.26}$$

Claramente,

$$\mathbb{P}[A_{n+1} \mid \text{Mateo no tuvo descendientes}] = 1,$$

ya que si Mateo no tuvo descendientes entonces su descendencia se extinguió dentro de las primeras $(n+1)$ generaciones.

Por otro lado, si Mateo tuvo exactamente un descendiente (al cual nos referiremos como Mateo Dos) entonces

$$\begin{aligned}
\mathbb{P}[A_{n+1} \mid \text{Mateo tuvo sólo un descendiente}] &= \mathbb{P}\begin{bmatrix}\text{la descendencia de Mateo} \\ \text{Dos se extinguió dentro de} \\ \text{las primeras } n \text{ generaciones}\end{bmatrix}, \\
&= \mathbb{P}\begin{bmatrix}\text{la descendencia de Mateo} \\ \text{se extinguió dentro de las} \\ \text{primeras } n \text{ generaciones}\end{bmatrix}, \\
&= \mathbb{P}(A_n) = p_n,
\end{aligned}$$

donde para la segunda igualdad hemos supuesto que la suerte de los descendientes de Mateo Dos es la misma que la de los de Mateo. Finalmente, supongamos que Mateo tuvo exactamente dos descendientes, llamémosles Mateo Dos y Mateo Tres. Tenemos entonces que

$$\begin{aligned}
&\mathbb{P}[A_{n+1} \mid \text{Mateo tuvo sólo dos descendientes}] \\
&= \mathbb{P}\begin{bmatrix}\text{la descendencia de Mateo Dos se extinguió dentro de } n \text{ generaciones, y} \\ \text{la descendencia de Mateo Tres se extinguió dentro de } n \text{ generaciones}\end{bmatrix}, \\
&= \mathbb{P}\begin{bmatrix}\text{la descendencia de Mateo} \\ \text{Dos se extinguió dentro} \\ \text{de } n \text{ generaciones}\end{bmatrix} \cdot \mathbb{P}\begin{bmatrix}\text{la descendencia de Mateo} \\ \text{Tres se extinguió dentro} \\ \text{de } n \text{ generaciones}\end{bmatrix}, \\
&= \mathbb{P}(A_n) \cdot \mathbb{P}(A_n) = p_n^2,
\end{aligned}$$

donde para la segunda igualdad hemos supuesto que sucesos asociados a los descendientes de Mateo Dos son independientes de sucesos asociados a los descendientes de Mateo Tres, y para la tercera igualdad que la suerte de los descendientes de Mateo Dos y Tres es la misma que la de los de Mateo.

Usando las identidades anteriores en la identidad (1.26) obtenemos la recursión

$$p_{n+1} = \frac{1}{4} + \frac{1}{12}p_n + \frac{2}{3}p_n^2 = f(p_n), \tag{1.27}$$

donde f es el polinomio cuadrático definido como $f(x) = 1/4 + x/12 + 2x^2/3$. Como el límite $\lim_{n\to\infty} p_n$ existe y, además, f es una función continua, deducimos que

$$\lim_{n\to\infty} p_{n+1} = \lim_{n\to\infty} f(p_n) = f\left(\lim_{n\to\infty} p_n\right).$$

Como $(p_{n+1})_{n\geq 0}$ es una subsecuencia de $(p_n)_{n\geq 0}$, la igualdad en (1.25) implica que

$$\mathbb{P}(A) = f\big(\mathbb{P}(A)\big).$$

Hemos así determinado, bajo los supuestos anteriores, que la probabilidad que la descendencia de Mateo desaparezca algún día es una solución de la ecuación $x = f(x)$

i.e. es un *punto fijo* de la función f. Como f es un polinomio cuadrático, esta ecuación tiene a lo más dos soluciones. En efecto, uno encuentra que

$$\begin{aligned} x = f(x) \quad &\Longleftrightarrow \quad 8x^2 - 11x + 3 = 0, \\ &\Longleftrightarrow \quad (x - 3/8) \cdot (x - 1) = 0. \end{aligned}$$

Por lo tanto, $\mathbb{P}(A) = 3/8$ o bien $\mathbb{P}(A) = 1$. Claramente estos dos escenarios son incompatibles y resta decidir cuál es el escenario correcto.

Justificaremos que $\mathbb{P}(A) = 3/8$. Para esto observe que $x = 3/8$ es la solución más pequeña de la ecuación $x = f(x)$. Por otro lado, como f es una función creciente y $p_0 \leq 3/8$ (recuerde que $p_0 = 0$) podemos usar (1.27) para concluir que $p_1 = f(p_0) \leq f(3/8) = 3/8$. Similarmente, $p_2 = f(p_1) \leq f(3/8) = 3/8$. El lector podrá corroborar inductivamente que $f(p_n) \leq 3/8$, para todo $n \geq 0$. Haciendo n tender a infinito, (1.25) implica que $f(\mathbb{P}(A)) \leq 3/8$ i.e. $\mathbb{P}(A) \leq 3/8$. Por lo tanto, $\mathbb{P}(A) = 3/8$ ya que $3/8$ es el punto fijo más pequeño de f. Esto demuestra nuestra afirmación, aunque y desafortunadamente para Mateo, existe una posibilidad de un 37.5 % que su descendencia desaparezca algún día.

1.6 Ejercicios propuestos

Sección 1.1.

Ejercicio 1. Considere el experimento que consiste en escoger un punto z al azar en el cuadrado con vértices $(0,0)$, $(1,0)$, $(1,1)$ y $(0,1)$. Usando el espacio muestral $\Omega = \{(x,y) \mid 0 \leq x, y \leq 1\}$ represente cada uno de los siguientes sucesos como un subconjunto de Ω:

$$\begin{aligned} A &= \text{“la distancia entre } z \text{ y el origen es menor o igual a uno”}, \\ B &= \text{“la distancia entre } z \text{ y el origen es mayor o igual a dos”}, \\ C &= \text{“la distancia entre } z \text{ y } (0,1) \text{ es la misma que entre } z \text{ y } (1,0)\text{”}, \\ D &= \text{“}z \text{ es equidistante a } (0,0), (1,0), (1,1) \text{ y } (0,1)\text{”}. \end{aligned}$$

Ejercicio 2. Considere el experimento que consiste primero en lanzar una moneda y luego tirar un dado. Representaremos los resultados posibles de este experimento como pares ordenados de la forma (x,y), donde $x \in \{cara, sello\}$ e $y \in \{1,2,3,4,5,6\}$. Considere los sucesos

$$\begin{aligned} A &= \text{“la moneda sale cara y el dado lanzado es un 4”}, \\ B &= \text{“la moneda sale cara”}, \\ C &= \text{“el dado lanzado es un 3”}, \\ D &= \text{“el dado lanzado es un número par”}. \end{aligned}$$

(a) Reescriba cada uno de los sucesos anteriores como un conjunto de resultados. (b) ¿Está A incluido en B? (c) ¿Está A incluido en C? (d) ¿Cuál es el complemento de A? (e) ¿Cuál es el complemento de B? (f) Muestre que $(A \cup B) = B$. (g) Explique en palabras lo que el suceso $(A \cup D)$ representa, ¿es cierto que $(A \cup D) = (B \cup D)$? (h) ¿Son A y C disjuntos?, ¿qué hay de A y D?

Ejercicio 3. En lo que sigue A, B y C denotarán sucesos asociados con un cierto experimento. Clasifique cada una de las siguientes afirmaciones como verdadera o falsa. En caso de verdadera, explique y/o demuestre la afirmación. En caso de falsa, entregue un contraejemplo.

(a) Si es posible que A y B ocurran simultáneamente entonces $A = B$.
(b) Si cada vez que A ocurre, B no lo hace, entonces $A \subset B^c$.
(c) Si A no ocurre entonces B tampoco ocurre, y viceversa, entonces $A = B$.
(d) Si A y B no pueden ocurrir simultáneamente entonces $(A \cap B) = \emptyset$.
(e) Si cada vez que A y B ocurren, C también ocurre, entonces $C \subset (A \cup B)$.
(f) Si cada vez que A ocurre, B y C ocurren, entonces $A \subset (B \cap C)$.

Sección 1.2.

Ejercicio 4. Una persona arroja al azar una pelota saltarina dentro de un cilindro cuya base tiene pintado los colores rojo, blanco y azul en las siguientes proporciones (ver Figura 1.2): 50 % rojo, 37.5 % blanco, y 12.5 % azul. ¿Cuál es la probabilidad de que la pelota quede estática sobre la zona blanca?, ¿cuál es la probabilidad de que la pelota repose sobre la zona roja o blanca? Después de 60 lanzamientos, ¿cuántas veces se espera que la bola termine reposando sobre la zona blanca?, ¿qué hay acerca de la zona roja o blanca?

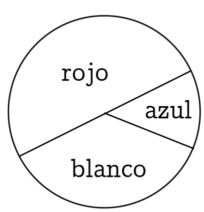


Figura 1.2

Ejercicio 5. Considere un colegio cuyos alumnos satisfacen las estadísticas entregadas en la Tabla 1.1. Suponiendo que un estudiante se escoge al azar, ¿cuál es la probabilidad de que éste sea varón?, ¿cuál es la probabilidad de que éste sea mujer y su madre tenga un grado profesional?, ¿cuál es la probabilidad de que el padre no tenga un grado profesional?

Ejercicio 6. Raúl tiene tres monedas de \$50 y dos de \$100 en su bolsillo derecho, y cinco monedas de \$50 y una de \$100 en su bolsillo izquierdo. (a) ¿Cuál es la probabilidad de que éste saque una moneda de \$100 de su bolsillo derecho?, ¿qué hay

	Padre con grado profesional	Padre sin grado profesional	Madre con grado profesional	Madre sin grado profesional
Estudiantes mujeres	25.6 %	8.4 %	16.8 %	17.2 %
Estudiantes varones	37.9 %	28.1 %	52.6 %	13.4 %

Tabla 1.1

de su bolsillo izquierdo? Asuma que Raúl saca las monedas al azar de sus bolsillos. (b) ¿Cuántas monedas de \$100 tiene Raúl que agregar a cada bolsillo, para que la probabilidad de sacar una moneda de \$100 sea la misma en ambos bolsillos?

Sección 1.3.

Ejercicio 7. En el Ejercicio 4: ¿cuál es la probabilidad de que la pelota termine reposando en la zona azul, dado que terminó reposando en la zona blanca o azul?, ¿cuál es la probabilidad de que ésta termine reposando en la zona roja, dado que no terminó reposando en la zona azul?

Ejercicio 8. Considere el experimento donde se lanza un dado equilibrado dos veces y el suceso A=“la primera cara es un 2”. Determine la probabilidad condicional $\mathbb{P}(A \mid B)$ para cada uno de los siguientes casos: (a) B=“la primera cara es un 2”, (b) B=“la primera cara es un 4”, (c) B=“la segunda cara es un 4”, (d) B=“la suma de la primera y segunda cara es un 4”.

Ejercicio 9. Devuelta en las estadísticas del colegio en la Tabla 1.1, considere los siguientes sucesos, relacionados con la selección de un estudiante al azar:

$$\begin{aligned} A &= \text{“el estudiante es mujer”;} \\ B &= \text{“el padre del estudiante tiene un grado profesional”;} \\ C &= \text{“la madre del estudiante tiene un grado profesional”.} \end{aligned}$$

(a) Determine la probabilidad condicional de B dado A, y de B dado A^c. Basado en sus cálculos, ¿diría usted que es más probable que una estudiante mujer tenga un padre con un grado profesional, que un estudiante varón? (b) Determine la probabilidad condicional de C dado A, y de C dado A^c, ¿diría usted que es más probable que una estudiante mujer tenga una madre con un grado profesional, que un estudiante varón? (c) Basado en lo anterior, ¿tiene alguna influencia el sexo de un alumno sobre el grado profesional del padre?, ¿qué ocurre en el caso de la madre?

Ejercicio 10. Juan tiene dos hermanos, ¿cuál es la probabilidad que su hermano menor sea mujer dado que al menos uno de sus hermanos es mujer?

Sección 1.4.

Ejercicio 11. Imagine un estudio cognitivo donde se le entrega a un chimpancé un teclado con los dígitos $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ y un punto decimal. Asumiendo que las teclas que marca el chimpancé son independientes de las que éste ha marcado en el pasado o marcará en el futuro, ¿cuál es la probabilidad que en un cierto experimento, éste comience escribiendo "3.1415"? Explique los supuestos hechos en su cálculo.

Ejercicio 12. Imagine que usted tira una moneda equilibrada tres veces. ¿Qué pares entre los siguientes sucesos son independientes?, ¿son estos tres sucesos independientes?

$$\begin{aligned} A &= \text{“la moneda sale cara en el segundo lanzamiento”}, \\ B &= \text{“la moneda sale cara en exactamente un lanzamiento”}, \\ C &= \text{“la moneda sale sello en el primer y tercer lanzamiento”}. \end{aligned}$$

Ejercicio 13. Considere nuevamente el juego discutido en el Ejemplo 1.12. (a) Si su amigo afirma primero que "el segundo lanzamiento saldrá sello" y al tirar la moneda usted obtiene el resultado $(cara, cara, sello)$ y reporta devuelta que "el primer y segundo lanzamiento son idénticos", ¿ha entregado usted información útil a su amigo? Explique. (b) Si su amigo afirma que "el primer y segundo lanzamiento son idénticos" y al tirar la moneda usted obtiene el resultado $(sello, cara, sello)$. ¿Qué afirmación podría usted reportar devuelta para no dar información útil a su amigo? Explique.

Sección 1.5.

Ejercicio 14. Considere sucesos $A_1, A_2, \ldots$ tales que $\mathbb{P}(A_i) = 1$, para cada $i \geq 1$. (a) Muestre que $\mathbb{P}(\cap_{i=1}^{n} A_i) = 1$, para todo $n \geq 1$. (b) Use lo anterior para concluir que $\mathbb{P}(\cap_{i=1}^{\infty} A_i) = 1$. (En palabras, la intersección de un número contable de sucesos seguros i.e. con probabilidad 1 cada uno, es también un suceso seguro.)

Ejercicio 15. Considere el experimento donde se lanza indefinida e independientemente una moneda con probabilidad p de salir cara. Muestre que si $p < 1$ entonces la probabilidad de que todos los lanzamientos salgan cara es cero. Muestre que si $p = 1$ entonces todos los lanzamientos saldrán cara con probabilidad uno.
Hint: Para la segunda parte utilice el Ejercicio 14.

Ejercicio 16. Considere el experimento donde se lanza indefinida e independientemente un dado equilibrado y el suceso A definido como "a partir de cierto momento la cara lanzada será 5 ". (a) Para cada $n \geq 1$ considere el suceso C_n definido como "el n-ésimo lanzamiento resulta en la cara 5". ¿Qué representa el suceso $\cap_{k=n}^{\infty} C_k$? (b) Muestre que $\mathbb{P}(\cap_{k=n}^{\infty} C_k) = 0$, para cada $n \geq 1$. (c) Muestre que $A = \cup_{n=1}^{\infty}(\cap_{k=n}^{\infty} C_k)$, y use el Teorema 1.7 para concluir que $\mathbb{P}(A) = 0$.

Problemas misceláneos Capítulo 1.

Ejercicio 17. Volviendo al Ejercicio 4, imagine ahora que la persona arroja dos veces la pelota saltarina adentro del cilindro: ¿cuál es la probabilidad de cada uno de los siguientes sucesos?

$$\begin{aligned}
A &= \text{“la primera vez, la pelota reposa en la zona roja y la segunda, en la azul”;}\\
B &= \text{“la primera vez, la pelota reposa en la zona azul y la segunda, en la roja”;}\\
C &= \text{“la pelota reposa una vez en la zona roja y una vez en la zona azul”;}\\
D &= \text{“en ambas instancias la pelota no reposa en la zona blanca”.}
\end{aligned}$$

Ejercicio 18. Explique y/o demuestre las siguientes equivalencias lógicas.

(a) $A = B$ si y solo si $A^c = B^c$,
(b) $A \subset B$ si y solo si $B^c \subset A^c$.
(c) $A \subset B$ si y solo si $A \setminus B = \emptyset$.

Ejercicio 19. Jaime tiene dos monedas de \$50 y tres de \$100 en su bolsillo derecho, y seis monedas de \$50 y dos de \$100 en su bolsillo izquierdo. Si Jaime mete la mano en uno de sus bolsillos y saca una moneda al azar, ¿cuál es la probabilidad de que éste haya sacado la moneda de su bolsillo izquierdo, dado que la moneda que sacó era de \$100?

Ejercicio 20. Considere un experimento que tiene asociado un cierto espacio muestral finito Ω. Muestre que si cada uno de los resultados del experimento es igualmente probable que cualquier otro entonces, dos sucesos A y B son independientes si y solo si $|\Omega| \cdot |A \cap B| = |A| \cdot |B|$.

Ejercicio 21. Si el arquero de un equipo de fútbol protege exitosamente el arco, el 84 % de las veces, ¿cuál es la probabilidad de que éste protega exitosamente el arco en tres ataques del equipo oponente?, ¿cuál es la probabilidad de que éste protega exitosamente el arco en sólo dos de tres ataques del equipo oponente? Dado que el arquero protegió exitosamente el arco en sólo dos de tres ataques del equipo oponente, ¿cuál es la probabilidad condicional de que éste haya protegido exitosamente el primer y segundo ataque?

Ejercicio 22. (Fórmula de Inclusión-Exclusión.) Considere una medida de probabilidad $\mathbb{P}$ y sucesos A, B y C. Muestre que: (i) $\mathbb{P}(A\cup B) = \mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cap B)$, (ii) $\mathbb{P}(A\cup B\cup C) = \mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)-\mathbb{P}(A\cap B)-\mathbb{P}(A\cap C)-\mathbb{P}(B\cap C)+\mathbb{P}(A\cap B\cap C)$. **Hint:** Para la propiedad (i) muestre primero que $(A \cup B) = A \cup (B \cap A^c)$ y que $B = (B \cap A^c) \cup (B \cap A)$. Use que $(A \cup B \cup C) = (A \cup B) \cup C$ y que $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ para demostrar la propiedad (ii) a partir de la propiedad (i).

Ejercicio 23. Paulette va a cumplir 30 años y tiene tres tías, Elsa, María y Carmen, a las que les encanta preparar tortas de lúcuma o de chocolate para su cumpleaños. La tía Elsa prepara una torta de lúcuma el 40 % de las veces, la tía María el 60 %, y la tía Carmen el 40 %. (a) ¿Cuál es la probabilidad que la torta de cumpleaños de

Paulette sea de lúcuma si cada una de sus tías puede perfectamente preparar la torta? (b) Si treinta años después Paulette sólo se acuerda que su torta fue de chocolate, ¿cuál es la probabilidad que la tía Elsa haya preparado su torta 30 años atrás?

Ejercicio 24. (a) Muestre que si A y B son sucesos independientes entonces A y B^c son también independientes i.e. $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) \cdot \mathbb{P}(B^c)$. Use esto para demostrar directamente que si A y B son independientes entonces A^c y B son independientes, y también lo son A^c y B^c. (b) Más generalmente, muestre que $A_1, \ldots, A_n$ son sucesos independientes si y solo si $\mathbb{P}(\cap_{i=1}^n B_i) = \prod_{i=1}^n \mathbb{P}(B_i)$, para $B_1, \ldots, B_n$ tal que $B_i = A_i$ ó $B_i = A_i^c$.

Ejercicio 25. Pedro y Jorge son dos gemelos pequeños y muy competitivos. Considere el experimento donde primero Pedro y luego Jorge escogen un número entero al azar entre 1 y 5. Un espacio muestral posible para este experimento es el conjunto Ω de todos los pares ordenados (x, y), con $1 \leq x, y \leq 5$, donde x representa el número escogido por Pedro e y el escogido por Jorge.

(a) ¿Qué representa en palabras el suceso $\{(1,1), (1,2), (1,3), (1,4), (1,5)\}$?

(b) Para cada $1 \leq i \leq 5$, considere los sucesos $P_i =$ "Pedro escoge el número i" y $J_i =$ "Jorge escoge el número i". Represente P_2 y J_3 como subconjuntos de Ω.

(c) Para cada $(x, y) \in \Omega$ defina

$$p_{(x,y)} = \begin{cases} \frac{1}{5(6-x)}, & 1 \leq x \leq y \leq 5; \\ 0, & 1 \leq y < x \leq 5. \end{cases}$$

Determine $p_{(2,2)}$, $p_{(2,3)}$ y $p_{(3,2)}$.

(d) Muestre que la función $\mathbb{P}(A) = \sum_{(x,y) \in A} p_{(x,y)}$, definida para $A \subset \Omega$, es una medida de probabilidad. Use esto para determinar la probabilidad del suceso P_2, y también la probabilidad condicional del suceso J_3 dado P_2.

(e) Más generalmente, para cada $1 \leq i \leq j \leq 5$, determine $\mathbb{P}(P_i)$ y $\mathbb{P}(J_j \mid P_i)$. Basado en su cálculo determine si la siguiente afirmación es verdadera o falsa: Pedro escoge su número al azar entre 1 y 5, mientras que Jorge lo escoge entre el número que escogió Pedro y 5.

Ejercicio 26. Considere una medida de probabilidad $\mathbb{P}$ y un suceso B tal que $\mathbb{P}(B) > 0$. Muestre que si $A_1, A_2, \ldots$ es una secuencia contable de sucesos disjuntos entonces $\mathbb{P}(\cup_i A_i \mid B) = \sum_i \mathbb{P}(A_i \mid B)$.

Ejercicio 27. La identidad dada en el Ejercicio 26, aplica solamente cuando el suceso sobre el cual uno condiciona, permanece fijo. Muestre que este es el caso definiendo un experimento y tres sucesos A, B y C, con B y C disjuntos, tales que $\mathbb{P}(A \mid (B \cup C)) \neq \mathbb{P}(A \mid B) + \mathbb{P}(A \mid C)$.

Ejercicio 28. Imagine que el personaje Mateo en el Ejemplo 1.19 es de apellido paterno González. (a) Determine para cada $i \in \{0, 1, 2\}$ la probabilidad que Mateo tenga exactamente i hijos varones. (b) Usando los supuestos discutidos en este ejemplo, determine la probabilidad que en la descendencia de Mateo continuará para siempre el apellido González.

Ejercicio 29. Para cada entero $i \geq 1$ considere el suceso $A_i = \emptyset$. Use la sigma aditividad y que para cada $n \geq 1$, $\emptyset = \cup_{i=1}^{n} A_i$, para concluir que $\mathbb{P}(\emptyset) = 0$.
Hint: Observe que $A_1, \ldots, A_n$ son sucesos disjuntos. Demuestre el resultado por contradicción asumiendo que $\mathbb{P}(\emptyset) > 0$. (Esto demuestra la identidad en (1.8).)

Ejercicio 30. Use la sigma aditividad para demostrar que $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$, para cada suceso A. Use que $\mathbb{P}(A) \geq 0$ y $\mathbb{P}(A^c) \geq 0$, para concluir que $0 \leq \mathbb{P}(A) \leq 1$. (Esto demuestra las identidades en (1.9) y (1.10).)

Ejercicio 31. Use que si $A \subset B$ entonces $B = (B \setminus A) \cup A$ para demostrar que $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$. Concluya a partir de esto que $\mathbb{P}(A) \leq \mathbb{P}(B)$. (Esto demuestra las identidades en (1.11) y (1.12).)

Comentarios de final de capítulo

En el formalismo de la teoría de la medida, una colección $\mathcal{F}$ de subconjuntos del espacio muestral Ω es llamada *sigma-álgebra* si:

(a) $\Omega \in \mathcal{F}$,
(b) si $A \in \mathcal{F}$ entonces $A^c \in \mathcal{F}$, y
(c) si $A_1, A_2, \ldots \in \mathcal{F}$ es una secuencia contable de sucesos, entonces $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

En palabras, una sigma-álgebra de Ω es cualquier colección de subconjuntos del espacio muestral que contiene a Ω, y que es cerrada bajo complementación y uniones contables de conjuntos en ella. ($\mathcal{F}$ es llamada *álgebra* cuando la propiedad (c) aplica, en principio, solamente para secuencias finitas de sucesos en $\mathcal{F}$.) Generalmente, hay muchas sigma-álgebras asociadas a un mismo espacio muestral. Por ejemplo, para todo suceso $B \subset \Omega$, $\mathcal{F} = \{\Omega, \emptyset, B, B^c\}$ es una sigma-álgebra. Por otro lado, *el conjunto* $\mathcal{P}(\Omega)$ *de las partes de* Ω, definido como $\mathcal{P}(\Omega) = \{A \mid A \subset \Omega\}$, es otro ejemplo de sigma-álgebra. Este conjunto también es llamado *conjunto potencia de* Ω.

Dada una sigma-álgebra $\mathcal{F}$, una transformación $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ se dice que es una *medida de probabilidad* si satisface las condiciones dadas en (1.2)-(1.4), aunque la condición $\mathbb{P}(A) \geq 0$ es exigida sólo para $A \in \mathcal{F}$, y la sigma aditividad solamente cuando $A_1, A_2, \ldots$ es una secuencia contable de sucesos disjuntos en $\mathcal{F}$.

Un *espacio probabilístico* es un trío-ordenado $(\Omega, \mathcal{F}, \mathbb{P})$, donde $\mathcal{F}$ es una sigma-álgebra de subconjuntos de Ω y $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ es una medida de probabilidad.

La definición de sigma-álgebra es en gran parte motivada por las propiedades que definen una medida de probabilidad. En el caso de un espacio muestral contable, la frecuencia asintótica de cada uno de los resultados posibles es suficiente para definir $\mathbb{P}$ sobre el conjunto de las partes de Ω, tal como vimos en el Ejemplo 1.3. Por otro lado, cuando Ω es infinito aunque no contable, por ejemplo, $\Omega = \mathbb{R}$, a veces no es posible definir $\mathbb{P}(A)$, para todo $A \in \mathcal{P}(\Omega)$, y a la vez satisfacer las propiedades (1.2)-(1.4). El uso de sigma-álgebras es, por lo tanto, mucho más que un formalismo matemático. En el caso de espacios muestrales infinitos no contables, la construcción, y de hecho la existencia y unicidad de una medida de probabilidad $\mathbb{P}$, con ciertas características, está generalmente dada por el siguiente resultado.

Teorema 1.9. Teorema de Carathéodory.
Si $\mathcal{A}$ *es un álgebra y* $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ *es una función que satisface: (i)* $\mathbb{P}(\Omega) = 1$*; (ii)* $\mathbb{P}(A) \geq 0$*, para todo* $A \in \mathcal{F}$*; y (iii)* $\mathbb{P}(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$*, para toda secuencia finita de sucesos disjuntos* $A_1, \ldots, A_n \in \mathcal{F}$*; entonces existe una sigma-álgebra más pequeña* $\mathcal{F} \subset \mathcal{P}(\Omega)$ *tal que* $\mathcal{A} \subset \mathcal{F}$*, y* $\mathbb{P}$ *admite una única extensión a una medida de probabilidad* $\mathbb{P} : \mathcal{F} \to \mathbb{R}$*.*

Capítulo 2: Variables aleatorias

En este capítulo introduciremos el concepto de variable aleatoria. También definiremos la función de distribución asociada con una variable aleatoria y veremos que esta función sintetiza toda la información relevante para determinar probabilidades de sucesos relacionados con dicha variable. Luego introduciremos los conceptos de variable aleatoria discreta y continua. Variables aleatorias de este tipo son comunes en la mayoría de los experimentos y a lo largo de la monografía veremos una serie de propiedades satisfechas por cada una de ellas. En el Capítulo 4 presentaremos (con nombre y apellido) algunas de las variables aleatorias discretas y continuas más sobresalientes. Para finalizar el capítulo introduciremos el concepto de igualdad en distribución y el de independencia entre variables aleatorias. Ambas nociones son fundamentales para el modelamiento probabilístico de experimentos y serán clave en dos de los teoremas más celebrados de la teoría de probabilidades: la Ley de los Grandes Números y el Teorema Central Límite.

2.1 ¿Qué es una variable aleatoria?

En un sentido amplio una variable aleatoria es una cantidad numérica asociada al resultado de un experimento. El término “variable aleatoria” quiere, por lo tanto, decir “cantidad incierta”. El interés en estos objetos radica en el hecho, que en la mayoría de los experimentos uno no está interesado en el resultado per se, pero sí en una o varias cantidades numéricas asociadas a este resultado.

En términos más precisos, si Ω es el espacio muestral asociado con un cierto experimento, entonces una *variable aleatoria* (de ahora en adelante v.a.) asociada con este experimento es cualquier función $X : \Omega \to \mathbb{R}$. Como el resultado ω del experimento es incierto, entonces $X(\omega)$ es también una cantidad incierta. Generalmente usaremos las letras X, Y, Z, o a veces X_1, X_2, X_3, etc para denotar variables aleatorias.

Dado un conjunto $I \subset \mathbb{R}$ (generalmente un intervalo), el suceso “el número obtenido al transformar el resultado del experimento a través de la función X pertenece a I” será denotado como $[X \in I]$. Este suceso es nada menos que a la pre-imagen del conjunto I a través de la función X i.e.

$$[X \in I] = \{\omega \in \Omega \mid X(\omega) \in I\}.$$

Usualmente denotaremos la probabilidad de este suceso como $\mathbb{P}[X \in I]$.

Ejemplo 2.1. Para fijar ideas, considere el experimento donde se registra la asistencia a clase en un curso de treinta alumnos. Una v.a. asociada con este experimento es el número de alumnos presente en clase. Si denotamos este número como X entonces el suceso "ningún alumno se apareció en clase" puede reescribirse como $[X = 0]$. Por otro lado, el suceso "al menos dos tercios de los alumnos asistió a clase" corresponde a $[20 \leq X \leq 30]$. También podemos reescribir este suceso como $[X \geq 20]$. Esto se debe a que la pre-imagen del intervalo cerrado $[20, 30]$ bajo la transformación X es la misma que el intervalo $[20, +\infty)$ en este caso (*¿por qué?*).

Para determinar las probabilidades de los sucesos anteriores necesitamos más información acerca de los alumnos inscritos en el curso. Pese a esto, las siguientes propiedades aplican para cualquier medida de probabilidad $\mathbb{P}$ y v.a. X asociada con este experimento:

(2.1) $\mathbb{P}[X \in \mathbb{R}] = 1$,

(2.2) para todo conjunto $I \subset \mathbb{R}$, $\mathbb{P}[X \in I] \geq 0$,

(2.3) si $I_1, I_2, \ldots \subset \mathbb{R}$ es una secuencia contable de conjuntos disjuntos entonces

$$\mathbb{P}(\text{existe } n \geq 1 \text{ tal que } X \in I_n) = \sum_{n=1}^{\infty} \mathbb{P}[X \in I_n].$$

Estas propiedades son consecuencia directa de las tres propiedades básicas que toda medida de probabilidad debe satisfacer (Sección 1.2). Por ejemplo, como $X : \Omega \to \mathbb{R}$ es una función entonces la pre-imagen de $\mathbb{R}$ a través de X es Ω i.e. $[X \in \mathbb{R}] = \Omega$. Por lo tanto, $\mathbb{P}[X \in \mathbb{R}] = 1$. Por otro lado, si $I_1, I_2, \ldots$ son subconjuntos disjuntos de $\mathbb{R}$ entonces también lo son los sucesos $[X \in I_1], [X \in I_2], \ldots$ (*¿por qué?*). En particular, la sigma aditividad implica que

$$\mathbb{P}(\text{existe } n \geq 1 \text{ tal que } X \in I_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} [X \in I_n]\right) = \sum_{n=1}^{\infty} \mathbb{P}[X \in I_n].$$

La discusión anterior se extiende de forma natural a más de una variable aleatoria. Por ejemplo, si $I_1, I_2, \ldots \subset \mathbb{R}$ es una secuencia contable de conjuntos y $X_1, X_2, \ldots$ son variables aleatorias entonces el suceso "para cada i, el número obtenido al transformar el resultado del experimento a través de la función X_i pertenece a I_i" será denotado como $[X_1 \in I_1, X_2 \in I_2, \ldots]$. Equivalentemente, lo siguiente aplica por definición:

$$[X_1 \in I_1, X_2 \in I_2, \ldots] = \bigcap_{n=1}^{\infty} [X_n \in I_n].$$

Ejemplo 2.2. Para fijar ideas retomemos el ejemplo donde se registra la asistencia a clase en un curso de treinta alumnos. Imagine que el curso está compuesto por doce niñas y dieciocho niños. El profesor del curso podría estar interesado e.g. en el potencial efecto del sexo, en la asistencia a clase. Dos variables naturales a considerar en este caso son

$$\begin{aligned} X &= \text{el número de niñas presente en la clase,} \\ Y &= \text{el número de niños presente en la clase.} \end{aligned}$$

El suceso "todas las niñas asistirán a clase" puede representarse como $[X = 12]$, mientras que el suceso "al menos la mitad de los niños asistirá a clase" puede representarse como $[Y \geq 9]$. En particular, $[X = 12, Y \geq 9]$ corresponde al suceso "todas las niñas y al menos la mitad de los niños asistirán a clase". Por otro lado, el suceso "el número de niñas excederá el número de niños presente en clase" corresponde al suceso $[X > Y]$. Nuevamente necesitamos más información acerca de los alumnos inscritos en el curso para determinar probabilidades. Sin embargo, cualquiera sea la medida de probabilidad $\mathbb{P}$ utilizada, la sigma aditividad implicará e.g. que

$$\mathbb{P}[X > Y] = \mathbb{P}\left(\bigcup_{(i,j)} [X = i, Y = j]\right) = \sum_{(i,j)} \mathbb{P}[X = i, Y = j].$$

Los indices (i, j) en la unión y suma anteriores son pares ordenados de números enteros tales que $0 \leq i \leq 12$, $0 \leq j \leq 18$, e $i > j$. Este conjunto es finito y la probabilidad de la unión corresponde a la suma de las probabilidades ya que, por ejemplo, los sucesos $[X = 3, Y = 2]$ y $[X = 4, Y = 1]$ no pueden suceder simultáneamente.

Ejemplo 2.3. Considere el experimento que consiste en lanzar dos dados equilibrados. Si apostáramos con un amigo al valor más alto entre las dos caras, *¿a qué valor debíeramos apostar para tener una posibilidad máxima de ganar?*

Denotaremos como X la cara más alta entre los dos dados. Por ejemplo, si ambos dados salen 4 entonces $X = 4$. Por otro lado, si un dado sale 3 y el otro 5 entonces $X = 5$. Claramente, X sólo puede tomar los valores $1, \ldots, 6$.

Para responder a la pregunta necesitamos determinar $k \in \{1, \ldots, 6\}$ para el cual, la probabilidad del suceso $[X = k]$ es máxima. Por ejemplo, el suceso $[X = 1]$ sólo puede ocurrir si el primer y segundo dado salen 1. Como la probabilidad de esto es $1/36$, $\mathbb{P}[X = 1] = 1/36$. Por otro lado, $[X = 2]$ sucede solamente si uno de los dados sale 2 y el otro sale 1, o ambos dados salen 2. Como cada uno de estos tres casos tiene probabilidad $1/36$ entonces $\mathbb{P}[X = 2] = 1/12$. El suceso $[X = 2]$ es por lo tanto tres veces más probable que el suceso $[X = 1]$.

Más generalmente, para $k \in \{1, \ldots, 6\}$, el suceso $[X = k]$ ocurrirá sólo si ambos dados toman el valor k, o uno de los dados sale k y el otro toma un cierto valor entre 1 y $(k - 1)$. Como en total hay $(1 + 2(k - 1)) = (2k - 1)$ casos diferentes a considerar, cada uno con una posibilidad de $1/36$, se tiene que

$$\mathbb{P}[X = k] = \frac{2k - 1}{36},$$

para cada $k \in \{1, \ldots, 6\}$. La posibilidad del suceso $[X = k]$ es, por lo tanto, máxima cuando $k = 6$ (*¿por qué?*). En particular, si apostáramos con un amigo al valor más alto entre las caras lanzadas de dos dados equilibrados, la estrategia óptima es apostar a la cara 6.

Ejemplo 2.4. Considere el experimento donde se registra el lapso entre un sismo y otro en la ciudad de Iquique. Una v.a. de interés asociada con este experimento es el tiempo transcurrido entre el último y el próximo sismo. Usando T para denotar esta v.a., el rango de T es el intervalo abierto $(0, +\infty)$. Debido a razones que justificaremos en la Sección 4.5, para cada $t > 0$, asignaremos al suceso "el siguiente sismo ocurrirá dentro de los próximos t días" la probabilidad

$$\mathbb{P}[T \leq t] = 1 - e^{-\lambda t}. \tag{2.4}$$

Arriba, el número λ es el promedio de sismos que se registran por día. Por ejemplo, si la variable t se mide en días, y en promedio se registran tres sismos cada dos semanas entonces $\lambda = 3/14$. (t y λ pueden medirse en otras unidades siempre y cuando la cantidad λt es un escalar sin unidades.)

Usando el modelo probabilístico anterior podemos determinar la probabilidad de cualquier suceso de interés asociado a la v.a. T. Por ejemplo, la probabilidad del suceso "no se registrará ningún sismo al menos una semana después del último" es

$$\mathbb{P}[T > 7] = 1 - \mathbb{P}[T \leq 7] = 1 - (1 - e^{-7\lambda}) = e^{-7\lambda}.$$

Por otro lado, la probabilidad del suceso "el próximo sismo se registrará dentro de las siguientes dos semanas, pero no antes de una semana a partir del último sismo" viene dada por $\mathbb{P}[7 < T \leq 14]$. Pero note que $[7 < T \leq 14] = [T \leq 14] \setminus [T \leq 7]$. Como $[T \leq 7] \subset [T \leq 14]$ (*¿por qué?*), la identidad en (1.11) implica que

$$\mathbb{P}[7 < T \leq 14] = \mathbb{P}[T \leq 14] - \mathbb{P}[T \leq 7] = (1 - e^{-14\lambda}) - (1 - e^{-7\lambda}) = e^{-7\lambda} - e^{-14\lambda}.$$

Intuitivamente, la probabilidad que el próximo sismo ocurra exactamente en 7 días, 3 horas, 17 minutos, 42 segundos, etc; debiera ser cero. Usando la misma lógica, para cada $t > 0$, el suceso $[T = t]$ debiera tener probabilidad cero. *¿Es esta intuición consistente con el modelo probilístico bajo uso?* Para responder la pregunta observe que $[T = t] \subset [t - 1/n < T \leq t + 1/n]$, para todo $n \geq 1$. Usando (1.3) y (1.12), esto implica que

$$\begin{aligned} 0 \leq \ \mathbb{P}[T = t] &\leq \mathbb{P}\left[t - 1/n < T \leq t + 1/n\right], \\ &= \mathbb{P}\left[T \leq t + 1/n\right] - \mathbb{P}\left[T \leq t - 1/n\right], \\ &= e^{-\lambda t}(e^{\lambda/n} - e^{-\lambda/n}). \end{aligned}$$

En particular, como (*¿por qué?*)

$$\lim_{n \to \infty} e^{-\lambda t}(e^{\lambda/n} - e^{-\lambda/n}) = 0,$$

las desigualdades anteriores implican que $\mathbb{P}[T = t] = 0$, para cada $t > 0$.

2.2 Función de distribución

En esta sección introduciremos el concepto de función de distribución asociado con una variable aleatoria. Veremos en la Sección 2.2.1, que la función de distribución de una v.a. condensa toda la información disponible acerca de su comportamiento probabilístico. En la Sección 5.1, veremos cómo simular cualquier v.a. a partir de su función de distribución.

Definición 2.1. La *función de distribución* de una v.a. X es la transformación $F_X : \mathbb{R} \to \mathbb{R}$ definida como $F_X(x) = \mathbb{P}[X \leq x]$, para todo $x \in \mathbb{R}$.

Para fijar ideas consideraremos algunos ejemplos.

Ejemplo 2.5. Imagine un cierto experimento que tiene asociado una v.a. X tal que $F_X(\ln(2)) = 3/7$. Determine $F_Y(2)$ si $Y = e^X$. En efecto, debido a la definición, se tiene que

$$F_Y(2) = \mathbb{P}[Y \leq 2] = \mathbb{P}[e^X \leq 2] = \mathbb{P}[X \leq \ln(2)] = F_X(\ln(2)) = 3/7.$$

Ejemplo 2.6. Considere el experimento que consiste en lanzar un dado equilibrado y la v.a. definida como

$$X = \begin{cases} 0, & \text{si el lado del dado es menor o igual a 2;} \\ 1, & \text{en caso contrario.} \end{cases}$$

Demostraremos que (ver Figura 2.1(a))

$$F_X(x) = \begin{cases} 0, & x < 0; \\ 1/3, & 0 \leq x < 1; \\ 1, & x \geq 1. \end{cases} \tag{2.5}$$

En efecto, X sólo puede tomar el valor 0 o 1. En particular, $\mathbb{P}[X \leq x] = 0$, para todo $x < 0$, y $\mathbb{P}[X \leq x] = 1$, para todo $x \geq 1$. Por otro lado, si $0 \leq x < 1$ entonces

$$\begin{aligned} \mathbb{P}[X \leq x] &= \mathbb{P}[X < 0] + \mathbb{P}[X = 0] + \mathbb{P}[0 < X \leq x], \\ &= 0 + P[\text{dado=1 ó 2}] + 0 = 1/3, \end{aligned}$$

lo que establece el resultado.

Ejemplo 2.7. Considere el experimento donde se escoge un número X equidistribuidamente en el intervalo $[-1, 1]$. Esto se interpreta como sigue: *si dos intervalos* $I, J \subset [-1, 1]$ *tienen el mismo largo entonces, sin importar su ubicación relativa, se tiene que* $\mathbb{P}[X \in I] = \mathbb{P}[X \in J]$. La probabilidad de que X pertenezca a un subintervalo de $[-1, 1]$ depende, por lo tanto, sólo de su largo. En particular, e.g. si duplicáramos el largo de un intervalo (sin salirnos de $[-1, 1]$) entonces la probabilidad de que X tome un valor en dicho intervalo también debiera duplicarse. Esto es posible sólo si la probabilidad de que X tome un valor dentro de un cierto intervalo es proporcional a su largo. Como X está obligada a tomar un valor en el intervalo $[-1, 1]$, la constante de proporcionalidad tiene que ser $1/2$. Por lo tanto

$$\mathbb{P}[X \in I] = \frac{largo(I)}{2}, \text{ para todo intervalo } I \subset [-1, 1].$$

Usando la identidad anterior mostraremos que (ver Figura 2.1(b))

$$F_X(x) = \begin{cases} 0, & x < -1; \\ \frac{x+1}{2}, & -1 \leq x < 1; \\ 1, & x \geq 1. \end{cases} \tag{2.6}$$

En efecto, observe que $\mathbb{P}[X \leq x] = 0$ si $x < -1$, y $\mathbb{P}[X \leq x] = 1$ si $x \geq 1$. Por otro lado, si $-1 \leq x < 1$ entonces

$$\mathbb{P}[X \leq x] = \mathbb{P}[X < -1] + \mathbb{P}[X \in [-1, x]] = 0 + \frac{largo[-1, x]}{2} = \frac{x+1}{2},$$

lo que establece el resultado.

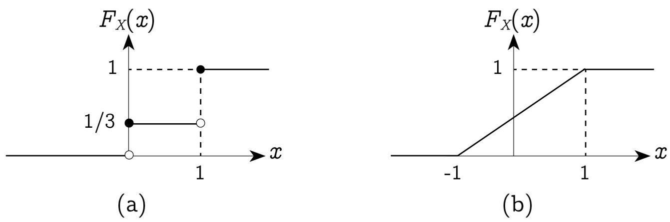


Figura 2.1. (a) Grafo asociado a la función de distribución de la v.a. X del Ejemplo 2.6. (b) Grafo de la función de distribución de la v.a. X del Ejemplo 2.7.

En lo que sigue diremos que una función $F : \mathbb{R} \to \mathbb{R}$ *tiene límite a la izquieda* si, para todo $x \in \mathbb{R}$, el límite $\lim_{y \to x^-} F(y)$ existe. Por otro lado, diremos que F es *continua a la derecha* si, para todo $x \in \mathbb{R}$, $\lim_{y \to x^+} F(y) = F(x)$.

Toda función continua tiene límite a la izquieda y es continua a la derecha. Esto se debe a que si F es continua entonces las siguientes identidades aplican:

$$\lim_{y \to x} F(y) = \lim_{y \to x^-} F(y) = \lim_{y \to x^+} F(y) = F(x).$$

La función de distribución de cualquier v.a. tiene cinco propiedades fundamentales que demostraremos a continuación. Las propiedades (d)-(e) establecen que F_X tiene límite a la izquierda y es continua a la derecha. Las propiedades (a)-(c) establecen que $F_X(x)$ es una función creciente que converge a 1 cuando $x \to (+\infty)$ y a 0 cuando $x \to (-\infty)$. La intuición detrás de esto es la siguiente. Claramente, si $X \leq -10$ entonces $X \leq 100$ y, por lo tanto, la posibilidad del suceso $[X \leq -10]$ debiera ser menor que la del suceso $[X \leq 100]$. Por lo mismo, los sucesos $[X \leq -10]$, $[X \leq -100]$, $[X \leq -1000]$, etc tiene una probabilidad cada vez más pequeña. Esta probabilidad tiene que aproximarse a cero ya que, en el caso contrario, X tendría una probabilidad estrictamente positiva de tomar el valor $(-\infty)$. Sin embargo, esto último no es posible ya que X sólo puede tomar valores reales (ver Sección 2.1). Un argumento similar muestra que los sucesos $[X \leq 10]$, $[X \leq 100]$, $[X \leq 1000]$, etc, tiene una probabilidad cada vez más cercana a uno. El lector podrá facilmente verificar todas éstas propiedades inspeccionando los grafos de la Figura 2.1.

De ahora en adelante usaremos el término *función de distribución* para referirnos a cualquier función que satisfaga las propiedades (a)-(e) que siguen. Veremos en la

Sección 5.1 que toda función de distribución es de hecho la función de distribución de una cierta variable aleatoria.

Teorema 2.2. *Si X es una v.a. entonces su función de distribución satisface las siguientes propiedades: (a) F_X es creciente, (b) $\lim_{x\to-\infty} F_X(x)=0$, (c) $\lim_{x\to+\infty} F_X(x)=1$, (d) F_X tiene límite a la izquierda, (e) F_X es continua a la derecha.*

Demostración. Para demostrar (a) observe que si $x_1 \leq x_2$ entonces $[X \leq x_1] \subset [X \leq x_2]$ y, por lo tanto, $F_X(x_1)=\mathbb{P}[X \leq x_1] \leq \mathbb{P}[X \leq x_2]=F_X(x_2)$. Esto muestra la parte (a) en el teorema.

En lo que sigue sólo demostraremos las partes (b) y (d), dejando las partes (c) y (e) como ejercicio al final del capítulo. Para demostrar la parte (b) y usando que F_X es creciente tenemos que

$$\lim_{x\to-\infty} F_X(x)=\lim_{n\to+\infty} F_X(-n)=\lim_{n\to+\infty} \mathbb{P}[X \leq -n].$$

Pero observe que los sucesos $[X \leq -n]$, con $n \geq 1$, decrecen con n al suceso [para todo $n \geq 1$, $X \leq -n$]. Como X solamente toma valores reales, este útimo suceso es vacío. Por lo tanto, y de acuerdo al Teorema 1.8, se tiene que

$$\lim_{n\to+\infty} \mathbb{P}[X \leq -n]=\mathbb{P}(\emptyset)=0,$$

lo que demuestra la parte (b). Finalmente, con relación a la parte (d) demostraremos más generalmente que

$$\lim_{y\to x^-} F_X(y)=\mathbb{P}[X<x], \tag{2.7}$$

para todo $x \in \mathbb{R}$. En efecto, usando que F_X es una función creciente se tiene que

$$\lim_{y\to x^-} F_X(y)=\lim_{n\to\infty} F_X(x-1/n)=\lim_{n\to\infty} \mathbb{P}[X \leq x-1/n]=\mathbb{P}[X<x],$$

donde para la última igualdad hemos usado que los sucesos $[X \leq x-1/n]$, con $n \geq 1$, crecen al suceso $[X<x]$ (*¿por qué?*). Esto demuestra que F_X tiene límite a la izquierda y concluye la demostración del teorema. □

2.2.1 Probabilidades computables a partir de F_X.

Una consecuencia importante de la demostración del Teorema 2.2 es que la función de distribución de una v.a. X, puede utilizarse para determinar la probabilidad que ésta tome un valor en cualquier intervalo $I \subset \mathbb{R}$ y no solamente intervalos de la forma $(-\infty,x]$, con $x \in \mathbb{R}$. En efecto, usando que el suceso $[X \leq x]$ se reescribe como la unión disjunta de los sucesos $[X<x]$ y $[X=x]$ encontramos que

$$\mathbb{P}[X \leq x]=\mathbb{P}[X<x]+\mathbb{P}[X=x].$$

Por lo tanto, y de acuerdo a la identidad en (2.7), deducimos que

$$\mathbb{P}[X=x]=F_X(x)-F_X(x^-), \tag{2.8}$$

para todo $x \in \mathbb{R}$, donde hemos definido

$$F_X(x^-) = \lim_{y \to x^-} F_X(y).$$

Finalmente, usando la definición de F_X y la identidad anterior, se puede demostrar para todo $a \leq b$ las siguiente identidades:

$$\begin{align*}
\mathbb{P}[X > a] &= 1 - F_X(a); \tag{2.9}\\
\mathbb{P}[a < X \leq b] &= F_X(b) - F_X(a); \tag{2.10}\\
\mathbb{P}[a \leq X \leq b] &= F_X(b) - F_X(a^-); \tag{2.11}\\
\mathbb{P}[a \leq X < b] &= F_X(b^-) - F_X(a^-); \tag{2.12}\\
\mathbb{P}[a < X < b] &= F_X(b^-) - F_X(a). \tag{2.13}
\end{align*}$$

Por ejemplo, la primera identidad se deduce directamente del hecho que el suceso $[X > a]$ es el complemento del suceso $[X \leq a]$. Por otro lado, la segunda identidad se deduce del hecho que $[a < X \leq b] = [X \leq b] \setminus [X \leq a]$ y que $[X \leq a] \subset [X \leq b]$. Un razonamiento similar demuestra las otras tres identidades.

En palabras, la identidad en (2.8) significa que la probabilidad de que X tome exactamente el valor x es igual al tamaño de la discontinuidad de su función de distribución sobre el punto x. En particular, *la función de distribución de X es continua sobre un punto x si y solo si la probabilidad del suceso $[X = x]$ es cero*. Por ejemplo, en la Figura 2.1(a) vemos que el suceso $[X = 4/5]$ tiene probabilidad cero, mientras que el suceso $[X = 1]$ tiene probabilidad $(1 - 1/3) = 2/3$.

Por otro lado, debido a (2.13), F_X *permanece constante sobre un intervalo abierto si y solo si la probabilidad de que X tome un valor en dicho intervalo es cero*. Por ejemplo, como la función de distribución en la Figura 2.1(a) permanece constante en el intervalo abierto $(1/5, 4/5)$ entonces el suceso $[1/5 < X < 4/5]$ tiene probabilidad cero.

Las observaciones anteriores resultan muy útiles para especificar, rápidamente, la función de distribución de una serie de variables aleatorias. Pondremos en práctica estos principios en el siguiente ejemplo.

Ejemplo 2.8. Si X denota la cara lanzada de un dado equilibrado entonces F_X tiene que ser una función discontinua en $x = 1, 2, 3, 4, 5$ ó 6. Más aún, el tamaño de la discontinuidad de F_X en cada uno de estos puntos tiene que ser $1/6$. Por otro lado, como estos son los únicos valores posibles para X entonces F_X tiene ser continua sobre cualquier $x \notin \{1, 2, 3, 4, 5, 6\}$. De hecho, F_X tiene que permanecer constante sobre cada uno de los intervalos abiertos $(-\infty, 1)$, $(1, 2)$, $(2, 3)$, $(3, 4)$, $(4, 5)$, $(5, 6)$ y $(6, \infty)$ ya que de lo contrario X tomaría valores dentro de estos intervalos. Usando que F_X es continua a la derecha con límite a la derecha, es fácil deducir de éstas observaciones que (ver Figura 2.2)

$$F_X(x) = \begin{cases} 0, & x < 1; \\ 1/6, & 1 \le x < 2; \\ 2/6, & 2 \le x < 3; \\ 3/6, & 3 \le x < 4; \\ 4/6, & 4 \le x < 5; \\ 5/6, & 5 \le x < 6; \\ 1, & 6 \le x. \end{cases}$$

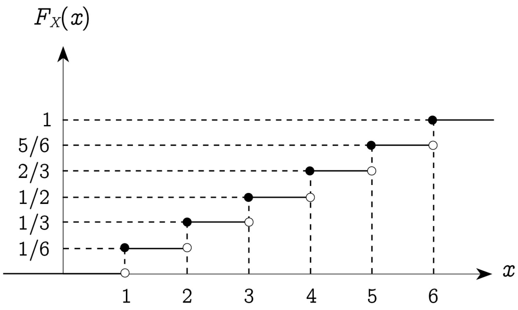


FIGURA 2.2. Función de distribución de la cara observada al lanzar un dado equilibrado.

2.3 Variables aleatorias discretas

Recordemos que un conjunto A se dice contable si es finito o enumerable. Esto último es equivalente a la existencia de una transformación $h : A \to \mathbb{N}$ que es biyectiva. Ejemplos de conjuntos contables no finitos incluyen $\mathbb{N}$ y $\mathbb{Z}$. Sin embargo, cualquier subconjunto de los números reales que contiene un intervalo de largo estrictamente positivo no es contable.

En términos generales una v.a. discreta es un número aleatorio que solamente puede tomar valores dentro de un cierto conjunto contable. La definición precisa es la siguiente.

Definición 2.3. Una v.a. X se dice *discreta* si existe un conjunto contable $A \subset \mathbb{R}$ tal que $\mathbb{P}[X \in A] = 1$.

Las variables aleatorias discutidas en los Ejemplos 2.3, 2.6 y 2.8 son ejemplos de variables aleatorias discretas. Otros ejemplos son los siguientes:

Considere el experimento que consiste en lanzar dos dados. El valor de la cara más baja es una v.a. discreta pues, con probabilidad uno, ésta tomará un valor en el conjunto finito $\{1, 2, 3, 4, 5, 6\}$. La suma de las caras de los dados es también una v.a.

discreta ya que sólo puede tomar valores en el conjunto $\{2, 3, \ldots, 12\}$, que es finito. La razón entre la cara más baja y la más alta es otro ejemplo de una v.a. discreta. Esta v.a. sólo puede tomar valores en el conjunto finito de valores $\{i/j : 1 \leq i \leq j \leq 6\}$.

Considere ahora el experimento que consiste en lanzar una moneda equilibrada. El número de lanzamientos hasta observar la primera cara es una v.a. discreta ya que ésta sólo puede tomar valores en el conjunto enumerable $\{1, 2, 3, \ldots\}$. Por otro lado, el número de sellos entre dos caras consecutivas es también un ejemplo de v.a. discreta. Esta variable sólo puede tomar valores en el conjunto $\{0, 1, 2, \ldots\}$ con probabilidad uno. Finalmente, el número de lanzamientos hasta observar por primera vez la secuencia: cara, sello, sello, sello, es también una v.a. discreta que sólo puede tomar valores en el conjunto enumerable $\{4, 5, 6, \ldots\}$.

Otros ejemplos de variables aleatorias discretas son el número de llamadas a una oficina en un intervalo de tiempo dado, o el número de gotitas de agua que sacudirán una cierta sección del suelo durante un chaparrón de agua.

¡No toda v.a. es discreta! Por ejemplo, la v.a. T que mide el tiempo transcurrido entre el último y próximo sismo en la ciudad de Iquique (Ejemplo 2.4) no es discreta. Esto se debe a que $\mathbb{P}[T = t] = 0$, para todo $t \in \mathbb{R}$. En particular, si $A \subset \mathbb{R}$ es un conjunto contable entonces la sigma aditividad implica que

$$\mathbb{P}[T \in A] = \sum_{t \in A} \mathbb{P}[T = t] = 0.$$

Por lo tanto, no existe un conjunto contable $A \subset \mathbb{R}$ tal que $\mathbb{P}[T \in A] = 1$ i.e. T no puede ser una v.a. discreta. T es, sin embargo, un caso especial de lo que se llama una v.a. continua, que introducimos a continuación.

2.4 Variables aleatorias continuas

Una v.a. continua es un número aleatorio cuya probabilidad de caer en un intervalo arbitrario, está dada por el área contenida entre dicho intervalo y el grafo de una cierta función. La definición precisa es la siguiente:

Definición 2.4. Una v.a. X se dice *continua* si existe una función $f : \mathbb{R} \to \mathbb{R}_+$ continua excepto sobre un número finito de puntos (i.e. existe un conjunto finito D y posiblemente vacío tal que f es continua sobre todo $x \notin D$) tal que

$$\mathbb{P}[X \in I] = \int_I f(x)\, dx, \tag{2.14}$$

para todo intervalo cerrado $I \subset \mathbb{R}$. En este caso decimos que f es una *función de densidad para X*.

Usaremos la notación f_X cuando queramos enfatizar que una v.a. X tiene función de densidad f_X. Observe que la misma función f puede intergrarse sobre diferentes intervalos para determinar la probabilidad que X tome valores en dichos intervalos.

Más aún, se tiene que

$$\int_{\mathbb{R}} f(x)\,dx = \mathbb{P}[X \in \mathbb{R}] = 1.$$

Esto motiva el término de *función de densidad* para referirnos a cualquier función $f : \mathbb{R} \to \mathbb{R}_+$, continua fuera de un cierto conjunto finito de puntos y tal que

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

.

Ejemplo 2.9. Considere una v.a. X con función de densidad

$$f(x) = \begin{cases} 2xe^{-x^2}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Observe que $f \geq 0$ y que f es continuamente diferenciable excepto sobre el punto $x = 0$. Más aún, $\int_{\mathbb{R}} f(x)\,dx = 1$; en particular, f es efectivamente una función de densidad. Usando esta función podemos determinar la probabilidad de cualquier suceso asociado a la v.a. X. Por ejemplo, se encuentra que

$$\begin{aligned} \mathbb{P}[X \leq -1] &= \int_{-\infty}^{-1} f(x)\,dx = \int_{-\infty}^{-1} 0\,dx = 0; \\ \mathbb{P}[-1 \leq X \leq 1/2] &= \int_{-1}^{1/2} f(x)\,dx = \int_{0}^{1/2} 2xe^{-x^2}\,dx = 1 - e^{-1/4}; \\ \mathbb{P}[X \geq 2] &= \int_{2}^{\infty} f(x)\,dx = \int_{2}^{\infty} 2xe^{-x^2}\,dx = e^{-4}. \end{aligned}$$

Ejemplo 2.10. Determine la función de distribución de una v.a. X con función de densidad

$$f_X(x) = \begin{cases} 0, & x < 0; \\ \frac{2x}{(1+x^2)^2}, & x \geq 0. \end{cases}$$

Observe primero que $F_X(t) = 0$, para todo $t \leq 0$, debido a que f_X es idénticamente igual a cero sobre el intervalo $(-\infty, 0]$. Por otro lado, si $t > 0$ entonces

$$\begin{aligned} F_X(t) = \mathbb{P}[X \leq t] &= \mathbb{P}[X \in [0, t]], \\ &= \int_0^t f_X(x)\,dx, \\ &= \int_0^t \frac{2x}{(1+x^2)^2}\,dx = \int_1^{1+t^2} \frac{1}{y^2}\,dy = \frac{t^2}{1+t^2}, \end{aligned}$$

donde para la penúltima identidad hemos susbtituido $y = (1 + x^2)$ en la integral. La función de distribución de X viene, por lo tanto, dada por la fórmula

$$F_X(t) = \begin{cases} 0, & t < 0; \\ \frac{t^2}{1+t^2}, & t \geq 0. \end{cases}$$

La terminología usada en la Definición 2.4 está en parte motivada por el siguiente resultado.

Teorema 2.5. *Si X es una v.a. continua con función de densidad f entonces $\mathbb{P}[X = t] = 0$, para todo $t \in \mathbb{R}$; en particular, F_X es una función continua. Más aún, F_X es continuamente diferenciable excepto sobre un número finito de puntos y*

$$F_X(t) = \int_{-\infty}^{t} f(x)\, dx.$$

Recíprocamente, si existe una función $f : \mathbb{R} \to \mathbb{R}_+$ continua excepto sobre un número finito de puntos, tal que la identidad anterior aplica para todo t, entonces X es una v.a. continua con función de densidad f.

Demostración. Si X tiene función de densidad f entonces, para cada $t \in \mathbb{R}$, tenemos que $\mathbb{P}[X = t] = \int_t^t f(x)\, dx = 0$. Por lo tanto, y de acuerdo a la propiedad dada en (2.8), $F_X(t) = F_X(t^-)$. Como F_X es continua a la derecha (Teorema 2.2), concluimos que F_X es continua. Finalmente observe que

$$F_X(t) = \mathbb{P}[X \in (\infty, t]] = \int_{-\infty}^{t} f(x)\, dx,$$

para todo $t \in \mathbb{R}$. En particular, y debido al Teorema Fundamental del Cálculo, la función F_X es diferenciable donde f es continua. Específicamente, $F_X'(t) = f(t)$ si f es continua sobre t. Como f es continua fuera de un cierto conjunto finito, F_X es continuamente diferenciable fuera de este conjunto, lo que demuestra la primera parte del teorema.

Recíprocamente, supongamos que existe una función f continua excepto por un cierto conjunto finito de puntos tal que la identidad anterior aplica. Entonces, F_X es una función continua y en particular $\mathbb{P}[X = t] = 0$, para cada $t \in \mathbb{R}$. Por lo tanto, si $a < b$ y debido a la identidad en (2.10) obtenemos que

$$\begin{aligned} \mathbb{P}[X \in [a, b]] &= \mathbb{P}[X \in (a, b]], \\ &= F_X(b) - F_X(a), \\ &= \int_{-\infty}^{b} f(x)\, dx - \int_{-\infty}^{a} f(x)\, dx = \int_a^b f(x)\, dx. \end{aligned}$$

X es, por lo tanto, una v.a. continua con función de densidad f, lo que completa la demostración del teorema. □

Una consecuencia importante del teorema es que X es una v.a. continua, cuando su función de distribución es continuamente diferenciable excepto sobre un número finito de puntos. En este caso su función de densidad puede obternerse directamente vía diferenciación:

$$f_X(x) = \begin{cases} F_X'(x), & \text{si } F_X' \text{ existe y es continua sobre } x; \\ 0, & \text{en cualquier otro caso.} \end{cases} \tag{2.15}$$

Ejemplo 2.11. Para fijar ideas retomemos el experimento donde se registran los tiempos de sismos en la ciudad de Iquique (Ejemplo 2.4). De acuerdo a lo discutido, si T denota el tiempo transcurrido entre el último y próximo sismo y en promedio se registran λ sismos por unidad de tiempo entonces $\mathbb{P}[T \leq t] = 1 - e^{-\lambda t}$, para todo $t \geq 0$. En particular,

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

La función anterior es continuamente diferenciable sobre $\mathbb{R}$ excepto por el punto $t = 0$. En particular, y debido a la identidad en (2.15), T es una v.a. continua con función de densidad dada por la fórmula

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t > 0; \\ 0, & t \leq 0. \end{cases}$$

Es importante destacar que la función de densidad de una v.a. continua no es única. Esto se debe a que es posible definir funciones de densidad diferentes que, sin embargo, integran lo mismo sobre cada intervalo cerrado. Por ejemplo, la función

$$g(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0; \\ 0, & t < 0; \end{cases}$$

es también una función de densidad para la v.a. T en el ejemplo anterior. Esto se debe a que $f(t) = g(t)$ para todo t, excepto por $t = 0$. En particular, $\int_I f(t)\,dt = \int_I g(t)\,dt$, para todo intervalo cerrado $I \subset \mathbb{R}$. Si cambiáramos el valor de la función g en un segundo punto entonces obtendríamos una segunda función de densidad. Por lo tanto, existe un sin fin de funciones de densidad asociadas a una misma v.a. continua. Uno puede demostrar, sin embargo, que si f y g son funciones de densidad asociadas a una misma v.a. entonces el conjunto de puntos $\{t \in \mathbb{R} : f(t) \neq g(t)\}$ es finito. Las funciones de densidad asociadas a una misma v.a. son, por lo tanto, esencialmente idénticas.

Ejemplo 2.12. Considere el experimento que consiste en lanzar al azar un dardo a un blanco redondo de radio 1 unidad. Una v.a. interesante asociada con este experimento es la distancia R entre el centro del blanco y el punto de impacto del dardo, *¿es R una v.a. continua?*

Claramente $0 \leq R \leq 1$; en particular, $F_R(r) = 0$ si $r < 0$, y $F_R(r) = 1$ si $r \geq 1$. Por otro lado, si (X, Y) denota el punto de impacto del dardo entonces para $0 \leq r < 1$ se tiene que

$$F_R(r) = \mathbb{P}[0 \leq R \leq r] = \mathbb{P}[(X, Y) \in C_r], \tag{2.16}$$

donde C_r es el círculo con centro $(0, 0)$ y de radio r.

El hecho que (X, Y) es un punto equidistribuido se interpreta como sigue: *si dos regiones A, B dentro del blanco tienen la misma área entonces, sin importar su ubicación relativa, los sucesos $[(X, Y) \in A]$ y $[(X, Y) \in B]$ tienen la misma probabilidad.* Esto significa que la probabilidad de que (X, Y) pertenezca a una cierta región dentro del blanco sólo depende del área de dicha región. En particular, si duplicáramos el área de una región (sin salirnos del blanco) entonces la probabilidad de que el dardo caiga en dicha región también debiera duplicarse i.e. si área$(A) = 2 \cdot$ área(B) entonces $\mathbb{P}[(X,Y) \in A] = 2 \cdot \mathbb{P}[(X,Y) \in B]$. Similarmente, si área$(A) =$ área$(B)/3$ entonces $\mathbb{P}[(X,Y) \in A] = \mathbb{P}[(X,Y) \in B]/3$. La probabilidad de que el dardo caiga en una cierta región A, adentro del blanco, es por lo tanto proporcional al área de dicha región. Como el área total del blanco es π, la constante de proporcionalidad tiene que ser $1/\pi$ para que las probabilidades estén acotadas entre cero y uno. Concluimos, por lo tanto, que

$$\mathbb{P}[(X,Y) \in A] = \frac{\text{área}(A)}{\pi}, \text{ para toda región } A \text{ contenida en el blanco.}$$

Devuelta en la identidad (2.16), la fórmula anterior implica que

$$F_R(r) = \frac{\text{área}(C_r)}{\pi} = \frac{\pi r^2}{\pi} = r^2.$$

R tiene, por lo tanto, una función de distribución dada por la fórmula

$$F_R(r) = \begin{cases} 0, & r < 0; \\ r^2, & 0 \leq r < 1; \\ 1, & r \geq 1. \end{cases}$$

La función anterior es continuamente diferenciable sobre el conjunto de puntos $\mathbb{R} \setminus \{0,1\}$. ($F_R$ es de hecho continuamente diferenciable en todo punto excepto $r = 1$.) Por lo tanto, y de acuerdo a la identidad en (2.15), R es una v.a. continua con función de densidad dada por la fórmula

$$f(r) = \begin{cases} 0, & r < 0; \\ 2r, & 0 \leq r < 1; \\ 0, & r \geq 1. \end{cases} = \begin{cases} 2r, & r \in [0,1); \\ 0, & r \notin [0,1). \end{cases}$$

Observe que esta función es creciente entre cero y uno; en particular, R es más probable de tomar un cierto valor cerca de 1 que de 0. Debido a esto, en el juego del lanzamiento de un dardo al blanco, el centro da mayor puntaje que el perímetro: la lógica es que un jugador que tira el dardo al azar, tendrá tendencia a impactar más bien cerca del perímetro que del centro del blanco.

2.4.1 Variables aleatorias uniformes.

Este tipo de v.a. es un caso especial de variable aleatoria continua. En términos generales, una v.a. uniforme sobre un intervalo dado es un número aleatorio que en principio puede tomar cualquier valor en dicho intervalo, pero sin preferencia por un cierto valor más que otro. Esto se interpreta matemáticamente como sigue.

Definición 2.6. Sean $a < b$ números reales dados. Diremos que X es una *v.a. uniforme en el intervalo* $[a,b]$ si X es una v.a. continua con función de densidad

$$f(x) = \begin{cases} 0, & x \notin [a,b]; \\ 1/(b-a), & x \in [a,b]. \end{cases}$$

En el caso anterior, anotaremos $X \sim Uniforme[a, b]$.

Claramente $\mathbb{P}[X \in [a, b]] = 1$ i.e. X siempre tomará un valor entre a y b. Otra consecuencia de la definición es que

$$\mathbb{P}[X \in I] = \int_I \frac{1}{b-a}\,dx = \frac{\text{largo}(I)}{b-a} = \frac{\text{largo}(I)}{\text{largo}[a, b]},$$

para todo intervalo cerrado $I \subset [a, b]$. Note, sin embargo, que si $I = [c, d]$ entonces $\mathbb{P}[X = c] = \mathbb{P}[X = d] = 0$ (*¿por qué?*). En particular, al remover uno o ambos de los extremos de I la probabilidad del suceso anterior no cambia. Como el largo de un intervalo no cambia si incluimos o no sus extremos, esta fórmula también aplica para todo intervalo $I \subset [a, b]$. La probabilidad del suceso $[X \in I]$ es, por lo tanto, proporcional al largo de I cuando $I \subset [a, b]$. (La constante de proporcionalidad tiene que ser $1/(b-a)$ para asegurar que las probabilidades anteriores estén acotadas entre cero y uno.)

La identidad anterior justifica el término de "uniforme" en la Definición 2.6: *si $I, J \subset [a, b]$ son intervalos del mismo largo entonces, sin importar su ubicación relativa, se tendrá que $\mathbb{P}[X \in I] = \mathbb{P}[X \in J]$*. La v.a. X no tiene, por lo tanto, preferencia de ubicación en el intervalo $[a, b]$.

Una consecuencia inmediata de la discusión anterior es que

$$\mathbb{P}[X \in I] = \mathbb{P}[X \in I \cap [a, b]] = \frac{\text{largo}(I \cap [a, b])}{\text{largo}[a, b]}, \tag{2.17}$$

para todo intervalo $I \subset \mathbb{R}$ (*¿por qué?*).

Ejemplo 2.13. Si $X \sim Uniforme[-1, 2]$ entonces

$$\begin{aligned}
\mathbb{P}[X \leq 0] &= \frac{\text{largo}[-1, 0]}{3} = \frac{1}{3}, \\
\mathbb{P}[X > 1/2] &= \frac{\text{largo}(1/2, 2]}{3} = \frac{1}{2}, \\
\mathbb{P}[-1/3 \leq X < 3/2] &= \frac{\text{largo}[-1/3, 3/2)}{3} = \frac{11}{18}.
\end{aligned}$$

Ejemplo 2.14. Mariel es una profesora de colegio muy metódica y que cree fervientemente en el dicho "es mejor prevenir que lamentar". Al principio de cada clase Mariel parte al azar—en dos pedazos—la tiza que utilizará para escribir en la pizarra y comienza a escribir con el pedazo más corto. (Al partir la tiza en dos, es menos probable que cada pedazo se rompa en trozos inútiles si la tiza se cae al suelo.) *¿Qué fracción del total de las clases, comenzará Mariel sus notas en el pizarrón con al menos $1/3$ del largo original de la tiza?*

Para responder la pregunta, definiremos el largo genérico de la tiza como una unidad. Si X denota la distancia entre el extremo izquierdo y el lugar donde Mariel partirá la tiza, entonces es razonable suponer que $X \sim Uniforme[0, 1]$.

El largo Y del pedazo de tiza con que Mariel comenzará su clase es el más pequeño entre X y $(1-X)$, o sea $Y = \text{mín}\{X, 1-X\}$. Observe que $Y \geq 1/3$ si y solo si $X \geq 1/3$ y $(1-X) \geq 1/3$. En particular, la respuesta a la pregunta planteada es

$$\begin{aligned}
\mathbb{P}[Y \geq 1/3] &= \mathbb{P}[X \geq 1/3, (1-X) \geq 1/3], \\
&= \mathbb{P}[1/3 \leq X \leq 2/3], \\
&= \frac{\text{largo}[1/3, 2/3]}{\text{largo}[0,1]} = 1/3.
\end{aligned}$$

En otras palabras, alrededor de un tercio de las clases, Mariel comenzará a escribir en el pizarrón con al menos un tercio del largo original de la tiza.

El siguiente resultado nos permitirá identificar más fácilmente cuando una v.a. es uniforme. En términos generales, podemos decir que una v.a. es uniforme sobre un intervalo $[a, b]$ si $F_X(a) = 0$ y $F_X(b) = 1$, y F_X crece linealmente entre a y b.

Teorema 2.7. *$X \sim Uniforme[a, b]$ si y solo si*

$$F_X(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x < b; \\ 1, & x \geq b. \end{cases} \tag{2.18}$$

Demostración. En efecto, si $X \sim Uniforme[a, b]$ entonces el suceso $[a \leq X \leq b]$ tiene probabilidad uno y, en particular, $F_X(x) = 0$, para $x < a$, y $F_X(x) = 1$, para $x \geq b$. Por otro lado, debido a la identidad en (2.17) obtenemos para $a \leq x < b$ que

$$F_X(x) = \mathbb{P}[X \leq x] = \frac{\text{largo}[a, x]}{b-a},$$

lo que demuestra la fórmula en (2.18). Recíprocamente, como la función en (2.18) es continuamente diferenciable excepto sobre los puntos $x = a$ y $x = b$, la identidad en (2.15) implica que

$$f_X(x) = \begin{cases} 0, & x \leq a; \\ \frac{1}{b-a}, & a < x < b; \\ 0, & x \geq b. \end{cases}$$

i.e. $X \sim Uniforme[a, b]$. Esto demuestra el teorema. □

Ejemplo 2.15. Para fijar ideas retomemos el ejemplo de la profesora Mariel, que parte la tiza en dos al comenzar cada clase (Ejemplo 2.14). Según vimos, el largo del trozo con que Mariel comenzará a escribir en el pizarrón es $Y = \text{mín}\{X, 1-X\}$, con $X \sim Uniforme[0, 1]$. Mostraremos que $Y \sim Uniforme[0, 1/2]$.

Primero note que $Y \leq X$ y $Y \leq (1-X)$; en particular, $Y \leq 1/2$. Por otro lado, como $0 \leq X \leq 1$ entonces $Y \geq 0$. Por lo tanto, $F_Y(y) = 0$ si $y < 0$, y $F_Y(y) = 1$ si $y \geq 1/2$. Más aún, si $0 \leq y < 1/2$ entonces

$$\begin{aligned}
F_Y(y) &= \mathbb{P}[\text{mín}\{X, 1-X\} \leq y], \\
&= 1 - \mathbb{P}[\text{mín}\{X, 1-X\} > y] = 1 - \mathbb{P}[y < X < 1-y],
\end{aligned}$$

donde, para la última identidad, hemos usado que $\min\{X, 1-X\} > y$ es equivalente a $X > y$ e $(1-X) > y$. Finalmente, como $X \sim Uniforme[0,1]$, la identidad en (2.17) nos permite determinar que $F_Y(y) = 1 - \text{largo}(y, 1-y) = 2y$. Obtenemos, por lo tanto, que

$$F_Y(y) = \begin{cases} 0, & y < 0; \\ 2y, & 0 \leq y < 1/2; \\ 1, & y \geq 1. \end{cases}$$

Como la función anterior crece linealmente entre entre 0 y 1/2, con $F_X(0) = 0$ y $F_X(1/2) = 1$, el Teorema 2.7 implica que $Y \sim Uniforme[0, 1/2]$.

Para terminar esta sección observe que si X es una v.a. uniforme en el intervalo $[a,b]$ entonces $\mathbb{P}[X = a] = 0$ y $\mathbb{P}[X = b] = 0$. En particular, los extremos del intervalo $[a,b]$ no juegan ningún rol en la definición de esta variable. Debido a esto, de ahora en adelante usaremos la terminología $Uniforme(a,b)$, $Uniforme(a,b]$ o incluso $Uniforme[a,b)$ como sinónimos de una v.a. uniforme en el intervalo $[a,b]$.

2.4.2 El método del Jacobiano

En la Sección 2.4 vimos cómo determinar la función de densidad de una v.a. continua a partir de su función de distribución. En esta sección, veremos que si transformamos una v.a. continua a través de una función continuamente diferenciable, entonces la v.a. que resulta también será continua. El primer método que presentaremos es un caso especial del segundo, sin embargo, es instructivo presentar cada método por separado.

Teorema 2.8. *Sea X una v.a. continua. Si $I \subset \mathbb{R}$ es un intervalo abierto tal que $\mathbb{P}[X \in I] = 1$ y $g : I \to \mathbb{R}$ es una transformación continuamente diferenciable tal que $g'(x) \neq 0$, para todo $x \in I$, entonces la v.a. $Y = g(X)$ es también continua y con función de densidad dada por la fórmula*

$$f_Y(y) = \begin{cases} 0, & \text{si } y \notin g(I); \\ \frac{f_X(x)}{|g'(x)|}, & \text{si } y \in g(I) \text{ e } y = g(x). \end{cases}$$

De las hipótesis del teorema se deduce que g es una función continua y estrictamente creciente o estrictamente decreciente sobre el intervalo I (*¿por qué?*). En particular, el rango de g i.e. el conjunto $g(I)$ es un intervalo abierto. Más aún, si $y \in g(I)$ entonces existe un único $x \in I$ tal que $y = g(x)$. Por lo tanto, si g^{-1} denota la función inversa de g, entonces la siguiente identidad también aplica

$$f_Y(y) = \begin{cases} 0, & \text{si } y \notin g(I), \\ \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}, & \text{si } y \in g(I). \end{cases} \tag{2.19}$$

Demostración. Como $g : I \to \mathbb{R}$ es continuamente diferenciable se tiene, de las hipótesis del teorema, que $g'(x) > 0$, para todo $x \in I$, o bien $g'(x) < 0$, para todo $x \in I$ (*¿por qué?*). Supondremos que $g' > 0$, dejando como ejercicio para el lector la demostración del otro caso.

En efecto, debido a la continuidad de g, existen $-\infty \leq a < b \leq \infty$ tales que $g(I) = (a,b)$. En particular, $\mathbb{P}[Y \in (a,b)] = 1$. Por lo tanto, $F_Y(y) = 0$ si $y \leq a$, y $F_Y(y) = 1$ si $y \geq b$. Por otro lado, para $y \in (a,b)$ y como g es estrictamente creciente, tenemos que

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[g(X) \leq y] = \mathbb{P}[X \leq g^{-1}(y)] = F_X(g^{-1}(y)).$$

Esto muestra que

$$F_Y(y) = \begin{cases} 0, & y \leq a; \\ F_X(g^{-1}(y)), & y \in (a,b); \\ 1, & y \geq b. \end{cases}$$

Pero recuerde que F_X es continuamente diferenciable excepto sobre un conjunto finito de puntos (Teorema 2.5). En particular, la función anterior es también continuamente diferenciable sobre $\mathbb{R}$ excepto por un número finito de puntos. Y es, por lo tanto, una v.a. continua. Más aún, debido a la regla de la cadena, tenemos que

$$F_Y'(y) = F_X'(g^{-1}(y)) \cdot \frac{d}{dy}[g^{-1}(y)] = \frac{F_X'(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))},$$

para todo $y \in (a,b)$ con la excepción de un número finito de puntos en este intervalo. Esto demuestra que la función definida en (2.19) es una función de densidad para Y y completa la demostración del teorema. $\square$

Ejemplo 2.16. En este ejemplo, usaremos el Método del Jacobiano para determinar la función de densidad de la v.a. $Y = X/(1-X)$, donde $X \sim Uniforme(0,1)$. Para esto note que $Y = g(X)$, con $g : (0,1) \to \mathbb{R}$ la función definida como $g(x) = x/(1-x)$.

Primero observe que

$$g'(x) = \frac{d}{dx}\left[\frac{x}{1-x}\right] = \frac{1}{(1-x)^2} > 0, \tag{2.20}$$

para todo $0 < x < 1$. Como $\mathbb{P}[X \in (0,1)] = 1$, el Teorema 2.8 implica que Y tiene una función de densidad dada por la fórmula (2.19). Para determinar f_Y explícitamente debemos caracterizar primero el rango de g. Para esto observe que $\lim_{x\to 0+} g(x) = 0$ y $\lim_{x\to\infty} g(x) = +\infty$. Como g es continua y creciente, deducimos que el rango de g es el intervalo abierto $(0,+\infty)$. En particular, $\mathbb{P}[Y > 0] = 1$.

Dado $y > 0$, para determinar $g^{-1}(y)$ debemos resolver la ecuación

$$g(x) = y \iff \frac{x}{1-x} = y \iff x(1+y) = y \iff x = \frac{y}{y+1}.$$

Por lo tanto, $g^{-1} : (0,+\infty) \to (0,1)$ está explícitamente dada por la fórmula

$$g^{-1}(y) = \frac{y}{y+1}.$$

Finalmente, y de acuerdo a las identidades en (2.19) y (2.20) tenemos que

$$f_Y(y) = \begin{cases} 0, & \text{si } y \leq 0; \\ \frac{1}{\left|g'\left(\frac{y}{y+1}\right)\right|}, & \text{si } y > 0; \end{cases} = \begin{cases} 0, & \text{si } y \leq 0; \\ \frac{1}{(1+y)^2}, & \text{si } y > 0. \end{cases}$$

El lector podrá corroborar por su cuenta que f_Y es efectivamente una función de densidad i.e. $f_Y \geq 0$ y $\int_{\mathbb{R}} f_Y(y)\,dy = 1$. Esta práctica resulta útil para identificar errores de cálculo.

Ejemplo 2.17. En el Ejemplo 2.12 vimos que si lanzamos un dardo al azar sobre un blanco redondo de radio 1 unidad, entonces la v.a. R definida como la distancia entre el centro y el punto de impacto tiene una función de densidad dada por la fórmula

$$f_R(r) = \begin{cases} 2r, & r \in [0,1]; \\ 0, & r \notin [0,1]. \end{cases} \tag{2.21}$$

Imaginemos ahora que el blanco es parte de un juego de azar donde un jugador recibe $Y = 2\pi(1-R^2)$ pesos por cada peso apostado. Por ejemplo, si el jugador apuesta \$100 pesos y al lanzar el dardo $R = 0{,}9$ entonces el jugador ganará $100Y \approx$ \$119,38 pesos en el juego. En particular, su ganancia neta será $100 \cdot (Y-1) \approx$ \$19,38 pesos. Por otro lado, si $R = 0{,}99$ entonces $100 \cdot (Y-1) \approx -\$87{,}5$ i.e. el jugador perderá aproximadamente \$87,5 pesos.

¿Cuál es la función de densidad de Y? Para responder a la pregunta observe primero que $R \in (0,1)$ con probabilidad uno. Por otro lado, $Y = g(R)$ donde $g : (0,1) \to \mathbb{R}$ es el polinomio definido como $g(r) = 2\pi(1-r^2)$. Como $g'(r) = -4\pi r < 0$, el Método del Jacobiano asegura que Y tiene una función de densidad. Para determinar ésta explícitamente, observe primero que $g(0,1) = (0,2\pi)$ (*¿por qué?*). Por lo tanto, $f_Y(y) = 0$ si $y \notin (0,2\pi)$. Por otro lado, se deduce de (2.19) y (2.21) que

$$f_Y(y) = \frac{f_R(g^{-1}(y))}{|g'(g^{-1}(y))|} = \frac{2 \cdot g^{-1}(y)}{|-4\pi \cdot g^{-1}(y)|} = \frac{1}{2\pi},$$

para todo $y \in (0,2\pi)$. En resumen,

$$f_Y(y) = \begin{cases} 0, & y \notin (0,2\pi); \\ 1/(2\pi), & y \in (0,2\pi); \end{cases}$$

i.e. $Y \sim Uniforme(0,2\pi)$.

¿Jugaría usted este juego de azar? Como Y puede tomar cualquier valor entre 0 y 2π, sin preferencia por ubicación, el jugador ganará en promedio $(\pi-1) \approx$ \$2,14 pesos, por cada peso apostado. Más aún, este juego es bastante excitante. Por ejemplo, la posibilidad de al menos triplicar lo apostado está dada por

$$\mathbb{P}[Y \geq 3] = \int_3^{2\pi} f_Y(y)\,dy = \frac{2\pi - 3}{2\pi} \approx 52{,}3\,\%,$$

i.e. en más de la mitad de los juegos triplicaremos lo apostado. Si usted tiene la suerte de encontrar un casino con este juego de azar, ¡no pierda la oportunidad de hacerse millonario!

El Método del Jacobiano puede extenderse para considerar transformaciones de variables aleatorias que no son estrictamente crecientes ni tampoco decrecientes.

Teorema 2.9. *(Método general del Jacobiano.) Sea X una v.a. continua y $n \geq 1$ un número entero. Si $I_1, \ldots, I_n \subset \mathbb{R}$ son intervalos abiertos y disjuntos tales que $\mathbb{P}[X \in \cup_{i=1}^n I_i] = 1$, y $g : \cup_{i=1}^n I_i \to \mathbb{R}$ es una función continuamente diferenciable tal que $g'(x) \neq 0$, para todo $x \in \cup_{i=1}^n I_i$, entonces la v.a. $Y = g(X)$ es también continua y con función de densidad dada por la fórmula*

$$f_Y(y) = \begin{cases} 0, & y \notin g(\cup_{i=1}^n I_i); \\ \sum\limits_{x:\, g(x)=y} \frac{f_X(x)}{|g'(x)|}, & y \in g(\cup_{i=1}^n I_i). \end{cases} \tag{2.22}$$

Es instructivo explicar la notación anterior. Primero observe que g tiene que ser estrictamente creciente o estrictamente decreciente sobre cada intervalo I_i (*¿por qué?*). En particular, y como $g(\cup_{i=1}^n I_i) = \cup_{i=1}^n g(I_i)$, para cada $y \in \mathbb{R}$, la ecuación $g(x) = y$ puede tener a lo más n soluciones. De hecho, si y está fuera del rango de g i.e. $y \notin g(\cup_{i=1}^n I_i)$ entonces esta ecuación no tiene solución. Por otro lado, si y pertenece al rango de g, esta ecuación tiene tantas soluciones como hay i's tales que $g(x_i) = y$ para un cierto $x_i \in I_i$. La demostración del Teorema 2.9 será omitida, sin embargo, hemos indicado en el Ejercicio 54 cómo demostrar este teorema para el caso especial de $n = 2$.

Ejemplo 2.18. Considere una v.a. continua X tal que $f_X(-3) = 21/2$ y $f_X(6) = 9/2$ y defina $Y = X - X^2/3$. *¿Es Y una v.a. continua? En caso afirmativo, determine* $f_Y(-6)$.

Primero observe que $Y = g(X)$, con $g(x) = x - x^2/3$; en particular, g es una función continuamente diferenciable sobre $\mathbb{R}$. Más aún, g es estrictamente creciente sobre el intervalo $I_1 = (-\infty, 3/2)$ y estrictamente decreciente sobre el intervalo $I_2 = (3/2, +\infty)$. Como $\mathbb{P}[X \in (I_1 \cup I_2)] = 1 - \mathbb{P}[X = 3/2] = 1$, el Método general del Jacobiano implica que Y es una v.a. continua.

Para determinar $f_Y(-6)$, necesitamos determinar las soluciones de la ecuación $g(x) = -6$, con $x \in (I_1 \cup I_2)$. Uno encuentra que $x = (-3)$ y $x = 6$. En particular, la identidad en (2.22) implica que

$$f_Y(-6) = \frac{f_X(-3)}{|g'(-3)|} + \frac{f_X(6)}{|g'(6)|} = \frac{21/2}{|3|} + \frac{9/2}{|-3|} = 5.$$

Ejemplo 2.19. Sea X una v.a. continua con función de densidad

$$f(x) = \begin{cases} 0, & x \notin (0,1); \\ x + 2x^3, & x \in (0,1). \end{cases}$$

¿Cuál es la función de densidad de $Y = (2X-1)^2$? Observe que $Y = g(X)$, donde g es el polinomio cuadrático $g(x) = (2x-1)^2$ (ver Figura 2.3). Note que g es estrictamente decreciente en el intervalo abierto $I_1 = (0, 1/2)$, y estrictamente creciente en el intervalo $I_2 = (1/2, 0)$. Como $\mathbb{P}[X \in (I_1 \cup I_2)] = 1$ (*¿por qué?*), el Teorema 2.9

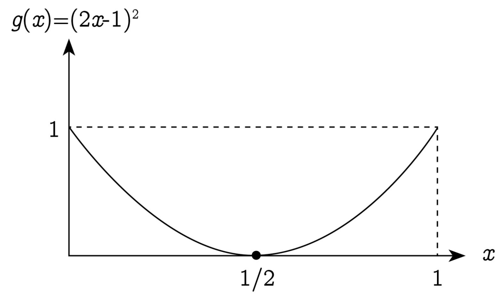


FIGURA 2.3

implica que Y es una v.a. continua. Más aún, como $g(I_1 \cup I_2) = (0,1)$ (*¿por qué?*) y, para cada $y \in (0,1)$ la ecuación $g(x) = y$, con $x \in (I_1 \cup I_2)$, tiene exactamente dos soluciones:

$$\begin{aligned} x_1(y) &= \frac{1-\sqrt{y}}{2}, \\ x_2(y) &= \frac{1+\sqrt{y}}{2}, \end{aligned}$$

la identidad en (2.22) implica que

$$f_Y(y) = \begin{cases} 0, & y \notin (0,1); \\ \frac{f(x_1(y))}{|g'(x_1(y))|} + \frac{f(x_2(y))}{|g'(x_2(y))|}, & y \in (0,1); \end{cases} = \begin{cases} 0, & y \notin (0,1); \\ \frac{3(1+y)}{8\sqrt{y}}, & y \in (0,1). \end{cases}$$

Nuevamente, una manera de corroborar la respuesta anterior es verificar que f_Y es efectivamente una función de densidad. En efecto, de la definición es claro que $f_Y(y) \geq 0$, para todo $y \in \mathbb{R}$ y, por otro lado:

$$\int_{\mathbb{R}} f_Y(y)\,dy = \int_0^1 \frac{3(1+y)}{8\sqrt{y}}\,dy = \left.\frac{3y^{1/2}+y^{3/2}}{4}\right|_{y=0}^{y=1} = 1.$$

2.5 Igualdad en distribución

Considere dos variables aleatorias X e Y asociadas a un mismo experimento. Diremos que X *es igual a* Y y anotaremos $X = Y$ si $X(\omega) = Y(\omega)$, para todo $\omega \in \Omega$. En palabras $X = Y$ cuando—sin importar el resultado del experimento—ambas variables aleatorias toman el mismo valor. En esta sección estudiaremos otra noción de igualdad que es menos restrictiva que la recién entregada, pero aún así fundamental para el desarrollo de la teoría.

Definición 2.10. Dos variables aleatorias X e Y se dicen *igualmente distribuidas* o que *tienen la misma distribución* si $\mathbb{P}[X \in I] = \mathbb{P}[Y \in I]$, para todo intervalo $I \subset \mathbb{R}$.

La definición anterior requiere, por ejemplo, que si X tiene una posibilidad de un 15 % de tomar un valor en un cierto intervalo I entonces Y tiene exactamente la misma oportunidad de tomar un valor en este intervalo. Por otro lado, si Y tiene una posibilidad de un 0 % de tomar un valor en un cierto intervalo J, entonces X también tiene una oportunidad de un 0 % de tomar un valor en este intervalo. En otras palabras: *X e Y tienen la misma distribución cuando la probabilidad de cualquier suceso asociado a la v.a. X no cambia cuando reemplazamos X por Y en dicho suceso.*

Ejemplo 2.20. Si X e Y son variables aleatorias con la misma distribución y X tiene función de densidad

$$f_X(x) = \frac{|x|}{2} \cdot \exp\left\{-\frac{x^2}{2}\right\},$$

¿cuál es la probabilidad del suceso $[-4 \leq Y \leq -2]$*?* Debido a la igualdad en distribución, los sucesos $[-4 \leq Y \leq -2]$ y $[-4 \leq X \leq -2]$ tienen la misma probabilidad. En particular, tenemos que:

$$\begin{aligned}
\mathbb{P}[-4 \leq Y \leq -2] &= \mathbb{P}[-4 \leq X \leq -2], \\
&= -\int_{-4}^{-2} \frac{x}{2} \cdot \exp\left\{-\frac{x^2}{2}\right\} dx, \\
&= \frac{1}{2} \cdot \exp\left\{-\frac{x^2}{2}\right\}\Bigg|_{x=-4}^{x=-2} = \frac{e^{-2} - e^{-8}}{2}.
\end{aligned}$$

Ejemplo 2.21. Para fijar ideas imagine que usted y luego un amigo, lanzan un mismo dado al azar. Si X denota el resultado de su lanzamiento e Y el de su amigo, entonces X e Y debieran tener la misma distribución, ya que el mismo dado es lanzado dos veces. Por ejemplo, si el dado es equilibrado entonces tanto para el primer como segundo lanzamiento, la probabilidad que la cara lanzada sea menor o igual a tres es $1/2$, la probabilidad que la cara lanzada sea un cinco es $1/6$, y la probabilidad que la cara lanzada sea un número mayor o gual a diez es 0. Note que la igualdad en distribución de X e Y aplica incluso si el dado no es equilibrado.

En principio, para verificar la igualdad en distribución de dos variables aleatorias X e Y tenemos que comparar la probabilidad del suceso $[X \in I]$ con la de $[Y \in I]$, para cada intervalo I. Sin embargo, y según hemos visto, usando la función de distribución de una v.a. podemos determinar la probabilidad de que ésta tome un valor en cualquier intervalo (fórmulas (2.8)-(2.13)). En particular, si dos variables aleatorias tienen la misma función de distribución, entonces no queda otra cosa que tengan la misma distribución. Recíprocamente, si X e Y tienen la misma distribución entonces

$$F_X(t) = \mathbb{P}[X \leq t] = \mathbb{P}[X \in (-\infty, t]] = \mathbb{P}[Y \in (-\infty, t]] = \mathbb{P}[Y \leq t] = F_Y(t),$$

para todo $t \in \mathbb{R}$. Hemos, por lo tanto, demostrado el siguiente resultado.

Teorema 2.11. *Dos variables aleatorias X e Y tienen la misma distribución si y solo si $F_X = F_Y$.*

Ejemplo 2.22. Considere el experimento donde se escoge un punto (X,Y) equidistribuido en el *círculo unitario* i.e. el conjunto de puntos (x,y) en el plano tales que $x^2+y^2 \leq 1$.

¿Son X e Y variables aleatorias con la misma distribución? El hecho que (X,Y) es escogido equidistribuidamente en el círculo se interpreta como sigue (Ejemplo 2.12): *la probabilidad que (X,Y) pertenezca a una cierta región contenida en el círculo es proporcional al área de dicha región.* En particular, como el área total del círculo unitario es π, la constante de proporcionalidad tiene que ser $1/\pi$. Equivalentemente,

$$\mathbb{P}[(X,Y)\in R] = \frac{\text{área}(R)}{\pi}, \text{ para toda región } R \text{ contenida en el círculo.} \tag{2.23}$$

Demostraremos que X e Y tienen la misma distribución usando la identidad recién vista. Para esto observe que tanto X como Y solamente pueden tomar valores en el intervalo $[-1,1]$. En particular,

$$F_X(t) = F_Y(t) = 0, \quad \text{cuando} \quad t < -1; \tag{2.24}$$

$$F_X(t) = F_Y(t) = 1, \quad \text{cuando} \quad t > 1. \tag{2.25}$$

Por otro lado, si $-1 \leq t \leq 1$ entonces

$$\begin{aligned} F_X(t) &= \mathbb{P}[X \leq t] = \mathbb{P}[(X,Y)\in A] = \frac{\text{área}(A)}{\pi}, \\ F_Y(t) &= \mathbb{P}[Y \leq t] = \mathbb{P}[(X,Y)\in B] = \frac{\text{área}(B)}{\pi}, \end{aligned}$$

donde se definen (ver Figura 2.4)

$$\begin{aligned} A &= \{(x,y) : x \leq t \text{ y } x^2+y^2 \leq 1\}, \\ B &= \{(x,y) : y \leq t \text{ y } x^2+y^2 \leq 1\}. \end{aligned}$$

Pero observe que A y B tienen la misma área ya que el conjunto B corresponde a una rotación del conjunto A y al rotar una región no cambiamos su área. En particular,

$$F_X(t) = F_Y(t), \text{ para todo } t \in [-1,1]. \tag{2.26}$$

Las identidades en (2.24)-(2.26) implican que $F_X = F_Y$, por lo tanto, X e Y tienen la misma distribución.

Para finalizar la primera parte de esta sección note que si X e Y tienen la misma distribución entonces X^2 e Y^2 también tienen la misma distribución ya que $F_{X^2}(t) = F_{Y^2}(t)$, para todo $t < 0$, y también

$$\begin{aligned} F_{X^2}(t) = \mathbb{P}[X^2 \leq t] &= \mathbb{P}[-\sqrt{t} \leq X \leq \sqrt{t}], \\ &= \mathbb{P}[-\sqrt{t} \leq Y \leq \sqrt{t}] = \mathbb{P}[Y^2 \leq t] = F_{Y^2}(t), \end{aligned}$$

para todo $t \geq 0$. El Teorema 2.11 implica que X^2 tiene la misma distribución que Y^2. Sorprendentemente, y según lo establece el siguiente resultado, e^{-X} y e^{-Y} también tienen la misma distribución, e incluso $\cos(X/(1+X^2))$ y $\cos(Y/(1+Y^2))$ tienen la misma distribución.

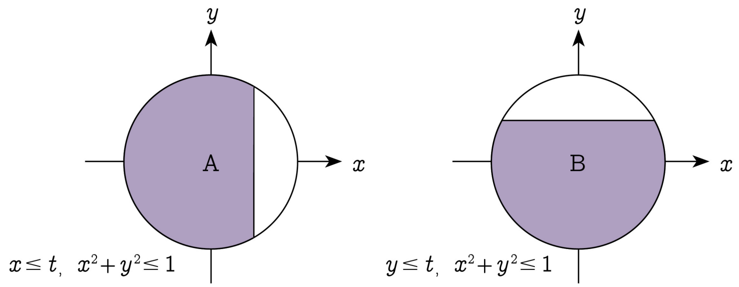


FIGURA 2.4. Representación de los sucesos $[X \leq t]$ e $[Y \leq t]$, con $0 \leq t \leq 1$, cuando (X, Y) es un punto equidistribuido en el círculo unitario.

Teorema 2.12. *Si X e Y son variables aleatorias con la misma distribución entonces, para toda función $g : \mathbb{R} \to \mathbb{R}$ continua excepto por un número finito de puntos, las variables aleatorias $g(X)$ y $g(Y)$ tienen la misma ditribución.*

Demostración. Para demostrar el resultado usaremos que si $g : \mathbb{R} \to \mathbb{R}$ es una función continua excepto sobre un cierto conjunto finito de puntos entonces, para todo intervalo abierto $I \subset \mathbb{R}$, la pre-imagen de I a través de g puede escribirse como una unión contable de intervalos disjuntos i.e. existe una secuencia contable $I_1, I_2, \ldots$ de intervalos disjuntos tales que $g^{-1}(I) = \cup_i I_i$. En particular, y debido a la sigma aditividad, tenemos que

$$\begin{aligned} \mathbb{P}[g(X) \in I] = \mathbb{P}[X \in g^{-1}(I)] &= \sum_i \mathbb{P}[X \in I_i], \\ &= \sum_i \mathbb{P}[Y \in I_i] = \mathbb{P}[Y \in g^{-1}(I)] = \mathbb{P}[g(Y) \in I]. \end{aligned}$$

Más generalmente, si $I \subset \mathbb{R}$ es un intervalo arbitrario, entonces existe una sequencia de intevalos abiertos $(I_n)_{n\geq 1}$ tal que $I_{n+1} \subset I_n$ e $I = \cap_{n=1}^{\infty} I_n$. En particular, los sucesos $[g(X) \in I_n]$ decrecen al suceso $[g(X) \in I]$. Por lo tanto, debido al Teorema 1.8 y a las identidades anteriores, encontramos que

$$\mathbb{P}[g(X) \in I] = \lim_{n\to\infty} \mathbb{P}[g(X) \in I_n] = \lim_{n\to\infty} \mathbb{P}[g(Y) \in I_n] = \mathbb{P}[g(Y) \in I],$$

i.e. X e Y tienen la misma distribución. □

Ejemplo 2.23. Si X e Y son variables aleatorias tales que X^3 e $(Y^5 + Y)$ tienen la misma distribución, *¿es cierto que X tiene la misma distribución que $\sqrt[3]{Y^5 + Y}$?* Observe que la función $g : \mathbb{R} \to \mathbb{R}$ definida como $g(t) = \sqrt[3]{t}$ es continua (*¿por qué?*). Por lo tanto, el Teorema 2.12 nos permite concluir que $g(X^3)$ i.e. X tiene la misma distribución que $g(Y^5 + Y)$ i.e. $\sqrt[3]{Y^3 + Y}$.

2.5.1 Igualdad en distribución de variables aleatorias discretas.

En esta sección especializaremos la discusión anterior en el contexto de variables aleatorias discretas. El resultado principal es el siguiente.

Teorema 2.13. *Si X e Y son variables aleatorias discretas entonces X e Y tienen la misma distribución si y solo si $\mathbb{P}[X = a] = \mathbb{P}[Y = a]$, para todo $a \in \mathbb{R}$.*

Demostración. Claramente, si X e Y tienen la misma distribución entonces, para todo $a \in \mathbb{R}$, $\mathbb{P}[X = a] = \mathbb{P}[X \in [a, a]] = \mathbb{P}[Y \in [a, a]] = \mathbb{P}[Y = a]$. Recíprocamente, como X e Y son variables aleatorias discretas, existe un conjunto contable A tal que $\mathbb{P}[X \in A] = \mathbb{P}[Y \in A] = 1$ (*¿por qué?*). En particular, tenemos que

$$\begin{aligned} F_X(t) = \mathbb{P}[X \leq t, X \in A] &= \sum_{a \in A\colon a \leq t} \mathbb{P}[X = a], \\ &= \sum_{a \in A\colon a \leq t} \mathbb{P}[Y = a] = \mathbb{P}[Y \leq t, Y \in A] = F_Y(t), \end{aligned}$$

para cada $t \in \mathbb{R}$, i.e. X e Y tienen la misma distribución (Teorema 2.11). Esto completa la demostración del teorema. □

Es importante enfatizar que la condición dada en el Teorema 2.13, no es suficiente para concluir que dos variables aleatorias arbitrarias tienen la misma distribución. Por ejemplo, si $X \sim Uniforme[0, 1]$ e $Y \sim Uniforme[2, 3]$ entonces $\mathbb{P}[X = a] = \mathbb{P}[Y = a] = 0$, para todo $a \in \mathbb{R}$ (*¿por qué?*). Sin embargo, X e Y no pueden tener la misma distribución ya que $\mathbb{P}[2 \leq X \leq 3] = 0 < 1 = \mathbb{P}[2 \leq Y \leq 3]$.

Ejemplo 2.24. Considere el experimento donde se escogen al azar dos profesores de un cierto colegio y se registran los números X e Y de libros de matemáticas en sus oficinas, respectivamente. Claramente X e Y son variables aleatorias discretas que sólo pueden tomar valores enteros no negativos. Para cada entero $k \geq 0$ note que $\mathbb{P}[X = k] = n_k/n$, donde n_k es el número de profesores con k libros en sus oficinas y n es el número de profesores en el colegio. Similarmente, $\mathbb{P}[Y = k] = n_k/n$. El Teorema 2.13 implica, por lo tanto, que X e Y tienen la misma distribución.

Ejemplo 2.25. Considere una bolsa con diez bolitas marcadas con los dígitos $0, \ldots, 9$ de modo que ningún dígito aparece repetido. Considere el siguiente experimento con dos fases. Una persona saca al azar una bolita de la bolsa marcada con un cierto dígito X. Luego, sin reponer la bolita, la persona mezcla las bolitas que quedan en la bolsa y saca una segunda bolita al azar esta vez marcada con un cierto dígito Y, *¿son X e Y igualmente distribuidas?*

Claramente X e Y son variables aleatorias discretas que sólo pueden tomar valores en el conjunto $\{0, \ldots, 9\}$. En particular, para determinar si tienen la misma distribución basta con comparar las probabilidades $\mathbb{P}[Y = k]$ y $\mathbb{P}[X = k]$, para cada $k \in \{0, \ldots, 9\}$. En efecto, usando la Fórmula de Probabilidades Totales (ver identidad

(1.16)) obtemos que

$$\begin{aligned}
\mathbb{P}[Y=k] &= \sum_{i=0,\ldots,9} \mathbb{P}[Y=k \mid X=i]\cdot\mathbb{P}[X=i],\\
&= \sum_{i=0,\ldots,9:\, i\neq k} \mathbb{P}[Y=k \mid X=i]\cdot\mathbb{P}[X=i],\\
&= \sum_{i=0,\ldots,9:\, i\neq k} \frac{1}{9}\cdot\frac{1}{10} = \frac{1}{10}.
\end{aligned}$$

Esto muestra que $\mathbb{P}[Y=k]=\mathbb{P}[X=k]$, para cada $k\in\{0,\ldots,9\}$. Por lo tanto, X e Y tienen la misma distribución. Observe, sin embargo, que $\mathbb{P}[X\neq Y]=1$ ya que la primera bolita extraída no es repuesta en la bolsa. Este ejemplo muestra que dos variables aleatorias pueden tener la misma distribución incluso si siempre toman valores diferentes.

2.5.2 Igualdad en distribución de variables aleatorias continuas.

En esta sección, veremos una caracterización alternativa de la igualdad en distribución de variables aleatorias continuas. El resultado principal es el siguiente.

Teorema 2.14. *Si X e Y son variables aleatorias continuas entonces, X e Y tienen la misma distribución si y solo si $f_X=f_Y$ excepto sobre un número finito de puntos.*

Demostración. Recordemos que debido al Teorema 2.11, X e Y tienen la misma distribución si y solo si $F_X=F_Y$. En particular, si X e Y son variables aleatorias continuas entonces la identidad en (2.15) implica que $f_X=f_Y$. Por otro lado, si $f_X=f_Y$ excepto sobre un número finito de puntos entonces, para cada $t\in\mathbb{R}$ encontramos que

$$F_X(t)=\mathbb{P}[X\leq t]=\int_{-\infty}^{t} f_X(s)\,ds=\int_{-\infty}^{t} f_Y(s)\,ds=\mathbb{P}[Y\leq t]=F_Y(t),$$

i.e. $F_X=F_Y$ y, por lo tanto, X e Y tienen la misma distribución. Esto demuestra el teorema. □

En otras palabras, la función de densidad de una v.a. continua, especifica completamente su distribución. Por ejemplo, si X e Y son variables aleatorias uniformes sobre un mismo intervalo $[a,b]$ entonces X e Y tienen la misma distribución debido a que la definición de v.a. uniforme fue dada en términos de su función de densidad. Aplicaremos esta misma idea para definir las llamadas variables aleatorias del tipo exponencial y del tipo Normal en las secciones 4.5 y 4.6.

2.6 Independencia de variables aleatorias

La noción de independencia entre variables aleatorias está íntimamente ligada a la noción de independencia de sucesos discutida en las secciones 1.3 y 1.4. En un sentido amplio, dos o más variables aleatorias son independientes si cualquier suceso asociado a una de ellas es independiente de cualquier otro suceso asociado al resto. La definición precisa es la siguiente.

Definición 2.15. Las variables aleatorias $X_1, \ldots, X_n$ se dirán *independientes* cuando dada cualquier secuencia de intervalos $I_1, \ldots, I_n$, los sucesos $[X_1 \in I_1], \ldots, [X_n \in I_n]$ son independientes i.e.

$$\mathbb{P}[X_1 \in I_1, \ldots, X_n \in I_n] = \mathbb{P}[X_1 \in I_1] \cdots \mathbb{P}[X_n \in I_n].$$

Para fijar ideas consideraremos algunos ejemplos.

Ejemplo 2.26. Si X e Y son variables aleatorias independientes tales que $X \sim Uniforme[1,7]$ e $Y \sim Uniforme[-2,2]$, *¿cuál es la probabilidad del suceso* $[X \geq 5, Y \leq 1]$*?*

Observe primero que $[X \geq 5, Y \leq 1] = [X \geq 5] \cap [Y \leq 1]$ y que los sucesos $[X \geq 5]$ e $[Y \leq 1]$ son independientes debido a que X e Y son variables aleatorias independientes. Obtenemos de esta manera que

$$\begin{aligned}
\mathbb{P}[X \geq 5, Y \leq 1] &= \mathbb{P}[X \geq 5] \cdot \mathbb{P}[Y \leq 1], \\
&= \mathbb{P}[5 \leq X \leq 7] \cdot \mathbb{P}[-2 \leq Y \leq 1], \\
&= \frac{2}{6} \cdot \frac{3}{4} = \frac{1}{4}.
\end{aligned}$$

Ejemplo 2.27. Imagine que la mayor de dos hermanas mandó X invitaciones para su matrimonio, mientras que la menor mandó Y. En principio, y sin saber más sobre las hermanas, $\mathbb{P}[Y = k]$ representa la fracción de parejas que envió k invitaciones para su matrimonio. Sin embargo, si sabemos que $X \geq 500$ entonces—con alta certeza—la probabilidad del suceso $[Y \geq 300]$ debiera ser mucho mayor que la de una persona típica i.e. $\mathbb{P}[Y \geq 300 \mid X \geq 500] > \mathbb{P}[Y \geq 300]$. Esto último es equivalente a tener que $\mathbb{P}[Y \geq 300, X \geq 500] > \mathbb{P}[Y \geq 300] \cdot \mathbb{P}[X \geq 500]$, por lo tanto, X e Y no son independientes.

Imagine ahora que, en dos días diferentes, escogemos al azar, en la calle, a una pareja de casados y le preguntamos el número total de invitaciones que mandaron para su matrimonio. Si X denota la respuesta de la primera pareja e Y la de la segunda entonces la probabilidad del suceso $[Y \geq 300]$ seguirá siendo igualmente pequeña sabiendo que $[X \geq 500]$. Más generalmente, la información que acarrea X no debiera tener ningún efecto sobre las probabilidades de cada uno de los sucesos $[Y = 0]$, $[Y = 1]$, $[Y = 2]$, etc. Por lo tanto, X e Y debieran ser independientes en este caso.

Ejemplo 2.28. Considere el siguiente experimento donde se escoge un punto (X, Y) equidistribuido dentro del círculo unitario, *¿es X independiente de Y?*

Para responder la pregunta, observe que Y puede en principio tomar cualquier valor en el intervalo $[-1,1]$. Por otro lado, si X toma un valor cerca de uno entonces Y está obligado a tomar un valor cerca de cero debido a que $X^2 + Y^2 \leq 1$. Esto indica

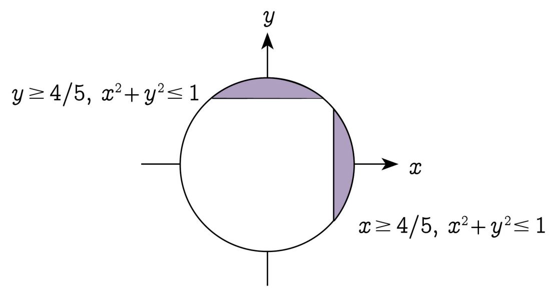


Figura 2.5. Representación de los sucesos $[X \geq 4/5]$ e $[Y \geq 4/5]$, cuando (X, Y) es un punto equidistribuido en el círculo unitario. El diagrama muestra que estos dos sucesos no pueden ocurrir simultáneamente.

que X e Y no son realmente independientes. Matemáticamente podemos demostrar esto notando que si $X \geq 4/5$ entonces $|Y| \leq 3/5$. En particular, tenemos que $\mathbb{P}[X \geq 4/5, Y \geq 4/5] = 0$ (ver Figura 2.5). Sin embargo, como $\mathbb{P}[X \geq 4/5] > 0$ y $\mathbb{P}[Y \geq 4/5] > 0$ deducimos que

$$\mathbb{P}[X \geq 4/5, Y \geq 4/5] \neq \mathbb{P}[X \geq 4/5] \cdot \mathbb{P}[Y \geq 4/5].$$

Por lo tanto, X e Y no pueden ser independientes.

Recordemos que para comprobar la independencia de n sucesos necesitamos verificar $(2^n - n - 1)$ identidades (ver Sección 1.4), lo que dificulta verificar la independencia de más que unos pocos sucesos. El siguiente resultado, sin embargo, nos permitirá verificar de forma relativamente simple la independencia de variables aleatorias en una variedad de situaciones.

Teorema 2.16. *Las variables aleatorias X e Y son independientes si y solo si $\mathbb{P}[X \leq x, Y \leq y] = F_X(x) \cdot F_Y(y)$, para todo $x, y \in \mathbb{R}$. Más generalmente, las variables aleatorias $X_1, \ldots, X_n$ son independientes si y solo si*

$$\mathbb{P}[X_1 \leq x_1, \ldots, X_n \leq x_n] = F_{X_1}(x_1) \cdots F_{X_n}(x_n),$$

para todo $x_1, \ldots, x_n \in \mathbb{R}$.

Demostración. Sólo explicaremos la validez del resultado para el caso de $n = 2$. Primero observe que si X e Y son idependientes, entonces los sucesos $[X \leq x]$ e $[Y \leq y]$ son por definición independientes para cada $x, y \in \mathbb{R}$. En particular,

$$\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x] \cdot \mathbb{P}[Y \leq y] = F_X(x) \cdot F_Y(y).$$

Recíprocamente, si la identidad anterior aplica para cada $x, y \in \mathbb{R}$ entonces para todo $-\infty \leq a < b \leq \infty$ y $c \in \mathbb{R}$ encontramos que (*¿por qué?*)

$$\begin{aligned} \mathbb{P}[X \in (a,b], Y \leq c] &= \mathbb{P}[X \leq b, Y \leq c] - \mathbb{P}[X \leq a, Y \leq c], \\ &= F_X(b) \cdot F_Y(c) - F_X(a) \cdot F_Y(c), \\ &= (F_X(b) - F_X(a)) \cdot F_Y(c) = \mathbb{P}[X \in (a,b]] \cdot F_Y(c). \end{aligned}$$

Más aún, como $\mathbb{P}[X \in (a,b], Y \in (c,d]] = \mathbb{P}[X \in (a,b], Y \leq d] - \mathbb{P}[X \in (a,b], Y \leq c]$, deducimos que

$$\mathbb{P}[X \in (a,b], Y \in (c,d]] = \mathbb{P}[X \in (a,b]] \cdot \mathbb{P}[Y \in (c,d]], \tag{2.27}$$

para $-\infty \leq a < b \leq \infty$ y $-\infty \leq c < d \leq \infty$. Esto muestra que $[X \in I]$ e $[Y \in J]$ son sucesos independientes para todo par de intervalos de la forma $I = (a,b]$ y $J = (c,d]$. El objetivo es ahora demostrar que la identidad anterior, también aplica para otros tipos de intervalos. Esto puede hacerse de manera similar a como demostramos el Teorema 2.2. Por ejemplo, usando que los sucesos $[X \in (a, b-1/n], Y \in (c,d]]$ crecen al suceso $[X \in (a,b), Y \in (c,d]]$ y que $[X \in (a, b-1/n]]$ crecen al suceso $[X \in (a,b)]$ (*¿por qué?*), la identidad en (2.27) nos permite concluir que

$$\begin{aligned} \mathbb{P}[X \in (a,b), Y \in (c,d]] &= \lim_{n\to\infty} \mathbb{P}[X \in (a, b-1/n], Y \in (c,d]], \\ &= \lim_{n\to\infty} \mathbb{P}[X \in (a, b-1/n]] \cdot \mathbb{P}[Y \in (c,d]], \\ &= \mathbb{P}[X \in (a,b)] \cdot \mathbb{P}[Y \in (c,d]]. \end{aligned}$$

Similarmente, se puede demostrar que $\mathbb{P}[X \in I, Y \in J] = \mathbb{P}[X \in I] \cdot \mathbb{P}[Y \in J]$ para todo par de intervalos $I, J \subset \mathbb{R}$, lo que completa la demostración. □

El siguiente ejemplo ilustra cómo aplicar el resultado anterior en una situación concreta.

Ejemplo 2.29. Considere el experimento donde se escoge un punto (X,Y) equidistribuido dentro del cuadrado con vértices $(1,1)$, $(-1,1)$, $(-1,-1)$ y $(1,-1)$. Demostraremos que X e Y son independientes usando el Teorema 2.16.

Claramente X e Y sólo pueden tomar valores en el intervalo $[-1,1]$. En particular, la identidad en el Teorema 2.16 es obvia si $x \notin [-1,1]$ o $y \notin [-1,1]$. Por ejemplo, si $x < -1$ entonces $\mathbb{P}[X \leq x, Y \leq y] = 0$ y $F_X(x) = 0$. Por otro lado, si $x > 1$ entonces $\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[Y \leq y] = F_Y(y)$ y $F_X(x) = 1$. En lo que sigue asumiremos, por lo tanto, que $x, y \in [-1,1]$.

En efecto, si $x, y \in [-1,1]$ entonces

$$\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[-1 \leq X \leq x, -1 \leq Y \leq y] = \mathbb{P}[(X,Y) \in R],$$

donde R es el rectángulo $\{(s,t) : -1 \leq s \leq x, -1 \leq t \leq y\}$. Como (X,Y) es escogido equidistribuidamente en el cuadrado, entonces la probabilidad de que (X,Y) pertenezca a una cierta región A dentro del cuadrado debiera ser proporcional al área de dicha región (ver ejemplos 2.12 y 2.22). Como el área total del cuadrado es 4, la constante de proporcionalidad tiene que ser 1/4, y obtenemos, por lo tanto, que

$$\mathbb{P}[X \leq x, Y \leq y] = \frac{\text{área}(R)}{4} = \frac{(x+1)(y+1)}{4}.$$

La identidad anterior nos permite determinar la función de distribución de X e Y directamente como sigue:

$$\begin{aligned} F_X(x) &= \mathbb{P}[X \leq x] = \mathbb{P}[X \leq x, Y \leq 1] = \frac{(x+1)\cdot 2}{4} = \frac{x+1}{2}, \\ F_Y(y) &= \mathbb{P}[Y \leq y] = \mathbb{P}[X \leq 1, Y \leq y] = \frac{2\cdot(y+1)}{4} = \frac{y+1}{2}. \end{aligned}$$

En particular, $\mathbb{P}[X \leq x, Y \leq y] = F_X(x) \cdot F_Y(y)$, para todo $x, y \in [-1, 1]$, y por lo tanto, X e Y son variables aleatorias independientes.

Para finalizar esta sección, note que si X e Y son variables aleatorias independientes, entonces también lo son X^2 y e^{-Y}. Para verificar esto usaremos el Teorema 2.16. En efecto, si $s \geq 0$ y $t > 0$ entonces

$$\begin{aligned} \mathbb{P}[X^2 \leq s, e^{-Y} \leq t] &= \mathbb{P}[-\sqrt{s} \leq X \leq \sqrt{s}, Y \geq -\ln(t)], \\ &= \mathbb{P}[-\sqrt{s} \leq X \leq \sqrt{s}] \cdot \mathbb{P}[Y \geq -\ln(t)], \\ &= \mathbb{P}[X^2 \leq s] \cdot \mathbb{P}[e^{-Y} \leq t]. \end{aligned}$$

Observe ahora que si $s < 0$ o $t \leq 0$ entonces tanto el lado derecho como el izquierdo anteriores son idénticamente cero (*¿por qué?*). Esto muestra que X e Y son independientes. Interesantemente, y según lo establece el siguiente resultado, $\cos(X^3)$ e $Y\ln(1+Y^2)$ son también independientes, al igual que lo son $(-1)^X$ e Y.

Teorema 2.17. *Si X e Y son independientes y $g, h : \mathbb{R} \to \mathbb{R}$ son funciones continuas excepto por un número finito de puntos, entonces $g(X)$ y $h(Y)$ son también independientes. Más generalmente, si $X_1, \ldots, X_n$ son variables aleatorias independientes y para cada i, $g_i : \mathbb{R} \to \mathbb{R}$ es una función continua excepto por un número finito de puntos, entonces $g_1(X_1), \ldots, g_n(X_n)$ son también independientes.*

Demostración. La razón de este resultado es similar a la del Teorema 2.12. Solamente justificaremos el caso con $n = 2$. En efecto, si $I, J \subset \mathbb{R}$ son intervalos abiertos entonces existen secuencias contables de intervalos disjuntos $I_1, I_2, \ldots$ y $J_1, J_2, \ldots$ tales que $g^{-1}(I) = \cup_i I_i$ y $h^{-1}(J) = \cup_j J_j$. En particular, y debido a la sigma aditividad, tenemos que

$$\begin{aligned}
\mathbb{P}[g(X) \in I, h(Y) \in J] &= \mathbb{P}[X \in g^{-1}(I), Y \in h^{-1}(J)], \\
&= \sum_i \sum_j \mathbb{P}[X \in I_i, Y \in J_j], \\
&= \sum_i \sum_j \mathbb{P}[X \in I_i] \cdot \mathbb{P}[Y \in J_j], \\
&= \sum_i \mathbb{P}[X \in I_i] \cdot \sum_j \mathbb{P}[Y \in J_j], \\
&= \mathbb{P}[X \in g^{-1}(I)] \cdot \mathbb{P}[Y \in h^{-1}(J)], \\
&= \mathbb{P}[g(X) \in I] \cdot \mathbb{P}[h(Y) \in J].
\end{aligned}$$

Más generalmente, si $I, J \subset \mathbb{R}$ son intervalos arbitrarios, entonces existen sequencias de intevalos abiertos $(I_n)_{n\geq 1}$ y $(J_n)_{n\geq 1}$ tales que $I_{n+1} \subset I_n$, $J_{n+1} \subset J_n$, $I = \cap_{n=1}^{\infty} I_n$ y $J = \cap_{n=1}^{\infty} J_n$. Como los sucesos $[g(X) \in I_n, h(Y) \in J_n]$ decrecen al suceso $[g(X) \in I, h(Y) \in J]$ (*¿por qué?*), obtenemos finalmente que (Teorema 1.8)

$$\begin{aligned}
\mathbb{P}[g(X) \in I, h(Y) \in J] &= \lim_{n\to\infty} \mathbb{P}[g(X) \in I_n, h(Y) \in J_n], \\
&= \lim_{n\to\infty} \mathbb{P}[g(X) \in I_n] \cdot \mathbb{P}[h(Y) \in J_n], \\
&= \lim_{n\to\infty} \mathbb{P}[g(X) \in I_n] \cdot \lim_{n\to\infty} \mathbb{P}[h(Y) \in J_n], \\
&= \mathbb{P}[g(X) \in I] \cdot \mathbb{P}[h(Y) \in J],
\end{aligned}$$

lo que demuestra que X e Y son independientes. □

2.6.1 Independencia de variables aleatorias discretas.

En esta sección veremos cómo comprobar más fácilmente la independencia de variables aleatorias discretas. El resultado principal es el siguiente.

Teorema 2.18. *Si X e Y son variables aleatorias discretas entonces X e Y son independientes si y solo si $\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y]$, para todo $x, y \in \mathbb{R}$. Más generalmente, las variables aleatorias discretas $X_1, \ldots, X_n$ son independientes si y solo si $\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n] = \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n]$, para todo $x_1, \ldots, x_n \in \mathbb{R}$.*

Demostración. Solo demostraremos el resultado para $n = 2$, dejando la demostración general como ejercicio para el lector. En efecto, si X e Y son independientes entonces los sucesos $[X = x]$ e $[Y = y]$ son independientes, para todo $x, y \in \mathbb{R}$, por lo tanto, $\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y]$. Recíprocamente, para demostrar que X e Y son independientes usaremos el Teorema 2.16. En efecto, como X e Y son discretas, entonces existen conjuntos contables $A, B \subset \mathbb{R}$ tales que $\mathbb{P}[X \in A] = \mathbb{P}[Y \in B] = 1$.

Dados $x, y \in \mathbb{R}$ obtenemos, por lo tanto, que

$$\begin{aligned}
\mathbb{P}[X \le x, Y \le y] &= \sum_{a\in A: a\le x} \sum_{b\in B: b\le y} \mathbb{P}[X = a, Y = b], \\
&= \sum_{a\in A: a\le x} \sum_{b\in B: b\le y} \mathbb{P}[X = a] \cdot \mathbb{P}[Y = b], \\
&= \sum_{a\in A: a\le x} \mathbb{P}[X = a] \cdot \sum_{b\in B: b\le y} \mathbb{P}[Y = b], \\
&= \mathbb{P}[X \in A, X \le x] \cdot \mathbb{P}[Y \in B, Y \le y], \\
&= \mathbb{P}[X \le x] \cdot \mathbb{P}[Y \le y] = F_X(x) \cdot F_Y(y).
\end{aligned}$$

Esto muestra que X e Y son independientes y completa la demostración. □

Ejemplo 2.30. Considere el experimento que consiste en lanzar dos dados equilibrados e independientemente. Denotaremos como X e Y la cara del primer y segundo dado lanzado, respectivamente. Por otro lado definiremos $Z = \text{máx}\{X, Y\}$ i.e. Z corresponde al valor más alto entre las caras de los dos dados. Observe que X, Y y Z solamente pueden tomar valores en el conjunto $A = \{1, 2, 3, 4, 5, 6\}$.

¿Son X e Y variables aleatorias independientes? Intuitivamente el valor que X toma no influencia el valor que Y toma y viceversa. Por lo tanto, X e Y debieran ser variables aleatorias independientes. Matemáticamente tenemos que verificar que $\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y]$, para todo $x, y \in \mathbb{R}$. Como tanto el lado derecho como el izquierdo en esta identidad son idénticamente cero si $x \notin A$ o $y \notin A$ (*¿por qué?*), supondremos que $x, y \in A$. En efecto, en este caso encontramos que

$$\mathbb{P}[X = x, Y = y] = 1/36 = 1/6 \cdot 1/6 = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y],$$

por lo tanto, X e Y son independientes.

¿Son X y Z variables aleatorias independientes? Intuitivamente no, ya que, por ejemplo, si sabemos que $X = 2$ entonces Z ya no puede tomar el valor 1. En particular, la probabilidad del suceso $[Z = 1]$ pasa de ser estrictamente positiva a cero cuando sabemos que $X = 2$. En términos más matemáticos tenemos que

$$\mathbb{P}[X = 2, Z = 1] = 0 \neq 1/6 \cdot 1/36 = \mathbb{P}[X = 2] \cdot \mathbb{P}[Z = 1].$$

Esto muestra que X y Z no son independientes.

2.6.2 Independencia de variables aleatorias continuas.

Para finalizar nuestra discusión sobre independencia presentaremos un último resultado que caracteriza la independencia de dos variables aleatorias continuas. El resultado principal requiere el uso de integrales dobles, por lo tanto, puede ser omitido si el lector no está familiarizado con este concepto.

Teorema 2.19. *Si X e Y son variables aleatorias continuas entonces X e Y son independientes si y solo si, para toda región $R \subset \mathbb{R}^2$ como las descritas a continuación,* *se tiene que*

$$\mathbb{P}[(X,Y)\in R] = \iint_R f_X(x) f_Y(y)\, dy\, dx. \tag{2.28}$$

Hay tres casos importantes de destacar acerca de la integral en el teorema. La integral doble se reduce a dos integrales unidimensionales cuando la región R corresponde al área contenida entre los grafos de dos funciones continuas. Específicamente, dadas constantes $-\infty \le a < b \le +\infty$ y dos funciones continuas g y h, las siguientes identidades aplican:

(i) si $R = \{(x,y) :\, a \le x \le b, g(x) \le y \le h(x)\}$ entonces

$$\iint_R f_X(x) f_Y(y)\, dy\, dx = \int_a^b \left\{ f_X(x) \cdot \int_{g(x)}^{h(x)} f_Y(y)\, dy \right\} dx,$$

(ii) si $R = \{(x,y) :\, g(y) \le x \le h(y), a \le y \le b\}$ entonces

$$\iint_R f_X(x) f_Y(y)\, dy\, dx = \int_a^b \left\{ f_Y(y) \cdot \int_{g(y)}^{h(y)} f_X(x)\, dx \right\} dy.$$

Las identidades anteriores también aplican, cuando las desigualdades que definen la región R son reemplazadas por desigualdades estrictas, e incluso si $g(x) = -\infty$ para todo $x \in [a,b]$, o $h(x) = +\infty$ para todo $x \in [a,b]$. Más aún, en cualquiera de los casos anteriores, y si $c \in \mathbb{R}$ es una constante dada entonces

$$\iint_R c\, dy\, dx = c \cdot \text{área}(R). \tag{2.29}$$

Debido a razones técnicas que exceden el nivel de la monografía no demostraremos el Teorema 2.19. Los siguientes ejemplos, sin embargo, mostrarán como aplicarlo en situaciones concretas.

Ejemplo 2.31. Si X e Y son variables aleatorias independientes tales que $X \sim Uniforme[0,1]$ e Y tiene función de densidad

$$f_Y(y) = \begin{cases} e^{-y}, & y \ge 0; \\ 0, & y < 0; \end{cases}$$

¿cuál es la probabilidad del suceso $[X \le Y]$*?*

Observe que el suceso $[0 \le X \le 1]$ tiene probabilidad uno. En particular, debido a la identidad en (1.13), $\mathbb{P}[X \le Y] = \mathbb{P}[0 \le X \le 1, X \le Y]$. Equivalentemente, si definimos $R = \{(x,y) :\, 0 \le x \le 1 \text{ y } x \le y\}$ entonces $\mathbb{P}[X \le Y] = \mathbb{P}[(X,Y) \in R]$. La propiedad (i) anterior nos permite, por lo tanto, concluir que

$$\begin{aligned}
\mathbb{P}[X \leq Y] &= \iint_R f_X(x) f_Y(y)\, dy\, dx, \\
&= \int_0^1 \left\{ \int_x^\infty e^{-y}\, dy \right\} dx, \\
&= \int_0^1 e^{-x}\, dx = 1 - e^{-1}.
\end{aligned}$$

Ejemplo 2.32. Imagine que la separación temporal entre sismos consecutivos es una v.a. con función de densidad $f(t) = e^{-t}$ si $t \geq 0$, y $f(t) = 0$ si $t < 0$. Mostraremos que la separación temporal entre el último sismo y el subsiguiente es una v.a. continua y determinaremos su función de densidad.

En efecto, denotemos como X la separación temporal entre el último y próximo sismo, y como Y la separación entre el próximo y el subsiguiente. De acuerdo al enunciado, $f_X = f_Y = f$. Supondremos que X e Y son independientes.

La v.a. $(X + Y)$ representa la separación temporal entre el último sismo y el subsiguiente. Para demostrar que $(X + Y)$ es continua, determinaremos primero su función de distribución. En efecto, note que $(X+Y)$ sólo puede tomar valores reales no negativos. Debido a esto, determinaremos $F_{X+Y}(t)$, para $t \geq 0$. Con este fin, observe que $F_{X+Y}(t) = \mathbb{P}[X + Y \leq t] = \mathbb{P}[(X, Y) \in R]$, donde

$$R = \{(x, y) : x \geq 0, y \geq 0, x + y \leq t\} = \{(x, y) : 0 \leq x \leq t, 0 \leq y \leq (t - x)\}.$$

Debido al Teorema 2.19 podemos determinar $F_{X+Y}(t)$ como sigue:

$$\begin{aligned}
F_{X+Y}(t) = \iint_R e^{-x} \cdot e^{-y}\, dydx &= \int_0^t e^{-x} \left\{ \int_0^{t-x} e^{-y}\, dy \right\} dx, \\
&= \int_0^t e^{-x}(1 - e^{x-t})\, dx = 1 - e^{-t}(1 + t).
\end{aligned}$$

Como $F_{X+Y}(t) = 0$ si $t < 0$, la identidad en (2.15) nos permite concluir que

$$f_{X+Y}(t) = \begin{cases} te^{-t}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

Esto demuestra que, la separación temporal entre el último sismo y el subsiguiente, es una v.a. continua.

Ejemplo 2.33. A la hija de Marcela le encanta leer libros de divulgación científica. Es así que, para su cumpleaños, Marcela ha decidido regalarle el libro "El gen egoísta" del controvertido escritor Richard Dawkins. Siendo éste un libro importado, Marcela cotizará el precio en dos librerías diferentes. Denotaremos como X e Y respectivamente el precio del libro en la primera y segunda librería. Marcela pagará por el libro la cantidad $Z = \text{mín}\{X, Y\}$.

Bajo el supuesto que X e Y son independientes y que $X \sim Uniforme[1, 2]$ e $Y \sim Uniforme[1, 2]$, mostraremos que Z es una v.a. continua. Aquí "1" denota el precio

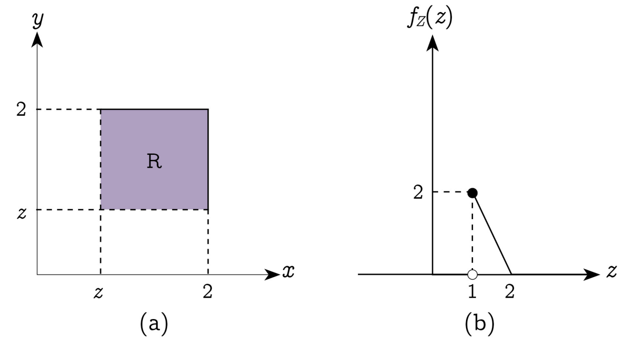


FIGURA 2.6. (a) Representación visual de la región del plano $R = \{(x, y) : z < x \leq 2, z < y \leq 2\}$ para un valor genérico de $z \in [0, 2]$. (b) Grafo de la función de densidad de la v.a. Z del Ejemplo 2.33.

razonable del libro y usualmente recomendado por la casa editora (renormalizado al valor uno). Supondremos, por lo tanto, que las librerías escogen el precio del libro al azar entre una y dos veces el precio recomendado.

Para determinar la función de densidad de Z, determinaremos primero su función de distribución. Como Z sólo puede tomar valores entre 1 y 2 (*¿por qué?*), determinaremos $F_Z(z)$ para $1 \leq z \leq 2$. Observe primero que el suceso $[Z > z]$ es igual al suceso $[(X, Y) \in R]$, donde $R = \{(x, y) : z < x \leq 2, z < y \leq 2\}$ (ver Figura 2.6(a)). En particular, y usando que X e Y son son independientes, el Teorema 2.19 implica que

$$\begin{aligned} F_Z(z) = \mathbb{P}[Z \leq z] &= 1 - \mathbb{P}[(X, Y) \in R], \\ &= 1 - \iint_R f_X(x) \cdot f_Y(y)\, dx\, dy, \\ &= 1 - \iint_R 1\, dx\, dy = 1 - \text{área}(R) = 1 - (2 - z)^2, \end{aligned}$$

donde, para la penúltima igualdad anterior, hemos usado la identidad en (2.29). Concluimos, por lo tanto, que

$$F_Z(z) = \begin{cases} 0, & z < 1; \\ 4z - z^2 - 3, & 1 \leq z \leq 2; \\ 1, & z > 2. \end{cases}$$

Como esta función es continuamente diferenciable, excepto sobre el punto $z = 1$ (*¿por qué?*), obtenemos finalmente que

$$f_Z(z) = \begin{cases} 0, & z \notin [1, 2]; \\ 4 - 2z, & z \in [1, 2]. \end{cases}$$

Observe que la función de densidad de Z decrece con el precio del libro (ver Figura 2.6(b)); en particular, es más probable que Marcela pague por el libro un precio cercano a 1 que a 2. En contraste, como la función de densidad de X permanece constante entre estos dos valores, si Marcela comprara el libro sin cotizarlo, entonces sería igualmente probable que ella pagara tanto un precio alto como uno bajo. ¡La estrategia de cotizar el libro favorece pagar un precio más bien bajo que alto!

2.7 Ejercicios propuestos

Sección 2.1.

Ejercicio 1. Considere cierto experimento que tiene asociado el espacio muestral $\Omega = \{(i,j) : i,j \in \{-2,-1,0,1,2\}\}$. Considere la v.a. definida como $X(i,j) = i \cdot j$. Represente cada uno de los siguientes sucesos como un subconjunto de Ω:
(a) $[X=-3]$, (b) $[X=2]$, (c) $[X=0]$.

Ejercicio 2. Marcela organiza un taller semanal de química, de cinco alumnos, en el colegio donde trabaja. Denotaremos como M el número de alumnos que traen sus materiales en un taller dado. (a) ¿Cuál es la probabilidad del suceso $[M \in \{0,1,2,3,4,5\}]$? (b) Si cada alumno olvida sus materiales en el 2 % de los talleres e independientemente de sus compañeros, ¿cuál es la probabilidad del suceso $[M=5]$? (c) ¿Qué fracción de las veces al menos un alumno olvidará sus materiales? (d) En un año académico de 34 semanas, estime el número de talleres en los que al menos un alumno olvidará sus materiales.

Ejercicio 3. En una cierta comisaría se registran 15 llamadas por hora. Suponga que la v.a. T, definida como la separación temporal entre la última y próxima llamada, satisface $\mathbb{P}[T \leq t] = 1 - e^{-15t}$, para todo $t \geq 0$. (t se mide en horas.) (a) ¿Cuál es la probabilidad de que la siguiente llamada ocurra dentro de los próximos 12 minutos? (b) Determine $\mathbb{P}[T \leq 1/5 \mid T \leq 1/2]$ i.e. la probabilidad condicional de que la siguiente llamada ocurra dentro de los próximos 12 minutos, dado que ocurrirá en los próximos 30 minutos.

Sección 2.2.

Ejercicio 4. Considere una v.a. X con función de distribución

$$F_X(x) = \begin{cases} 0, & x < -1; \\ (3x+5)/6, & -1 \leq x < 0; \\ 5/6, & 0 \leq x < 2/3; \\ (x+1)/2, & 2/3 \leq x < 1; \\ 1, & x \geq 1. \end{cases}$$

Determine la probabilidad de cada uno de los sucesos $[X=-4/3]$, $[X=-1]$, $[X=-2/3]$, $[X=-1/3]$, $[X=0]$, $[X=1/3]$, $[X=2/3]$, $[X=1]$, $[X=4/3]$, $[X<-4/3]$, $[-2/3 < X \leq -1/3]$, $[-2/3 \leq X < -1/3]$, $[-2/3 \leq X \leq 1/3]$, $[-2/3 < X < 1/3]$, $[-1 < X \leq 2/3]$, $[-1 \leq X < 2/3]$, $[X > -1]$, $[X \geq -1]$, $[X \geq 4/3]$.

Ejercicio 5. Determine la función de distribución de la siguiente v.a., definida en términos del lanzamiento de un dado equilibrado:

$$X = \begin{cases} 1, & \text{si la cara sale impar;} \\ 2, & \text{si la cara sale par.} \end{cases}$$

Ejercicio 6. Considere una v.a. X tal que F_X crece linealmente sobre el intervalo $[1,4]$ i.e. existen constantes $a, b \in \mathbb{R}$, con $a \geq 0$, tales que $F_X(x) = ax + b$, cuando $1 \leq x \leq 4$. Si $\mathbb{P}[X \leq 2] = 1/3$ y $\mathbb{P}[2 < X \leq 3] = 1/6$, ¿cuál es la probabilidad del suceso $[1 < X < 4]$?, ¿qué hay del suceso $[X = 4]$? Explique por qué no hay información suficiente para determinar la probabilidad del suceso $[X = 1]$.

Sección 2.3.

Ejercicio 7. Determine cuáles de las siguientes variables aleatorias son discretas: $X_1 =$"el número de granos de arroz en una bolsa comprada en el supermercado", $X_2 =$"la duración de su próxima llamada de teléfono", $X_3 =$"el número de e-mails que usted recibirá en su casilla de correo electrónico el próximo martes", $X_4 =$"el promedio de errorres ortográficos por página en un libro escogido al azar", X_5="el número de veces que el patrón ACGATA ocurre en el DNA de una persona escogida al azar".

Ejercicio 8. La *función de masa de probabilidad* de una v.a. X es la función $p_X : \mathbb{R} \to [0,1]$ definida como $p_X(x) = \mathbb{P}[X = x]$, para cada $x \in \mathbb{R}$. Muestre que si X es una v.a. discreta y $A \subset \mathbb{R}$ un conjunto contable tal que $\mathbb{P}[X \in A] = 1$, entonces

$$F_X(x) = \sum_{a \in A:\, a \leq x} p_X(a).$$

Ejercicio 9. Determine la función de masa de probabilidad de la v.a. X que corresponde a la cara más baja entre tres lanzamientos independientes de un dado equilibrado (ver Ejercicio 8).

Sección 2.4.

Ejercicio 10. Muestre que si X es una v.a. tal que $F_X(x) = 2 \cdot \arctan(x)/\pi$, para todo $x \geq 0$, entonces X es una v.a. continua y determine su función de densidad.

Ejercicio 11. Considere el experimento donde se escoge un punto (X, Y) equidistribuidamente en el cuadrado con vértices $(1,3)$, $(2,3)$, $(2,5)$ y $(1,5)$. Determine la función de distribución de X y la función de distribución de Y.
Hint: Ver ejemplos 2.12 y 2.22.

Ejercicio 12. ¿La siguiente proposición es verdadera o falsa?

"Si $X \sim Uniforme[0,1]$ entonces $(1 - X) \sim Uniforme[0,1]$".

Ejercicio 13. Considere una v.a. $X \sim Uniforme[0,1]$ y defina $Y = 1/(1 + X)$. Muestre que Y es una v.a. continua y determine explícitamente su densidad. Use la función de densidad de Y para determinar directamente la probabilidad del suceso $4/7 < Y \leq 1$. Corrobore su respuesta, determinando la probabilidad de este último suceso a partir de la v.a. X.

Ejercicio 14. Considere una v.a. $X \sim Uniforme[-1, 2]$. Determine la función de distribución de $Y = X^2$ y utilice ésta para mostrar que Y es una v.a. continua, determinando su función de densidad. Verifique que la densidad obtenida es efectivamente una función de densidad.

Ejercicio 15. Aplique el Método del Jacobiano para determinar la función de densidad de la v.a. $Y = e^{-X}$, donde $X \sim Uniforme(0, 1)$. Compruebe su respuesta calculando directamente y luego diferenciando la función F_Y.

Ejercicio 16. Considere una v.a. X con función de densidad $f_X(x) = e^{-x}$ si $x > 0$, y la transformación definida como

$$g(x) = \begin{cases} \frac{3-x}{3}, & 0 < x < 3; \\ \frac{x-3}{x-2}, & x \geq 3. \end{cases}$$

(a) Aplique el Método general del Jacobiano para determinar directamente la función de densidad de la v.a. $Y = g(X)$. Verifique que f_Y es efectivamente una función de densidad. (b) Compruebe nuevamente su respuesta determinando directamente y luego diferenciando la función F_Y.

Ejercicio 17. Considere la siguiente v.a. definida a partir del lanzamiento de una moneda equilibrada: si la moneda sale cara entonces definimos $X = 0$ pero, si la moneda sale sello, entonces X es un número selecionado al azar y equidistribuido en el intervalo $[1, 3]$. ¿Es X una v.a. discreta?, ¿es X una v.a. continua? Justifique sus respuestas.

Sección 2.5.

Ejercicio 18. Muestre que si existe $t \in \mathbb{R}$ tal que $\mathbb{P}[X = t] \neq \mathbb{P}[Y = t]$ entonces la distribución de X es diferente de la de Y.

Ejercicio 19. Gloria y Loreto son estudiantes de educación básica de un mismo colegio. Considere el número de monedas X e Y de cien pesos que Gloria y Loreto tienen en sus delantales, respectivamente. (a) Si la situación económica de la familia de Gloria es similar a la de Loreto, ¿es razonable suponer que X e Y tienen la misma distribución? Explique su respuesta intuitivamente. (b) ¿Cómo respondería a la pregunta previa si la situación económica de Gloria es muy diferente de la de Loreto? Explique su respuesta intuitivamente.

Ejercicio 20. Considere el experimento donde se extrae una carta al azar de una baraja inglesa. Asociaremos a este experimento, las siguientes variables aleatorias:

$$\begin{aligned} X &= \begin{cases} 1, & \text{si la pinta de la carta extraída es un corazón o un trébol;} \\ 0, & \text{en cualquier otro caso;} \end{cases} \\ Y &= (1 - X). \end{aligned}$$

Muestre que X e Y tienen la misma distribución. ¿Es posible que X e Y tomen el mismo valor simultáneamente?, ¿contradice esto la igualdad en distribución de estas variables aleatorias?

Ejercicio 21. Usando lanzamientos de dados como experimento defina tres variables aleatorias diferentes que, sin embargo, tienen la misma distribución.

Ejercicio 22. Considere dos variables aleatorias: $X \sim Uniforme[-1,1]$ e $Y \sim Uniforme[-3,2]$. Determine constantes $a,b \in \mathbb{R}$ tales que Y tiene la misma distribución que $(aX+b)$.

Sección 2.6.

Ejercicio 23. Si X e Y son variables aleatorias independientes, ¿cuál es la probabilidad del suceso $[X \geq 3 \text{ ó } Y < 0]$, si $\mathbb{P}[X \geq 3] = 0{,}35$ y $\mathbb{P}[Y < 0] = 0{,}70$?, ¿qué hay del suceso $[X < 3 \text{ ó } Y < 0]$?
Hint: Use la Fórmula de Inclusión-Exclusión (Ejercicio 22 en el Capítulo 1).

Ejercicio 24. Considere un punto (X,Y) escogido al azar y equidistribuido en el triángulo con vértices $(0,0)$, $(1,0)$ y $(0,1)$. ¿Son X e Y variables aleatorias independientes? Justifique su respuesta con un argumento matemático.

Ejercicio 25. Muestre que si X e Y son variables aleatorias independientes, entonces para todo par de intervalos I,J se tienen las identidades

$$\begin{aligned}
\mathbb{P}[X \in I, Y \notin J] &= \mathbb{P}[X \in I] \cdot \mathbb{P}[Y \notin J];\\
\mathbb{P}[X \notin I, Y \in J] &= \mathbb{P}[X \notin I] \cdot \mathbb{P}[Y \in J];\\
\mathbb{P}[X \notin I, Y \notin J] &= \mathbb{P}[X \notin I] \cdot \mathbb{P}[Y \notin J].
\end{aligned}$$

Ejercicio 26. Muestre que si I, J son intervalos tales que $\mathbb{P}[X \in I] = 1$, $\mathbb{P}[Y \in J] = 1$ y, para todo $x \in I$ e $y \in J$, $\mathbb{P}[X \leq x, Y \leq y] = F_X(x) \cdot F_Y(y)$, entonces X e Y son variables aleatorias independientes.

Ejercicio 27. Considere el experimento donde se escoge un punto (X,Y) equidistribuidamente en el rectángulo con vértices $(0,0)$, $(2,0)$, $(2,1)$ y $(0,1)$. ¿Son X e Y variables aleatorias independientes?, ¿qué puede decir acerca de X y $(X+Y)$?

Problemas misceláneos Capítulo 2.

Ejercicio 28. (a) Muestre que si X es una v.a. continua y $A \subset \mathbb{R}$ un conjunto contable, entonces $\mathbb{P}[X \in A] = 0$. (Esto muestra que una v.a. continua no puede ser discreta.) (b) Muestre que si X es una v.a. discreta entonces X no puede ser una v.a. continua.

Ejercicio 29. Muestre que si $X \sim Uniforme[0,1]$ y definimos $Y = X$ entonces X e Y no son independientes y, sin embargo, para todo $x,y \in \mathbb{R}$ se tiene que $\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y]$. (Esto muestra que la condición dada en el Teorema 2.18 aplica específicamente para variables aleatorias discretas y no es suficiente para concluir que dos o más variables en general son independientes.)

Ejercicio 30. Muestre que si $U \sim Uniforme[0,1]$ entonces $Z = \max\{U, 1-U\}$ es uniforme en el intervalo $[1/2, 1]$.
Hint: $\max\{a,b\} \leq c$ si y solo si $a \leq c$ y $b \leq c$.

Ejercicio 31. Considere dos constantes $a, b \in \mathbb{R}$ tales que $a > 0$. Muestre que si X es una v.a. continua con función de densidad f entonces $Y = (aX + b)$ es también una v.a. continua con función de densidad:

$$f_Y(y) = \frac{1}{a} \cdot f\left(\frac{y-b}{a}\right).$$

Ejercicio 32. Muestre que si $X \sim Uniforme[a,b]$ y $c > 0$ y d son constantes dadas entonces $(cX+d) \sim Uniforme[ca+d, cb+d]$. ¿Qué conclusión se tiene cuando $c < 0$? (Esto demuestra que una transformación lineal afín de una v.a. uniforme, es también una v.a. uniforme aunque posiblemente en otro intervalo.)

Ejercicio 33. Muestre las identidades (2.9)-(2.13).

Ejercicio 34. Considere una v.a. $U \sim Uniforme[0,1]$ y defina

$$X = \begin{cases} 0, & U \leq 1/3; \\ U, & 1/3 < U \leq 2/3; \\ 1, & U > 2/3; \end{cases}$$

¿Es X una v.a. discreta? ¿Es X una v.a. continua? Demuestre sus respuestas.

Ejercicio 35. Considere dos variables aleatorias X e Y. Muestre que $F_X = F_Y$ si y solo si, para todo par de constantes reales $a \leq b$, $\mathbb{P}[X \in [a,b]] = \mathbb{P}[Y \in [a,b]]$. (Esto demuestra, que dos o más variables aleatorias tienen la misma función de distribución si y solo si tienen la misma probabilidad de tomar un valor sobre cada intervalo cerrado de largo finito.)

Ejercicio 36. Imagine que usted y un cierto amigo lanzan un par de dados equilibrados simultáneamente. Denotaremos como X el valor de la cara que usted lanzó, y como Y la de su amigo. (a) Explique en palabras lo que representa el suceso $[X \leq Y]$, ¿cuál es la probabilidad de este suceso? (b) Determine la probabilidad del suceso $[2 \leq (X+Y) \leq 10]$. (c) Para cada $k \in \{2, \ldots, 12\}$ muestre que

$$\mathbb{P}[X+Y=k] = \frac{1 + \min\{6, k-1\} - \max\{1, k-6\}}{36}.$$

Verifique que la suma de estas probabilidades es 1.

Ejercicio 37. Considere el experimento que consiste en escoger un punto (X,Y) equidistribuidamente en el cuadrado con vértices $(1,0)$, $(0,1)$, $(-1,0)$, y $(0,-1)$. Muestre que X e Y tienen la misma distribución. ¿Es X una v.a. uniforme en el intervalo $[-1,1]$?, ¿qué puede decir a este respecto acerca de la v.a. Y?

Ejercicio 38. Considere dos sucesos diferentes A y B que tienen la misma posibilidad de ocurrir en un cierto experimento. (Por ejemplo, si el experimento consiste en lanzar un dado equilibrado entonces podríamos considerar A=“la cara lanzada es un número primo" y B="la cara lanzada es mayor o igual a tres".) Considere las variables aleatorias

$$X = \begin{cases} 1, & \text{si } A \text{ ocurre} \\ 0, & \text{si } A \text{ no ocurre} \end{cases} ; \qquad Y = \begin{cases} 1, & \text{si } B \text{ ocurre} \\ 0, & \text{si } B \text{ no ocurre} \end{cases} .$$

¿Es necesariamente cierto que $X = Y$?, ¿tienen X e Y la misma distribución? Demuestre sus respuestas o entregue contraejemplos.

Ejercicio 39. Muestre la identidad en (2.17).

Ejercicio 40. Considere variables aleatorias $X : \Omega \to \mathbb{R}$ e $Y : \Omega \to \mathbb{R}$ asociadas a un mismo experimento. (a) Muestre que si $I_1, I_2 \subset \mathbb{R}$ son intervalos disjuntos entonces los sucesos $[X \in I_1]$ y $[X \in I_2]$ son también disjuntos. (b) Muestre que si $I_3 \subset \mathbb{R}$ es un intervalo entonces $[X \in I_1, Y \in I_3]$ y $[X \in I_2, Y \in I_3]$ son sucesos disjuntos. (c) Muestre usando un contraejemplo que los sucesos $[X \in I_1]$ e $[Y \in I_2]$ no son necesariamente disjuntos.

Ejercicio 41. Muestre que si $X \sim Uniforme[0, 1]$ entonces

$$2 \cdot \text{mín}\{X, 1 - X\} \sim Uniforme[0, 1].$$

Ejercicio 42. Considere dos variables aleatorias X e Y. Demuestre que si $\mathbb{P}[X = Y] = 1$ entonces X e Y tienen la misma distribución. ¿Es la implicación opuesta, en general, cierta? Justifique su respuesta con una demostración o un contraejemplo.

Ejercicio 43. Aplique el Método general del Jacobiano para determinar la densidad de la v.a. $Y = \text{sen}(\pi X)$, con $X \sim Uniforme(-1, 1)$. Verifique que f_Y es efectivamente una función de densidad.

Ejercicio 44. Si $U \sim Uniforme[0, 1]$ entonces $\mathbb{P}[U \in [0, 1]] = 1$ y, para todo $x \in [0, 1]$, $\mathbb{P}[U \in \{x\}] = 0$ (*¿por qué?*). Explique por qué estos dos hechos no se contradicen entre sí pese a que $[0, 1] = \cup_{0 \le x \le 1}\{x\}$.

Ejercicio 45. Considere constantes reales $a < c < d < b$. Muestre que si $X \sim Uniforme[a, b]$ entonces $X \sim Uniforme[c, d]$ si condicionamos sobre el suceso $X \in [c, d]$ i.e. muestre que $\mathbb{P}[X \in I \mid X \in [c, d]] = (\text{largo de } I)/(d - c)$, para todo intervalo $I \subset [c, d]$.

Ejercicio 46. Considere una v.a. T tal que $\mathbb{P}[T \ge 0] = 1$ y $\mathbb{P}[T = t] = 0$, para todo $t \in \mathbb{R}$. (En el Ejemplo 2.4 describimos una v.a. con estas características). (a) Muestre que $\mathbb{P}[T = t] = \mathbb{P}[(-T) = t]$, para todo $t \in \mathbb{R}$. (b) Muestre que $\mathbb{P}[T \in [0, +\infty)] \neq \mathbb{P}[(-T) \in [0, +\infty)]$; en particular, T y $(-T)$ no tienen la misma distribución. (c) Explique por qué las partes (a) y (b) no se contradicen entre sí.

Ejercicio 47. Considere una moneda con probabilidad p de salir cara y variables aleatorias $X \sim Uniforme[0, 1]$ e $Y \sim Uniforme[1, 4]$, independientes de la moneda.

Muestre que la v.a. Z definida como sigue es una v.a. continua y determine su función de densidad:

$$Z = \begin{cases} X, & \text{si la moneda sale cara;} \\ Y, & \text{si la moneda sale sello.} \end{cases}$$

¿Para qué valores de p es la v.a. Z uniforme sobre el intervalo $[0, 4]$?

Ejercicio 48. Considere el experimento donde se escoge un punto (X, Y) equidistribuido en el círculo unitario. Usaremos (R, Θ) para denotar las coordenadas polares de (X, Y) i.e. $R = \sqrt{X^2 + Y^2}$ y $\Theta = \arctan(Y/X)$ (ver Figura 2.7). Observe que $0 \leq R \leq 1$, $-\pi < \Theta \leq \pi$, $X = R \cdot \cos(\Theta)$ y $Y = R \cdot \operatorname{sen}(\Theta)$. (a) Muestre que $F_R(r) = r^2$, para todo $0 \leq r \leq 1$. (b) Muestre que $F_\Theta(\theta) = (\theta + \pi)/(2\pi)$, para todo $-\pi < \theta < \pi$. (c) Use el Teorema 2.16 para demostrar que R y Θ son variables aleatorias independientes.

Ejercicio 49. Considere un punto (X, Y) escogido al azar y equidistribuido en el sector anular $\{(x, y) : 1 \leq x^2 + y^2 \leq 4\}$. Explique primero intuitivamente y luego demuestre que si (R, Θ) son las coordenadas polares de (X, Y) entonces R y Θ son variables aleatorias independientes.

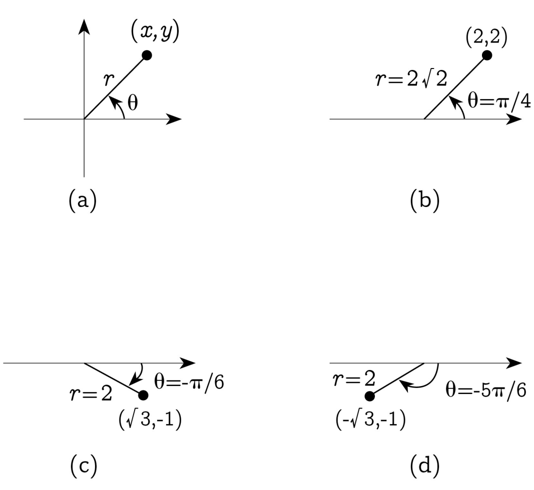


Figura 2.7. (a) Coordenadas polares (r, θ) de un punto genérico (x, y) en el plano. (b) Coordenadas polares del punto $(2, 2)$. Note cómo el ángulo θ se mide en radianes. (c) Coordenadas polares del punto $(\sqrt{3}, -1)$. (d) Coordenadas polares del punto $(-\sqrt{3}, -1)$.

Ejercicio 50. Use que los sucesos $[X \leq n]$, con $n \geq 1$, crecen con n al suceso [existe $n \geq 1$ tal que $X \leq n$] para demostrar la parte (c) en el Teorema 2.2. Dado

$x \in \mathbb{R}$, use que los sucesos $[X \leq x + 1/n]$, con $n \geq 1$, decrecen con n al suceso $[X \leq x]$ para demostrar la parte (e) en el Teorema 2.2.

Ejercicio 51. Muestre que si X es una v.a. discreta y $g : \mathbb{R} \to \mathbb{R}$ es una transformación dada entonces $g(X)$ es también una v.a. discreta. Más generalmente, muestre que si $X_1, \ldots, X_n$ son variables aleatorias discretas y $g : \mathbb{R}^n \to \mathbb{R}$ es una transformación dada entonces $g(X_1, \ldots, X_n)$ es también una v.a. discreta. En particular, $\sum_{i=1}^{n} X_i^3$, $\prod_{i=1}^{n} X_i$, y $(1 + \sum_{i=1}^{n-1} \cos^2(X_i X_{i+1}))^{-1}$ son ejemplos de variables aleatorias discretas cuando $X_1, \ldots, X_n$ lo son.

Ejercicio 52. Si X_1, X_2, Y_1, Y_2 son variables aleatorias discretas tales que X_1 y X_2 tienen la misma distribución, e Y_1 e Y_2 también tienen la misma distribución, responda: ¿es necesariamente cierto que $g(X_1, Y_1)$ tiene la misma distribución que $g(X_2, Y_2)$ con $g : \mathbb{R}^2 \to \mathbb{R}$ una función dada? Demuestre su aseveración o entregue un contraejemplo.

Ejercicio 53. Sea X una v.a. continua e $I \subset \mathbb{R}$ un intervalo abierto tal que $\mathbb{P}[X \in I] > 0$. Defina Y como la v.a. que tiene la misma distribución que X pero condicionada a caer en el intervalo I i.e. para cada intervalo $J \subset I$ aplica que $\mathbb{P}[Y \in J] = \mathbb{P}[X \in J \mid X \in I]$. Demuestre que Y es también una v.a. continua pero con función de densidad dada por la fórmula

$$f_Y(y) = \begin{cases} 0, & y \notin I; \\ \frac{f_X(y)}{\mathbb{P}[X \in I]}, & y \in I. \end{cases}$$

Aplique lo anterior para determinar f_Y en cada uno de los siguientes casos:

(a) $I = (1/3, 2/3)$ y $X = U$, con $U \sim Uniforme[0, 1]$,
(b) $I = (5/7, 6/7)$ y $X = U^2$, con $U \sim Uniforme[0, 1]$.

Ejercicio 54. Considere una v.a. continua X y dos intervalos abiertos y disjuntos $I_1, I_2 \subset \mathbb{R}$ tales que $\mathbb{P}[X \in I_i] > 0$ y $\mathbb{P}[X \in (I_1 \cup I_2)] = 1$. Sea $Y = g(X)$ donde $g : (I_1 \cup I_2) \to \mathbb{R}$ es una función continuamente diferenciable tal que $g'(x) \neq 0$, para todo $x \in (I_1 \cup I_2)$. Para cada i, sea X_i una v.a. que tiene la misma distribución que X pero condicionada a tomar un valor en el intervalo I_i. (a) Use el Ejercicio 53 y el Método del Jacobiano para determinar la función de densidad de $g(X_i)$. (b) Use que para cada intervalo cerrado $I \subset \mathbb{R}$, $\mathbb{P}[Y \in I] = \sum_{i=1}^{2} \mathbb{P}[g(X_i) \in I \mid X \in I_i] \cdot \mathbb{P}[X \in I_i]$ para determinar la función de densidad de Y. (Esto demuestra el Teorema 2.9 para el caso especial de $n = 2$.)

Comentarios de final de capítulo

La *sigma-álgebra de Borel de* $\mathbb{R}$, que denotaremos como $\mathcal{B}(\mathbb{R})$, es la sigma-álgebra más pequeña que contiene todos los conjuntos abiertos de $\mathbb{R}$. Los conjuntos de $\mathcal{B}(\mathbb{R})$ son llamados *Borelianos*. Algunos ejemplos de conjuntos Borelianos son: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{I}$, $\mathbb{R}$, $[0, 1]$, $[0, +\infty)$, y el conjunto de Cantor, entre muchos otros. *Enfatizamos, sin embargo, que no todo subconjunto de* $\mathbb{R}$ *es Boreliano i.e. la inclusión* $\mathcal{B}(\mathbb{R}) \subset \mathcal{P}(\mathbb{R})$ *es estricta.* En lo que sigue, $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio probabilístico dado.

En el formalismo de la teoría de la medida, una v.a. es una función $X : \Omega \to \mathbb{R}$ que es *medible*. Ésto significa que $X^{-1}(A) \in \mathcal{F}$, para todo $A \in \mathcal{B}(\mathbb{R})$. Por ejemplo, una función $g : \mathbb{R} \to \mathbb{R}$ es medible si y solo si $g^{-1}(A) \in \mathcal{B}(\mathbb{R})$, para todo $A \in \mathcal{B}(\mathbb{R})$.

Las nociones de igualdad en distribución e independencia para variables aleatorias son análogas a la que dimos, aunque los sucesos bajo consideración pertenecen a $\mathcal{F}$. Específicamente, dos variables aleatorias X e Y se dicen *igualmente distribuidas* cuando $\mathbb{P}[X \in A] = \mathbb{P}[Y \in A]$, para todo $A \in \mathcal{B}(\mathbb{R})$. Por otro lado, X e Y se dicen *independientes* si $\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A] \cdot \mathbb{P}[Y \in B]$, para todo $A, B \in \mathcal{B}(\mathbb{R})$. En este marco teórico, se tiene lo siguiente:

(a) *Si X tiene la misma distribución que Y entonces, para toda función medible $g : \mathbb{R} \to \mathbb{R}$, $g(X)$ también tiene la misma distribución que $g(Y)$, y*

(b) *Si X e Y son independientes entonces, para todo par $g : \mathbb{R} \to \mathbb{R}$ y $h : \mathbb{R} \to \mathbb{R}$ de funciones medibles, $g(X)$ y $h(Y)$ son también independientes.*

La *distribución* de una v.a. X es la medida de probabilidad $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$ definida como $\mathbb{P}_X(A) = \mathbb{P}[X \in A]$, para todo $A \in \mathcal{B}(\mathbb{R})$. Note que X e Y tienen la misma distribución si y solo si $\mathbb{P}_X = \mathbb{P}_Y$. Una v.a. se dice *discreta* si existe un conjunto contable $A \subset \mathbb{R}$ tal que $\mathbb{P}_X(A) = 1$. Por otro lado, X se dice *continua* si existe una función no negativa y medible $f : \mathbb{R} \to \mathbb{R}$, tal que $\mathbb{P}_X(A) = \int_A f(x)\,dx$, para todo $A \in \mathcal{B}(\mathbb{R})$.

Alternativamente, X e Y son independientes si y solo si

$$\mathbb{P}[(X, Y) \in C] = (\mathbb{P}_X \otimes \mathbb{P}_Y)(C), \text{ para todo } C \in \mathcal{B}(\mathbb{R}^2).$$

En la identidad anterior, $\mathcal{B}(\mathbb{R}^2)$ es la *sigma-álgebra Boreliana de* $\mathbb{R}^2$ i.e. la sigma-álgebra más pequeña que contiene todos los subconjuntos abiertos de $\mathbb{R}^2$. Por otro lado, $(\mathbb{P}_X \otimes \mathbb{P}_Y)$ es la *medida producto de* $\mathbb{P}_X$ *con* $\mathbb{P}_Y$. Dadas dos medidas de probabilidad μ y ν definidas sobre $\mathcal{B}(\mathbb{R})$, $(\mu \otimes \nu)$ es la única medida de probabilidad sobre $\mathcal{B}(\mathbb{R}^2)$, que satisface la identidad: $(\mu \otimes \nu)(A \times B) = \mu(A) \cdot \nu(B)$, para todo $A, B \in \mathcal{B}(\mathbb{R})$.

Capítulo 3: La Ley de los Grandes Números

En este capítulo definiremos la esperanza de una variable aleatoria. La definición que daremos, será motivada a partir de la interpretación frecuentista de probabilidades i.e. que la probabilidad de un suceso representa su fracción asintótica de ocurrencia. Usando este principio interpretaremos la esperanza de una variable aleatoria, como el valor asintótico del promedio muestral de realizaciones independientes de dicha variable. El resultado matemático preciso a este respecto es lo que se conoce como la Ley de los Grandes Números. En la última sección del capítulo introduciremos el concepto de varianza de una variable aleatoria. Esta cantidad medirá cuán dispersa es la variable alrededor de su esperanza. Finalmente, usando la Ley de los Grandes Números, veremos que la varianza de una variable aleatoria puede interpretarse como el valor asintótico de la varianza muestral asociada a realizaciones independientes de dicha variable.

3.1 Esperanza de una variable aleatoria

La definición de medida de probabilidad se basa en la interpretación frecuentista de probabilidades: la probabilidad de un suceso en un cierto experimento, representa su frecuencia de ocurrencia asintótica al repetir el experimento indefinidamente y bajo circunstancias similares. ¡Esta interpretación popularmente aceptada conlleva, sin embargo, otro hecho insospechado!

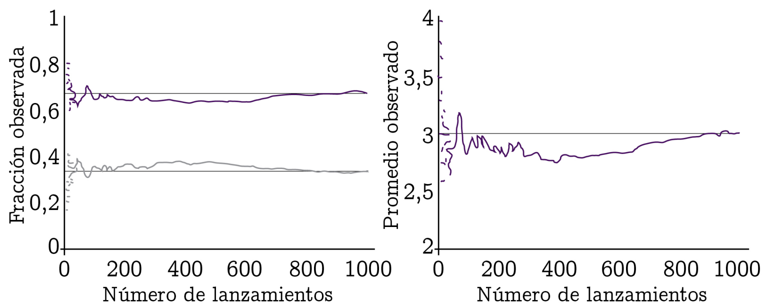
Para fijar ideas, imagine un dado equilibrado en el cual dos caras tienen marcado el número (-1) y las cuatro restantes el número 5. Imagine que usted lanza el dado n veces y que los valores de las caras observadas son $X_1, \ldots, X_n$. El promedio de las caras observadas entonces está dado por

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{(-1)\cdot Y + 5\cdot Z}{n} = (-1)\cdot\frac{Y}{n} + 5\cdot\frac{Z}{n},$$

donde Y denota el número de veces que un (-1) es observado mientras que Z denota el número de veces que un 5 es observado. Observe que Y/n es la fracción de veces que el valor (-1) es observado en los primeros n-lanzamientos. Por lo tanto, cuando n es grande, la interpretación frecuentista implica que $Y/n \approx 1/3$. Similarmente, $Z/n \approx 2/3$. Así concluimos que

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \approx (-1)\cdot\frac{1}{3} + 5\cdot\frac{2}{3} = 3. \tag{3.1}$$

La aproximación anterior es realmente notable: *si n es grande entonces $\bar{X} \approx 3$ pese a que $\bar{X}$ es una cantidad que no conoceremos con exactitud hasta lanzar el dado n veces.*



(a) (b)

FIGURA 3.1. (a) Curvas asociadas a las fraciones de veces, que las caras (-1) y 5 son observadas en mil simulaciones del lanzamiento de un dado equilibrado, en el cual dos caras tienen marcado el número (-1) y las cuatro restantes el número 5. Observe cómo la fracción de veces que la cara (-1) es observada en función del número de lanzamientos (curva inferior) es asintótica al valor $1/3$ a medida que más lanzamientos son tomados en cuenta. Similarmente, la fracción de veces que la cara 5 es observada (curva superior) es asintótica al valor $2/3$. (b) Promedio de las caras observadas en función del número de lanzamientos. Observe cómo la curva es asintótica al valor 3.

Claramente el valor al lado derecho en la aproximación en (3.1) depende de los números marcados sobre las caras del dado. Cabe entonces preguntarse: *¿qué valor debiera $\bar{X}$ tener aproximadamente cuando n es grande pero $X_1, \ldots, X_n$ pueden tomar valores arbitrarios?*

Antes de responder la pregunta es conveniente introducir la siguiente notación. Dada una proposición p, usaremos la notación $[\![\, p \,]\!]$ para denotar la *función indicadora de p* i.e. la cantidad definida como sigue:

$$[\![\, p \,]\!] = \begin{cases} 1, & \text{si la proposición } p \text{ es verdadera;} \\ 0, & \text{si la proposición } p \text{ es falsa.} \end{cases} \tag{3.2}$$

(En la literatura, $[\![\, p \,]\!]$ se llama el *paréntesis de Iverson* asociado con p.) Por ejemplo, $[\![5 \geq 4]\!] = 1$ y $[\![5 \leq 3]\!] = 0$. Por otro lado, para $x \in \mathbb{R}$, el valor de $[\![x \geq 2]\!]$ depende de x: si $x \geq 2$ entonces $[\![x \geq 2]\!] = 1$, pero si $x < 2$ entonces $[\![x \geq 2]\!] = 0$.

Si A es un suceso asociado con un cierto experimento, con espacio muestral Ω, entonces $[\![A \text{ ocurre}]\!]$ define lo que comúnmente se llama la *función indicadora de A*:

si ω denota el resultado del experimento entonces

$$[\![A \text{ ocurre}]\!] = \begin{cases} 1, & \text{si } \omega \in A; \\ 0, & \text{si } \omega \notin A. \end{cases}$$

(La notación 1_A ó $1_A(\omega)$ es, a veces, utilizada en la literatura para denotar la función indicadora de un suceso.) Por ejemplo, si $X : \Omega \to \mathbb{R}$ es una v.a. y $I \subset \mathbb{R}$ un intervalo, entonces

$$[\![X \in I]\!] = \begin{cases} 1, & \text{si } X(\omega) \in I; \\ 0, & \text{si } X(\omega) \notin I. \end{cases}$$

El siguiente resultado será clave para responder a la pregunta anterior con amplia generalidad.

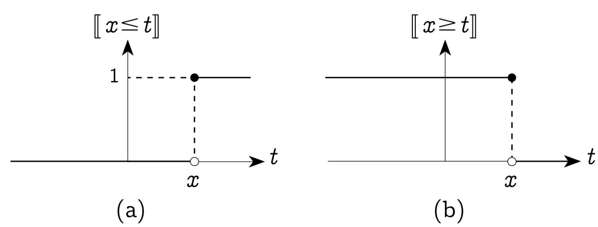


FIGURA 3.2. Grafos asociados a las funciones $t \longrightarrow [\![x \leq t]\!]$, y $t \longrightarrow [\![x \geq t]\!]$, cuando $x \geq 0$.

Lema 3.1. *Para todo $x \in \mathbb{R}$, $x = \int_0^\infty [\![x \geq t]\!]\, dt - \int_{-\infty}^0 [\![x \leq t]\!]\, dt$.*

Demostración. Para un $x \geq 0$ fijo, la función $t \to [\![x \leq t]\!]$ es idénticamente igual a cero para todo $t < 0$ (ver Figura 3.2(a)). En particular, la segunda integral de arriba se anula. Por otro lado, la función $t \to [\![x \geq t]\!]$ es idéticamente igual a cero cuando $t > x$ (ver Figura 3.2(b)). Por lo tanto, se tiene que

$$\int_0^\infty [\![x \geq t]\!]\, dt - \int_{-\infty}^0 [\![x \leq t]\!]\, dt = \int_0^x [\![x \geq t]\!]\, dt = \int_0^x 1\, dt = x.$$

Por otro lado, si $x < 0$ entonces

$$\begin{aligned} \int_0^\infty [\![x \geq t]\!]\, dt &= \int_0^\infty 0\, dt = 0; \\ \int_{-\infty}^0 [\![x \leq t]\!]\, dt &= \int_x^0 [\![x \leq t]\!]\, dt = \int_x^0 1\, dt = -x; \end{aligned}$$

y, por lo tanto, tenemos que

$$\int_0^\infty [\![x \geq t]\!]\, dt - \int_{-\infty}^0 [\![x \leq t]\!]\, dt = 0 - (-x) = x.$$

Esto completa la demostración del lema. $\square$

Considere ahora variables aleatorias $X_1, \dots, X_n$ independientes y con la misma distribución que una cierta v.a. X. Debido al Lemma 3.1, para cada i tenemos que

$$X_i = \int_0^\infty [\![X_i \geq t]\!]\, dt - \int_{-\infty}^0 [\![X_i \leq t]\!]\, dt.$$

En particular, la siguiente identidad aplica

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n X_i &= \frac{1}{n}\sum_{i=1}^n \int_0^\infty [\![X_i \geq t]\!]\, dt - \int_{-\infty}^0 [\![X_i \leq t]\!]\, dt, \\ &= \int_0^\infty \frac{1}{n}\sum_{i=1}^n [\![X_i \geq t]\!]\, dt - \int_{-\infty}^0 \frac{1}{n}\sum_{i=1}^n [\![X_i \leq t]\!]\, dt, \end{aligned}$$

donde para la última identidad hemos intercambiado la sumatoria con la integral. Note, sin embargo, que para un número t dado, y cuando n es grande, la interpretación frecuentista de probabilidades implica que

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n [\![X_i \geq t]\!] &= \begin{array}{l}\text{fracción de datos}\\ \text{mayor o igual a } t\end{array} \approx \mathbb{P}[X \geq t], \\ \frac{1}{n}\sum_{i=1}^n [\![X_i \leq t]\!] &= \begin{array}{l}\text{fracción de datos}\\ \text{menor o igual a } t\end{array} \approx \mathbb{P}[X \leq t]. \end{aligned}$$

En particular, para un valor grande de n, es plausible esperar lo siguiente:

$$\frac{1}{n}\sum_{i=1}^n X_i \approx \int_0^\infty \mathbb{P}[X \geq t]\, dt - \int_{-\infty}^0 \mathbb{P}[X \leq t]\, dt. \tag{3.3}$$

Esta aproximación es nuevamente sorprendente: *el promedio de un número grande de variables aleatorias independientes y con la misma distribución puede adivinarse sin la necesidad de observar cada una de las variables que participan en dicho promedio.* El lado derecho en (3.3) es una cantidad determinista y únicamente definida por las probabilidades de los sucesos $[X \geq t]$, con $t \geq 0$, y $[X \leq t]$, con $t \leq 0$. Más aún, el lado derecho de la aproximación anterior es de alguna manera el valor "esperado" para el promedio de $X_1, \dots, X_n$. Esto motiva la siguiente definición.

Definición 3.2. La *esperanza* o *valor esperado* de una v.a. X es la cantidad definida como

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}[X \geq t]\, dt - \int_{-\infty}^0 \mathbb{P}[X \leq t]\, dt,$$

cuando al menos una de las integrales anteriores es finita. En este caso diremos que *X tiene una esperanza bien definida*. Si ambas integrales son finitas diremos que *X tiene esperanza finita*.

La definición anterior no es la estándar, sin embargo, es equivalente a la usualmente entregada en los textos clásicos de probabilidades. En las secciones 3.2 y 3.3 veremos maneras alternativas de determinar la esperanza de variables aleatorias discretas y continuas, recuperandando así las definiciones estándares. ¡La definición anterior aplica, sin embargo, para todo tipo de variable aleatoria! Para fijar ideas consideraremos algunos ejemplos.

Ejemplo 3.1. Retomemos el experimento donde se lanza un dado equilibrado en el cual dos caras tienen marcado el número (-1) y las cuatro restantes el número 5. Si X denota el valor de la cara lanzada entonces $\mathbb{P}[X \geq t] = 0$ cuando $t > 5$ y $\mathbb{P}[X \leq t] = 0$ cuando $t < (-1)$. Obtenemos, por lo tanto, que

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}[X \geq t]\,dt - \int_{-\infty}^0 \mathbb{P}[X \leq t]\,dt = \int_0^5 \mathbb{P}[X \geq t]\,dt - \int_{-1}^0 \mathbb{P}[X \leq t]\,dt.$$

Ahora, observe que si $0 \leq t \leq 5$ entonces $\mathbb{P}[X \geq t] = \mathbb{P}[X = 5] = 2/3$ ya que X sólo puede tomar los valores (-1) o 5. Similarmente, si $-1 \leq t \leq 0$ entonces $\mathbb{P}[X \leq t] = \mathbb{P}[X = -1] = 1/3$. Concluimos de este modo que

$$\mathbb{E}(X) = \int_0^5 \frac{2}{3}\,dt - \int_{-1}^0 \frac{1}{3}\,dt = 3.$$

La v.a. X tiene, por lo tanto, un valor esperado finito. Note cómo la aproximación dada en (3.1) es consistente con la obtenida más generalmente en (3.3).

Es importante enfatizar que *la esperanza de una variable aleatoria asociada con un cierto experimento es una cantidad no aleatoria i.e. una cantidad determinística que no depende del valor que la v.a. tomó o tomará al realizar el experimento.* En el ejemplo anterior, $\mathbb{E}(X) = 3$ tanto antes como después de lanzar el dado. Más aún, observe que el suceso $[X = 3]$ tiene probabilidad cero.

Ejemplo 3.2. En el Ejemplo 2.12 vimos que la distancia R, entre el centro de un blanco redondo de radio 1 unidad y el punto de impacto de un dardo lanzado al azar, es una v.a. continua con función de distribución $F_R(r) = r^2$, cuando $0 \leq r \leq 1$.

¿Cuál es el valor esperado de R? Para calcular $\mathbb{E}(R)$ necesitamos determinar $\int_0^\infty \mathbb{P}[R \geq t]\,dt$ y $\int_{-\infty}^0 \mathbb{P}[R \leq t]\,dt$. Como R sólo puede tomar valores entre cero y uno, $\mathbb{P}[R \leq t] = 0$ cuando $t < 0$; en particular, la segunda integral es nula. Por otro lado, para la primera integral, observe que $\mathbb{P}[R \geq t] = 0$ cuando $t > 1$ y, como R es una variable aleatoria continua, $\mathbb{P}[R \geq t] = 1 - F_R(t) = 1 - t^2$ cuando $0 \leq t \leq 1$. Concluimos, por lo tanto, que

$$\begin{aligned}
\mathbb{E}(R) &= \int_0^\infty \mathbb{P}[R \geq t]\,dt - \int_{-\infty}^0 \mathbb{P}[R \leq t]\,dt, \\
&= \int_0^1 (1 - t^2)\,dt - 0, \\
&= \left(t - \frac{t^3}{3}\right)\Big|_0^1 = \frac{2}{3}.
\end{aligned}$$

Ejemplo 3.3. *Determine* $\mathbb{E}(S)$*, si* $\mathbb{P}[S \le t] = -1/t$*, para cada* $t \le -1$*.*
Observe que $\mathbb{P}[S \le -1] = 1$. Por lo tanto, $\mathbb{P}[S > -1] = 0$. En particular, $\mathbb{P}[S \ge t] = 0$ cuando $t \ge 0$, y $\mathbb{P}[S \le t] = 1$ cuando $-1 \le t \le 0$ (*¿por qué?*). Debido a estas consideraciones tenemos que

$$\begin{aligned}
\mathbb{E}(S) &= \int_0^\infty \mathbb{P}[S \ge t]\,dt - \int_{-\infty}^0 \mathbb{P}[S \le t]\,dt,\\
&= \int_0^\infty 0\,dt - \int_{-\infty}^{-1} \mathbb{P}[S \le t]\,dt - \int_{-1}^0 \mathbb{P}[S \le t]\,dt,\\
&= 0 + \int_{-\infty}^{-1} \frac{1}{t}\,dt - \int_{-1}^0 1\,dt,\\
&= \ln|t|\Big|_{t=-\infty}^{t=-1} - 1 = -\infty.
\end{aligned}$$

La v.a. S tiene, por lo tanto, una esperanza bien definida aunque su esperanza es $(-\infty)$.

Ejemplo 3.4. Considere una v.a. $X \sim Uniforme[-1,2]$. Observe que si $t \ge 2$ entonces $\mathbb{P}[X \ge t] = 0$. Por otro lado, si $(-1) \le t \le 2$ entonces $\mathbb{P}[X \ge t] = (2-t)/3$. Por lo tanto,

$$\int_0^\infty \mathbb{P}[X \ge t]\,dt = \int_0^2 \mathbb{P}[X \ge t]\,dt = \int_0^2 \frac{2-t}{3}\,dt = \frac{2}{3}.$$

Similarmente, si $t \le (-1)$ entonces $\mathbb{P}[X \le t] = 0$, y si $(-1) \le t \le 2$ entonces $\mathbb{P}[X \le t] = (t+1)/3$. Por lo tanto,

$$\int_{-\infty}^0 \mathbb{P}[X \le t]\,dt = \int_{-1}^0 \mathbb{P}[X \le t]\,dt = \int_{-1}^0 \frac{t+1}{3}\,dt = \frac{1}{6}.$$

Concluimos que $\mathbb{E}(X) = 2/3 - 1/6 = 1/2$ i.e. el valor esperado de X corresponde al punto medio del intervalo $[-1,2]$ en este caso.

3.2 Esperanza de variables aleatorias discretas

En esta sección especializaremos la definición de valor esperado, para facilitar su cálculo en el contexto de variables aleatorias discretas. El resultado principal es el siguiente.

Teorema 3.3. *Considere una v.a. discreta* X *y un conjunto contable* $A \subset \mathbb{R}$ *tal que* $\mathbb{P}[X \in A] = 1$*. Si* $g : \mathbb{R} \to \mathbb{R}$ *es una función tal que* $\sum_{a \in A} |g(a)| \cdot \mathbb{P}[X = a] < +\infty$ *entonces* $g(X)$ *tiene un valor esperado finito dado por la fórmula*

$$\mathbb{E}(g(X)) = \sum_{a \in A} g(a) \cdot \mathbb{P}[X = a].$$

Un caso especial, pero importante del teorema, es cuando $g(x) = x$, para todo $x \in \mathbb{R}$. En este caso, si $\sum_{a \in A} |a| \cdot \mathbb{P}[X = a] < +\infty$ entonces

$$\mathbb{E}(X) = \sum_{a \in A} a \cdot \mathbb{P}[X = a]. \tag{3.4}$$

Antes de presentar la demostración del Teorema 3.3 es instructivo considerar algunos ejemplos. La demostración del teorema será entregada al final de esta sección.

Ejemplo 3.5. Considere el experimento donde se lanza un dado equilibrado. Si X denota el número de la cara lanzada entonces X es una v.a. discreta y, por lo tanto,

$$\mathbb{E}(X) = \sum_{k=1}^{6} k \cdot \mathbb{P}[X = k] = \sum_{k=1}^{6} \frac{k}{6} = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}.$$

Por otro lado, también tenemos que

$$\mathbb{E}(X^2) = \sum_{k=1}^{6} k^2 \cdot \mathbb{P}[X = k] = \sum_{k=1}^{6} \frac{k^2}{6} = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}.$$

Finalmente observe que

$$\mathbb{E}(\text{mín}\{X, 6-X\}) = \sum_{k=1}^{6} \text{mín}\{k, 6-k\} \cdot \mathbb{P}[X = k] = \frac{1+2+3+2+1+0}{6} = \frac{3}{2}.$$

Ejemplo 3.6. Considere el experimento donde se lanza una moneda sesgada con probabilidad $2/3$ de salir cara en cada lanzamiento, y el número de lanzamientos X necesarios hasta observar por primera vez una cara. Claramente, X solamente puede tomar valores en el conjunto $\{1, 2, \ldots\}$. De acuerdo al Teorema 3.3, tenemos que

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \mathbb{P}[X = k].$$

Observe que el suceso $[X = k]$ corresponde a una secuencia de $(k-1)$ sellos que es inmediatamente seguida por una cara. Como los lanzamientos de la moneda son independientes, la probabilidad de este suceso es $(1/3)^{k-1} 2/3$. La esperanza de X viene, por lo tanto, dada por la serie

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{3}\right)^{k-1} \left(\frac{2}{3}\right) = \frac{2}{3} \cdot \sum_{k=1}^{\infty} k \left(\frac{1}{3}\right)^{k-1}.$$

Usando que

$$\sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}, \text{ para todo } |x| < 1, \tag{3.5}$$

finalmente obtenemos que

$$\mathbb{E}(X) = \frac{2}{3} \cdot \sum_{k=1}^{\infty} k \left(\frac{1}{3}\right)^{k-1} = \frac{2}{3} \cdot \frac{1}{(1-1/3)^2} = \frac{3}{2}.$$

Demostración Teorema 3.3. Primero demostraremos la identidad en (3.4). En efecto, considere el conjunto definido como $A_+ = A \cap [0, +\infty)$. Dado $t \geq 0$ y como $\mathbb{P}[X \in A] = 1$ obtenemos que

$$\mathbb{P}[X \geq t] = \mathbb{P}[X \geq t, X \in A] = \sum_{a \in A : a \geq t} \mathbb{P}[X = a] = \sum_{a \in A_+} \mathbb{P}[X = a] \cdot [\![a \geq t]\!],$$

donde $[\![a \geq t]\!] = 1$ si $a \geq t$, y $[\![a \geq t]\!] = 0$ si $a < t$. En particular,

$$\begin{aligned} \int_0^{+\infty} \mathbb{P}[X \geq t]\, dt &= \int_0^{+\infty} \sum_{a \in A_+} \mathbb{P}[X = a] \cdot [\![a \geq t]\!]\, dt, \\ &= \sum_{a \in A_+} \mathbb{P}[X = a] \cdot \int_0^{+\infty} [\![a \geq t]\!]\, dt. \end{aligned}$$

Observe, sin embargo, que $\int_0^{+\infty} [\![a \geq t]\!]\, dt = \int_0^a 1\, dt = a$, cuando $a \geq 0$. Concluimos, por lo tanto, que

$$\int_0^{+\infty} \mathbb{P}[X \geq t]\, dt = \sum_{a \in A_+} a \cdot \mathbb{P}[X = a].$$

Note que la serie, al lado superior derecho, es finita debido a la hipótesis que $\sum_{a \in A} |a| \cdot \mathbb{P}[X = a] < +\infty$. Un argumento análogo al anterior muestra por otro lado que

$$\int_{-\infty}^{0} \mathbb{P}[X \leq t]\, dt = \sum_{a \in A_-} (-a) \cdot \mathbb{P}[X = a] < +\infty,$$

donde $A_- = A \cap (-\infty, 0]$.

El valor esperado de X viene, por lo tanto, dado por

$$\begin{aligned} \mathbb{E}(X) &= \sum_{a \in A_+} a \cdot \mathbb{P}[X = a] - \sum_{a \in A_-} (-a) \cdot \mathbb{P}[X = a], \\ &= \sum_{a \in A_+} a \cdot \mathbb{P}[X = a] + \sum_{a \in A_-} a \cdot \mathbb{P}[X = a] = \sum_{a \in A} a \cdot \mathbb{P}[X = a], \end{aligned}$$

donde para la última identidad hemos usado que $(A_+ \cup A_-) = A$ y $(A_+ \cap A_-) = \{0\}$; en particular, el término $0 \cdot \mathbb{P}[X = 0]$ no contribuye a ninguna de las sumas al lado superior derecho. Esto completa la demostración de (3.4).

Completaremos ahora la demostración del teorema. Considere la v.a. $Y = g(X)$. Observe que $\mathbb{P}[Y \in B] = 1$, donde $B = \{g(a) : a \in A\}$. Como B es un conjunto contable (*¿por qué?*), Y es una v.a. discreta. Más aún, como para cada $b \in B$ el conjunto $g^{-1}(b) = \{a \in A : g(a) = b\}$ es contable, tenemos debido a la sigma aditividad que

$$\mathbb{P}[Y = b] = \sum_{a \in g^{-1}(b)} \mathbb{P}[X = a].$$

Concluimos de este modo que

$$\begin{aligned}\sum_{b\in B}|b|\cdot\mathbb{P}[Y=b] &= \sum_{b\in B}|b|\cdot\sum_{a\in g^{-1}(b)}\mathbb{P}[X=a],\\ &= \sum_{b\in B}\sum_{a\in g^{-1}(b)}|g(a)|\cdot\mathbb{P}[X=a]=\sum_{a\in A}|g(a)|\cdot\mathbb{P}[X=a]<+\infty.\end{aligned}$$

En particular, y de acuerdo a la identidad en (3.4), Y es una v.a. con esperanza finita. Más aún, repitiendo el argumento anterior obtenemos finalmente que

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{b\in B}b\cdot\mathbb{P}[Y=b],\\ &= \sum_{b\in B}b\cdot\sum_{a\in g^{-1}(b)}\mathbb{P}[X=a],\\ &= \sum_{b\in B}\sum_{a\in g^{-1}(b)}g(a)\cdot\mathbb{P}[X=a]=\sum_{a\in A}g(a)\cdot\mathbb{P}[X=a].\end{aligned}$$

Esto completa la demostración del teorema.

3.3 Esperanza de variables aleatorias continuas

En esta sección caracterizaremos de manera alternativa el valor esperado de una v.a. continua. El resultado principal es el siguiente.

Teorema 3.4. *Considere una v.a. continua X con función de densidad f. Si $g:\mathbb{R}\to\mathbb{R}$ es una función continuamente diferenciable excepto sobre un número finito de puntos y $\int_{\mathbb{R}}|g(x)|\cdot f(x)\,dx<\infty$ entonces $g(X)$ tiene un valor esperado finito dado por la fórmula*

$$\mathbb{E}(g(X))=\int_{\mathbb{R}}g(x)\cdot f(x)\,dx.$$

Un caso importante del teorema anterior es cuando $g(x)=x$. En este caso, si $\int_{\mathbb{R}}|x|\cdot f(x)\,dx<\infty$ entonces X tiene un valor esperado finito dado por la fórmula

$$\mathbb{E}(X)=\int_{\mathbb{R}}x\cdot f(x)\,dx. \tag{3.6}$$

Demostraremos el Teorema 3.4 al final de esta sección. En el intertanto veremos algunos ejemplos que clarificarán la importancia de este teorema.

Ejemplo 3.7. Si $X\sim Uniforme[a,b]$ entonces X es una v.a. continua con función de densidad $f(x)=1/(b-a)$, cuando $a\leq x\leq b$, y $f(x)=0$ en cualquier otro caso (Definición 2.6). En particular, para determinar la integral en (3.6) basta con integrar entre a y b i.e.

$$\mathbb{E}(X)=\int_a^b x\cdot f(x)\,dx=\frac{1}{b-a}\int_a^b x\,dx=\left.\frac{x^2}{2(b-a)}\right|_{x=a}^{x=b}=\frac{b^2-a^2}{2(b-a)}=\frac{a+b}{2},$$

lo que corresponde al punto medio del intervalo $[a,b]$.

Una de las ventajas del Teorema 3.4 es que podemos determinar el valor esperado de cualquier función de X, sin tener que caracterizar la distribución de la nueva variable aleatoria. Por ejemplo, si $0 \notin [a,b]$ entonces, usando el Método del Jacobiano, se encuentra que la v.a. $Y = 1/(1+X^2)$ tiene una función de densidad dada por la fórmula

$$f_Y(y) = \frac{1}{2(b-a)y^2}\sqrt{\frac{y}{1-y}},$$

siempre y cuando y pertenezca al rango de la función $g(x) = 1/(1+x^2)$, con $x \in [a,b]$. El cálculo de $\mathbb{E}(Y)$, usando la definición, requeriría integrar la función $y \cdot f_Y(y)$, lo que resulta un poco tedioso. Sin embargo, debido al Teorema 3.4, tenemos más directamente que

$$\mathbb{E}\left(\frac{1}{1+X^2}\right) = \frac{1}{b-a}\int_a^b \frac{1}{1+x^2}\,dx = \frac{\arctan(b)-\arctan(a)}{b-a}.$$

Ejemplo 3.8. Considere una v.a. $X \sim Uniforme[0,1]$ y la variable aleatoria

$$Y = \begin{cases} 0, & \text{si } X < 1/3; \\ 1, & \text{si } 1/3 \leq X < 2/3; \\ 2, & \text{si } 2/3 \leq X. \end{cases}$$

Recordemos que la función de densidad de X viene dada por $f(x) = 1$ si $x \in [0,1]$, y $f(x) = 0$ en cualquier otro caso. Para determinar la esperanza de Y observe que $Y = g(X)$, donde g es la función definida como

$$g(x) = \begin{cases} 0, & \text{si } x < 1/3; \\ 1, & \text{si } 1/3 \leq x < 2/3; \\ 2, & \text{si } 2/3 \leq x. \end{cases}$$

Observe que g es continuamente diferenciable sobre $\mathbb{R}$, excepto sobre los puntos $x = 1/3$ y $x = 2/3$. El Teorema 3.4 implica, por lo tanto, que

$$\mathbb{E}(Y) = \int_{\mathbb{R}} g(x)\cdot f(x) = \int_0^1 g(x)\,dx = \int_0^{1/3} 0\,dx + \int_{1/3}^{2/3} 1\,dx + \int_{2/3}^{1} 2\,dx = 1.$$

Una forma de corroborar el cálculo anterior, es notar que Y es una v.a. discreta que sólo puede tomar los valores 0, 1 o 2 con probabilidad $1/3$ cada uno (*¿por qué?*). En particular, debido al Teorema 3.3, podemos determinar su esperanza alternativamente como sigue:

$$\mathbb{E}(Y) = 0\cdot\mathbb{P}[Y=0] + 1\cdot\mathbb{P}[Y=1] + 2\cdot\mathbb{P}[Y=2] = 1.$$

Demostración Teorema 3.4. Primero demostraremos la identidad en (3.6). En efecto, dado $t \geq 0$ y como X tiene función de densidad f tenemos que

$$\mathbb{P}[X \geq t] = \int_t^\infty f(x)\,dx = \int_0^\infty f(x)\cdot [\![x \geq t]\!]\,dx.$$

En particular, la siguientes identidades aplican

$$\begin{aligned}
\int_0^\infty \mathbb{P}[X \geq t]\,dt &= \int_0^\infty \int_0^\infty f(x) \cdot [\![x \geq t]\!]\,dx\,dt, \\
&= \int_0^\infty \int_0^\infty f(x) \cdot [\![x \geq t]\!]\,dt\,dx, \\
&= \int_0^\infty f(x) \int_0^\infty [\![x \geq t]\!]\,dt\,dx, \\
&= \int_0^\infty f(x) \int_0^x 1\,dt\,dx = \int_0^\infty x \cdot f(x)\,dx.
\end{aligned}$$

En las identidades que mostramos, para la segunda, hemos intercambiado el orden de integración. Esto es posible debido a lo que se llama el *Teorema de Fubini*, y por la misma razón que las series $\sum_i \sum_j a_{i,j} = \sum_j \sum_i a_{i,j}$ pueden intercambiarse cuando los coeficientes $a_{i,j}$ son no negativos. Para la tercera identidad hemos usado que $f(x)$ se comporta como un factor constante cuando t es la variable de integración. Finalmente, para la cuarta identidad, hemos usado que $[\![x \geq t]\!] = 1$ cuando $t \leq x$, y $[\![x \geq t]\!] = 0$ cuando $t > x$.

Un argumento parecido al recién utilizado muestra por otro lado que

$$\int_{-\infty}^0 \mathbb{P}[X \geq t]\,dt = -\int_{-\infty}^0 x \cdot f(x)\,dx.$$

En particular, el valor esperado de X viene dado por

$$\mathbb{E}(X) = \int_0^\infty x \cdot f(x)\,dx + \int_{-\infty}^0 x \cdot f(x)\,dx = \int_{\mathbb{R}} x \cdot f(x)\,dx.$$

Esto demuestra la identidad en (3.6).

Para completar la demostración del teorema supondremos que g es continuamente diferenciable sobre $\mathbb{R}$ y que $g'(x) > 0$, para todo $x \in \mathbb{R}$. En este caso, de acuerdo al Método del Jacobiano (Teorema 2.8), la v.a. $Y = g(X)$ es también una v.a. continua, con función de densidad $f_Y(y) = f(g^{-1}(y))/g'(g^{-1}(y))$. En particular, debido a la identidad en (3.6) tenemos que

$$\mathbb{E}(g(X)) = \mathbb{E}(Y) = \int_{\mathbb{R}} y \cdot \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}\,dy = \int_{\mathbb{R}} g(x) \cdot \frac{f(x)}{g'(x)} \cdot g'(x)\,dx = \int_{\mathbb{R}} g(x) \cdot f(x)\,dx.$$

Para la tercera identidad que mostramos, hemos substituido $y = g(x)$ en la integral. Esto completa la demostración del teorema.

3.4 La Ley de los Grandes Números

De acuerdo a la discusión dada en la Sección 3.1, la interpretación frecuentista de probabilidades sugiere que

$$\frac{1}{n}\sum_{i=1}^n X_i \approx \mathbb{E}(X_1),$$

cuando n es suficientemente grande. Esta aproximación asume que, las variables aleatorias que participan en el promedio, son independientes e igualmente distribuidas. Abreviadamente diremos que $X_1, X_2, \ldots$ *son i.i.d.*

De acuerdo al siguiente resultado, la aproximación anterior es correcta y de hecho, exacta en el límite cuando n tiende a infinito.

Teorema 3.5. *(Ley de los Grandes Números.) Si $X_1, X_2, \ldots$ son i.i.d. y X_1 tiene un valor esperado bien definido entonces, con probabilidad uno, sucede lo siguiente:*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}(X_1).$$

La Ley de los Grandes Números muestra que la interpretación frecuentista, es consistente con la teoría de probabilidades motivada a partir de ella. Para fijar ideas, considere el experimento donde se lanza una moneda que tiene probabilidad p de salir cara. De acuerdo a la interpretación frecuentista, si lanzáramos la moneda un número grande de veces, entonces ésta debiera salir cara aproximadamente en una fracción p de los lanzamientos. Podemos ahora demostrar esto matemáticamente como sigue. Para cada $i \geq 1$ defina la variable aleatoria

$$\begin{aligned} X_i &= [\![\text{la moneda sale cara en el i-ésimo lanzamiento}]\!], \\ &= \begin{cases} 1, & \text{si la moneda sale cara en el i-ésimo lanzamiento;} \\ 0, & \text{si la moneda sale sello en el i-ésimo lanzamiento.} \end{cases} \end{aligned}$$

Claramente X_1 es una v.a. discreta. Más aún, tenemos que (Teorema 3.3):

$$\mathbb{E}(X_1) = 0 \cdot \mathbb{P}[X_1 = 0] + 1 \cdot \mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = 1] = p.$$

Como $X_1, X_2, \ldots$ son i.i.d., el Teorema 3.5 implica que

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = p.$$

Observe que $\sum_{i=1}^{n} X_i/n$ es la fracción de veces que la moneda salió cara en los primeros n lanzamientos. La identidad anterior muestra, por lo tanto, que la fracción asintótica del suceso "la moneda sale cara" es igual a su probabilidad.

Ejemplo 3.9. Si en promedio ocurren λ sismos por unidad de tiempo, entonces la separación temporal T entre el último y el próximo sismo tiene una función de densidad dada por la fórmula (Ejemplo 2.11):

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

Debido al Teorema 3.4, tenemos, por lo tanto, que

$$\mathbb{E}(T) = \int_{\mathbb{R}} t \cdot f(t)\, dt = \int_0^{\infty} t \cdot \lambda e^{-\lambda t}\, dt.$$

Substituyendo $s = \lambda t$ y luego usando el Método de Integración por Partes (ver la identidad en (4.15) en la Sección 4.5), obtenemos que

$$\mathbb{E}(T) = \frac{1}{\lambda}\int_0^\infty se^{-s}\,ds = \frac{1}{\lambda}\int_0^\infty e^{-s}\,ds = \frac{1}{\lambda}.$$

¿Cómo se interpreta lo anterior? Imagine que medimos el tiempo T_1 entre el último y el próximo sismo, luego el tiempo T_2 entre el próximo sismo y el subsiguiente, etc. Bajo el supuesto que $T_1, T_2, \ldots$ tienen la misma distribución y son independientes e.g. debido a que los sismos son detectados en una zona de actividad geológica moderada, la Ley de los Grandes Números implica que

$$\frac{1}{n}\sum_{i=1}^n T_i \approx \frac{1}{\lambda},$$

cuando n es grande. Esto es consistente con la interpretación original de λ: *si en promedio ocurren λ sismos por unidad de tiempo, entonces la separación temporal promedio entre sismos consecutivos debiera ser $1/\lambda$.*

Ejemplo 3.10. Imagine que usted y un cierto oponente lanzan simultáneamente dados equilibrados en un juego de azar. Denotaremos como X e Y el valor de la cara que usted y su oponente lanzarán, respectivamente. Las ganancias y pérdidas de este juego de azar están definidas por la siguiente regla: usted recibirá k pesos por cada peso apostado si $X > Y$, y perderá todo lo apostado si $X \leq Y$. En particular, su ganancia neta por juego, por cada peso apostado, será la variable aleatoria

$$Z = k \cdot [\![X > Y]\!] - 1 = \begin{cases} k-1, & \text{si } X > Y; \\ (-1), & \text{si } X \leq Y. \end{cases}$$

Arriba $k > 1$ es una constante dada, *¿para qué valores de k es este juego más conveniente para usted que su oponente?* Para responder a la pregunta determinaremos primero el valor esperado de Z. Como esta v.a. sólo puede tomar los valores $(k-1)$ y (-1) obtenemos que (Teorema 3.3):

$$\begin{aligned} \mathbb{E}(Z) &= (k-1)\cdot\mathbb{P}[Z = k-1] + (-1)\cdot\mathbb{P}[Z = 1], \\ &= (k-1)\cdot\mathbb{P}[X > Y] + (-1)\cdot\mathbb{P}[X \leq Y], \\ &= (k-1)\cdot\frac{15}{36} - \frac{21}{36} = \frac{5}{12}\left(k - \frac{12}{5}\right). \end{aligned}$$

Para que el juego sea más conveniente para usted, basta que $\mathbb{E}(Z) > 0$ i.e. $k > 12/5$. Esto se debe a la Ley de los Grandes Números: si $Z_1, \ldots, Z_n$ denotan las ganancias en el primer juego, en el segundo, etc entonces $\sum_{i=1}^n Z_i \approx n \cdot \mathbb{E}(Z)$, cuando n es grande. En este caso, la ganancia total después de un número grande n de juegos será aproximadamente $n \cdot \mathbb{E}(Z)$, la que tiende a infinito a medida que n tiende a infinito. Por otro lado, si $k < 12/5$ e.g. $k = 2$ entonces el juego será más conveniente para su oponente en el largo plazo.

3.5 Propiedades de la esperanza

En esta sección discutiremos cuatro propiedades fundamentales acerca de la esperanza. La primera propiedad establece que el valor esperado de una v.a. constante es la misma constante.

Teorema 3.6. *(Esperanza de una constante.) Si X es una v.a. constante i.e. existe una constante $c \in \mathbb{R}$ tal que $\mathbb{P}[X = c] = 1$ entonces $\mathbb{E}(X) = c$.*

Demostración. Como $\mathbb{P}[X \in \{c\}] = 1$, X es una v.a. discreta. En particular, $\mathbb{E}(X) = c \cdot \mathbb{P}[X = c] = c$, lo que demuestra el teorema. □

La segunda propiedad importante establece que la esperanza de una variable aleatoria, está únicamente determinada por su función de distribución. En particular, variables aleatorias con la misma distribución tienen el mismo valor esperado.

Teorema 3.7. *Si X e Y tienen misma distribución y X tiene una esperanza bien definida entonces $\mathbb{E}(X) = \mathbb{E}(Y)$.*

Demostración. Debido a que X e Y tienen la misma distribución, se tiene directamente de la definición que $\mathbb{P}[X \geq t] = \mathbb{P}[Y \geq t]$ y $\mathbb{P}[X \leq t] = \mathbb{P}[Y \leq t]$, para todo $t \in \mathbb{R}$. En particular, las siguientes identidades aplican

$$\begin{aligned}
\int_0^{\infty} \mathbb{P}[X \geq t]\, dt &= \int_0^{\infty} \mathbb{P}[Y \geq t]\, dt, \\
\int_{-\infty}^{0} \mathbb{P}[X \leq t]\, dt &= \int_{-\infty}^{0} \mathbb{P}[Y \leq t]\, dt.
\end{aligned}$$

Como al menos una de las integrales al lado superior izquierdo es finita, Y tiene una esperanza bien definida y

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^{\infty} \mathbb{P}[X \geq t]\, dt - \int_{-\infty}^{0} \mathbb{P}[X \leq t]\, dt, \\
&= \int_0^{\infty} \mathbb{P}[Y \geq t]\, dt - \int_{-\infty}^{0} \mathbb{P}[Y \leq t]\, dt = \mathbb{E}(Y).
\end{aligned}$$

Esto demuestra el teorema. □

Una consecuencia del teorema anterior es que si $X_1, \ldots, X_n$ tienen la misma distribución entonces $\mathbb{E}(X_1) = \cdots = \mathbb{E}(X_n)$. En particular, debido al Teorema 2.12, si $g : \mathbb{R} \to \mathbb{R}$ es una función continua excepto sobre un conjunto finito de puntos y $g(X_1)$ tiene esperanza finita entonces $\mathbb{E}(g(X_1)) = \cdots = \mathbb{E}(g(X_n))$. Para fijar ideas, si $X \sim Uniforme[-1, 1]$ e $Y \sim Uniforme[-1, 1]$ entonces $\mathbb{E}(X^2) = \mathbb{E}(Y^2)$ y $\mathbb{E}(e^X) = \mathbb{E}(e^Y)$ (ver Ejercicio 5). Similarmente, si $X_1, \ldots, X_n$ son i.i.d. entonces $\mathbb{E}(\text{sen}(X_1)) = \cdots = \mathbb{E}(\text{sen}(X_n))$ y $\mathbb{E}(\cos(X_1)) = \cdots = \mathbb{E}(\cos(X_n))$.

Ejemplo 3.11. Imagine un profesor que corre apuradamente al kiosko del colegio donde enseña, para comprar un par de golosinas antes de su próxima clase. Supongamos que entre sus bolsillos cuenta con cinco monedas de \$50 y tres de \$100 pesos. Denotaremos como X e Y el valor de la primera y segunda moneda que sacará entre sus bolsillos, respectivamente. Claramente, $\mathbb{P}[X=50]=5/8$ y $\mathbb{P}[X=100]=3/8$, y como X e Y tienen la misma distribución (*¿por qué?*), el teorema anterior implica que

$$\begin{aligned}\mathbb{E}(Y)=\mathbb{E}(X) &= 50\cdot\mathbb{P}[X=50]+100\cdot\mathbb{P}[X=100],\\ &= 50\cdot 5/8+100\cdot 3/8=275/4.\end{aligned}$$

Una tercera propiedad importante, acerca de la esperanza de variables aleatorias, es qué factores constantes, dentro o fuera de un valor esperado, pueden intercambiarse, y más relevante aún, que el valor esperado de una suma corresponde a la suma de los valores esperados. Esta propiedad se llama *linealidad de la esperanza* y establece lo siguiente.

Teorema 3.8. *(Linealidad de la esperanza.) Si α,β son constantes dadas y X e Y variables aleatorias con esperanza finita entonces $\mathbb{E}(\alpha X)=\alpha\cdot\mathbb{E}(X)$ y $\mathbb{E}(X+Y)=\mathbb{E}(X)+\mathbb{E}(Y)$. Más generalmente, si $\alpha_1,\ldots,\alpha_n\in\mathbb{R}$ son constantes dadas y $X_1,\ldots,X_n$ tienen esperanzas finitas entonces*

$$\mathbb{E}\left(\sum_{i=1}^n \alpha_i X_i\right)=\sum_{i=1}^n \alpha_i\cdot\mathbb{E}(X_i).$$

Enfatizamos que la identidad en el teorema, aplica, más generalmente, cuando $X_1,\ldots,X_n$ tienen valores esperados bien definidos (ver Definición 3.2), y la sumatoria, al lado derecho, es una cantidad bien definida. Esto último requiere que los términos $(+\infty)-(+\infty)$, y $(-\infty)-(-\infty)$, no ocurran al simplificar la sumatoria.

Para demostrar el teorema basta con mostrar que

$$\mathbb{E}(\alpha X+Y)=\alpha\cdot\mathbb{E}(X)+\mathbb{E}(Y), \tag{3.7}$$

cuando X e Y tienen valores esperados finitos (*¿por qué?*). En vez de demostrar esto, sólo justificaremos por qué debería ser cierto. Imagine que X e Y son variables aleatorias asociadas a un mismo experimento el cual repetimos indefinida e independientemente. Si X_i e Y_i denotan los valores observados para X e Y en la i-ésima realización del experimento, entonces $(\alpha X_1+Y_1),(\alpha X_2+Y_2),\ldots$ son variables aleatorias i.i.d. y con la misma distribución que $(\alpha X+Y)$. En particular, y de acuerdo a la Ley de los Grandes Números, tenemos que

$$\begin{aligned}\mathbb{E}(\alpha X+Y) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n(\alpha X_i+Y_i),\\ &= \alpha\cdot\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i+\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n Y_i=\alpha\cdot\mathbb{E}(X)+\mathbb{E}(Y).\end{aligned}$$

Esto justifica (4.24) y, más generalmente, la linealidad de la esperanza.

Ejemplo 3.12. Considere variables aleatorias X e Y, cada una con función de densidad dada por la fórmula:

$$\begin{aligned} f_X(x) &= \begin{cases} 3x^2/2, & x \in [-1,1]; \\ 0, & x \notin [-1,1]; \end{cases} \\ f_Y(y) &= \begin{cases} 4y^3, & y \in [0,1]; \\ 0, & y \notin [0,1]; \end{cases} \end{aligned}$$

Determine el valor esperado de $(X+Y)$ usando la linealidad de la esperanza.

En efecto, debido al Teorema 3.4, tenemos que

$$\begin{aligned} \mathbb{E}(X) &= \int_{-1}^{1} x \cdot \frac{3x^2}{2}\, dx = \frac{3x^4}{8}\bigg|_{x=-1}^{x=1} = 0; \\ \mathbb{E}(Y) &= \int_0^1 y \cdot 4y^3\, dy = \frac{4y^5}{5}\bigg|_{y=0}^{y=1} = \frac{4}{5}. \end{aligned}$$

Así, debido a la linealidad de la esperanza, obtenemos que

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 0 + \frac{4}{5} = \frac{4}{5}.$$

Ejemplo 3.13. Retomemos el ejemplo del profesor que corre apuradamente al kiosko para comprar golosinas y que entre sus bolsillos tiene cinco monedas de \$50 y tres de \$100 pesos (Ejemplo 3.11). Si con la primera moneda el profesor comprará caramelos, y cada caramelo cuesta \$25 pesos, mientras que con la segunda moneda él comprará chicles, y cada chicle cuesta \$50 pesos, entonces el número total de golosinas que comprará en el kiosko será $X/25 + Y/50$. Debido a la linealidad de la esperanza, el valor esperado del número de golosinas que el profesor comprará es

$$\mathbb{E}\left(\frac{X}{25} + \frac{Y}{50}\right) = \frac{1}{25} \cdot \mathbb{E}(X) + \frac{1}{50} \cdot \mathbb{E}(Y) = \left(\frac{1}{25} + \frac{1}{50}\right) \cdot \mathbb{E}(X) = \frac{33}{8},$$

donde para la penúltima identidad hemos usado que X e Y tienen la misma distribución. Una consecuencia de lo anterior es la siguiente: si el profesor aplica cada día la misma estrategia entonces, de acuerdo a la Ley de los Grandes Números, en un periodo de 40 semanas (i.e. aproximadamente dos semestres) el habrá consumido alrededor de $40 \cdot 5 \cdot 33/8 = 825$ golosinas. Si en promedio, cada golosina contiene 10 gramos de azúcar, el profesor consumirá durante el año académico alrededor de 8 kilos de azúcar, ¡por concepto de golosinas solamente!

Ejemplo 3.14. Considere el experimento donde uno de los 16 triángulos en la Figura 3.3(a) se escoge equiprobablemente al azar y luego un punto (X,Y) se escoge equidistribuidamente dentro de dicho triángulo.

Si quisiéramos determinar la esperanza de $(8X+3Y)$ usando la definición, entonces necesitaríamos determinar la integral $\int_0^\infty \mathbb{P}[8X+3Y \geq t]\, dt$. Esto, sin embargo, es tedioso. Por ejemplo, sólo para $t = 5$ necesitaríamos determinar la probabilidad que el punto (X, Y) pertenezca a la región oscura de puntos en la Figura 3.3(b).

Por otro lado, usando la linealidad de la esperanza tenemos que

$$\mathbb{E}(8X + 3Y) = 8\mathbb{E}(X) + 3\mathbb{E}(Y),$$

lo que nos permite analizar X e Y por separado. ¡Esta es una de las ventajas de la linealidad de la esperanza!

Para determinar $\mathbb{E}(X)$ mostraremos primero que $X \sim Uniforme[0, 1]$. Para esto imagine que trasladamos verticalmente los triángulos sobre el intervalo $[0, 1/4]$ a modo de formar el rectángulo con vértices $(0, 0)$, $(1/4, 0)$, $(1/4, 1/2)$ y $(0, 1/2)$. Similarmente, trasladamos verticalmente los cuatro triángulos sobre el intervalo $[1/4, 1/2]$ para formar el rectángulo con vértices $(1/4, 0)$, $(1/2, 0)$, $(1/2, 1/2)$ y $(1/4, 1/2)$. Un proceso similar es aplicado para los triángulos sobre el intervalo $[1/2, 3/4]$ y el intervalo $[3/4, 1]$.

El resultado final del proceso anterior será el rectángulo R con vértices $(0, 0)$, $(1, 0)$, $(1, 1/2)$ y $(0, 1/2)$. Más aún, como al trasladar los triángulos verticalmente sólo habremos trasladado el punto (X, Y) verticalmente, al final del proceso éste se ubicará en un cierto punto de la forma (X, Y'). Debido a la manera como el punto (X, Y) es escogido, el punto (X, Y') podrá ubicarse en cualquier parte del rectángulo sin preferencia de ubicación por un cierto lugar más que otro i.e. (X, Y') es un punto equidistribuido en el rectángulo R. En particular, si $0 \leq t \leq 1$ y R_t es el rectángulo definido como $\{(x, y) \mid 0 \leq x \leq t \text{ y } 0 \leq y \leq 1/2\}$ entonces

$$F_X(t) = \mathbb{P}[X \leq t] = \mathbb{P}[(X, Y') \in R_t] = \frac{\text{área}(R_t)}{\text{área}(R)} = \frac{t/2}{1/2} = t.$$

Diferenciado con respecto a la variable t obtenemos, por lo tanto, que X es una v.a. continua con función de densidad

$$f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1; \\ 0, & \text{en cualquier otro caso.} \end{cases}$$

Lo anterior muestra que $X \sim Uniforme[0, 1]$; en particular, se tiene que $\mathbb{E}(X) = 1/2$ (Ejemplo 3.7). Un proceso similar, muestra que $Y \sim Uniforme[0, 1]$ y, por lo tanto,

$$\mathbb{E}(8X + 3Y) = 8 \cdot 1/2 + 3 \cdot 1/2 = 11/2.$$

Para finalizar esta sección, veremos cómo determinar el valor esperado de un producto de variables aleatorias independientes. Una de las maneras más fáciles de construir variables independientes es usando sucesos independientes. En efecto, considere un cierto experimento que tiene asociado un espacio muestral Ω y dos sucesos independientes $A, B \subset \Omega$. Considere las variables aleatorias discretas definidas como

$$\begin{aligned} X &= [\![\text{el suceso } A \text{ ocurre}]\!], \\ Y &= [\![\text{el suceso } B \text{ ocurre}]\!]. \end{aligned}$$

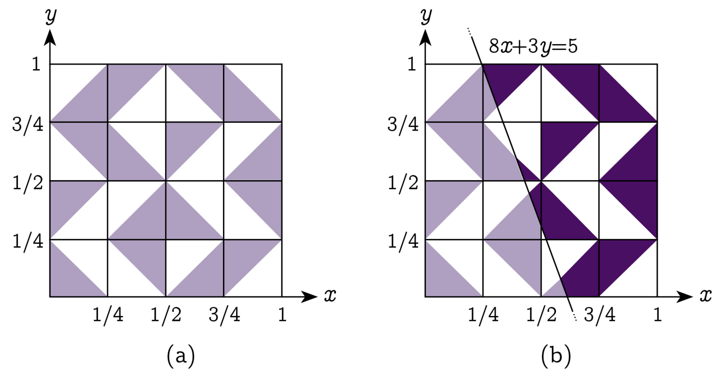


Figura 3.3

El lector podrá verificar fácilmente usando el Teorema 2.18 que X e Y son independientes. Note que X e Y sólo pueden tomar los valores cero o uno. Por ejemplo, $X = 1$ cada vez que A ocurre y $X = 0$ en el caso contrario. En particular, se tiene que (Teorema 3.3)

$$\begin{aligned}\mathbb{E}(X) &= 1 \cdot \mathbb{P}[X = 1] = \mathbb{P}(A);\\ \mathbb{E}(Y) &= 1 \cdot \mathbb{P}[Y = 1] = \mathbb{P}(B).\end{aligned}$$

Por otro lado, XY también toma solamente los valores cero o uno. Más aún, $XY = 1$ sólo cuando A y B ocurren simultáneamente. Como A y B son sucesos independientes, esto implica que

$$\mathbb{E}(XY) = \mathbb{P}[X = 1, Y = 1] = \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

i.e. $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$. Sorprendentemente, y según lo establece el siguiente resultado, dado cualquier par de variables aleatorias independientes y con esperanza finita, el valor esperado de su producto corresponderá siempre al producto de sus valores esperados.

Teorema 3.9. *(Esperanza de factores independientes.) Si X e Y son variables aleatorias independientes con esperanza finita entonces $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$. Más generalmente, si $X_1, \ldots, X_n$ son variables aleatorias independientes con esperanza finita entonces*

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

Ejemplo 3.15. Considere el experimento donde se dibuja un cuadrado como sigue: escogemos al azar e independientemente dos puntos X e Y uniformemente en el intervalo $[0, 1]$ y usamos $(X + Y)$ como el largo de la diagonal del cuadrado. *¿Cuál es el valor esperado del área del cuadrado?*

El área S de un cuadrado con una diagonal de largo $(X+Y)$ viene dada por $S=(X+Y)^2/2=(X^2+2XY+Y^2)/2$ (*¿por qué?*). En particular, la linealidad de la esperanza nos permite concluir que

$$\mathbb{E}(S)=\frac{\mathbb{E}(X^2)+2\cdot\mathbb{E}(XY)+\mathbb{E}(Y^2)}{2}.$$

Observe, sin embargo, que $\mathbb{E}(XY)=\mathbb{E}(X)\cdot\mathbb{E}(Y)$, debido a que X e Y son factores independientes (Teorema 3.9). Más aún, $\mathbb{E}(X)=\mathbb{E}(Y)$ debido a que X e Y tienen la misma distribución (Teorema 3.7). Similarmente, $\mathbb{E}(X^2)=\mathbb{E}(Y^2)$. Así obtenemos que

$$\mathbb{E}(S)=\frac{2\cdot\mathbb{E}(X^2)+2\cdot\{\mathbb{E}(X)\}^2}{2}=\mathbb{E}(X^2)+\{\mathbb{E}(X)\}^2.$$

Note que, en general, $\mathbb{E}(X^2)\neq\{\mathbb{E}(X)\}^2$; en particular, para determinar el lado superior derecho no basta con determinar el valor esperado de X. De hecho, como $X\sim Uniforme[0,1]$, encontramos que

$$\begin{aligned}\mathbb{E}(X^2) &= \int_0^1 x^2\cdot 1\,dx=\frac{1}{3};\\ \mathbb{E}(X) &= \int_0^1 x\cdot 1\,dx=\frac{1}{2}.\end{aligned}$$

Por lo tanto, $\mathbb{E}(S)=1/3+1/4=7/12$.

Es importante enfatizar que la linealidad de la esperanza, aplica para cualquier colección de variables aleatorias—independientes o no. Sin embargo, el Teorema 3.9 sólo aplica cuando los factores que participan del producto son independientes. El siguiente ejemplo aclarará la importancia de esta hipótesis.

Ejemplo 3.16. Retomemos el ejemplo del profesor que corre al kiosko para comprar golosinas y que entre sus bolsillos porta cinco monedas de \$50 y tres de \$100 pesos (ejemplos 3.11 y 3.13). Si X e Y denotan el valor de la primera y segunda moneda que sacará entre sus bolsillos, respectivamente, entonces X e Y no son variables aleatorias independientes. Por ejemplo, se tiene que

$$\mathbb{P}[Y=100\mid X=100]=\frac{2}{7}<\frac{3}{8}=\mathbb{P}[X=100].$$

Veremos que $\mathbb{E}(X \cdot Y) \neq \mathbb{E}(X) \cdot \mathbb{E}(Y)$. En efecto, la v.a. XY sólo puede tomar los valores 2500, 5000 o 10000 con las siguientes probabilidades:

$$
\begin{aligned}
\mathbb{P}[XY = 2500] &= \mathbb{P}[X = 50, Y = 50], \\
&= \mathbb{P}[X = 50] \cdot \mathbb{P}[Y = 50 \mid X = 50] = 5/14, \\
\mathbb{P}[XY = 5000] &= \mathbb{P}[X = 50, Y = 100] + \mathbb{P}[X = 100, Y = 50], \\
&= \mathbb{P}[X=50] \cdot \mathbb{P}[Y=100 \mid X=50] + \mathbb{P}[X=100] \cdot \mathbb{P}[Y=50 \mid X=100], \\
&= 5/8 \cdot 3/7 + 3/8 \cdot 5/7 = 15/28, \\
\mathbb{P}[XY = 10000] &= \mathbb{P}[X = 100, Y = 100], \\
&= \mathbb{P}[X = 100] \cdot \mathbb{P}[Y = 100 \mid X = 100] = 3/28.
\end{aligned}
$$

Por lo tanto,

$$\mathbb{E}(XY) = 2500 \cdot 5/14 + 5000 \cdot 15/28 + 10000 \cdot 3/28 = 32500/7.$$

Por otro lado, vimos en el Ejemplo 3.11 que $\mathbb{E}(X) = \mathbb{E}(Y) = 275/4$ y, por lo tanto, $\mathbb{E}(XY) < \mathbb{E}(X) \cdot \mathbb{E}(Y)$. La orientación de esta desigualdad es consistente con la intuición. De hecho, si $X = 50$, entonces $Y = 100$ con probabilidad $3/7$, y si $X = 100$, entonces $Y = 100$ con probabilidad $2/7$. En cualquier caso, Y tenderá a tomar valores menores o iguales a X la mayoría de las veces. Por lo tanto, el valor esperado para XY debiera ser menor que $\mathbb{E}(X) \cdot \mathbb{E}(Y)$ ya que esto último asume que el valor de X no influencia el valor de Y.

3.6 Varianza de variables aleatorias

Considere el experimento donde se lanza un dado equilibrado y la v.a. definida como

$$X = \begin{cases} 1, & \text{si la cara del dado es par;} \\ (-1), & \text{si la cara del dado es impar.} \end{cases}$$

Esta v.a. puede interpretarse como la ganancia neta, por cada peso apostado, en un juego donde el apostador duplica su apuesta si al tirar un dado éste sale par. En este caso, como la probabilidad que el dado salga par es la misma que salga impar, $\mathbb{E}(X) = 0$. En particular, si definimos $Y = X/1000$ y $Z = 1000X$, entonces $\mathbb{E}(Y) = \mathbb{E}(X)/1000 = 0$ y $\mathbb{E}(Z) = 1000 \cdot \mathbb{E}(X) = 0$. Vemos, por lo tanto, que tanto Y como Z tienen valor esperado cero, sin embargo, Y sólo puede tomar los valores $\pm 0{,}001$, mientras que Z sólo puede tomar los valores ± 1000. Claramente, Y es una v.a. mucho más concentrada alrededor de cero (su valor esperado) que Z.

La varianza de una v.a. es una cantidad que mide cuán concentrada está una v.a. alrededor de su valor esperado. La definición precisa es la siguiente.

Definición 3.10. Si X es una v.a. con esperanza finita, entonces su *varianza* es la cantidad definida como

$$\mathbb{V}(X) = \mathbb{E}\left(\{X - \mathbb{E}(X)\}^2\right).$$

La idea detrás de la definición es la siguiente: la v.a. X toma valores cerca de su valor esperado con alta probabilidad si y solo si, $\{X - \mathbb{E}(X)\}^2$ toma valores cerca de cero con alta probabilidad. Como $\{X - \mathbb{E}(X)\}^2 \geq 0$, el valor esperado de esta última v.a. es pequeño solamente cuando X está bien concentrada alrededor de su valor esperado. (Cuantificaremos este hecho de forma más precisa usando la *desigualdad de Chebyshev* en el Ejercicio 46.)

La *desviación estándar* de X se define como la raíz cuadrada positiva de su varianza. A diferencia de la varianza, la desviación estándar se mide en las mismas unidades que la variable aleatoria. La varianza, sin embargo, se mide usando unidades-cuadradas e.g. cm^2 cuando X se mide en centímetros, o seg^2 cuando X se mide en segundos.

Observe que la varianza de una v.a. corresponde a la esperanza de otra variable aleatoria aunque no negativa. En particular, la varianza de una v.a. es una cantidad no negativa (Ejercicio 5), aunque podría ser igual a infinito. La varianza al igual que la esperanza tiene una serie de propiedades interesantes que discutiremos en esta sección.

Teorema 3.11. *Si X es una v.a. con esperanza finita, entonces*

(a) $\mathbb{V}(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2$,
(b) $\mathbb{V}(X) = \mathbb{E}(X \cdot (X-1)) - \mathbb{E}(X) \cdot \{\mathbb{E}(X) - 1\}$, *y*
(c) *para todo par de constantes $\alpha, \beta \in \mathbb{R}$ se tiene que $\mathbb{V}(\alpha X + \beta) = \alpha^2 \cdot \mathbb{V}(X)$.*

Demostración. Denotaremos como μ el valor esperado de X. Debido a la linealidad de la esperanza, tenemos que

$$\begin{aligned}
\mathbb{V}(X) = \mathbb{E}(\{X-\mu\}^2) &= \mathbb{E}(X^2 - 2X\mu + \mu^2), \\
&= \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2, \\
&= \mathbb{E}(X^2) - 2\mu^2 + \mu^2, \\
&= \mathbb{E}(X^2) - \mu^2, \\
&= \mathbb{E}(X^2) - \mathbb{E}(X) + \mu - \mu^2, \\
&= \mathbb{E}(X^2 - X) - \mu(\mu-1) = \mathbb{E}(X(X-1)) - \mu(\mu-1).
\end{aligned}$$

La quinta identidad demuestra la parte (a), y la última identidad demuestra la parte (b). Finalmente, y nuevamente debido a la linealidad de la esperanza, encontramos que

$$\begin{aligned}
\mathbb{V}(\alpha X + \beta) &= \mathbb{E}\left(\{\alpha X + \beta - \mathbb{E}(\alpha X + \beta)\}^2\right), \\
&= \mathbb{E}\left(\{\alpha X + \beta - \alpha \cdot \mathbb{E}(X) - \beta\}^2\right), \\
&= \mathbb{E}(\alpha^2 \cdot \{X - E(X)\}^2), \\
&= \alpha^2 \cdot \mathbb{E}(\{X - \mathbb{E}(X)\}^2), \\
&= \alpha^2 \cdot \mathbb{V}(X),
\end{aligned}$$

lo que demuestra la parte (c). Esto completa la demostración del teorema. □

Ejemplo 3.17. Retomemos el experimento donde tiramos un dado equilibrado al azar y definimos

$$X = \begin{cases} 1, & \text{si la cara del dado es par;} \\ (-1), & \text{si la cara del dado es impar.} \end{cases}$$

Note que $X^2 = 1$ i.e. X^2 es una v.a. constante. Por lo tanto, $\mathbb{E}(X^2) = 1$ debido al Teorema 3.6. Como $\mathbb{E}(X) = 0$, la identidad en la parte (a) del Teorema 3.11 nos permite concluir que $\mathbb{V}(X) = 1 - 0^2 = 1$. En particular, si definimos $Y = 10^{-3} \cdot X$ y $Z = 10^3 \cdot X$, entonces la parte (c) en el teorema nos permite concluir que

$$\begin{aligned} \mathbb{V}(Y) &= \mathbb{V}\left(10^{-3} \cdot X\right) = (10^{-3})^2 \cdot \mathbb{V}(X) = 10^{-6}, \\ \mathbb{V}(Z) &= \mathbb{V}\left(10^{3} \cdot X\right) = (10^{3})^2 \cdot \mathbb{V}(X) = 10^{6}. \end{aligned}$$

Vemos que la varianza de Y es mucho menor que la varianza de Z. Esto significa que Y está mucho más concentrado alrededor de su valor esperado que Z. Lo anterior resulta bastante intuitivo, ya que $\mathbb{E}(Y) = \mathbb{E}(Z) = 0$, sin embargo, Y sólo puede tomar los valores $\pm 10^{-3}$, mientras que Z sólo puede tomar los valores $\pm 10^3$.

Ejemplo 3.18. *Determine la varianza de una v.a. X tal que* $\mathbb{P}[X = 0] = 1/7$, $\mathbb{P}[X = 1] = 2/7$ *y* $\mathbb{P}[X = 2] = 4/7$. Observe que el suceso $[X \in \{0, 1, 2\}]$ tiene probabilidad uno; en particular, X es una v.a. discreta. Usando el Teorema 3.3 obtenemos, por lo tanto, que

$$\begin{aligned} \mathbb{E}(X) &= 0 \cdot 1/7 + 1 \cdot 2/7 + 2 \cdot 4/7 = 10/7; \\ \mathbb{E}(X^2) &= 0^2 \cdot 1/7 + 1^2 \cdot 2/7 + 2^2 \cdot 4/7 = 18/7. \end{aligned}$$

Finalmente, el Teorema 3.11 implica que

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2 = 18/7 - 100/49 = 26/49.$$

Ejemplo 3.19. Si $X \sim Uniforme[a, b]$, entonces vimos en el Ejemplo 3.7 que $\mathbb{E}(X) = (a + b)/2$. Por otro lado,

$$\mathbb{E}(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a}\, dx = \left.\frac{x^3}{3(b-a)}\right|_{x=a}^{x=b} = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}.$$

El Teorema 3.11 implica, por lo tanto, que

$$\mathbb{V}(X) = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}.$$

La linealidad de la esperanza establece que el valor esperado de una suma corresponde a la suma de los valores esperados siempre y cuando esta última cantidad esté bien definida. En el caso de la varianza, no es en general cierto que la varianza de una suma corresponde a la suma de las varianzas (Ejercicio 44). Sin embargo, cuando las variables en cuestión son independientes, la varianza de una suma será siempre igual a la suma de las varianzas. Demostraremos este resultado a continuación.

Teorema 3.12. *Si $X_1, \ldots, X_n$ son variables aleatorias independientes con esperanza finita, entonces*

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

Demostración. Si μ_i denota el valor esperado de X_i, entonces tenemos lo siguiente

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \mathbb{E}\left(\left\{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right\}^2\right) = \mathbb{E}\left(\left\{\sum_{i=1}^{n} (X_i - \mu_i)\right\}^2\right). \tag{3.8}$$

Pero observe que si $a_1, \ldots, a_n$ son números reales, entonces

$$\left\{\sum_{i=1}^{n} a_i\right\}^2 = 2 \cdot \sum_{1 \le i < j \le n} a_i \cdot a_j + \sum_{i=1}^{n} a_i^2. \tag{3.9}$$

En particular, aplica que

$$\left\{\sum_{i=1}^{n} (X_i - \mu_i)\right\}^2 = 2 \cdot \sum_{1 \le i < j \le n} (X_i - \mu_i) \cdot (X_j - \mu_j) + \sum_{i=1}^{n} \{X_i - \mu_i\}^2. \tag{3.10}$$

Si $i < j$ y como X_i y X_j son variables aleatorias independientes, el Teorema 2.17 implica que $(X_i - \mu_i)$ y $(X_j - \mu_j)$ son también independientes. Por lo tanto, de acuerdo al Teorema 3.9, tenemos que

$$\begin{aligned}
\mathbb{E}((X_i - \mu_i) \cdot (X_j - \mu_j)) &= \mathbb{E}(X_i - \mu_i) \cdot \mathbb{E}(X_j - \mu_j), \\
&= (\mathbb{E}(X_i) - \mu_i) \cdot (\mathbb{E}(X_j) - \mu_j), \\
&= (\mu_i - \mu_i) \cdot (\mu_j - \mu_j) = 0.
\end{aligned}$$

Como $\mathbb{E}(\{X_i - \mu_i\}^2) = \mathbb{V}(X_i)$, usando las identidades en (3.8) y (3.10), concluimos finalmente que

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = 2 \cdot \sum_{1 \le i < j \le n} 0 + \sum_{i=1}^{n} \mathbb{V}(X_i) = \sum_{i=1}^{n} \mathbb{V}(X_i).$$

Esto completa la demostración del teorema. □

Ejemplo 3.20. Un profesor pregunta a sus alumnos "¿será cierto que si X e Y son variables aleatorias independientes y uniformes en el intervalo $[0, 1]$, entonces $(X+Y)$ es una v.a. uniforme en el intervalo $[0, 2]$?". Siguiendo esto, el profesor hace una pausa y continúa diciendo "debido a que tanto X como Y pueden tomar cualquier valor entre 0 y 1, parece razonable que $(X + Y)$ pueda tomar cualquier valor entre 0 y 2, sin preferencia de ubicación entre estos dos valores. ¿Es esta lógica correcta?". *Decida si la lógica expuesta por el profesor es correcta o no.*

Hay muchas formas de afrontar la pregunta del profesor e.g. determinar la función de distribución de $(X + Y)$ usando un método análogo al Ejemplo 2.32. Una forma más directa es la siguiente. Recordemos que si $Z \sim Uniforme[a,b]$, entonces $\mathbb{V}(Z) = (b-a)^2/12$ (Ejemplo 3.19). En particular, si la lógica del profesor es correcta y como X e Y son independientes, entonces debiera ternerse que

$$\frac{1}{3} = \frac{(2-0)^2}{12} = \mathbb{V}(X+Y) \stackrel{?}{=} \mathbb{V}(X) + \mathbb{V}(Y) = \frac{(1-0)^2}{12} + \frac{(1-0)^2}{12} = \frac{1}{6}.$$

Concluimos así que la lógica del profesor es incorrecta. De hecho, el lector podrá determinar por su cuenta la función de distribución de $(X+Y)$ y concluir a partir de ella, que esta v.a. tiene una función de densidad dada por la fórmula

$$f_{X+Y}(t) = \begin{cases} t, & t \in [0,1]; \\ 2-t, & t \in [1,2]; \\ 0, & t \notin [0,2]. \end{cases}$$

$(X+Y)$ no es, por lo tanto, uniforme sobre el intervalo $[0,2]$ (*¿por qué?*).

Ejemplo 3.21. Considere las notas $X_1, \ldots, X_n$ de un alumno de un cierto curso. El promedio final de notas $\bar{X}$ se usa como una medida objetiva del entendimiento y habilidad en la materia del curso por parte del alumno, *¿en qué medida es esto respaldado por la teoría de probabilidades?*

Para responder a la pregunta supondremos que $X_1, \ldots, X_n$ son variables aleatorias i.i.d. Observe que cada X_i sólo puede tomar valores entre 1 y 7. Por lo tanto, la esperanza y la varianza de estas variables aleatorias son finitas (Ejercicio 5). Más aún, como las variables aleatorias tienen la misma distribución, existen constantes μ y σ^2 tales que $\mathbb{E}(X_1) = \cdots = \mathbb{E}(X_n) = \mu$ y $\mathbb{V}(X_1) = \cdots = \mathbb{V}(X_n) = \sigma^2$.

La constante μ representa la nota intrínseca del alumno en el curso. En particular, debido a la Ley de los Grandes Números, $\bar{X}$ será una buena aproximación de la nota intrínseca cuando un número suficientemente grande de exámenes es usado para calcular el promedio final de notas. ¡Pero hay más! Observe que

$$\begin{aligned} \mathbb{E}(\bar{X}) &= \mathbb{E}\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \sum_{i=1}^n \mathbb{E}\left(\frac{X_i}{n}\right) = \sum_{i=1}^n \frac{\mu}{n} = \mu, \\ \mathbb{V}(\bar{X}) &= \mathbb{V}\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \sum_{i=1}^n \mathbb{V}\left(\frac{X_i}{n}\right) = \sum_{i=1}^n \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

Para el cálculo del valor esperado, hemos usado la linealidad de la esperanza, y para el cálculo de la varianza, hemos usado que $X_1, \ldots, X_n$ son independientes (Teorema 3.12).

Vemos que el valor esperado de $\bar{X}$ es el mismo que cada una de las variables $X_1, \ldots, X_n$, sin embargo, la varianza de $\bar{X}$ es mucho más pequeña que la varianza de cada una de estas variables (debido al factor $1/n$). En particular, $\bar{X}$ es más probable de tomar un valor cerca de μ que cualquiera de las variables X_i. Esto justifica la estrategia de adivinar μ usando $\bar{X}$.

3.6.1 Interpretación de la varianza

De acuerdo a la Ley de los Grandes Números, la esperanza de una v.a. representa el valor aproximado del promedio de un número grande de realizaciones independientes de dicha variable aleatoria. Cabe entonces preguntarse: *¿qué representa la varianza?*

Para responder a la pregunta considere una v.a. X con esperanza finita μ. (Recuerde que μ es una constante que no depende del valor que X tomó o tomará en el experimento.) Si $X_1, X_2, X_3, \ldots$ son realizaciones independientes de X, entonces $(X_1-\mu)^2, (X_2-\mu)^2, (X_3-\mu)^2, \ldots$ son variables aleatorias independientes y con la misma distribución que $(X-\mu)^2$ (*¿por qué?*). La Ley de los Grandes Números implica, por lo tanto, que

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \mathbb{E}((X-\mu)^2) = \mathbb{V}(X), \tag{3.11}$$

con probabilidad uno. Como $\bar{X} \approx \mu$ cuando n es grande, la identidad anterior sugiere que

$$\lim_{n\to\infty} S_X^2 = \mathbb{V}(X), \tag{3.12}$$

donde se define

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

La cantidad de arriba se llama la *varianza muestral* asociada con $X_1, \ldots, X_n$ y mide la fluctuación cuadrática promedio de los datos alrededor de $\bar{X}$. La utilización de cuadrados tiene dos efectos: los términos en la suma son todos de signo positivo (en particular, fluctuaciones positivas y negativas contribuyen igualmente), y fluctuaciones significativas son exageradas en la suma. Por ejemplo, si para un cierto par de datos, $(X_i-\bar{X}) = 2(X_j-\bar{X})$, entonces la fluctuación de X_i alrededor del promedio contribuye a la varianza muestral 4-veces más (en vez de solamente el doble!) que la fluctuación de X_j alrededor de $\bar{X}$. Por otro lado, si $(X_i-\bar{X}) = 5(X_j-\bar{X})$, entonces el término $(X_i-\bar{X})^2$ contribuye 25-veces más que el término $(X_j-\bar{X})^2$.

Comentamos al pasar que en la Sección 4.7 usaremos la varianza muestral para construir un intervalo que, con alta probabilidad, contendrá el valor esperado μ cuando este parámetro es desconocido.

La demostración de (3.12) es como sigue:

$$\begin{aligned}
S_X^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2, \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 + 2(\mu-\bar{X})\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu) + (\mu-\bar{X})^2, \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 + 2(\mu-\bar{X})(\bar{X}-\mu) + (\mu-\bar{X})^2, \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\mu-\bar{X})^2.
\end{aligned}$$

Finalmente, como $\lim_{n\to\infty}\bar{X}=\mu$, la identidad en (3.11) nos permite concluir con relación a lo anterior que $\lim_{n\to\infty} S_X^2 = \mathbb{V}(X)$. Esto demuestra (3.12) y concluimos que: *la varianza de una v.a. X representa la fluctuación cuadrática promedio asintótica de $X_1,\ldots,X_n$ alrededor de $\bar{X}$ i.e. el valor asintótico de la varianza muestral.*

Ejemplo 3.22. Considere el experimento donde se lanza un dado equilibrado donde dos caras tienen marcado el número 2 y las cuatro restantes el número 5. Si X denota el valor de la cara lanzada, entonces $\mathbb{E}(X)=2\cdot 1/3+5\cdot 2/3=4$. En particular, debido a la identidad en la parte (a) del Teorema 3.11, tenemos que $\mathbb{V}(X)=\mathbb{E}(X^2)-16$. Como X^2 es una v.a. que sólo puede tomar los valores 4 y 25 con las probabilidades 1/3 y 2/3 respectivamente, obtenemos que (Teorema 3.3)

$$\mathbb{E}(X^2) = 4\cdot 1/3 + 25\cdot 2/3 = 18.$$

Por lo tanto, $\mathbb{V}(X)=18-16=2$. *¿Qué interpretación tiene este número?* Si lanzaramos el dado independientemente un número grande n de veces y observaramos las caras $X_1,\ldots,X_n$, entonces

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-4)^2 \approx 2,$$

i.e. la distancia cuadrática promedio de los datos alrededor de 4 será aproximadamente igual a 2.

3.7 Ejercicios propuestos

Sección 3.1

Ejercicio 1. Imagine a cierto profesor que al corregir examenes asigna 0, 2, 5, 7 ó 10 puntos por pregunta, y a cierto alumno que a lo largo del semestre ha obtenido la distribución de puntos resumida en la Tabla 3.1. Si el examen final del curso consistirá de 18 preguntas: ¿cuál es el puntaje promedio por pregunta que se espera para el alumno en el examen final?

Puntos en pregunta	0	2	5	7	10
Porcentaje de preguntas	15 %	17 %	40 %	24 %	4 %

Tabla 3.1

Ejercicio 2. Considere una v.a. X con función de distribución dada por la fórmula $F_X(t) = t/(t+1)$ si $t \geq 0$, y $F_X(t) = 0$ si $t < 0$. Determine $\mathbb{E}(X)$ usando la Definición 3.2.

Ejercicio 3. Usando la Definición 3.2, muestre lo siguiente:

(a) Si $\mathbb{P}[X \geq 0] = 1$, entonces $\mathbb{E}(X) = \int_0^\infty \mathbb{P}[X \geq t]\, dt \geq 0$.

(b) Si $\mathbb{P}[X \leq 0] = 1$, entonces $\mathbb{E}(X) = -\int_{-\infty}^0 \mathbb{P}[X \leq t]\, dt \leq 0$.

Ejercicio 4. Demuestre que si X tiene un valor esperado bien definido, entonces $\mathbb{E}(\alpha X) = \alpha \cdot \mathbb{E}(X)$, para toda constante α.
Hint: Use la Definición 3.2 y substituya $t = \alpha \cdot s$.

Ejercicio 5. Usando la Definición 3.2, muestre que si X es una v.a. con valor esperado bien definido y $\mathbb{P}[X \geq c] = 1$, para una cierta constante c, entonces $\mathbb{E}(X) \geq c$. Más generalmente, muestre que si existen constantes reales $a \leq b$ tales que $\mathbb{P}[a \leq X \leq b] = 1$, entonces X tiene un valor esperado finito y $a \leq \mathbb{E}(X) \leq b$.
Hint: Para la primera parte, distinga los casos $c \geq 0$ y $c < 0$ separadamente.

Sección 3.2.

Ejercicio 6. Considere un dado tal que, las caras impares son dos veces más probables que las pares. ¿Cuál es la probabilidad de que el dado salga 1?, ¿qué hay de la cara 2? Use el Teorema 3.3 para determinar el valor esperado de la cara obtenida al lanzar el dado, ¿qué intepretación tiene el número encontrado?

Ejercicio 7. Muestre que si $\mathbb{P}[X{=}i]{=}1/n$, para todo $i \in \{1, \ldots, n\}$, donde $n \geq 1$ es un número entero dado, entonces $\mathbb{E}(X) = (n+1)/2$.
Hint: Use el Teorema 3.3, y que $\sum_{i=1}^n i = n(n+1)/2$.

Ejercicio 8. Determine el valor esperado de la v.a. X tal que

$$F_X(x) = \begin{cases} 0, & x < -3/2; \\ 1/9, & -3/2 \leq x < 0; \\ 2/3, & 0 \leq x < 1/2; \\ 1, & x \geq 1/2. \end{cases}$$

Sección 3.3.

Ejercicio 9. Considere una v.a. X con función de distribución $F(t) = 1 - 1/t^2$ si $t \geq 1$, y $F(t) = 0$ si $t < 1$. Muestre que X es una v.a. aleatoria continua y determine su función de densidad. Use el Teorema 3.4, para determinar el valor esperado de X^k, para cada número entero $k \geq 0$.

Ejercicio 10. Considere una v.a. X con función de densidad $f(x) = |x|e^{-x^2}$. Use el Teorema 3.4 para determinar el valor esperado de $e^{X^2}/(1+X^2)$.

Ejercicio 11. Considere una v.a. continua X con función de densidad

$$f(x) = \begin{cases} 1/(2c), & 0 \leq x \leq c; \\ c^2/x^3, & x > c; \end{cases}$$

donde $c > 0$ es una cierta constante, ¿para qué valor de c se tiene que $\mathbb{E}(X) = 25/4$?

Sección 3.4.

Ejercicio 12. El test de laboratorio para detectar el Virus de Inmuno Deficiencia Humana (VIH), tiene una posibilidad de un 5 % de entregar un resultado Falso-Negativo. (El resultado es Falso-Negativo cuando se reporta, incorrectamente, que la persona testeada no está infectada con el VIH.) Aproximadamente, ¿cuántas personas infectadas con el VIH, entre 1000 que toman este test, obtendrá un resultado Falso-Negativo? Explique los supuestos detrás de su respuesta y en qué medida su respuesta está respaldada por la teoría de probabilidades.

Ejercicio 13. Considere el experimento donde se lanzan independiente e indefinidamente dos dados equilibrados. Use la Ley de los Grandes Números para demostrar que la fracción asintótica de veces que los dados saldrán con la misma cara es $1/6$ con probabilidad uno.

Ejercicio 14. Considere una moneda sesgada que tiene probabilidad p de salir cara. Imagine que lanzamos la moneda indefinidamente. Para cada $i \geq 0$, defina X_{i+1} como el número adicional de lanzamientos a partir de i-ésima cara hasta observar una cara nuevamente. Por ejemplo, si los primeros lanzamientos de la moneda son: sello, sello, sello, sello, cara, sello, sello, cara, cara, etc., entonces $X_1 = 5$, $X_2 = 3$, $X_3 = 1$, etc. Use la Ley de los Grandes Números para mostrar que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{p},$$

con probabilidad uno. ¿Qué representa $\sum_{i=1}^{n} X_i/n$? Basado en su respuesta, ¿es la identidad anterior del todo sorprendente? Explique.

Sección 3.5.

Ejercicio 15. Considere variables aleatorias X e Y tales que $\mathbb{E}(X-Y) = 1$ y $\mathbb{E}(X+2Y) = 4$. Determine el valor esperado de X y de Y.

Ejercicio 16. Muestre que si $\mathbb{P}[X = Y] = 1$ y X tiene una esperanza bien definida, entonces Y también tiene una esperanza bien definida y $\mathbb{E}(Y) = \mathbb{E}(X)$.

Ejercicio 17. Considere una urna donde tres bolitas tienen marcado el número 0, dos el número 1, y una el número 10. (a) Imagine que una bolita se saca al azar de la urna y tiene marcado el número X, y que luego de reponer la bolita en la urna, una segunda bolita se saca también al azar y tiene marcado el número Y. ¿Cuál es el valor esperado de $(X+Y)$? (b) Por otro lado, ¿cuál sería el valor esperado de $(X+Y)$ si la primera bolita no es repuesta en la urna? (c) ¿Son las respuestas a las dos preguntas previas del todo sorprendente? Explique.

Ejercicio 18. Considere el siguiente experimento con dos fases. Primero se lanza un dado equilibrado y se registra el número X de la cara lanzada. Luego, se escoge al azar un número entero Y entre X y 6. Por ejemplo, si $X = 4$, entonces Y puede tomar cualquiera de los valores 4, 5 ó 6, con probabilidad 1/3 cada uno. Sin embargo, si $X = 6$, entonces Y tiene que tomar el valor 6. Determine $\mathbb{E}(X)$, $\mathbb{E}(Y)$ y $\mathbb{E}(XY)$, ¿cómo se compara el producto $\mathbb{E}(X) \cdot \mathbb{E}(Y)$ con $\mathbb{E}(X \cdot Y)$?, ¿contradice lo encontrado el resultado del Teorema 3.9?

Sección 3.6.

Ejercicio 19. Considere una v.a. X tal que $\mathbb{P}[X = 1] = 1 - \mathbb{P}[X = -1] = p$, donde $0 < p < 1$ es una constante dada. Determine $\mathbb{E}(X)$ y $\mathbb{V}(X)$.

Ejercicio 20. Determine la varianza de la v.a. en el Ejercicio 8.

Ejercicio 21. Considere una v.a. $U \sim Uniforme[0,1]$. Determine la varianza de U^k, para cada número entero $k \geq 0$.

Ejercicio 22. Considere dos variables aleatorias X, Y i.i.d. con esperanza (-1) y varianza 3. Determine el valor esperado de $(2X + Y) \cdot (X - 3Y)$.

Problemas misceláneos Capítulo 3.

Ejercicio 23. Considere dos variables aleatorias X e Y tales que

$$\begin{array}{ll} \mathbb{P}[X = 1, Y = 0] = 1/8; & \mathbb{P}[X = 1, Y = 1] = 1/8; \\ \mathbb{P}[X = 2, Y = 0] = 1/4; & \mathbb{P}[X = 2, Y = 1] = 1/2. \end{array}$$

(a) Muestre que $\mathbb{P}[X = 1] = 1/4$, y $\mathbb{P}[X = 2] = 3/4$. Use esto para determinar $\mathbb{E}(X)$.
(b) Muestre que $\mathbb{E}(Y) = 5/8$.

Ejercicio 24. Muestre que para toda constante $c > 0$, la función definida como $f(x) = c/(c+x)^2$ si $x \geq 0$, y $f(x) = 0$ si $x < 0$, es una función de densidad. Muestre que si X es una v.a. con función de densidad f, entonces $\mathbb{E}(X) = +\infty$.

Ejercicio 25. Considere variables aleatorias $X_1, \ldots, X_n$ i.i.d. y uniformes en el intervalo $[0,1]$, y defina $Y = \text{mín}\{X_1, \ldots, X_n\}$.

(a) Muestre que para cada $t \in [0,1]$, $\mathbb{P}[Y \geq t] = (1-t)^n$. Use la Definición 3.2 para determinar el valor esperado de Y.
(b) Determine la función de densidad de Y. Corrobore el valor calculado para $\mathbb{E}(Y)$, usando el Teorema 3.4.

Ejercicio 26. Muestre que si X e Y son variables aleatorias con la misma distribución y X tiene esperanza finita, entonces Y tiene esperanza finita y $\mathbb{V}(X) = \mathbb{V}(Y)$. (Esto muestra que la varianza al igual que la esperaza de una v.a. está solamente determinada por su distribución.)

Ejercicio 27. Considere un juego de azar donde usted y un oponente, escogen independientemente un número uniformemente distribuido en el intervalo $[0, 1]$. Denotaremos como X e Y el número que usted y su oponente escogerán respectivamente. Suponga que usted recibirá $Z = k \cdot |X - Y| - 1$ pesos por cada peso apostado, donde k es una cierta constante.

(a) Muestre y/o explique que, para simular este juego, es suficiente simular un punto (X, Y) uniformemente distribuido en el cuadrado con vértices $(0, 0)$, $(1, 0)$, $(1, 1)$ y $(0, 1)$.

(b) Use lo anterior para mostrar que si $-1 \leq z \leq (k - 1)$, entonces

$$F_Z(z) = 1 - \left(\frac{k-1-z}{k}\right)^2,$$

y deduzca que

$$f_Z(z) = \begin{cases} 0, & z \notin [-1, k-1]; \\ \frac{2(k-1-z)}{k^2}, & z \in [-1; k-1]. \end{cases}$$

(c) ¿Para qué valores de k es este juego más conveniente para usted que para su oponente?

Ejercicio 28. Use la Definición 3.2 para mostrar que si X es una v.a. tal que $\mathbb{P}[X \in \{0, 1, \ldots\}] = 1$, entonces $\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{P}[X \geq n]$.

Ejercicio 29. Use diferenciación y la serie geométrica $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ para demostrar que $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$, para $|x| < 1$. (Esto demuestra la identidad en (3.5).)

Ejercicio 30. Considere variables aleatorias X_1, X_2, X_3 i.i.d. con la misma distribución que una cierta v.a. X tal que $\mathbb{P}[X = i/4] = 1/5$, para cada $i = 0, \ldots, 4$. Determine $\mathbb{E}(\text{mín}\{X_1, X_2, X_3\})$.

Ejercicio 31. Muestre que si X e Y son variables aleatorias independientes con esperanza finita, entonces $\mathbb{V}(XY) = \mathbb{E}(X^2) \cdot \mathbb{E}(Y^2) - \{\mathbb{E}(X)\}^2 \cdot \{\mathbb{E}(Y)\}^2$.

Ejercicio 32. Considere un experimento que tiene asociado un cierto espacio muestral Ω y un suceso $A \subset \Omega$ tal que $\mathbb{P}(A) > 0$. Dada una v.a. $X : \Omega \to \mathbb{R}$ con esperanza finita, la *esperanza condicional de X dado (el suceso) A* es la cantidad definida como (ver identidad en (3.2)):

$$\mathbb{E}(X \mid A) = \frac{\mathbb{E}(X \cdot [\![\text{el suceso } A \text{ ocurre}]\!])}{\mathbb{P}(A)}.$$

Bajo los supuestos anteriores, muestre que

$$\mathbb{E}(X \mid A) = \int_0^\infty \mathbb{P}[X \geq t \mid A]\, dt - \int_{-\infty}^0 \mathbb{P}[X \leq t \mid A]\, dt.$$

Más aún, muestre que si $A_1, \ldots, A_n$ son sucesos disjuntos tales que $\mathbb{P}(A_i) > 0$ y $\Omega = \cup_{i=1}^n A_i$, entonces

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X \mid A_i) \cdot \mathbb{P}(A_i).$$

Ejercicio 33. Considere variables aleatorias discretas X e Y. Muestre que si $A, B \subset \mathbb{R}$ son conjuntos contables tales que $\mathbb{P}[X \in A] = \mathbb{P}[Y \in B] = 1$ y $g : \mathbb{R}^2 \to \mathbb{R}$ es una función dada tal que $\sum_{(a,b)\in A\times B} |g(a,b)| \cdot \mathbb{P}[X = a, Y = b] < \infty$, entonces $g(X, Y)$ es una v.a. discreta con valor esperado finito dado por la fórmula

$$\mathbb{E}(g(X,Y)) = \sum_{(a,b)\in A\times B} g(a,b) \cdot \mathbb{P}[X = a, Y = b].$$

Hint: $A \times B = \{(a,b) : \, a \in A, b \in B\}$.

Ejercicio 34. Demuestre, usando el Ejercicio 33, que si X e Y son variables aleatorias discretas con esperanza finita, entonces $\mathbb{E}(\alpha \cdot X + \beta \cdot Y) = \alpha \cdot \mathbb{E}(X) + \beta \cdot \mathbb{E}(Y)$, para todo par de constantes $\alpha, \beta \in \mathbb{R}$. (Esto demuestra la linealidad de la esperanza en el contexto de variables aleatorias discretas.)

Ejercicio 35. Demuestre usando el Ejercicio 33 que si X e Y son variables aleatorias discretas e independientes, con esperanza finita cada una, entonces $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$. (Esto demuestra, en el contexto de variables aleatorias discretas, que el valor esperado de un producto de variables aleatorias independientes, corresponde al producto de sus valores esperados.)

Ejercicio 36. Considere el siguiente experimento donde un número aleatorio X será escogido en el intervalo $[0,1]$: primero se lanza una moneda equilibrada. Si la moneda sale cara, entonces X se escoge de acuerdo a una v.a. uniforme en el intervalo $[0, 2/3]$. En cambio, si la moneda sale sello, entonces X se escoge de acuerdo a una v.a. uniforme en el intervalo $[1/3, 1]$.

(a) Determine las probabilidades condicionales $\mathbb{P}[X \geq t \mid \text{la moneda salió cara}]$ y $\mathbb{P}[X \geq t \mid \text{la moneda salió sello}]$, para cada $t \in [0,1]$.
(b) Use la Fórmula de Probabilidades Totales para determinar la probabilidad del suceso $\mathbb{P}[X \geq t]$. Aplique esto para determinar el valor esperado de X.
(c) Muestre que $\mathbb{E}(X) = \mathbb{E}(X_1)/2 + \mathbb{E}(X_2)/2$, donde $X_1 \sim Uniforme[0, 2/3]$ y $X_2 \sim Uniforme[1/3, 1]$. Explique por qué esta identidad es, de alguna manera, intuitiva.

Ejercicio 37. Muestre que si $X_1, \ldots, X_n$ son variables aleatorias i.i.d. con valor esperado finito μ y varianza finita σ^2, entonces para toda secuencia de números reales

$\lambda_1, \ldots, \lambda_n$ tales que $\sum_{i=1}^n \lambda_i = 1$ se tiene que

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^n \lambda_i X_i\right) &= \mu, \\
\mathbb{V}\left(\sum_{i=1}^n \lambda_i X_i\right) &= \left(\sum_{i=1}^n \lambda_i^2\right)\sigma^2.
\end{aligned}$$

Se puede demostrar que el factor $\sum_{i=1}^n \lambda_i^2$ anterior es minimizado solamente cuando $\lambda_1 = \cdots = \lambda_n = 1/n$. Por lo tanto, el promedio $\bar{X}$ es la combinación lineal de $X_1, \ldots, X_n$ con mínima varianza.

Ejercicio 38. Imagine un juego de azar donde—por cada peso apostado—su ganancia neta por juego es $(Z-1)$ pesos, donde Z es una cierta v.a. definida por las reglas del juego. Explique por qué la condición $\mathbb{P}[Z > 0] = 0{,}99$ no es suficiente para asegurar que este juego es más conveniente para usted que para su oponente.

Ejercicio 39. Sea X una v.a. tal que $\mathbb{P}[X = a] + \mathbb{P}[X = b] = 1$, y considere variables aleatorias $X_1, X_2, \ldots$ independientes y con la misma distribución que X. Para cada $n \geq 1$ defina A_n (respectivamente B_n) como el número de variables aleatorias $X_1, \ldots, X_n$ que toma el valor a (respectivamente b) i.e. $A_n = \sum_{i=1}^n [\![X_i = a]\!]$ y $B_n = \sum_{i=1}^n [\![X_i = b]\!]$. (a) Use la Ley de los Grandes Números para demostrar que $\lim_{n\to\infty} A_n/n = \mathbb{P}[X = a]$ y $\lim_{n\to\infty} B_n/n = \mathbb{P}[X = b]$. (b) Aplique nuevamente la Ley de los Grandes Números para demostrar alternativamente que $\mathbb{E}(X) = a \cdot \mathbb{P}[X = a] + b \cdot \mathbb{P}[X = b]$.

Ejercicio 40. Considere una v.a. X tal que $\mathbb{P}[X \in \{0, 1, 2, \ldots\}] = 1$. La *función generadora de momentos de* X es la función definida como $\varphi_X(z) = \mathbb{E}(z^X)$, para cada $0 \leq z \leq 1$. Muestre las siguientes propiedades: (i) $\varphi_X(1) = 1$. (ii) Si $a, b \in \mathbb{R}$ son constantes dadas, entonces $\varphi_{aX+b}(z) = z^b \cdot \varphi_X(z^a)$. (iii) Si X e Y tienen la misma distribución, entonces $\varphi_X = \varphi_Y$. (iv) Si $X_1, \ldots, X_n$ son variables aleatorias independientes, entonces $\varphi_{\sum_{i=1}^n X_i} = \prod_{i=1}^n \varphi_{X_i}$. Use esto para concluir que si $X_1, \ldots, X_n$ son i.i.d., entonces $\varphi_{\sum_{i=1}^n X_i} = (\varphi_{X_1})^n$.

Ejercicio 41. Muestre que si $\mathbb{P}[X \in \{0, 1, 2, \ldots\}] = 1$, entonces la función generadora de momento de X satisface la identidad $\varphi_X(z) = \sum_{k=0}^\infty z^k \cdot \mathbb{P}[X = k]$ (ver Ejercicio 40). Concluya que $\mathbb{E}(X) = \varphi_X'(1)$.

Ejercicio 42. Si X e Y son dos variables aleatorias con esperanza finita, entonces su *covarianza* es la cantidad definida como

$$\operatorname{cov}(X, Y) = \mathbb{E}(\{X - \mathbb{E}(X)\} \cdot \{Y - \mathbb{E}(Y)\}),$$

siempre y cuando la v.a. $\{X - \mathbb{E}(X)\} \cdot \{Y - \mathbb{E}(Y)\}$ tenga un valor esperado bien definido. Muestre las siguientes propiedades: (a) Si X e Y tienen varianza finita, entonces la covarianza entre X e Y es también finita. (b) Si X e Y tienen una covarianza bien definida, entonces $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$ y, para toda constante real α, $\operatorname{cov}(\alpha \cdot X, Y) =$

$\mathrm{cov}(X, \alpha \cdot Y) = \alpha \cdot \mathrm{cov}(X, Y)$. (c) Si X e Y son variables aleatorias independientes con esperanza finita, entonces $\mathrm{cov}(X, Y) = 0$.
Hint: Para demostrar la parte (a) use que $|x \cdot y| \leq (x^2 + y^2)/2$, para todo $x, y \in \mathbb{R}$.

Ejercicio 43. (a) Muestre que si X, Y, Z son variables aleatorias tales que $\mathrm{cov}(X, Z)$, $\mathrm{cov}(Y, Z)$, y $\mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$ son cantidades bien definidas, entonces $\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$. Más generalmente, muestre que si $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ son variables aleatorias tales que la covarianza entre X_i e Y_j es finita para todo i, j, entonces

$$\mathrm{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{cov}(X_i, Y_j).$$

Ejercicio 44. Muestre que si $X_1, \ldots, X_n$ son variables aleatorias con esperanza y varianza finita, entonces

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2 \cdot \sum_{1 \leq i < j \leq j} \mathrm{cov}(X_i, X_j).$$

Explique por qué el Teorema 3.12 es un caso especial de este resultado.
Hint: Use la identidad en (3.9) y los ejercicios 42 y 43.

Ejercicio 45. Use la Ley de los Grandes Números para justificar que

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}) \cdot (Y_i - \bar{Y}) = \mathrm{cov}(X, Y),$$

donde $X_1, X_2, \ldots$ e $Y_1, Y_2, \ldots$ son realizaciones independientes de X e Y, respectivamente.

Ejercicio 46. La *desigualdad de Chebyshev* establece que si X es una v.a. con esperanza finita μ y varianza finita σ^2, entonces

$$\mathbb{P}\big[|X - \mu| \geq t \cdot \sigma\big] \leq \frac{1}{t^2},$$

para todo $t > 0$. Use esto para demostrar que

$$\mathbb{P}\big[|X - \mu| < t \cdot \sigma\big] \geq \left(1 - \frac{1}{t^2}\right).$$

(En palabras, la probabilidad de que X tome un cierto valor a t desviaciones estándares de su valor esperado es al menos $(1 - 1/t^2)$. En particular, si X es una v.a. con una desviación estándar pequeña, entonces X es altamente probable de tomar un cierto valor cerca de su valor esperado. La desviación estándar de una variable aleatoria mide, por lo tanto, cuán concentrada es la v.a. alrededor de su valor esperado.)

Ejercicio 47. (a) Considere variables aleatorias $X_1, \ldots, X_n$ con esperanza μ y varianza σ^2 finitas. ¿Cuál es el t más pequeño para el cual usted puede garantizar que el intervalo $[\bar{X} - t \cdot \sigma, \bar{X} + t \cdot \sigma]$ contiene el valor μ con una probabilidad mayor o igual a 0,95? **Hint:** Aplique la desigualdad de Chebyshev (Ejercicio 46) con la v.a. $\bar{X}$. (b) Para apreciar una aplicación concreta de lo anterior, imagine un alumno que obtuvo el promedio final de notas $\bar{X} = 5{,}9764$ en cierto curso. Suponiendo que las notas del alumno son variables aleatorias independientes e idénticamente distribuidas con varianza $\sigma^2 = 0{,}09$, ¿cuál es el número mínimo de notas consideradas en el promedio para el cual el profesor del curso estaría al menos un 95 % seguro que la nota intrínseca del alumno en su curso es mayor o igual a 6.0?

Ejercicio 48. (a) Use la desigualdad de Chebyshev (Ejercicio 46) para demostrar que si X es una v.a. con valor esperado y varianza finita, entonces para todo $\epsilon > 0$, la siguiente desigualdad aplica:

$$\mathbb{P}\Big[|X - \mathbb{E}(X)| \geq \epsilon\Big] \leq \frac{\mathbb{V}(X)}{\epsilon^2}.$$

(b) Aplique lo anterior para demostrar el siguiente resultado. Note que, de acuerdo al teorema, para cualquier margen de error $\epsilon > 0$, la probabilidad de que $\bar{X}$ tome un valor a más de ϵ-unidades de su valor esperado, es cada vez más pequeña a medida que se incluyen más datos en el promedio.

Teorema 3.13. *(Ley Débil de los Grandes Números.) Si $X_1, X_2, \ldots$ son variables aleatorias i.i.d. con valor esperado μ y varianza σ finita, entonces, para todo $\epsilon > 0$, se tiene lo siguiente:*

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right] = 0.$$

Comentarios de final de capítulo

En lo que sigue, $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio probabilístico dado.

En cierto sentido, la teoría de la medida es una teoría sobre integración en espacios abstractos. De hecho, el *valor esperado* de una v.a. $X : \Omega \to \mathbb{R}$ corresponde a la integral $\int_\Omega X\, d\mathbb{P}$, siempre y cuando esta integral esté bien definida. Dos maneras alternativas de reescribir esta integral son:

$$\int_\Omega X\, d\mathbb{P} = \int_{\mathbb{R}} t\, d\mathbb{P}_X = \int_{\mathbb{R}} t\, dF_X(t).$$

Por ejemplo, si X es una v.a. discreta y $A \subset \mathbb{R}$ es un conjunto contable tal que $\mathbb{P}[X \in A] = 1$, entonces $\int_\Omega X\, d\mathbb{P} = \sum_{a\in A} a \cdot \mathbb{P}[X = a]$, siempre y cuando la serie sea absolutamente convergente. Cuando A es finito, X es lo que se llama una *función simple*, y la identidad dada aplica por definición. El caso con A contable, es consecuencia de uno de los teoremas destacados de la teoría de la medida: el *Teorema de Convergencia Dominada.*

Por otro lado, si X es una v.a. continua con función de densidad f entonces, el Teorema de Fubini para integrales con respecto a medidas, implica que $\int_\Omega X\, d\mathbb{P} = \int_{\mathbb{R}} t \cdot f(t)\, dt$, siempre y cuando esta última integral sea absolutamente convergente.

Las propiedades de la esperanza son propiedades heredadas a partir de la definición de integral. Por ejemplo, en términos de integrales, la linealidad de la esperanza es una consecuencia inmediata de la linealidad de la integral: si $\alpha, \beta \in \mathbb{R}$ son constantes dadas, entonces

$$\int_\Omega (\alpha \cdot X + \beta \cdot Y)\, d\mathbb{P} = \alpha \cdot \int_\Omega X\, d\mathbb{P} + \beta \cdot \int_\Omega Y\, d\mathbb{P},$$

siempre y cuando, la cantidad al lado derecho de lo anterior esté bien definida; en particular, las integrales $\int_\Omega X\, d\mathbb{P}$ y $\int_\Omega Y\, d\mathbb{P}$ tienen que estar bien definidas. Más aún, si $X(\omega) \leq Y(\omega)$, para todo $\omega \in \Omega$, entonces $\int_\Omega X\, d\mathbb{P} \leq \int_\Omega Y\, d\mathbb{P}$. Finalmente, si X e Y son independientes y además $X \geq 0$ e $Y \geq 0$, o bien $\int_\Omega |X \cdot Y|\, d\mathbb{P} < +\infty$, entonces $\int_\Omega X \cdot Y\, d\mathbb{P} = \int_\Omega X\, d\mathbb{P} \cdot \int_\Omega Y\, d\mathbb{P}$. La Ley de los Grandes Números es en gran parte una consecuencia de las tres propiedades anteriores.

Capítulo 4: Distribuciones Destacadas y el Teorema Central Límite

En esta sección introduciremos y estudiaremos las propiedades principales de seis tipos de variables aleatorias (distribuciones) que surgen recurrentemente en experimentos. Las primeras tres que estudiaremos i.e. del tipo Bernoulli, geométricas, y binomiales, son casos especiales de variables aleatorias discretas que describen, respectivamente, la ocurrencia, primera ocurrencia, y el número total de ocurrencias de un cierto suceso al repetir independientemente un experimento. El cuarto y último tipo de variable aleatoria discreta que estudiaremos es la del tipo Poisson. Esta distribución describe el número de ocurrencias de un suceso extremadamente improbable al repetir muchas veces e independientemente un cierto experimento. Veremos e.g. que el número de gotitas de agua que sacudirán una sección del suelo durante un chaparrón de agua es una variable aleatoria del tipo Poisson.

Entre las variables aleatorias continuas que estudiaremos se encuentran las del tipo exponencial y del tipo normal. La distribución exponencial usualmente describe la separación temporal entre eventos de la misma naturaleza e.g. llamadas consecutivas de telefóno o el tiempo de espera entre el último y el próximo e-mail que una persona recibirá. Por otro lado, la distribución normal es particularmente importante debido al Teorema Central Límite. De acuerdo a este celebrado resultado, el promedio de un número grande de variables aleatorias independientes y con la misma distribución (i.e. variables i.i.d.) tendrá aproximadamente una distribución normal. Esto aplica sorprendentemente, tanto para un promedio de variables aleatorias discretas, como continuas.

4.1 Distribución de Bernoulli

En esta sección p es un número real dado en el intervalo $[0,1]$.

Definición 4.1. Una v.a. X se dice tener una *distribución del tipo Bernoulli* con parámetro p si $\mathbb{P}[X = 1] = p$ y $\mathbb{P}[X = 0] = (1-p)$. En este caso, anotaremos $X \sim Bernoulli(p)$.

Observe que

$$\mathbb{P}[X \in \{0,1\}] = \mathbb{P}[X = 0] + \mathbb{P}[X = 1] = (1-p) + p = 1.$$

X es, por lo tanto, una v.a. discreta. En particular, la esperanza y la varianza de X vienen dadas por (Teorema 3.3)

$$\mathbb{E}(X) = 1 \cdot \mathbb{P}[X = 1] + 0 \cdot \mathbb{P}[X = 0] = p, \tag{4.1}$$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2 = \mathbb{E}(X) - p^2 = p(1 - p). \tag{4.2}$$

Para el cálculo de la varianza, hemos usado que $X^2 = X$ debido a que $X \in \{0, 1\}$ con probabilidad uno.

Ejemplo 4.1. Un ejemplo típico de una v.a. del tipo Bernoulli es la función indicadora de un suceso A, definida como (ver la identidad en (3.2)):

$$X = [\![\text{el suceso } A \text{ ocurre}]\!] = \begin{cases} 1, & \text{si el suceso } A \text{ ocurre;} \\ 0, & \text{en caso contrario.} \end{cases}$$

En este caso, $X \sim \mathit{Bernoulli}(p)$, con $p = \mathbb{P}(A)$. Por ejemplo, si el experimento consiste en escoger una persona al azar y A es el suceso "la persona nació en un año bisiesto", entonces la función indicadora de A es una v.a. del tipo Bernoulli con parámetro $p = 1/4$.

Por otro lado, si B es el suceso "la persona nació un 29 de febrero", entonces la función indicadora de B es una v.a. del tipo Bernoulli con parámetro

$$\begin{aligned} p &= \mathbb{P}\left[\begin{array}{c} \text{la persona nació} \\ \text{un 29 de febrero} \end{array} \middle| \begin{array}{c} \text{la persona nació} \\ \text{un año bisiesto} \end{array} \right] \cdot \mathbb{P}\left[\begin{array}{c} \text{la persona nació} \\ \text{un año bisiesto} \end{array} \right], \\ &= 1/366 \cdot 1/4 = 1/1464. \end{aligned}$$

Ejemplo 4.2. Considere una v.a. X que sólo puede tomar los valores (-1) y 7. *Reescriba X en términos de una v.a. del tipo Bernoulli.*

Observe que $(-1) \cdot [\![X = -1]\!] + 7 \cdot [\![X = 7]\!]$ es igual a (-1) cuando $X = -1$ y a 7 cuando $X = 7$. Por lo tanto, $X = (-1) \cdot [\![X = -1]\!] + 7 \cdot [\![X = 7]\!]$. Pero note que $[\![X = 7]\!] = 1 - [\![X = -1]\!]$ (*¿por qué?*). Así, obtenemos que

$$\begin{aligned} X &= (-1) \cdot [\![X = -1]\!] + 7 \cdot [\![X = 7]\!], \\ &= (-1) \cdot [\![X = -1]\!] + 7 \cdot (1 - [\![X = -1]\!]) = 7 - 8 \cdot [\![X = -1]\!]. \end{aligned}$$

Para corroborar la última identidad anterior, note que el lado derecho es $(7 - 8)$ cuando $X = -1$, y $(7 - 0)$ cuando $X = 7$. Esto concluye el ejemplo ya que $[\![X = -1]\!]$ i.e. la indicadora del suceso $[X = -1]$ es del tipo Bernoulli.

Ejemplo 4.3. Si $X \sim \mathit{Bernoulli}(7/10)$, $Y \sim \mathit{Bernoulli}(1/4)$ y la probabilidad del suceso $[X = 1 \text{ ó } Y = 1]$ es $7/8$, responda: *¿cuál es la distribución de XY?*

Recuerde que tanto X como Y sólo pueden tomar el valor 0 ó 1; en particular, $XY = 0$ cuando $X = 0$ ó $Y = 0$, y $XY = 1$ cuando $X = 1$ e $Y = 1$. La v.a. XY es consecuentemente del tipo Bernoulli con parámetro $p = \mathbb{P}[X = 1, Y = 1]$. Para determinar p note que de acuerdo a la Fórmula de Inclusión-Exclusión tenemos (Ejercicio 22 en el Capítulo 1):

$$\begin{aligned} \mathbb{P}[X = 1 \text{ ó } Y = 1] &= \mathbb{P}[X = 1] + \mathbb{P}[Y = 1] - \mathbb{P}[X = 1, Y = 1], \\ 7/8 &= 7/10 + 1/4 - p. \end{aligned}$$

Por lo tanto, $p = 3/40$ i.e. $XY \sim Bernoulli(3/40)$.

4.2 Distribución Geométrica

En esta sección p es un número real dado en el intervalo $(0, 1]$.

Definición 4.2. Una v.a. X se dice tener una *distribución geométrica* con parámetro p si $\mathbb{P}[X = k] = (1-p)^{k-1}p$, para $k = 1, 2, \ldots$ En este caso, anotaremos $X \sim Geométrica(p)$.

Se deduce de la definición que una v.a. geométrica sólo puede tomar valores enteros mayores o iguales a uno. Esto se debe al siguiente cálculo basado en la serie geométrica (ver indentidad en (1.7)):

$$\begin{aligned}\mathbb{P}[X \in \{1,2,\ldots\}] = \sum_{k=1}^{\infty} \mathbb{P}[X = k] &= p \cdot \sum_{k=1}^{\infty} (1-p)^{k-1}, \\ &= p \cdot \sum_{k=0}^{\infty} (1-p)^{k} = p \cdot \frac{1}{1-(1-p)} = 1.\end{aligned}$$

Ejemplo 4.4. Considere una v.a. $X \sim Geométrica(2/5)$, *¿cuál es la probabilidad del suceso* $[X \geq 11]$*?* Como X sólo puede tomar valores enteros no negativos, la respuesta a la pregunta viene dada por

$$\begin{aligned}\mathbb{P}[X \geq 11] = \sum_{k=11}^{\infty} \left(\frac{3}{5}\right)^{k-1} \cdot \left(\frac{2}{5}\right) &= \left(\frac{2}{5}\right) \cdot \sum_{k=10}^{\infty} \left(\frac{3}{5}\right)^{k}, \\ &= \left(\frac{2}{5}\right) \cdot \sum_{k=0}^{\infty} \left(\frac{3}{5}\right)^{k+10}, \\ &= \left(\frac{2}{5}\right) \cdot \left(\frac{3}{5}\right)^{10} \cdot \sum_{k=0}^{\infty} \left(\frac{3}{5}\right)^{k}, \\ &= \left(\frac{2}{5}\right) \cdot \left(\frac{3}{5}\right)^{10} \cdot \frac{1}{1-3/5} = \left(\frac{3}{5}\right)^{10}.\end{aligned}$$

Las variables aleatorias geométricas ocurren en una amplia variedad de situaciones debido a lo siguiente: *si repetimos un experimento independientemente hasta observar un cierto suceso que tiene probabildad p de ocurrir, entonces el número de experimentos necesarios hasta observar dicho suceso tiene una distribución geométrica con parámetro p.*

Ejemplo 4.5. Para fijar ideas, considere el experimento que consiste en lanzar un dado hasta observar por primera vez la cara 3 ó 4, y la v.a. X que registra el número de lanzamientos hasta que esto ocurre. Por ejemplo, si la secuencia de caras es $5, 3, \ldots$ entonces $X = 2$. Por otro lado, si la secuencia de caras es $6, 6, 2, 6, 4, \ldots$ entonces $X = 5$. Debido al comentario anterior, debiera tenerse que $X \sim Geométrica(1/3)$.

Para justificar esto considere las variables aleatorias definidas como sigue

$$X_i = [\![\text{el dado sale un 3 ó un 4 en el } i\text{-ésimo lanzamiento}]\!].$$

Por ejemplo, si la secuencia de caras es $5, 3, \ldots$ entonces $X_1 = 0$ y $X_2 = 1$. Por otro lado, si la secuencia de caras es $6, 6, 2, 6, 4, \ldots$ entonces $X_1 = X_2 = X_3 = X_4 = 0$ y $X_5 = 1$. En ambos casos, X corresponde a índice i más pequeño tal que $X_i = 1$. Equivalentemente, la siguiente identidad aplica

$$X = \text{mín}\{i \geq 1 : X_i = 1\}.$$

Observe que $X_1, X_2, \ldots$ son del tipo Bernoulli con parámetro $1/3$ (*¿por qué?*). Como $X_1, X_2, \ldots$ son independientes, la demostración que X es una v.a. geométrica es ahora una consecuencia del siguiente resultado.

Teorema 4.3. *Si $X_1, X_2, \ldots$ son variables aleatorias i.i.d. del tipo Bernoulli con parámetro común $p > 0$, entonces $X = \text{mín}\{i \geq 1 : X_i = 1\} \sim Geométrica(p)$.*

Demostración. En efecto, dado $k \geq 1$ y usando que $X_1, \ldots, X_k$ son variables aleatorias independientes, se tiene que

$$\begin{aligned}
\mathbb{P}[X = k] &= \mathbb{P}[X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1], \\
&= \mathbb{P}[X_1 = 0] \cdots \mathbb{P}[X_{k-1} = 0] \cdot \mathbb{P}[X_k = 1], \\
&= (1-p) \cdots (1-p) \cdot p, \\
&= (1-p)^{k-1} p,
\end{aligned}$$

lo que demuestra el teorema. $\square$

Ejemplo 4.6. Considere el experimento donde se lanzan simultáneamente dos dados idénticos y equilibrados, hasta que los dados forman una serie aritmética al ser ordenados de menor a mayor, *¿cuál es el mínimo número de lanzamientos para el cual hay una probabilidad de al menos un 50 % de observar una serie aritmética?*

Denotemos como X el número de lanzamientos necesarios hasta obtener una serie aritmética. Para fijar ideas note que si los lanzamientos de los dados son de acuerdo a la Tabla 4.1, entonces $X = 5$ i.e. $X = \text{mín}\{i \geq 1 : X_i = 1\}$, donde $X_i = [\![\text{una serie aritmética es observada en el } i\text{-ésimo lanzamiento}]\!]$. Claramente $X_1, X_2, \ldots$ son del tipo Bernoulli. Más aún, tenemos que

$$\begin{aligned}
\mathbb{P}[X_i = 1] &= \sum_{j=1}^{5} \mathbb{P}\left[\begin{array}{c} \text{dado } 1 = j, \\ \text{dado } 2 = j+1 \end{array}\right] + \sum_{j=1}^{5} \mathbb{P}\left[\begin{array}{c} \text{dado } 1 = j+1, \\ \text{dado } 2 = j \end{array}\right], \\
&= \sum_{j=1}^{5} \frac{1}{36} + \sum_{j=1}^{5} \frac{1}{36} = \frac{5}{18}.
\end{aligned}$$

Por lo tanto, cada $X_i \sim Bernoulli(5/18)$. Como $X_1, X_2, \ldots$ son independientes, el Teorema 4.3 implica que $X \sim Geométrica(5/18)$. En particular, por ejemplo, la probabilidad que una serie aritmética sea observada por primera vez en el quinto lanzamiento está dada por

$$\mathbb{P}[X=5]=\left(\frac{13}{18}\right)^4\cdot\frac{5}{18}=\frac{142805}{1889568}\approx 7{,}5\,\%.$$

lanzamiento	#1	#2	#3	#4	#5
dado 1	2	4	6	3	5
dado 2	6	1	6	5	4
¿serie aritmética?	no	no	no	no	sí

Tabla 4.1. Resultados posibles al lanzar simultáneamente dos dados y hasta observar una serie aritmética entre sus caras.

Para responder a la pregunta que motivó el análisis anterior, necesitamos determinar el número entero $k\geq 1$ más pequeño tal que $\mathbb{P}[X\leq k]\geq 1/2$. Como $[X\leq k]$ corresponde a la unión finita y disjunta de los sucesos $[X=1],\ldots,[X=k]$, obtenemos que

$$\mathbb{P}[X\leq k]=\sum_{n=1}^{k}\left(\frac{13}{18}\right)^{n-1}\cdot\frac{5}{18}=\frac{5}{18}\sum_{n=0}^{k-1}\left(\frac{13}{18}\right)^{n}=\frac{5}{18}\cdot\frac{1-\left(\frac{13}{18}\right)^k}{1-\frac{13}{18}}=1-\left(\frac{13}{18}\right)^k.$$

Como la función de distribución de toda v.a. es una función creciente, la probabilidad anterior es una función creciente de k. Como $\mathbb{P}[X\leq 2]=0{,}478...$ y $\mathbb{P}[X\leq 3]=0{,}623...$, concluimos que al menos 3 lanzamientos son necesarios para observar una serie aritmética con una chance de, al menos, un 50 %.

Si $X\sim \mathit{Geométrica}(p)$, entonces encontramos que

$$\mathbb{E}(X)=\frac{1}{p}; \tag{4.3}$$

$$\mathbb{V}(X)=\frac{1}{p}\left(\frac{1}{p}-1\right). \tag{4.4}$$

Para demostrar las identidades anteriores usaremos (ver identidad en (3.5)):

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

Observe que, si diferenciamos ambos lados de arriba con respecto a la variable x, obtenemos que

$$\sum_{k=2}^{\infty}k(k-1)x^{k-2}=\frac{2}{(1-x)^3},\ \text{para todo } |x|<1.$$

Usando la primera serie, obtenemos que

$$\mathbb{E}(X)=\sum_{k=1}^{\infty}k\cdot\mathbb{P}[X=k]=p\cdot\sum_{k=1}^{\infty}k\cdot(1-p)^{k-1}=p\cdot\frac{1}{(1-(1-p))^2}=\frac{1}{p},$$

lo que demuestra (4.3). Por otro lado, la segunda serie, implica que

$$\begin{aligned}\mathbb{E}(X\cdot(X-1)) &= \sum_{k=1}^{\infty}k(k-1)\cdot\mathbb{P}[X=k],\\ &= p(1-p)\sum_{k=2}^{\infty}k(k-1)(1-p)^{k-2},\\ &= p(1-p)\cdot\frac{2}{(1-(1-p))^3}=\frac{2}{p}\left(\frac{1}{p}-1\right).\end{aligned}$$

En particular, debido a la identidad $\mathbb{V}(X)=\mathbb{E}(X\cdot(X-1))-\mathbb{E}(X)\cdot\{\mathbb{E}(X)-1\}$ (Teorema 3.11), obtenemos finalmente que

$$\mathbb{V}(X)=\frac{2}{p}\left(\frac{1}{p}-1\right)-\frac{1}{p}\left(\frac{1}{p}-1\right)=\frac{1}{p}\left(\frac{1}{p}-1\right).$$

Esto demuestra la identidad en (4.4).

Ejemplo 4.7. Un niño colecciona las estampillas de las cartas que llegan a su casa. Asumiendo que hay en total diez tipos de estampillas diferentes y que cada carta tiene la misma probabilidad de ser estampada con una de éstas, *¿cuál es el valor esperado del número de cartas despachadas a la casa del niño hasta que éste coleccione todas las estampillas?*

Denotemos como X la v.a. que cuenta el número de cartas despachadas a la casa del niño hasta que éste forme la colección completa de estampillas. Para determinar $\mathbb{E}(X)$ explotaremos la linealidad de la esperanza. Para esto observe que si X_i es el número adicional de estampillas para que el niño aumente su colección de $(i-1)$ a i estampillas, entonces

$$X=\sum_{i=1}^{10}X_i.$$

Para fijar ideas indexaremos las diferentes estampillas usando las letras a, b, c, d, e, f, g, h, i, j. Si las estampillas que el niño colecciona llegaran a su casa en el siguiente orden (letras subrayadas corresponden a estampillas que aumentan la colección): $\underline{c}$, c, $\underline{a}$, $\underline{b}$, a, c, b, $\underline{f}$, $\underline{g}$, a, f, $\underline{d}$, f, f, f, c, $\underline{h}$, a, b, g, g, $\underline{j}$, g, a, $\underline{e}$, c, a, f, g, a, $\underline{i}$, entonces $X=31$ y, por otro lado, $X_1=1$, $X_2=2$, $X_3=1$, $X_4=4$, $X_5=1$, $X_6=3$, $X_7=5$, $X_8=5$, $X_9=3$, y $X_{10}=6$. El lector podrá fácilmente verificar que la identidad anterior aplica en este caso.

Observe que $X_1=1$ siempre; en particular, $X\sim \mathit{Geométrica}(1)$.

Por otro lado, $X_2\sim \mathit{Geométrica}(9/10)$. Esto se debe a que cada carta despachada, después de la primera, tiene una chance de $9/10$ de contribuir con una estampilla nueva a la colección del niño. Similarmente, una vez que el niño ha aumentado su colección de una a dos estampillas, cada carta despachada a partir de ese momento, tiene una chance de 8/10 de contribuir con una estampilla nueva. Por lo tanto, $X_3 \sim \mathit{Geométrica}(8/10)$.

Siguiendo la misma lógica anterior, se tiene que $X_i \sim \mathit{Geométrica}((11-i)/10)$. Esto se debe a que, una vez que la colección del niño consiste de $(i-1)$ estampillas, cada carta adicional tiene una chance de $(10-(i-1))/10$ de contribuir con una estampilla nueva a la colección. El hecho que X_i es una v.a. geométrica es ahora una consecuencia directa del Teorema 4.3.

Usando (4.3) y la linealidad de la esperanza, finalmente obtenemos que

$$\mathbb{E}(X) = \sum_{i=1}^{10} \mathbb{E}(X_i) = \sum_{i=1}^{10} \frac{10}{11-i} = \frac{7381}{252} \approx 29,$$

i.e. en promedio cada niño que colecciona estampillas requerirá alrededor de 29 despachos para completar la colección.

4.3 Distribución Binomial

Recuerde que si $n \geq 0$ es un número entero, entonces n *factorial* es la cantidad definida como $n! = 1$ si $n = 0$, y $n! = \prod_{i=1}^{n} i$ si $n > 0$. Por ejemplo, $3! = 1 \cdot 2 \cdot 3 = 6$ y $5! = 3! \cdot 4 \cdot 5 = 6 \cdot 20 = 120$.

Por otro lado, los *coeficientes binomiales* son los números definidos como

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

donde $0 \leq k \leq n$ son números enteros. Para un k y un n dado, este coeficiente representa el número de subconjuntos de cardinalidad k que pueden construirse a partir de un conjunto dado de cardinalidad n. Por ejemplo, $\binom{4}{1} = 4$ ya que $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, son los únicos subconjuntos de cardinalidad 1 del conjunto $\{a, b, c, d\}$. Por otro lado, $\binom{4}{2} = 6$ ya que el conjunto $\{a, b, c, d\}$ contiene solamente los siguientes subconjuntos de cardinalidad 2: $\{a, b\}$, $\{a, c\}$, $\{a, d\}$, $\{b, c\}$, $\{b, d\}$, y $\{c, d\}$. (El lector podrá repasar el Capítulo 4 en [**4**] para una discusión más detallada de estos coeficientes).

En esta sección $n \geq 1$ es un número entero y $p \in [0, 1]$ es un número real dado.

Definición 4.4. Una v.a. X se dice tener una *distribución binomial* con parámetros (n, p), si $\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$, para cada $k = 0, 1, \ldots, n$. En este caso, anotaremos $X \sim \mathit{Binomial}(n, p)$.

Usando para todo $a, b \in \mathbb{R}$ y número entero $n \geq 0$ (*Fórmula Binomial*):

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

obtenemos que

$$\mathbb{P}[X \in \{0,1,\ldots,n\}] = \sum_{k=0}^{n} \mathbb{P}[X=k] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \{p + (1-p)\}^n = 1.$$

En particular, X es una v.a. discreta.

La variables aleatorias del tipo binomial, al igual que las geométricas, están íntimamente ligadas con las variables del tipo Bernoulli: *si un cierto suceso tiene probabilidad p en un experimento y repetimos el experimento independientemente n veces, entonces el número total X de veces que el suceso será observado tiene una distribución del tipo binomial con parámetros (n, p).*

Ejemplo 4.8. Para fijar ideas considere el experimento que consiste en escoger 10 personas al azar en el centro de Santiago, a las cuales se les pregunta el día de su nacimiento. El número de personas nacidas un día 29 de febrero es una v.a. binomial con parámetros $(10, p)$, donde p es la probabilidad que una persona escogida al azar haya nacido un día 29 de febrero. Note que $p = 1/4 \cdot 1/366 = 1/1464$, ya que sólo uno de cada cuatro años es bisiesto y un año bisiesto consiste de 366 días.

Por otro lado, si escogiéramos 100 personas al azar entonces el número de personas nacidas un día 29 de febrero es una v.a. binomial con parámetros $(100, p)$. Más generalmente, si escogemos un número fijo, pero predeterminado n de personas al azar, entonces el número X de ellas nacidas un día 29 de febrero tendrá una distribución binomial con parámetros (n, p). Esto se debe a la siguiente identidad y al Teorema 4.5:

$$X = \sum_{i=1}^{n} X_i,$$

donde $X_1, \ldots, X_n$ son las variables aleatorias independientes del tipo Bernoulli, con parámetro común $p = 1/1464$, definidas como

$$X_i = [\![\text{la } i\text{-ésima persona escogida nació un 29 de febrero}]\!].$$

Teorema 4.5. *Si $X_1, \ldots, X_n$ son variables aleatorias i.i.d. del tipo Bernoulli con parámetro p, entonces $X = \sum_{i=1}^{n} X_i \sim Binomial(n, p)$.*

Demostración. Claramente, $X \in \{0, 1, \ldots, n\}$. Para determinar la probabilidad del suceso $[X = k]$, con $0 \leq k \leq n$ enteros, observe que

$$\mathbb{P}[X = k] = \sum_{(x_1,\ldots,x_n)} \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n],$$

donde los índices $(x_1, \ldots, x_n)$ en la sumatoria son n-tuplas de ceros y unos tales que $\sum_{i=1}^{n} x_i = k$. Note, sin embargo, que el factor p ocurre en el producto $\mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n]$ tantas veces como un uno aparece en $(x_1, \ldots, x_n)$ i.e. el factor p ocurre k veces. Como el factor $(1-p)$ tiene entonces que ocurrir $(n-k)$ veces en dicho producto, obtenemos que

$$\mathbb{P}[X=k] = p^k(1-p)^{n-k} \cdot \sum_{(x_1,\ldots,x_n)} 1 = \binom{n}{k} p^k(1-p)^{n-k}.$$

Para la última identidad hemos usado que hay $\binom{n}{k}$ n-tuplas de ceros y unos tales que $\sum_{i=1}^{n} x_i = k$. Esto se debe a que la transformación $h(x_1,\ldots,x_n) = \{i : x_i = 1\}$ es una bijección entre las n-tuplas que satisfacen esta propiedad y los subconjuntos de cardinalidad k del conjunto $\{1,\ldots,n\}$ (*¿por qué?*). Esto completa la demostración del teorema. □

Usando el Teorema 4.5 podemos determinar fácilmente el valor esperado y la varianza de una v.a. del tipo binomial. En efecto, usando (4.1) y (4.2) deducimos que si $X \sim Binomial(n,p)$, entonces

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} p = np;\\
\mathbb{V}(X) &= \mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p).
\end{aligned}$$

Ejemplo 4.9. Considere un cierto colegio con 70 alumnos en Educación Básica y 50 en Educación Media. Si la probabilidad de que un alumno de Básica olvide su corbata es 1/100, mientras que esta misma probabilidad para un alumno de Media es 1/60, *¿cuál es la probabilidad que en un cierto día ningún alumno olvide su corbata? Más aún, determine la esperanza y varianza del número de alumnos que olvidará su corbata ese día.*

Denotaremos como X e Y, respectivamente, el número de alumnos en Básica y en Media que olvidará su corbata un cierto día. Debido al Teorema 4.5, y bajo el supuesto que los diferentes alumnos olvidan sus corbatas de forma independiente, tenemos que $X \sim Binomial(70, 1/100)$ e $Y \sim Binomial(50, 1/60)$. Note, sin embargo, que $(X+Y)$ no es una v.a. binomial ya que la probabilidad del suceso "el alumno olvidará su corbata" no es la misma entre los alumnos de Básica que de Media.

La respuesta a la pregunta anterior viene dada por la probabilidad del suceso $[X+Y=0]$. Como X e Y sólo pueden tomar valores no negativos, esta probabilidad corresponde a la probabilidad del suceso $[X=0, Y=0]$. En particular, debido al supuesto que los alumnos olvidan sus corbatas, independientemente el uno del otro, tenemos que

$$\begin{aligned}
\mathbb{P}[X+Y=0] &= \mathbb{P}[X=0, Y=0],\\
&= \mathbb{P}[X=0] \cdot \mathbb{P}[Y=0],\\
&= \binom{70}{0}\left(\frac{1}{100}\right)^0\left(\frac{99}{100}\right)^{70} \cdot \binom{50}{0}\left(\frac{1}{60}\right)^0\left(\frac{59}{60}\right)^{50} = 0{,}21\ldots
\end{aligned}$$

En particular, alrededor de cuatro de cada cinco días, al menos un alumno olvidará su corbata. Esta probabilidad es mucho más grande que la probabilidad que un alumno específico olvide su corbata, debido al gran número de alumnos en el colegio. De hecho, y para completar el ejemplo, observe que

$$\begin{aligned}\mathbb{E}(X+Y) &= \mathbb{E}(X)+\mathbb{E}(Y) = 70\cdot 1/100 + 50\cdot 1/60 \approx 1{,}53;\\ \mathbb{V}(X+Y) &= \mathbb{V}(X)+\mathbb{V}(Y) = 70\cdot 1/100\cdot 99/100 + 50\cdot 1/60\cdot 59/60 \approx 1{,}51;\end{aligned}$$

donde, para la primera identidad de la segunda fila, hemos usado que X e Y son independientes (Teorema 3.12). Así tenemos que en un mes típico, alrededor de 30 ($\approx 1{,}53\cdot 20$) alumnos olvidarán su corbata.

Ejemplo 4.10. El consejo de profesores de un colegio, compuesto por 6 docentes, decidirá la suerte de aquellos alumnos con un promedio final de notas inferior pero cercano a 4.0: cada alumno necesitará al menos 5 votos a favor para pasar de curso. Si cada profesor toma la decisión correcta sobre un alumno el 75 % de las veces, y el 90 % de los alumnos en esta situación debiera repetir de curso, *¿cuál es la probabilidad de que el consejo tome la decisión correcta sobre un alumno?*

La respuesta a la pregunta es la probabilidad

$$\begin{aligned}&\mathbb{P}\begin{bmatrix}\text{el consejo toma la}\\ \text{decisión correcta}\end{bmatrix}\\ &\quad= \mathbb{P}\begin{bmatrix}\text{el alumno debiera}\\ \text{pasar de curso}\end{bmatrix}\cdot\mathbb{P}\left[\begin{matrix}\text{el consejo decide}\\ \text{aprobar al alumno}\end{matrix}\,\middle|\,\begin{matrix}\text{el alumno debiera}\\ \text{pasar de curso}\end{matrix}\right]\\ &\qquad+\mathbb{P}\begin{bmatrix}\text{el alumno debiera}\\ \text{repetir de curso}\end{bmatrix}\cdot\mathbb{P}\left[\begin{matrix}\text{el consejo decide}\\ \text{reprobar al alumno}\end{matrix}\,\middle|\,\begin{matrix}\text{el alumno debiera}\\ \text{repetir de curso}\end{matrix}\right],\\ &\quad= \frac{1}{10}\cdot\mathbb{P}\left[\begin{matrix}\text{el consejo decide}\\ \text{aprobar al alumno}\end{matrix}\,\middle|\,\begin{matrix}\text{el alumno debiera}\\ \text{pasar de curso}\end{matrix}\right]\\ &\qquad+\frac{9}{10}\cdot\mathbb{P}\left[\begin{matrix}\text{el consejo decide}\\ \text{reprobar al alumno}\end{matrix}\,\middle|\,\begin{matrix}\text{el alumno debiera}\\ \text{repetir de curso}\end{matrix}\right].\end{aligned}$$

Denotemos como X la v.a. que cuenta el número de profesores que piensan que el alumno debe ser promovido, cuando éste realmente tendría que pasar de curso. Como cada docente toma la decisión correcta el 75 % de las veces, el Teorema 4.5 implica que $X\sim Binomial(6,3/4)$. Aquí estamos suponiendo que cada docente decide independientemente del resto. Más aún, observe que

$$\mathbb{P}\left[\begin{matrix}\text{el consejo decide}\\ \text{aprobar al alumno}\end{matrix}\,\middle|\,\begin{matrix}\text{el alumno debiera}\\ \text{pasar de curso}\end{matrix}\right] = \mathbb{P}[X\geq 5].$$

Por otra parte, si Y denota el número de profesores que piensan que el alumno debe reprobar, cuando éste debiera repetir de curso, entonces, bajo los mismos supuestos anteriores, tendremos que $Y\sim Binomial(6;3/4)$ y que

$$\mathbb{P}\left[\begin{matrix}\text{el consejo decide}\\ \text{reprobar al alumno}\end{matrix}\,\middle|\,\begin{matrix}\text{el alumno debiera}\\ \text{repetir de curso}\end{matrix}\right] = \mathbb{P}[Y\geq 2].$$

Obtenemos, por lo tanto, que

$$\mathbb{P}\left[\begin{array}{l}\text{el consejo toma la}\\ \text{decisión \quad correcta}\end{array}\right] = \frac{1}{10}\cdot\sum_{k=5}^{6}\binom{6}{k}\left(\frac{3}{4}\right)^k\left(\frac{1}{4}\right)^{6-k} + \frac{9}{10}\cdot\sum_{k=2}^{6}\binom{6}{k}\left(\frac{3}{4}\right)^k\left(\frac{1}{4}\right)^{6-k},$$
$$= 243/256 \approx 95\,\%.$$

La probabilidad calculada resulta interesante: aunque cada profesor toma la decisión correcta sobre un alumno, sólo el 75 % de las veces, la decisión colectiva del consejo es prácticamente infalible.

Ejemplo 4.11. Remontémonos a los tiempos de la Colonia donde los rumores se propagaban verbalmente a través de la sociedad. Imagine un rumor que en un cierto momento es sólo conocido por el $100p\,\%$ de la sociedad, donde $0 < p < 1$ es una constante dada. Por ejemplo, si $p = 1/100$, entonces $100p\,\% = 1\,\%$; y si $p = 1/10$, entonces $100p\,\% = 10\,\%$. *Si Martín El Chismoso ya sabe del rumor, en promedio, ¿cuántas personas se enterarán del rumor directamente a través Martín y después que éste se ha encontrado con n personas al azar?*

Si X denota el número de personas que se enterarán del rumor por primera vez debido a los n encuentros de Martín, y bajo el supuesto que éste compartirá el rumor con cada persona que se encuentra, tenemos que $X \sim Binomial(n, 1-p)$ (*¿por qué?*). En particular, en promedio, $n(1-p)$ personas se enterarán del rumor gracias a Martín y después que éste se ha encontrado con n personas al azar.

Ejemplo 4.12. En el Ejemplo 4.11, *¿cómo cambia la respuesta si Martín El Chismoso no sabía orgininalmente del rumor?* (Si Martín se entera del rumor en los n encuentros, entonces lo incluiremos en el cálculo.)

Denotaremos como Y el número de personas que se enterarán del rumor debido a los n encuentros de Martín. La respuesta a la pregunta viene dada por $\mathbb{E}(Y)$. Claramente, $Y \in \{0, \ldots, n\}$. Por lo tanto,

$$\mathbb{E}(Y) = \sum_{k=0}^{n} k\cdot\mathbb{P}[Y = k] = \sum_{k=1}^{n} k\cdot\mathbb{P}[Y = k]. \tag{4.5}$$

La gran diferencia entre este ejemplo y el anterior es que el número de personas que se enterarán del rumor directamente a través de Martín, depende del número de encuentros que le tomará a éste enterarse del rumor. Si N denota este número y $1 \leq k \leq n$, entonces $1 \leq N \leq (n - k + 1)$ cuando $Y = k$. Usando la Fórmula de Probabilidades Totales obtenemos, por lo tanto, que

$$\mathbb{P}[Y = k] = \sum_{m=1}^{n-k+1} \mathbb{P}[Y = k \mid N = m]\cdot\mathbb{P}[N = m].$$

Pero, observe que $\mathbb{P}[N = m] = (1-p)^{m-1}\cdot p$ (*¿por qué?*). Por otro lado, si $N = m$ entonces, para tener $Y = k$, exactamente $(k-1)$ de las últimas $(n-m)$ personas que se encontrarán con Martín, escucharán del rumor por primera vez gracias a él. En particular, cuando $N = m$, la probabilidad condicional del suceso $[Y = k]$ es igual a la probabilidad que una v.a. binomial con parámetros $(n-m, 1-p)$ tome el valor $(k-1)$. En símbolos, $\mathbb{P}[Y = k \mid N = m] = \binom{n-m}{k-1}(1-p)^{k-1}p^{n-m-k+1}$. Así encontramos a partir de la identidad en (4.5) que

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{k=1}^{n} k \cdot \sum_{m=1}^{n-k+1} \binom{n-m}{k-1}(1-p)^{k-1}p^{n-m-k+1} \cdot (1-p)^{m-1} \cdot p, \\ &= \sum_{k=1}^{n} \sum_{m=1}^{n-k+1} k \cdot \binom{n-m}{k-1}(1-p)^{k+m-2}p^{n-m-k+2}.\end{aligned}$$

Aunque no simplificaremos la expresión anterior, es interesante cuantificar lo que establece. En la Figura 4.1 presentamos $\mathbb{E}(Y-1)$ y $\mathbb{E}(X)$ como funciones de p cuando $n = 10$. (Substraer uno tiene sentido porque Y, a diferencia de X, incluye a Martín en el cálculo.) Como es de esperarse $\mathbb{E}(Y-1) < \mathbb{E}(X)$, para todo valor de $p < 1$. Sin embargo, para valores pequeños de p, note que la velocidad de propagación del rumor es substancialmente menor cuando Martín no lo conoce desde un principio. Por ejemplo, si $p = 0{,}05$, entonces $\mathbb{E}(Y-1) \approx 1{,}55$, mientras que $\mathbb{E}(X) = 9{,}5$; en particular, el rumor se propagará alrededor de seis veces más rápido una vez que llegue a los oídos de un chismoso.

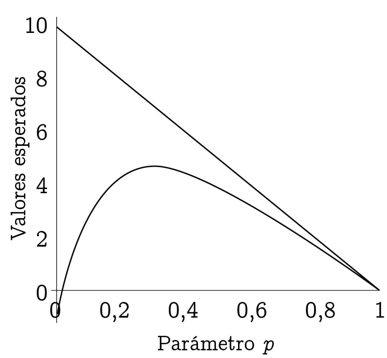


FIGURA 4.1. Valores esperados en función de p de las variables aleatorias X e $(Y-1)$ de los ejemplos 4.11 y 4.12 cuando $n = 10$. La función lineal corresponde a $\mathbb{E}(X)$.

4.4 Distribución de Poisson

En esta sección $\lambda > 0$ es un número real dado.

Definición 4.6. Una v.a. X se dice tener una *distribución del tipo Poisson* con parámetro λ si $\mathbb{P}[X = k] = \lambda^k e^{-\lambda}/k!$, para $k = 0, 1, 2, \ldots$ En este caso, anotaremos $X \sim \mathit{Poisson}(\lambda)$.

Usando la serie de Taylor de la función exponencial:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad x \in \mathbb{R}, \tag{4.6}$$

obtenemos que

$$\mathbb{P}[X \in \{0,1,2,\ldots\}] = \sum_{k=0}^{\infty} \mathbb{P}[X=k] = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Por lo tanto, X es una v.a. discreta, ya que sólo puede tomar valores enteros no-negativos. Más aún, su esperanza y varianza vienen dadas por

$$\mathbb{E}(X) = \lambda, \tag{4.7}$$

$$\mathbb{V}(X) = \lambda. \tag{4.8}$$

Ambas identidades resultan de la aplicación de la serie de Taylor de la función exponencial. En efecto,

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \cdot \mathbb{P}[X=k] = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda,$$

lo que demuestra la identidad en (4.7). Por otro lado, tenemos que (Teorema 3.11):

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X \cdot (X-1)) - \mathbb{E}(X) \cdot \{\mathbb{E}(X) - 1\}, \\
&= \sum_{k=0}^{\infty} k(k-1) \cdot \mathbb{P}[X=k] - \lambda(\lambda-1), \\
&= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} - \lambda(\lambda-1), \\
&= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} - \lambda(\lambda-1), \\
&= \lambda^2 e^{-\lambda} e^{\lambda} - \lambda(\lambda-1) = \lambda.
\end{aligned}$$

Esto demuestra la identidad en (4.8).

Ejemplo 4.13. *Determine el valor esperado de λ^{-X} si $X \sim Poisson(\lambda)$.* En efecto, como $\mathbb{P}[X \in \{0,1,2,\ldots\}] = 1$, tenemos que (Teorema 3.3)

$$\mathbb{E}(\lambda^{-X}) = \sum_{k=0}^{\infty} \lambda^{-k} \cdot \mathbb{P}[X=k] = \sum_{k=0}^{\infty} \lambda^{-k} \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} = e^{1-\lambda},$$

donde para la última identidad hemos usado la serie en (4.6).

Ejemplo 4.14. Considere una v.a. $X \sim Poisson(1)$, *¿cuál de los siguientes sucesos es más probable: $[X \text{ es par}]$ o $[X \text{ es impar}]$?*

Note primero que $\mathbb{P}[X = k] = e^{-1}/k!$, para $k = 0, 1, 2, \ldots$ Más aún, como estos son los únicos valores que X puede tomar, entonces

$$\begin{aligned}
\mathbb{P}[X \text{ es par}] &= \sum_{k=0}^{\infty} \mathbb{P}[X = 2k] = \sum_{k=0}^{\infty} \frac{1}{e \cdot (2k)!}; \\
\mathbb{P}[X \text{ es impar}] &= \sum_{k=0}^{\infty} \mathbb{P}[X = 2k+1], \\
&= \sum_{k=0}^{\infty} \frac{1}{e \cdot (2k+1)!} = \sum_{k=0}^{\infty} \frac{1}{e \cdot (2k)!} \cdot \frac{1}{2k+1}.
\end{aligned}$$

Como cada término en la serie de la primera fila es mayor que en la serie de la segunda fila, el suceso $[X \text{ es par}]$ es el más probable.

Las variables aleatorias del tipo Poisson son usadas para modelar en una serie de situaciones de la vida cotidiana. Por ejemplo, el número de gotitas de agua que sacuden una cierta sección del suelo durante un chaparrón de lluvia, o el número de llamadas telefónicas a una cierta oficina en un intervalo de tiempo dado, se modelan usando esta distribución. Para entender las razones detrás de esto, veremos primero cómo aproximar las probabilidades de sucesos asociados a una variable del tipo binomial usando una v.a. del tipo Poisson. El resultado principal es el siguiente [**6**].

Teorema 4.7. *Si $X \sim Binomial(n, p)$ entonces, para todo conjunto $A \subset \mathbb{N}$, se tiene que $|\mathbb{P}[X \in A] - \mathbb{P}[Z \in A]| \leq np^2$, donde $Z \sim Poisson(np)$. Equivalentemente, aplica que*

$$\left| \mathbb{P}\left[X \in A\right] - \sum_{k \in A} \frac{(np)^k}{k!} \cdot e^{-np} \right| \leq np^2.$$

Una consecuencia interesante del teorema anterior, es el siguiente resultado, el que será útil para aproximar probabilidades asociadas a una variable con distribución binomial, usando una variable del tipo Poisson.

Teorema 4.8. *Si $X_1, X_2, \ldots$ son variables aleatorias independientes tales que $X_n \sim Binomial(n, p_n)$ y $\lim_{n \to \infty} np_n = \lambda$, donde $0 < \lambda < +\infty$ es una constante, entonces, para todo número entero $k \geq 0$, se tiene que*

$$\lim_{n \to \infty} \mathbb{P}[X_n = k] = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Demostración. De acuerdo al Teorema 4.7, si $A = \{k\}$, entonces

$$\left| \mathbb{P}[X_n = k] - \frac{(np_n)^k}{k!} \cdot e^{-np_n} \right| \leq np_n^2,$$

para cada n. El teorema es ahora una consecuencia directa del siguiente hecho:

$$\lim_{n \to \infty} np_n^2 = \lim_{n \to \infty} \frac{(np_n)^2}{n} = \lambda^2 \cdot \lim_{n \to \infty} \frac{1}{n} = 0,$$

lo que completa la demostración. □

Note que bajo los supuestos del teorema anterior se tiene que

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} \frac{n \cdot p_n}{n} = \lambda \cdot \lim_{n\to\infty} \frac{1}{n} = 0.$$

En particular, para aproximar probabilidades asociadas a una variable del tipo binomial, usando una variable del tipo Poisson, necesitamos un n grande y un p pequeño. Bajo este supuesto, la probabilidad del suceso $[X = k]$, con $X \sim Binomial(n, p)$, puede aproximarse razonablemente por la probabilidad del suceso $[Z = k]$, con $Z \sim Poisson(np)$. El error de esta aproximación es np^2, para todo k según lo establecido en el Teorema 4.7.

Ejemplo 4.15. El profesor jefe de un cierto curso, traspasará la lista de teléfonos de emergencia de sus 25 alumnos a una planilla electrónica en su computador. Si cada teléfono consiste de nueve dígitos (el pimer dígito distingue los números fijos de los móbiles) y la probabilidad que el profesor ingrese incorrectamente un cierto dígito es 1/50, *¿cuál es, aproximadamente, la probabilidad que éste traspase correctamente todos los dígitos?*

Note que el profesor tendrá que ingresar 225 $(= 9 \cdot 25)$ dígitos en el computador. La respuesta a la pregunta viene dada por la probabilidad del suceso $[X = 0]$, donde X es el número de dígitos que el profesor ingresará incorrectamente. Más aún, bajo el supuesto que los errores de tipeo ocurren independientemente, tenemos que $X \sim Binomial(225, 1/50)$ (*¿por qué?*). Por lo tanto, de acuerdo al Teorema 4.7, si $Z \sim Poisson(9/2)$, entonces

$$\mathbb{P}[X = 0] \approx \mathbb{P}[Z = 0] = \frac{(9/2)^0}{0!} e^{-9/2} = e^{-9/2} = 0{,}011...$$

El error máximo en esta aproximación es $np^2 = 0{,}09$, en particular, al menos el primer dígito en la aproximación, es correcto. De hecho, uno encuentra que la probabilidad exacta viene dada por

$$\mathbb{P}[X = 0] = \binom{225}{0} \left(\frac{1}{50}\right)^0 \left(\frac{49}{50}\right)^{225} = \left(\frac{49}{50}\right)^{225} = 0{,}010...$$

En particular, la aproximación es correcta en los primeros dos dígitos.

Ejemplo 4.16. Si mil personas se escogen al azar, *¿cuál es, aproximadamente, la probabilidad que exactamente una o dos de ellas hayan nacido un día 29 de febrero?*

Si denotamos como X la v.a. que cuenta el número de las personas nacidas un día 29 de febrero, entonces $X \sim Binomial(1000, 1/1464)$. La respuesta a la pregunta anterior viene entonces dada por

$$\begin{aligned} \mathbb{P}[X \in \{1, 2\}] &= \mathbb{P}[X = 1] + \mathbb{P}[X = 2], \\ &= \sum_{k=1}^{2} \binom{1000}{k} \cdot \left(\frac{1}{1464}\right)^k \cdot \left(1 - \frac{1}{1464}\right)^{1000-k}. \end{aligned}$$

Es poco claro en esta expresión, cuán probable o improbable es el suceso $[X \in \{1,2\}]$. Observe, sin embargo, que X tiene una distribución binomial con un parámetro $n = 1000$ grande y $p = 1/1464$ pequeño. Por lo tanto, y de acuerdo al Teorema 4.8, si aproximamos probabilidades asociadas a la v.a. X usando una variable aleatoria $Z \sim Poisson(1000/1464)$, entonces incurriremos en un error menor o igual a

$$np^2 = 0{,}0004...$$

Específicamente, aplica que $|\mathbb{P}[X \in \{1,2\}] - \mathbb{P}[Z \in \{1,2\}]| \leq 0{,}0004...$ i.e. las probabilidades $\mathbb{P}[X \in \{1,2\}]$ y $\mathbb{P}[Z \in \{1,2\}]$ tienen que ser idénticas en los primeros tres dígitos. Como

$$\mathbb{P}[Z \in \{1,2\}] = \mathbb{P}[Z=1] + \mathbb{P}[Z=2] = \sum_{k=1}^{2} \frac{(1000/1464)^k}{k!} e^{-1000/1464} = 0{,}4628...$$

concluimos que la probabilidad de que, exactamente, una o dos personas entre mil escogidas al azar hayan nacido un día 29 de febrero es 46,2... %.

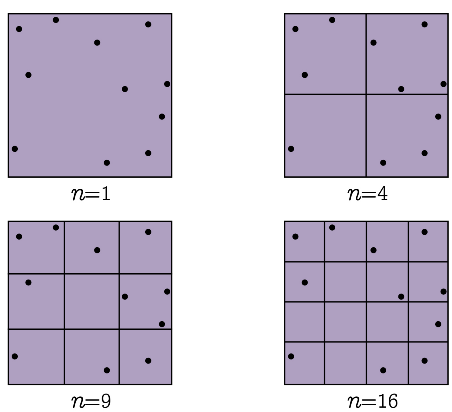


Figura 4.2. División de una sección del suelo en $n = 1, 4, 9, 16$ secciones más pequeñas. Observe cómo los rectángulos de la división contendrán, a lo más, una gotita de agua cuando $n \geq 16$ en este caso.

El siguiente y último ejemplo justificará el uso de la distribución de Poisson para modelar el número de gotitas de agua que sacudirán una cierta sección del suelo durante un chaparrón de agua.

Ejemplo 4.17. Considere el experimento donde se cuenta el número de gotitas N de agua que sacudirán una cierta sección rectangular del suelo de área A, durante un intervalo de tiempo de duración τ. Si en promedio caen g gotitas de agua por unidad de área y de tiempo, entonces

$$N \sim Poisson(g \cdot A \cdot \tau). \tag{4.9}$$

Por ejemplo, si $A = 10\ cm^2$, $\tau = 2\ seg$ y $g = 0{,}1\ gotitas/cm^2/seg$ (i.e. en promedio, se necesitan 10 segundos para que una gotita de agua sacuda una sección dada del suelo de área 1 cm^2), entonces $N \sim Poisson(2)$. Sin embargo, si $g = 3\ gotitas/cm^2/seg$, entonces $N \sim Poisson(60)$. Destacamos que los parámetros A, τ y g pueden medirse en otras unidades siempre y cuando la cantidad $gA\tau$ sea un escalar sin unidades.

La justificación del modelo probabilístico en (4.9) se basa en los siguientes supuestos intuitivos:

(1) para todo $n \geq 1$, si $R_1, \ldots, R_n$ son secciones disjuntas en el suelo, entonces los sucesos "al menos una gotita de agua sacudió la sección R_i", con $i \in \{1, \ldots, n\}$, son independientes, y
(2) si R es una sección del suelo, entonces la probabilidad del suceso "al menos una gotita de agua sacudió a R" depende solamente del área de R.

En lo que sigue, supondremos que n es un cuadrado perfecto i.e. $n = 1, 4, 9, 16, \ldots$ Supongamos que dividimos la sección rectangular en el suelo en n regiones rectangulares más pequeñas y disjuntas $R_{1,n}, \ldots, R_{n,n}$ con área A/n, respectivamente (ver Figura 4.2). Para cada $1 \leq i \leq n$ considere la v.a. del tipo Bernoulli definida como

$$Y_{i,n} = [\![\text{al menos una gotita de agua sacudió la sección } R_{i,n}]\!].$$

Si n es suficientemente grande entonces $\sum_{i=1}^{n} Y_{i,n}$ será una buena aproximación de N. De hecho, si la distancia más pequeña entre dos gotitas que sacudieron la sección del suelo es $d > 0$ y el diámetro de cada rectángulo $R_{i,n}$ es estrictamente menor que d (lo que aplica para todo n suficientemente grande), entonces cada rectángulo $R_{1,n}, \ldots, R_{n,n}$ contendrá, a lo más, una gotita. Por lo tanto, la siguiente identidad aplica:

$$N = \sum_{i=1}^{n} Y_{i,n}, \text{ para todo } n \text{ suficientemente grande.}$$

En particular, resulta intuitivo que

$$\mathbb{P}[N = k] = \lim_{n\to\infty} \mathbb{P}\left[\sum_{i=1}^{n} Y_{i,n} = k\right].$$

Ahora observe que, debido al supuesto (1), $Y_{1,n}, \ldots, Y_{n,n}$ son independientes. Por otro lado, debido al supuesto (2), la probabilidad de que una cierta región sea sacudida por al menos una gotita de agua depende solamente del área de dicha región. En particular, como $R_{1,n}, \ldots, R_{n,n}$ son regiones de áreas idénticas, $Y_{1,n}, \ldots, Y_{n,n}$ son variables aleatorias del tipo Bernoulli con un mismo parámetro p_n. Por lo tanto, su suma es una v.a. del tipo binomial con parámetros (n, p_n). El Teorema 4.8 y la identidad anterior implican, por lo tanto, que

$$\mathbb{P}[N = k] = \frac{L^k}{k!} \cdot e^{-L}, \tag{4.10}$$

siempre y cuando exista una constante $0 < L < +\infty$ tal que

$$\lim_{n\to\infty} np_n = L. \tag{4.11}$$

Demostraremos que éste es efectivamente el caso. En efecto, recuerde que p_n es la probabilidad que una cierta sección rectangular de área A/n es sacudida por, al menos, una gotita de agua en un intervalo de tiempo de duración τ. En particular, se tiene que

$$\begin{aligned} p_1 = \mathbb{P}\left(\bigcup_{i=1}^{n}[Y_{i,n}=1]\right) &= 1-\mathbb{P}\left(\bigcap_{i=1}^{n}[Y_{i,n}=0]\right), \\ &= 1-\prod_{i=1}^{n}\mathbb{P}[Y_{i,n}=0] = 1-\prod_{i=1}^{n}(1-p_n) = 1-(1-p_n)^n. \end{aligned}$$

Para la tercera identidad, hemos usado $Y_{1,n},\ldots,Y_{n,n}$ son variables aleatorias independientes y, para la cuarta identidad, que cada una de las regiones $R_{1,n},\ldots,R_{n,n}$ tienen área A/n. Obtenemos, de este modo, que $p_n = 1-(1-p_1)^{1/n}$, y en particular:

$$\lim_{n\to\infty} np_n = \lim_{n\to\infty} n(1-(1-p_1)^{1/n}) = \lim_{n\to\infty}\frac{1-(1-p_1)^{1/n}}{1/n} = \lim_{x\to 0^+}\frac{1-(1-p_1)^x}{x}.$$

Finalmente, usando la Fórmula de L'Hopital, obtenemos que

$$\begin{aligned} \lim_{n\to\infty} np_n &= \lim_{x\to 0^+}\frac{\frac{d}{dx}[1-(1-p_1)^x]}{\frac{d}{dx}[x]}, \\ &= \lim_{x\to 0^+}\frac{-\ln(1-p_1)\cdot(1-p_1)^x}{1} = -\ln(1-p_1). \end{aligned}$$

El cálculo anterior demuestra (4.11), ya que $0 < p_1 < 1$; en particular, y de acuerdo a la identidad en (4.10), $N \sim Poisson(L)$. Pero observe que $\mathbb{E}(N) = L$. Como en promedio caen g gotitas de agua por unidad de área y tiempo, el valor más plausible para L es $gA\tau$, lo que justifica la identidad en (4.9).

El modelo dado en (4.9) aplicará a cualquier sección del suelo de área A y en cualquier intervalo de tiempo de duración τ durante el cual llueva con la misma firmeza i.e. durante el cual λ permanezca constante. Una consecuencia interesante de esta observación es la siguiente: *la probabilidad de que una sección del suelo de área A no sea sacudida por ninguna gotita de agua, durante los primeros t segundos de un chaparrón, corresponde a la probabilidad del suceso $[Z_t = 0]$, con $Z_t \sim Poisson(g\cdot A\cdot t)$.* En particular,

$$\mathbb{P}[Z_t = 0] = \frac{(gAt)^0}{0!}e^{-gAt} = e^{-gAt}.$$

Vemos que la probabilidad de que una cierta sección del suelo no sea sacudida por la lluvia decrece exponencialmente a medida que el tiempo pasa. Esto explica por qué durante un chaparrón suficientemente largo, casi todas las secciones del suelo serán sacudidas por agua.

La discusión del Ejemplo 4.17 aplica para otros fenómenos donde la distribución de Poisson también ocurre naturalmente. Por ejemplo, si en una cierta oficina se registran en promedio λ llamadas telefónicas por unidad de tiempo, entonces, en un intervalo de tiempo de duración t, debieran registrarse λt llamadas en promedio. En particular, bajo los supuestos que

(1) el número de llamadas en intervalos de tiempo disjuntos son variables aleatorias independientes, y
(2) la probabilidad que al menos una llamada suceda en un cierto intervalo de tiempo depende solamente de su duración,

el número aleatorio N_t de llamadas telefónicas en un cierto intervalo de tiempo de duración t, será una v.a. del tipo Poisson con parámetro λt.

4.5 Distribución Exponencial

En esta sección $\lambda > 0$ es un número real dado.

Definición 4.9. Una v.a. X se dice tener una *distribución del tipo exponencial* con parámetro λ, si es una v.a. continua con función de densidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

En este caso, anotaremos $X \sim Exponencial(\lambda)$.

Ejemplo 4.18. Considere una v.a. $X \sim Exponencial(2)$. Dado $t \geq 0$, *¿cuál es la probabilidad que X tome un cierto valor mayor o igual a t?* La respuesta a la pregunta corresponde a la probabilidad

$$\mathbb{P}[X \geq t] = \int_t^{\infty} 2e^{-2x}\,dx = -e^{-2x}\Big|_{x=t}^{x=+\infty} = e^{-2t}.$$

Note, más generalmente, que si $X \sim Exponencial(\lambda)$, entonces

$$\mathbb{P}[X \geq t] = e^{-\lambda t}. \tag{4.12}$$

En particular, la probabilidad que X exceda un cierto valor decae exponencialmente con dicho valor. La identidad anterior será clave para establecer la llamada propiedad desmemoriada de la distribución exponencial, que discutiremos al final de la sección.

Ejemplo 4.19. Si X es una cierta variable aleatoria, tal que $\mathbb{P}[X = 1] = 1/3$ y $\mathbb{P}[X = 2] = 2/3$, e $Y \sim Exponencial(1)$ es independiente de X, responda: *¿cuál es la probabilidad del suceso $[Y \geq X]$?* Para determinar la probabilidad de este suceso aplicaremos primero la Fórmula de Probabilidades Totales (Teorema 1.3) condicionando sobre el valor que X tomará, y luego usaremos la independencia entre X e Y:

$$\begin{aligned} \mathbb{P}[Y \geq X] &= \mathbb{P}[Y \geq X, X = 1] + \mathbb{P}[Y \geq X, X = 2] \\ &= \mathbb{P}[Y \geq 1, X = 1] + \mathbb{P}[Y \geq 2, X = 2] \\ &= \mathbb{P}[Y \geq 1] \cdot \mathbb{P}[X = 1] + \mathbb{P}[Y \geq 2] \cdot \mathbb{P}[X = 2], \\ &= e^{-1} \cdot 1/3 + e^{-2} \cdot 2/3 \approx 21{,}3\,\%, \end{aligned}$$

donde, para la última igualdad, hemos usado la identidad en (4.12).

La esperanza y varianza de una v.a. exponencial X con parámetro λ vienen dadas por

$$\mathbb{E}(X) = 1/\lambda, \tag{4.13}$$

$$\mathbb{V}(X) = 1/\lambda^2. \tag{4.14}$$

Para demostrar las identidades anteriores, usaremos el *Método de Integración por Partes*. Para esto recuerde que si $a < b$ son números reales dados y $f, g : [a, b] \to \mathbb{R}$ son funciones continuas en el intervalo cerrado $[a, b]$, y continuamente diferenciables en el intervalo abierto (a, b), entonces

$$\int_a^b f(x) \cdot g'(x)\, dx = f(x) \cdot g(x)\Big|_a^b - \int_a^b f'(x) \cdot g(x)\, dx. \tag{4.15}$$

Para demostrar (4.13) observe que

$$\begin{aligned}
\mathbb{E}(X) = \int_0^\infty x \cdot \lambda e^{-\lambda x} dx &= \int_0^\infty x \cdot (-e^{-\lambda x})' dx, \\
&= (-xe^{-\lambda x})\Big|_0^{+\infty} + \int_0^\infty x' \cdot e^{-\lambda x} dx = \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda},
\end{aligned}$$

lo que muestra (4.13). Por otro lado, tenemos que

$$\begin{aligned}
\mathbb{E}(X^2) = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x} dx &= \int_0^\infty x^2 \cdot (-e^{-\lambda x})' dx, \\
&= \int_0^\infty (x^2)' \cdot e^{-\lambda x} dx, \\
&= 2 \int_0^\infty x \cdot e^{-\lambda x} dx, \\
&= \frac{2}{\lambda} \int_0^\infty x \cdot \lambda e^{-\lambda x} dx = \frac{2}{\lambda} \cdot \mathbb{E}(X) = \frac{2}{\lambda^2}.
\end{aligned}$$

Concluimos, por lo tanto, que

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2},$$

lo que demuestra (4.14).

Las variables aleatorias del tipo exponencial están íntimamente ligadas a las variables aleatorias del tipo Poisson. Para entender esta conexión, y de acuerdo a la discusión al final de la Sección 4.4, recordemos que si en promedio ocurren $\lambda > 0$ eventos (e.g. sismos o llamadas telefónicas) por unidad de tiempo, entonces el número de eventos en un cierto intervalo de duración $t > 0$ tiene una distribución del tipo Poisson con parámetro λt.

Mientras la distribución de Poisson describe el número de eventos que ocurren en un intervalo de tiempo dado, la distribución exponencial describe la separación temporal entre dos eventos sucesivos (ver Figura 4.3). En efecto, si usamos T para

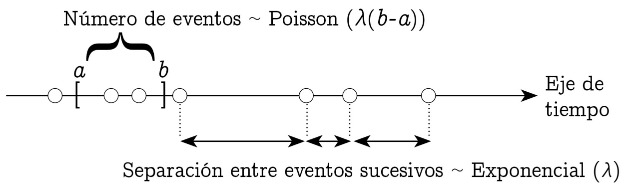


FIGURA 4.3. La distribución de Poisson, describe el número de eventos que ocurren en un intervalo de tiempo dado. Por ejemplo, la probabilidad de observar sólo dos eventos en el intervalo de tiempo $[a, b]$ es $\lambda^2(b-a)^2e^{-\lambda(b-a)}/2!$, cuando en promedio ocurren λ eventos por unidad de tiempo. Por otro lado, la distribución exponencial describe la separación temporal entre eventos consecutivos.

denotar la separación temporal entre el último y el próximo evento, entonces, dado un $t > 0$, se tiene que

$$F_T(t) = \mathbb{P}[T \leq t] = 1 - \mathbb{P}[T > t] = 1 - \mathbb{P}[N = 0],$$

donde N es el número de eventos que ocurrirán en el intervalo de tiempo $[0, t]$. Usando que $N \sim Poisson(\lambda t)$, se deduce de la identidad anterior que

$$F_T(t) = 1 - \frac{\lambda^0 e^{-\lambda t}}{0!} = 1 - e^{-\lambda t}.$$

Por lo tanto,

$$F_T(t) = \begin{cases} 0, & t < 0; \\ 1 - e^{-\lambda t}, & t \geq 0. \end{cases} \tag{4.16}$$

Como la función anterior es continuamente diferenciable sobre $\mathbb{R}$, excepto en el punto $t = 0$, T tiene una función de densidad dada por la fórmula

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0; \\ 0, & t < 0; \end{cases}$$

i.e. $T \sim Exponencial(\lambda)$; en particular, $\mathbb{E}(T) = 1/\lambda$. Retrospectivamente, esto es bastante intuitivo: *como en promedio ocurren λ eventos por unidad de tiempo, la separación temporal entre dos eventos consecutitvos debiera ser $1/\lambda$ en promedio.* El parámetro λ, por lo tanto, representa la *intensidad de los eventos* i.e. el número promedio de eventos por unidad de tiempo.

Ejemplo 4.20. Considere un cierto colegio donde la jornada de clases se extiende desde las 8am hasta las 4pm. Si en promedio dos estudiantes son expulsados de clases en cada jornada, responda: *si un cierto estudiante es expulsado y llevado a la inspectoría a las 10:24am, ¿cuál es la probabilidad que ningún otro estudiante sea expulsado de clases en los siguientes* 90 *minutos?*

Para responder a la pregunta necesitamos determinar la probabilidad del suceso $[T \geq 1{,}5]$, donde T denota el tiempo (medido en horas) transcurrido entre las 10:24am y la próxima expulsión de clases. Debido a la discusión anterior es razonable suponer que $T \sim \mathit{Exponencial}(0{,}25)$. Para corroborar que el parámetro de la distribución es correcto note que $\mathbb{E}(T) = 1/(0{,}25) = 4$. Esto es consistente con la información dada: como en promedio ocurren dos expulsiones en una jornada de 8 horas, la separación típica entre dos expulsiones consecutivas es 4 horas. Así encontramos, a partir de la identidad en (4.12), que

$$\mathbb{P}[T \geq 1{,}5] = e^{-0{,}25 \cdot 1{,}5} = 0{,}6872... \approx 68{,}7\,\%.$$

Una propiedad importante de las variables aleatorias del tipo exponencial es que son desmemoriadas. Para clarificar la terminología retomemos el ejemplo anterior e imagine que no han habido más expulsiones entre las 10:24am y las 11:54am, *¿cuál es la probabilidad que la siguiente suspensión ocurra antes de las 12:24pm?* Para responder a la pregunta observe que no ha habido expulsiones de clases en un periodo de 90 minutos (i.e. entre las 10:24am y las 11:54am). Dado este hecho, necesitamos determinar la probabilidad de que al menos una suspensión ocurrirá dentro de los próximos 30 minutos (i.e. entre las 11:54am y 12:24pm). Como T se mide en horas, esta última probabilidad corresponde a la probabilidad condicional

$$\begin{aligned} \mathbb{P}\left[(T - 1{,}5) \leq 0{,}5 \middle| T > 1{,}5\right] &= \frac{\mathbb{P}[1{,}5 < T \leq 2]}{\mathbb{P}[T > 1{,}5]}, \\ &= \frac{\int_{1,5}^{2} 0{,}25\, e^{-0{,}25t}\, dt}{\int_{1,5}^{+\infty} 0{,}25\, e^{-0{,}25t}\, dt}, \\ &= \frac{e^{-3/8} - e^{-1/2}}{e^{-3/8}} = 1 - e^{-1/8}. \end{aligned}$$

Sin embargo, observe que

$$\mathbb{P}\left[T \leq 0{,}5\right] = \int_0^{0,5} 0{,}25\, e^{-0{,}25t}\, dt = 1 - e^{-1/8}.$$

Concluimos, por lo tanto, que

$$\mathbb{P}\left[(T - 1{,}5) \leq 0{,}5 \middle| T > 1{,}5\right] = \mathbb{P}\left[T \leq 0{,}5\right].$$

En palabras: la probabilidad de que la próxima suspensión ocurra dentro de los próximos 30 minutos, dado que han pasado 90 minutos desde la última suspensión, es igual a la probabilidad de que la próxima suspensión ocurra dentro de los 30 minutos siguientes a la última suspensión. En particular, el hecho que 90 minutos han pasado desde la última suspensión, no retarda ni acelera la ocurrencia de la próxima. Esto es lo que queremos decir por "propiedad desmemoriada". El siguiente resultado formaliza esta idea de forma general.

Teorema 4.10. *Si $X \sim Exponencial(\lambda)$, entonces, para todo $s, t \geq 0$ se tiene que $\mathbb{P}[(X-s) \leq t \mid X > s] = \mathbb{P}[X \leq t]$ i.e. si condicionamos sobre el suceso $[X > s]$, la v.a. $(X - s)$ es también del tipo exponencial con parámetro λ.*

Demostración. Recordemos que $\mathbb{P}[X \geq u] = e^{-\lambda u}$, para todo $u \geq 0$, debido a la identidad en (4.12). En particular, como la función exponencial transforma sumas en productos, encontramos que

$$\mathbb{P}[X > s+t] = \mathbb{P}[X \geq s+t] = \mathbb{P}[X \geq s] \cdot \mathbb{P}[X \geq t] = \mathbb{P}[X > s] \cdot \mathbb{P}[X > t].$$

Dividiendo ambos lados de esta identidad por la probabilidad del suceso $[X > s]$, encontramos que $\mathbb{P}[X > s+t \mid X > s] = \mathbb{P}[X > t]$. Esto último es equivalente a tener que $\mathbb{P}[X \leq s+t \mid X > s] = \mathbb{P}[X \leq t]$ (*¿por qué?*), lo que demuestra que

$$\mathbb{P}[(X-s) \leq t \mid X > s] = \mathbb{P}[X \leq t],$$

para todo $t \geq 0$. El resultado es ahora una consecuencia directa del Teorema 2.11: el lado derecho es la función de distribución de X, y el lado izquierdo es la función de distribución de $(X - s)$ cuando condicionamos sobre el evento $[X > s]$. □

El siguiente ejemplo muestra que la propiedad desmemoriada, no es en general cierta para variables aleatorias no exponenciales. De hecho, se puede demostrar que las variables aleatorias exponenciales son las únicas variables aleatorias continuas y no negativas con esta propiedad.

Ejemplo 4.21. Imagine una persona de provincia, que visita la ciudad de Santiago por primera vez, y que sólo sabe que para llegar a su destino final debe tomar el bus #322, después de llegar al Terminal Central de Buses. Asumiendo que el bus #322 pasa aproximadamente cada 30 minutos, responda: *¿cuál es la función de distribución del tiempo de espera de la persona en el paradero?* y *¿cuál es la función de distribución del tiempo adicional de espera, dado que el bus no ha pasado en los primeros 5 minutos de espera?*

Denotemos como T el tiempo de espera (medido en minutos) de la persona en el paradero. Para responder a la primera pregunta debemos calcular la probabilidad del suceso $[T \leq t]$, para cada $0 < t < 30$. Para este efecto, denotaremos como X el momento de llegada de la persona al paradero en relación el ciclo del bus. La falta de información acerca del inicio y término de cada ciclo justifica el modelo $X \sim Uniforme[0, 30]$. (Aquí se asume que el bus pasa exactamente al principio y término de este intervalo.) Observe que si $X < (30 - t)$, entonces la probabilidad de que el bus pase en los próximos t minutos es cero (*¿por qué?*). Sin embargo, si $X \geq (30 - t)$, entonces la probabilidad de que el bus pase en los próximos t minutos es 1. En particular, usando la Fórmula de Probabilidades Totales, obtenemos que

$$\begin{aligned} \mathbb{P}[T \leq t] &= \mathbb{P}(\text{la persona espera a lo más } t \text{ minutos}), \\ &= 0 \cdot \mathbb{P}[X < 30 - t] + 1 \cdot \mathbb{P}[X \geq 30 - t], \\ &= 0 \cdot \frac{30 - t}{30} + 1 \cdot \frac{t}{30} = \frac{t}{30}. \end{aligned} \tag{4.17}$$

Para responder a la segunda pregunta debemos determinar para cada $0 < t < 25$ la probabilidad del suceso $[(T-5) \leq t]$ pero condicionando sobre el suceso $[T \geq 5]$. En efecto, usando la definición de probabilidad condicional, y la identidad anterior, obtenemos que

$$\mathbb{P}[(T-5) \leq t \mid T \geq 5] = \frac{\mathbb{P}[5 \leq T \leq (t+5)]}{\mathbb{P}[T \geq 5]} = \frac{t/30}{25/30} = \frac{t}{25}. \tag{4.18}$$

Se deduce de (4.17) y de (4.18) que la distribución del tiempo adicional de espera en el paradero cambia después de haber esperado el bus por 5 minutos. De hecho, diferenciado con respecto a la variable t en (4.17), vemos que el tiempo de espera del bus tiene una distribución uniforme en el intervalo $[0, 30]$. Sin embargo, el tiempo adicional después de 5 minutos de espera es uniformemente distribuido en el intervalo $[0, 25]$. Por lo tanto, la distribución del tiempo adicional de espera cambia a medida que esperamos más y más tiempo el bus. ¡Esto no sucedería si los ciclos del bus fueran variables aleatorias exponenciales!

4.6 Distribución Normal

En esta sección $\mu \in \mathbb{R}$ y $\sigma > 0$ son números reales dados.

Definición 4.11. Una v.a. X se dice tener una *distribución Normal* con media μ y varianza σ^2, si es una variable aleatoria continua con función de densidad dada por la fórmula

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}.$$

En este caso, anotaremos $X \sim Normal(\mu, \sigma^2)$.

La demostración que la función anterior, es efectivamente una función de densidad i.e. que $\int_{-\infty}^{\infty} f(x)\,dx = 1$, es intrincada y será dejada como ejercicio para el lector (Ejercicio 23).

Un caso especial de la definición anterior, es cuando $\mu = 0$ y $\sigma = 1$. En este caso, diremos que X tiene una distribución *Normal estándar*. La terminolgía usada en la definición anterior se basa en las siguientes identidades

$$\mathbb{E}(X) = \mu, \tag{4.19}$$

$$\mathbb{V}(X) = \sigma^2. \tag{4.20}$$

Para demostrar estas identidades usaremos los siguientes resultados.

Teorema 4.12. *Si* $Z \sim Normal(0,1)$, *entonces* $\mathbb{E}(Z) = 0$ *y* $\mathbb{V}(Z) = 1$.

Demostración. Observe primero que

$$\int_{-\infty}^{0} z e^{-z^2/2}\,dz = -\int_{0}^{+\infty} z e^{-z^2/2}\,dz,$$

debido a que la función $z \longrightarrow ze^{-z^2/2}$ es antisimétrica. En particular, se tiene que

$$\mathbb{E}(Z) = \int_{-\infty}^{+\infty} z \cdot \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = \frac{1}{\sqrt{2\pi}} \left\{ \int_{-\infty}^{0} ze^{-z^2/2} dz + \int_{0}^{+\infty} ze^{-z^2/2} dz \right\} = 0.$$

Por otro lado, usando la identidad $\mathbb{V}(Z) = \mathbb{E}(Z^2) - \{\mathbb{E}(Z)\}^2$, y el Método de Integración por Partes, obtenemos que

$$\begin{aligned}
\mathbb{V}(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-z^2/2} dz, \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \left(-e^{-z^2/2}\right)' dz, \\
&= -\frac{ze^{-z^2/2}}{\sqrt{2\pi}} \Bigg|_{z=-\infty}^{z=+\infty} + \int_{-\infty}^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = 1,
\end{aligned}$$

donde hemos usado que $\lim_{z\to\pm\infty} ze^{-z^2/2} = 0$ (*¿por qué?*) y que $z \longrightarrow e^{-z^2/2}/\sqrt{2\pi}$ es la función de densidad de una v.a. Normal estándar. Esto completa la demostración del teorema. □

Teorema 4.13. *Si $X \sim Normal(\mu, \sigma^2)$ y $a \neq 0$ y b son constantes dadas, entonces $(aX + b) \sim Normal(a\mu + b, a^2\sigma^2)$. En particular, $(X - \mu)/\sigma$ tiene una distribución Normal estándar.*

Demostración. De acuerdo al Método del Jacobiano (Teorema 2.8), si X es una v.a. continua con función de densidad f, entonces $(aX + b)$ es también una v.a. continua pero con función de densidad $\frac{1}{a} f\left(\frac{x-b}{a}\right)$. En particular, la función de densidad de $(aX + b)$ está dada por

$$\frac{1}{a} \cdot \frac{e^{-\frac{1}{2}\left(\frac{\frac{x-b}{a}-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} = \frac{e^{-\frac{1}{2}\left(\frac{x-(a\mu+b)}{a\sigma}\right)^2}}{a\sigma\sqrt{2\pi}},$$

lo que corresponde a la función de densidad de una v.a. Normal con media $(a\mu + b)$ y varianza $(a\sigma)^2$. Esto completa la demostración del teorema. □

Ahora podemos fácilmente demostrar que, si $X \sim Normal(\mu, \sigma^2)$, entonces $\mathbb{E}(X) = \mu$ y $\mathbb{V}(X) = \sigma^2$. En efecto, de acuerdo al Teorema 4.13, tenemos que

$$Z = \frac{X - \mu}{\sigma} \sim Normal(0, 1).$$

Como $X = \sigma Z + \mu$, y usando el Teorema 4.12, concluimos que

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(\sigma Z + \mu) = \sigma \cdot \mathbb{E}(Z) + \mu = \sigma \cdot 0 + \mu = \mu, \\
\mathbb{V}(X) &= \mathbb{V}(\sigma Z + \mu) = \sigma^2 \cdot \mathbb{V}(Z) = \sigma^2 \cdot 1 = \sigma^2,
\end{aligned}$$

lo que demuestra (4.19) y (4.20).

El Teorema 4.13 es útil para reescribir probabilidades relacionados con variables aleatorias normales arbitrarias en términos de Normales estándares. Una aplicación interesante de esta técnica es la siguiente. En lo que sigue usaremos Φ para denotar la *función de distribución de una v.a. Normal estándar* i.e.

$$\Phi(t) = \int_{-\infty}^{t} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, dz,$$

para todo $t \in \mathbb{R}$. Observe que $\Phi(0) = 1/2$ debido a la simetría de la función de densidad alrededor de $z = 0$ (ver Figura 4.4). Para valores generales de t no es posible calcular $\Phi(t)$ explícitamente y esto obliga el uso de métodos de integración numérica para aproximar su valor. En la Tabla 4.2 se muestran los primeros cuatro dígitos significativos de $\Phi(t)$ para ciertos valores de t.

Tampoco es posible calcular explícitamente la función de distribución de una variables aleatoria Normal con parámetros arbitrarios. Afortunadamente, usando el Teorema 4.13, podemos relacionar la función de distribución de toda v.a. Normal con la de una Normal estándar. En efecto, si $X \sim Normal(\mu, \sigma^2)$ y si definimos $Z = (X - \mu)/\sigma$, entonces

$$F_X(t) = \mathbb{P}\Big[\mu + \sigma \cdot Z \le t\Big] = \mathbb{P}\left[Z \le \frac{t-\mu}{\sigma}\right] = \Phi\left(\frac{t-\mu}{\sigma}\right), \tag{4.21}$$

para todo $t \in \mathbb{R}$.

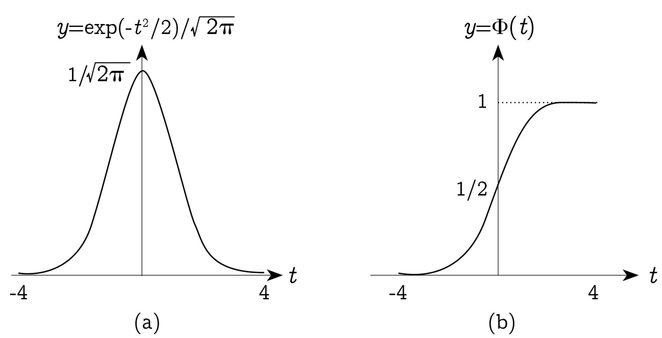


Figura 4.4. (a) Grafo de la función de densidad de una v.a. Normal estándar, para valores de t entre (-4) y 4. La probabilidad de que una de estas variables tome un valor entre (-4) y 4 es aproximadamente 99,993666 %. Debido a esto, los extremos del grafo parecen tocar el eje horizontal. (b) Grafo de la función de distribución de una v.a. Normal estándar.

Ejemplo 4.22. Para fijar ideas considere una v.a. X con una distribución Normal con media (-1) y varianza 4. *Determine la probabilidad* $\mathbb{P}[0{,}62 < X \le 1{,}9]$*, usando la Tabla 4.2.*

Primero observe que $\mathbb{P}[0{,}62 < X \leq 1{,}9] = \mathbb{P}[X \leq 1{,}9] - \mathbb{P}[X \leq 0{,}62]$ (*¿por qué?*). Por otro lado, debido al Teorema 4.13 tenemos que $(X+1)/2$ es una v.a. Normal estándar. Concluimos así que

$$\begin{aligned}\mathbb{P}[X \leq 1{,}9] &= \mathbb{P}\left[\frac{X+1}{2} \leq \frac{1{,}9+1}{2}\right] = \mathbb{P}\left[\frac{X+1}{2} \leq 1{,}45\right] = \Phi(1{,}45),\\ \mathbb{P}[X \leq 0{,}62] &= \mathbb{P}\left[\frac{X+1}{2} \leq \frac{0{,}62+1}{2}\right] = \mathbb{P}\left[\frac{X+1}{2} \leq 0{,}81\right] = \Phi(0{,}81).\end{aligned}$$

Pero note que $\Phi(1{,}45) = 0{,}9264...$ y $\Phi(0{,}81) = 0{,}7910...$ debido a la Tabla 4.2. Obtenemos, de esta manera, que $\mathbb{P}[0{,}62 < X \leq 1{,}9] = 0{,}9264... - 0{,}7910... = 0{,}135...$ i.e. hay una chance aproximada de un 13,5 % que X tome un valor estrictamente mayor que 0,62, pero menor o igual a 1,9.

Ejemplo 4.23. Considere una variable $X \sim Normal(\mu, \sigma^2)$, *¿cuál es la probabilidad que X tome un cierto valor t desviaciones estándares a partir de su valor esperado?* En otras palabras, debemos determinar la probabilidad del suceso $[|X-\mu| \leq t\cdot\sigma]$. En efecto, si definimos $Z = (X-\mu)/\sigma$, entonces

$$\mathbb{P}[|X-\mu| \leq t\cdot\sigma] = \mathbb{P}\left[\left|\frac{X-\mu}{\sigma}\right| \leq t\right] = \mathbb{P}[|Z| \leq t].$$

Como $Z \sim Normal(0,1)$, y como las variables aleatorias normales son continuas, tenemos que $\mathbb{P}[|Z| \leq t] = \mathbb{P}[-t \leq Z \leq t] = \mathbb{P}[-t < Z \leq t] = \Phi(t) - \Phi(-t)$. Sin embargo, debido a que la función de densidad de Z es simétrica alrededor de $z = 0$ (ver Figura 4.4), también tenemos que

$$\begin{aligned}\Phi(-t) = \int_{-\infty}^{-t} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz &= \int_{t}^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz,\\ &= \mathbb{P}[Z > t],\\ &= 1 - \mathbb{P}[Z \leq t] = 1 - \Phi(t),\end{aligned} \tag{4.22}$$

para todo $t > 0$.

Concluimos, de este modo, que si $X \sim Normal(\mu, \sigma^2)$, entonces

$$\mathbb{P}[|X-\mu| \leq t\cdot\sigma] = 2\cdot\Phi(t) - 1. \tag{4.23}$$

Observe cómo el lado derecho no depende ni de μ ni de σ. Por ejemplo, usando los valores tabulados en la Tabla 4.2, obtenemos que

$$\begin{aligned}\mathbb{P}[|X-\mu| \leq 1\sigma] &= 2\cdot 0{,}8413... - 1 = 0{,}682... \approx 68{,}2\,\%;\\ \mathbb{P}[|X-\mu| \leq 2\sigma] &= 2\cdot 0{,}9772... - 1 = 0{,}954... \approx 95{,}4\,\%;\\ \mathbb{P}[|X-\mu| \leq 3\sigma] &= 2\cdot 0{,}9986... - 1 = 0{,}997... \approx 99{,}7\,\%;\\ \mathbb{P}[|X-\mu| \leq 4\sigma] &= 2\cdot 0{,}9999... - 1 = 0{,}999... \approx 99{,}9\,\%.\end{aligned}$$

En palabras, la probabilidad de que X tome un cierto valor dentro de una, dos, tres o cuatro desviaciones estándar, a partir de su valor esperado, es aproximadamente 68,2 %, 95,4 %, 99,7 % ó 99,9 %, respectivamente (ver Figura 4.5). En particular, por ejemplo, la probabilidad del suceso $[|X-\mu|>3\sigma]$ es 0,3 % aproximadamente i.e. si simuláramos X mil veces, entonces solamente alrededor de tres de las simulaciones debieran tomar un valor a más de tres desviaciones estándares, a partir de su valor esperado. Similarmente, alrededor de una entre mil simulaciones, debiera tomar un valor a más de cuatro desviaciones estándares de μ. Éstas observaciones resultan útiles para testear la eficacia de un simulador de variables aleatorias normales. También resultan útiles en Estadística para testear hipótesis acerca del parámetro μ.

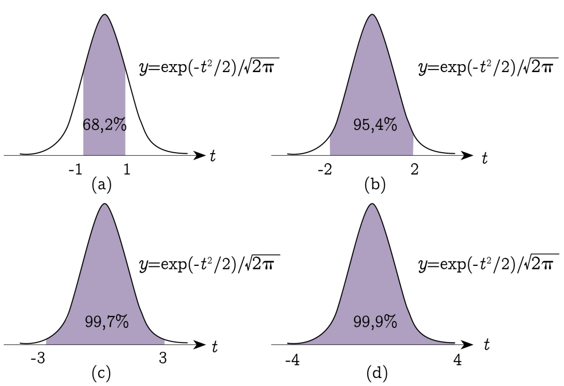


FIGURA 4.5. Probabilidades aproximadas que una v.a. Normal estándar tome un cierto valor a una, dos, tres o cuatro desviaciones estándar de su valor esperado.

Recordemos que la varianza de una v.a. mide cuán concentrada es dicha variable alrededor de su valor esperado. En particular, mientras más pequeño sea el parámetro σ de una v.a. Normal, más probable debiera ser el suceso que ésta tome un valor cerca de μ. Esta intuición es correcta de acuerdo al siguiente resultado.

Teorema 4.14. *Si $X_1 \sim Normal(\mu,\sigma_1^2)$ y $X_2 \sim Normal(\mu,\sigma_2^2)$, con $\sigma_1<\sigma_2$, entonces $\mathbb{P}[|X_1-\mu|\leq\delta]>\mathbb{P}[|X_2-\mu|\leq\delta]$, para todo $\delta>0$.*

Demostración. Debido a la identidad en (4.23), si $X\sim Normal(\mu,\sigma^2)$, entonces

$$\mathbb{P}\left[|X-\mu|\leq t\sigma\right]=2\cdot\Phi(t)-1.$$

La función $t\longrightarrow(2\cdot\Phi(t)-1)$ es creciente, toma el valor 0 cuando $t=0$, y converge hacia 1 a medida que t tiende a infinito (*¿por qué?*). En particular, dado $\delta>0$ y como $\delta/\sigma_1>\delta/\sigma_2$, la siguiente desigualdad aplica

$$\mathbb{P}[|X_1-\mu|\leq\delta]=2\cdot\Phi\left(\frac{\delta}{\sigma_1}\right)-1>2\cdot\Phi\left(\frac{\delta}{\sigma_2}\right)-1=\mathbb{P}[|X_2-\mu|\leq\delta].$$

Esto demuestra el teorema. $\square$

$t_1 \setminus t_2$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5039	0.5079	0.5119	0.5159	0.5199	0.5239	0.5279	0.5318	0.5358
0.1	0.5398	0.5438	0.5477	0.5517	0.5556	0.5596	0.5635	0.5675	0.5714	0.5753
0.2	0.5792	0.5831	0.5870	0.5909	0.5948	0.5987	0.6025	0.6064	0.6102	0.6140
0.3	0.6179	0.6217	0.6255	0.6293	0.6330	0.6368	0.6405	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6843	0.6879
0.5	0.6914	0.6949	0.6984	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7290	0.7323	0.7356	0.7389	0.7421	0.7453	0.7485	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7733	0.7763	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7938	0.7967	0.7995	0.8023	0.8051	0.8078	0.8105	0.8132
0.9	0.8159	0.8185	0.8212	0.8238	0.8263	0.8289	0.8314	0.8339	0.8364	0.8389
1.0	0.8413	0.8437	0.8461	0.8485	0.8508	0.8531	0.8554	0.8576	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8707	0.8728	0.8749	0.8769	0.8790	0.8810	0.8829
1.2	0.8849	0.8868	0.8887	0.8906	0.8925	0.8943	0.8961	0.8979	0.8997	0.9014
1.3	0.9032	0.9049	0.9065	0.9082	0.9098	0.9114	0.9130	0.9146	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9250	0.9264	0.9278	0.9292	0.9305	0.9318
1.5	0.9331	0.9344	0.9357	0.9369	0.9382	0.9394	0.9406	0.9417	0.9429	0.9440
1.6	0.9452	0.9463	0.9473	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9544
1.7	0.9554	0.9563	0.9572	0.9581	0.9590	0.9599	0.9608	0.9616	0.9624	0.9632
1.8	0.9640	0.9648	0.9656	0.9663	0.9671	0.9678	0.9685	0.9692	0.9699	0.9706
1.9	0.9712	0.9719	0.9725	0.9732	0.9738	0.9744	0.9750	0.9755	0.9761	0.9767
2.0	0.9772	0.9777	0.9783	0.9788	0.9793	0.9798	0.9803	0.9807	0.9812	0.9816
2.1	0.9821	0.9825	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9853	0.9857
2.2	0.9861	0.9864	0.9867	0.9871	0.9874	0.9877	0.9880	0.9884	0.9887	0.9889
2.3	0.9892	0.9895	0.9898	0.9901	0.9903	0.9906	0.9908	0.9911	0.9913	0.9915
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9937	0.9939	0.9941	0.9943	0.9944	0.9946	0.9947	0.9949	0.9950	0.9952
2.6	0.9953	0.9954	0.9956	0.9957	0.9958	0.9959	0.9960	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9972	0.9973
2.8	0.9974	0.9975	0.9976	0.9976	0.9977	0.9978	0.9978	0.9979	0.9980	0.9980
2.9	0.9981	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986
3.0	0.9986	0.9986	0.9987	0.9987	0.9988	0.9988	0.9988	0.9989	0.9989	0.9990
3.1	0.9990	0.9990	0.9991	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992
3.2	0.9993	0.9993	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995
3.3	0.9995	0.9995	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996
3.4	0.9996	0.9996	0.9996	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997
3.5	0.9997	0.9997	0.9997	0.9997	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.7	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
4.0	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

TABLA 4.2. Valores tabulados de la función de distribución de una v.a. con distribución Normal estándar. El valor en la fila t_1 y columna t_2 corresponde a los primeros cuatro dígitos significativos de $\Phi(t_1 + t_2)$. Por ejemplo, si $t_1 = 0{,}1$ y $t_2 = 0{,}02$, entonces $\Phi(0{,}12) = 0{,}5477...$

4.6.1 Combinaciones lineales.

El siguiente resultado establece que la combinación lineal de variables aleatorias normales e independientes, es también una v.a. con distribución Normal.

Teorema 4.15. *Si $X_1, \ldots, X_n$ son variables aleatorias independientes y tales que $X_i \sim Normal(\mu_i, \sigma_i)$, para cada $i = 1, \ldots, n$, entonces*

$$\sum_{i=1}^{n} X_i \sim Normal\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

Para demostrar el teorema basta con mostrar que

$$X + Y \sim Normal(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2), \tag{4.24}$$

cuando $X \sim Normal(\mu_1, \sigma_1^2)$ e $Y \sim Normal(\mu_2, \sigma_2^2)$ son independientes (*¿por qué?*). Aunque no demostraremos esto, es importante entender por qué los parámetros de la v.a. $(X + Y)$ son como son. En efecto, observe que

$$\begin{aligned} \mathbb{E}(X+Y) &= \mathbb{E}(X) + \mathbb{E}(Y) = \mu_1 + \mu_2; \\ \mathbb{V}(X+Y) &= \mathbb{V}(X) + \mathbb{V}(Y) = \sigma_1^2 + \sigma_2^2; \end{aligned}$$

donde, para el cálculo de la varianza, hemos usado que X e Y son independientes. En particular, una vez que asumimos que $(X + Y)$ es una v.a. Normal, sus parámetros no pueden ser otros que $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Ejemplo 4.24. Si cada uno de los doscientos pasajeros de un avión acarrea en sus maletas un peso neto con una distribución Normal con media 24 kilogramos y una desviación estándar de 10 kilogramos, responda: *¿cuál es la probabilidad que el peso neto de todas las maletas exceda 5019 kilogramos?*

Si X_i denota el peso neto de las maletas del i-ésimo pasajero entonces $X_i \sim Normal(24, 100)$ de acuerdo a la información dada. En particular, asumiendo que $X_1, \ldots, X_{200}$ son variables aleatorias independientes, el Teorema 4.15 nos permite concluir que

$$X = \sum_{i=1}^{200} X_i \sim Normal(4800, 20000).$$

La respuesta a la pregunta anterior viene, por lo tanto, dada por la probabilidad del suceso $[X \geq 5019]$. Usando el Teorema 4.13 obtenemos, de este modo, que

$$\mathbb{P}[X \geq 5019] = \mathbb{P}\left[\frac{X - 4800}{\sqrt{20000}} \geq \frac{5019 - 4800}{\sqrt{20000}}\right] = 1 - \Phi\left(\frac{219}{100\sqrt{2}}\right) = 1 - \Phi(1{,}548...).$$

Finalmente, debido a la Tabla 4.2, tenemos que

$$\mathbb{P}[X \geq 5019] \approx 1 - \Phi(1{,}55) = 0{,}0606... \approx 6\,\%.$$

Una consecuencia importante del Teorema 4.15 es el siguiente resultado.

Corolario 4.16. *Si $X_1, \ldots, X_n$ son variables aleatorias i.i.d. del tipo normal con valor esperado μ y varianza σ^2, entonces*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim Normal\left(\mu, \frac{\sigma^2}{n}\right).$$

Demostración. Debido al Teorema 4.13, $X_i/n \sim Normal(\mu/n, \sigma^2/n^2)$, para cada i. En particular, y de acuerdo al Teorema 4.15, la variable aleatoria

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \frac{X_i}{n}$$

tiene una distribución Normal. Los parámetros de esta distribución están dados en el Teorema 4.15. Sin embargo, podemos determinar alternativamente estos parámetros computando directamente el valor esperado y varianza de $\bar{X}$. En efecto, observe que

$$\begin{aligned}\mathbb{E}(\bar{X}) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu;\\ \mathbb{V}(\bar{X}) &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.\end{aligned}$$

Esto demuestra el teorema. □

De acuerdo al corolario anterior, $\bar{X}$ tiene el mismo valor esperado que cada una de las variables aleatorias normales que participan del promedio, sin embargo, su varianza es más pequeña debido al factor $1/n$. $\bar{X}$ es, por lo tanto, mucho más probable de tomar un valor cerca de μ que cualquiera de las variables aleatorias $X_1, \ldots, X_n$ (Teorema 4.14). Debido a esto, la mejor estrategia para estimar μ cuando este parámetro es desconocido es a través del promedio de $X_1, \ldots, X_n$. El siguiente ejemplo clarificará la importancia de este hecho.

Ejemplo 4.25. Primero demostraremos que si $\alpha \in (0,1)$ y $z \in \mathbb{R}$ son tales que $\Phi(z) = 1 - \alpha/2$, entonces

$$\mathbb{P}\left[\bar{X} - \frac{z\cdot\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{z\cdot\sigma}{\sqrt{n}}\right] = (1-\alpha). \tag{4.25}$$

En palabras, el intervalo con extremos aleatorios $[\bar{X} - z\sigma/\sqrt{n}, \bar{X} + z\sigma/\sqrt{n}]$, contendrá el valor μ un $100(1-\alpha)\,\%$ de las veces. Por lo mismo, nos referiremos a éste como un *intervalo de confianza para* μ. Esto es particularmente útil cuando μ es un parámetro desconocido.

Observe que el centro del intervalo de confianza i.e. $\bar{X}$ se aproximará a μ a medida que n tiende a infinito, debido a la Ley de los Grandes Números. Por otro lado, el radio del intervalo i.e. $z\sigma/\sqrt{n}$ decrece a cero cuando $n \to \infty$. Por lo tanto, el intervalo de confianza estará cada vez más concentrado alrededor de μ a medida que n crece i.e. a medida que el número de datos aumenta.

Típicamente uno escoge $\alpha = 0{,}01$ ó $\alpha = 0{,}05$. Por ejemplo, si $\alpha = 0{,}01$, entonces necesitamos determinar z, tal que $\Phi(z) = 0{,}995$. Usando la Tabla 4.2 encontramos que $z \approx 2{,}58$. Por lo tanto, el intervalo

$$\left[\bar{X} - \frac{2{,}58\cdot\sigma}{\sqrt{n}}, \bar{X} + \frac{2{,}58\cdot\sigma}{\sqrt{n}}\right] \tag{4.26}$$

contendrá a μ con una probabilidad de un $99\,\%$.

La demostración (4.25) es como sigue. Note primero que $\bar{X} \sim Normal(\mu, \sigma^2/n)$ debido al Corolario 4.16. En particular, y de acuerdo al Teorema 4.13, la variable aleatoria $(\bar{X} - \mu)/\sqrt{\sigma^2/n} = \sqrt{n}\cdot(\bar{X}-\mu)/\sigma$ tiene una distribución Normal estándar.

Finalmente, usando la identidad en (4.23), obtenemos que

$$\mathbb{P}\left[\left|\sqrt{n}\cdot\frac{\bar{X}-\mu}{\sigma}\right|\leq z\right] = 2\cdot\Phi(z)-1 = 2\left(1-\frac{\alpha}{2}\right)-1 = (1-\alpha).$$

La identidad en (4.25) es ahora una consecuencia directa de la siguiente igualdad de sucesos que dejaremos como ejercicio para el lector:

$$\left[\left|\sqrt{n}\cdot\frac{\bar{X}-\mu}{\sigma}\right|\leq z\right] = \left[\bar{X}-\frac{z\cdot\sigma}{\sqrt{n}}\leq\mu\leq\bar{X}+\frac{z\cdot\sigma}{\sqrt{n}}\right].$$

Una aplicación concreta de la discusión anterior es la siguiente. Imagine que $X_1,\ldots,X_n$ representan las notas de un alumno en el curso de un cierto profesor. Asumiendo que $X_1,\ldots,X_n$ son i.i.d. con una distribución Normal con esperanza desconocida μ y varianza conocida σ^2, el profesor puede utilizar el intervalo en (4.26) para decidir si pasa o no a un alumno con un promedio de notas 3,91. Por ejemplo, si $n=10$, $\sigma=0{,}1$ y $\bar{X}=3{,}91$, el intervalo de confianza para μ es $[3{,}828...;3{,}991...]$. Como la nota 4.0 no pertenece a este intervalo, el profesor estaría 99 % seguro que la nota intrínseca del alumno es menor que 4.0 y reprobar al alumno parece razonable en este caso. Por otro lado, si $n=4$ entonces el intervalo de confianza para μ es $[3{,}781...;4{,}039...]$. Como la nota 4.0 pertenece a este intervalo, el profesor podría dar el beneficio de la duda al alumno y pasarlo con un 4.0.

Note que la decisión del profesor no sólo depende de $\bar{X}$, sino también de n (i.e. del número de notas) y de σ (i.e. la desviación estándar de las notas). Más aún, como el largo del intervalo de confianza es proporcional a $\sigma/\sqrt{n}$, para un valor dado de σ, el profesor tendrá mayor certeza en sus decisiones al considerar un número más bien grande que pequeño de notas. Esto explica la costumbre de utilizar varias calificaciones para determinar notas finales.

4.7 Teorema Central Límite.

La variables aleatorias normales están íntimamente ligadas al cálculo de promedios. Una primera indicación de este hecho fue discutida en el Ejemplo 4.25 donde usamos que si $X_1,\ldots,X_n$ son variables aleatorias normales con esperanza μ y varianza σ^2, entonces

$$\sqrt{n}\cdot\frac{\bar{X}-\mu}{\sigma}\sim Normal(0,1).$$

Sorprendentemente, si $X_1,\ldots,X_n$ son i.i.d., pero no necesariamente Normales, la v.a. al lado izquierdo de lo anterior, tendrá aproximadamente una distribución Normal estándar cuando n sea grande. Este resultado, al igual que la Ley de los Grandes Números, es uno de los resultados más celebrados e importantes de la teoría de probabilidades y establece más precisamente lo siguiente.

Teorema 4.17. *(Teorema Central Límite.) Si $X_1,X_2,\ldots$ son variables aleatorias i.i.d. con esperanza μ y varianza $\sigma^2>0$ finitas, entonces para todo par de constantes*

$-\infty \leq a < b \leq +\infty$, *se tiene que*

$$\lim_{n\to\infty} \mathbb{P}\left[a \leq \sqrt{n}\cdot\frac{\bar{X}-\mu}{\sigma} \leq b\right] = \Phi(b)-\Phi(a).$$

Dos casos especiales de lo anterior, son cuando $a=-\infty$ i.e. $\Phi(a)=0$, o $b=+\infty$ i.e. $\Phi(b)=1$. En estos casos el Teorema Central Límite establece que

$$\lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}\cdot\frac{\bar{X}-\mu}{\sigma} \leq b\right] = \Phi(b), \tag{4.27}$$

$$\lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}\cdot\frac{\bar{X}-\mu}{\sigma} \geq a\right] = 1-\Phi(a). \tag{4.28}$$

La demostración del Teorema 4.17 es bastante técnica y va más allá del nivel de esta monografía. Sin embargo, en un primer estudio de probabilidades, es mucho más importante entender las implicaciones de este resultado que su demostración.

Observe que si $Z \sim Normal(0,1)$, entonces $\mathbb{P}[a \leq Z \leq b] = \Phi(b)-\Phi(a)$. El Teorema Central Límite establece, por lo tanto, que $\sqrt{n}\cdot(\bar{X}-\mu)/\sigma$ se comporta como si fuera una v.a. Normal estándar, cuando un número grande n de variables aleatorias i.i.d. participan del promedio.

Ejemplo 4.26. Considere variables aleatorias $X_1,\ldots,X_{100}$ i.i.d. con valor esperado $\mu=1$ y varianza $\sigma^2=4$, *¿cuál es aproximadamente la probabilidad del suceso* $\bar{X} \leq 1{,}466$*?*

Como $\bar{X}$ es el promedio de un número grande de variables aleatorias i.i.d. el Teorema 4.17 y la Tabla 4.2 implican que

$$\begin{aligned}
\mathbb{P}\left[\bar{X} \leq 1{,}466\right] &= \mathbb{P}\left[\bar{X}-1 \leq 1{,}466-1\right],\\
&= \mathbb{P}\left[\frac{\bar{X}-1}{2} \leq \frac{0{,}466}{2}\right],\\
&= \mathbb{P}\left[10\cdot\frac{\bar{X}-1}{2} \leq 2{,}33\right],\\
&\approx \Phi(2{,}33) = 0{,}9901...
\end{aligned}$$

i.e. el suceso $\bar{X} \leq 1{,}466$ tiene una posibilidad aproximada de un 99 % debido al Teorema Central Límite.

Ejemplo 4.27. Considere una moneda con probabilidad $1/3$ de salir cara, *¿cuál es aproximadamente la probabilidad de que la moneda salga cara en exactamente un tercio de* 144 *lanzamientos?*

Si X_i es la función indicadora del suceso "la moneda salió cara en el i-ésimo lanzamiento", entonces la respuesta a la pregunta viene dada por la probabilidad del suceso $[\sum_{i=1}^{144} X_i = 48]$. Ésta probabilidad es exactamente (*¿por qué?*)

$$\binom{144}{48}\cdot\left(\frac{1}{3}\right)^{48}\cdot\left(\frac{2}{3}\right)^{96} = 0{,}070...$$

Usando el Teorema Central Límite veremos cómo aproximar fácilmente la probabilidad anterior, evitando así el uso de coeficientes binomiales y las potencias de los números 1/3 y 2/3. El truco consiste en reescribir el suceso de interés en términos del promedio de $X_1, \ldots, X_{144}$ como sigue:

$$\begin{aligned} \mathbb{P}\left[\sum_{i=1}^{144} X_i = 48\right] &= \mathbb{P}\left[48 - \frac{1}{2} \leq \sum_{i=1}^{144} X_i \leq 48 + \frac{1}{2}\right], \\ &= \mathbb{P}\left[\frac{95}{2} \leq \sum_{i=1}^{144} X_i \leq \frac{97}{2}\right] = \mathbb{P}\left[\frac{95}{288} \leq \bar{X} \leq \frac{97}{288}\right], \end{aligned}$$

donde para la primera identidad hemos usado que $\sum_{i=1}^{144} X_i$ es un número entero. Como $\bar{X}$ es el promedio de un número grande $n = 144$ de variables aleatorias i.i.d. con valor esperado $\mu = 1/3$ y varianza $\sigma^2 = 1/3 \cdot 2/3 = 2/9$ (*¿por qué?*), el Teorema 4.17 implica que

$$\begin{aligned} \mathbb{P}\left[\sum_{i=1}^{144} X_i = 48\right] &= \mathbb{P}\left[\sqrt{144} \cdot \frac{\frac{95}{288} - 1/3}{\sqrt{2/9}} \leq \sqrt{144} \cdot \frac{\bar{X} - 1/3}{\sqrt{2/9}} \leq \sqrt{144} \cdot \frac{\frac{97}{288} - 1/3}{\sqrt{2/9}}\right], \\ &= \mathbb{P}\left[-\frac{\sqrt{2}}{16} \leq \sqrt{144} \cdot \frac{\bar{X} - 1/3}{\sqrt{2/9}} \leq \frac{\sqrt{2}}{16}\right], \\ &\approx \Phi\left(\frac{\sqrt{2}}{16}\right) - \Phi\left(-\frac{\sqrt{2}}{16}\right) = 2 \cdot \Phi\left(\frac{\sqrt{2}}{16}\right) - 1, \end{aligned}$$

donde para la última igualdad hemos usado la identidad en (4.22). Pero, observe que $\sqrt{2}/16 \approx 0{,}09$; en particular, debido a la segunda fila y última columna de la Tabla 4.2, tenemos que $\Phi(\sqrt{2}/16) \approx 0{,}5358$. Por lo tanto, la probabilidad del suceso $[\sum_{i=1}^{144} X_i = 48]$ es aproximadamente $(2 \cdot 0{,}5358 - 1) = 0{,}0716$, la que es correcta en los primeros dos dígitos.

Ejemplo 4.28. En el Ejemplo 4.25 vimos que si las notas $X_1, \ldots, X_n$ de un alumno en un cierto curso son variables aleatorias i.i.d. del tipo normal, entonces el intervalo de confianza

$$\left[\bar{X} - \frac{2{,}58 \cdot \sigma}{\sqrt{n}}, \bar{X} + \frac{2{,}58 \cdot \sigma}{\sqrt{n}}\right]$$

contrendrá a μ con una posibilidad de un 99 %. Esto se debió a que

$$\mathbb{P}\left[\left|\sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma}\right| \leq 2{,}58\right] = 0{,}99, \tag{4.29}$$

cuando $X_1, \ldots, X_n$ son normales i.i.d.

Aunque en la práctica es poco razonable suponer que las notas de un alumno son realizaciones de una v.a. Normal, el Teorema Central Límite establece que

$$\mathbb{P}\left[\left|\sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma}\right| \leq 2{,}58\right] \approx 0{,}99,$$

cuando un número grande de notas es considerado en el promedio. En este caso el intervalo de confianza anterior, contendrá a μ con una probabilidad aproximada (en vez de exacta) de un 99 %. Más aún, debido a que

$$S_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \approx \sigma$$

cuando n es grande (ver Sección 3.6.1), el intervalo

$$\left[\bar{X} - \frac{2{,}58 \cdot S_X}{\sqrt{n}}, \bar{X} + \frac{2{,}58 \cdot S_X}{\sqrt{n}}\right]$$

también contedrá a μ con una probabilidad aproximada de un 99 %. La gracia de esto último es que no necesitamos conocer σ para construir un intervalo de confianza para μ.

Ejemplo 4.29. Considere variables aleatorias $X_1, X_2, \ldots$ i.i.d. del tipo exponencial con parámetro λ. Estas variables podrían representar e.g. el tiempo que un cierto alumno invierte en cada uno de los ejercicios asignados como tarea en un curso. En este caso, tenemos que $\mu = 1/\lambda$ y $\sigma^2 = 1/\lambda^2$. Por lo tanto, y de acuerdo al Teorema Central Límite, a medida que n crece la variable aleatoria

$$\sqrt{n} \cdot \frac{\bar{X} - \frac{1}{\lambda}}{\frac{1}{\lambda}} = \sqrt{n} \cdot (\lambda\bar{X} - 1) \tag{4.30}$$

se comportará más y más como si fuera del tipo Normal estándar. En la Figura 4.6 hemos presentado la función de densidad de esta v.a. cuando para diferentes valores de n.

Para apreciar las posibles implicaciones de lo anterior, supongamos que un alumno típico invierte 6 minutos en cada ejercicio de tarea. Imagine ahora que uno de los alumnos en el curso—el cual no ha obtenido notas muy impresionantes en el pasado—obtuvo un 6.5 en la última tarea y dice haber invertido entre 2 y 3 horas en ella, *¿cuál es aproximadamente la probabilidad de este suceso, si la tarea consistió en* 36 *problemas?*

Midiendo el tiempo en minutos, la respuesta corresponde a la probabilidad del suceso $[2 \cdot 60 \leq \sum_{i=1}^{36} X_i \leq 3 \cdot 60]$. Dividiendo ambos lados de la desigualdad por 36, esto es equivalente a determinar la probabilidad del suceso $[10/3 \leq \bar{X} \leq 5]$. Para aplicar el Teorema Central Límite observe que $\mathbb{E}(X_i) = 6$. Por lo tanto, $\lambda = 1/6$. En

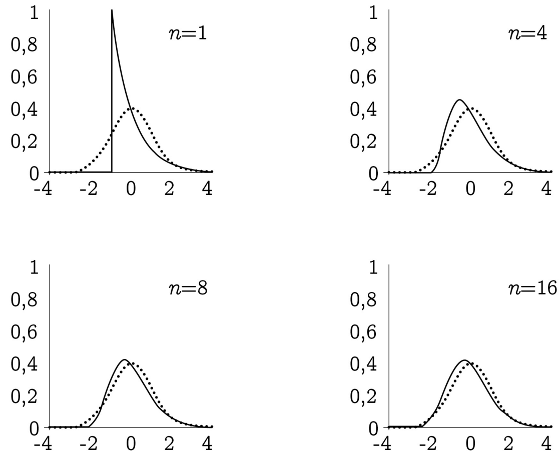


FIGURA 4.6. Funciones de densidad de $\sqrt{n}(\lambda\bar{X} - 1)$, cuando $X_1, \ldots, X_n$ son i.i.d. del tipo exponencial con un mismo parámetro λ y $n = 1, 4, 8, 16$. Debido al Ejercicio 39, estas funciones dependen de n pero no de λ. Observe cómo éstas se parecen más y más a la densidad de una v.a. Normal estándar, mostrada en línea punteada, a medida que n crece.

particular, si definimos $n = 36$, entonces

$$\begin{aligned}
\mathbb{P}\left[10/3 \leq \bar{X} \leq 5\right] &= \mathbb{P}\left[\frac{10\lambda}{3} \leq \lambda\bar{X} \leq 5\lambda\right], \\
&= \mathbb{P}\left[\frac{10\lambda}{3} - 1 \leq \lambda\bar{X} - 1 \leq 5\lambda - 1\right], \\
&= \mathbb{P}\left[\sqrt{n} \cdot \left(\frac{10\lambda}{3} - 1\right) \leq \sqrt{n} \cdot (\lambda\bar{X} - 1) \leq \sqrt{n} \cdot (5\lambda - 1)\right], \\
&\approx \Phi\left(\sqrt{n} \cdot (5\lambda - 1)\right) - \Phi\left(\sqrt{n} \cdot \left(\frac{10\lambda}{3} - 1\right)\right), \\
&\approx \Phi(-1) - \Phi(-8/3).
\end{aligned}$$

Usando la identidad en (4.22), finalmente obtenemos que

$$\begin{aligned}
\mathbb{P}\left[10/3 \leq \bar{X} \leq 5\right] &\approx (1 - \Phi(1)) - (1 - \Phi(8/3)), \\
&= \Phi(8/3) - \Phi(1), \\
&= 0{,}9961... - 0{,}8413..., \\
&= 0{,}154... \approx 15{,}4\,\%.
\end{aligned}$$

Basado en el cálculo anterior, es un poco sospechoso que el alumno haya invertido entre entre 2 y 3 horas en la tarea y quizás el profesor debiera indagar más en el asunto.

Ejemplo 4.30. Considere variables aleatorias $X_1, X_2, \ldots$ i.i.d. del tipo Bernoulli con parámetro $0 < p < 1$. En este caso, $\mu = p$ y $\sigma^2 = p(1-p)$. Por lo tanto, para valores grande de n, la variable aleatoria

$$\sqrt{n} \cdot \frac{\bar{X} - p}{\sqrt{p(1-p)}} \tag{4.31}$$

se comportará como si tuviera una distribución Normal estándar. En la Figura 4.7 hemos presentado la función de distribución de esta v.a. cuando $p = 49/100$, para diferentes valores de n.

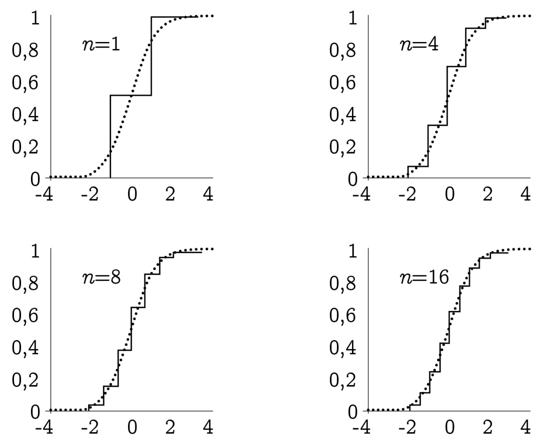


FIGURA 4.7. Función de distribución de la variable aleatoria discreta $\sqrt{n}(\bar{X} - p)/\sqrt{p(1-p)}$, cuando $X_1, \ldots, X_n$ son i.i.d. del tipo Bernoulli con parámetro $p = 0{,}49$ y $n = 1, 4, 8, 16$. Observe cómo las funciones de distribución se parecen más y más a la de una v.a. Normal estándar, mostrada en línea punteada, a medida que n crece.

Imagine que X_i es la indicadora del suceso "un jugador gana la i-ésima apuesta" en un cierto juego de azar donde se tiene una posibilidad $49/100$ de ganar; en particular, $X_1, X_2, \ldots$ son i.i.d. del tipo *Bernoulli*$(49/100)$. Supongamos ahora que, durante varios años, un jugador fanático de este juego lo ha practicado alrededor de 900 veces, *¿cuál es aproximadamente la probabilidad que el jugador haya perdido al menos la mitad de estos juegos?*

La respuesta a la pregunta es la probabilidad del suceso $[\bar{X} \leq 1/2]$. El cálculo exacto de esta probabilidad es bastante complicado. Por otro lado, si $p = 49/100$ y

$n = 900$, entonces

$$\begin{aligned}
\mathbb{P}\left[\bar{X} \leq 1/2\right] &= \mathbb{P}\left[\bar{X} - p \leq \frac{1}{2} - p\right], \\
&= \mathbb{P}\left[\frac{\bar{X} - p}{\sqrt{p(1-p)}} \leq \frac{\frac{1}{2} - p}{\sqrt{p(1-p)}}\right], \\
&= \mathbb{P}\left[\sqrt{n} \cdot \frac{\bar{X} - p}{\sqrt{p(1-p)}} \leq \sqrt{n} \cdot \frac{\frac{1}{2} - p}{\sqrt{p(1-p)}}\right].
\end{aligned}$$

En particular, y debido al comentario en (4.31), tenemos que

$$\mathbb{P}\left[\bar{X} \leq 1/2\right] \approx \Phi\left(\sqrt{n} \cdot \frac{\frac{1}{2} - p}{\sqrt{p(1-p)}}\right) = \Phi\left(\frac{10}{119}\sqrt{51}\right).$$

Como $10\sqrt{51}/119 \approx 0{,}6$, los valores tabulados en la Tabla 4.2 implican que

$$\mathbb{P}\left[\bar{X} \leq 1/2\right] \approx 0{,}7257 \approx 72{,}6\,\%. \tag{4.32}$$

Esto responde la pregunta planteada. También indica que alrededor de 73 jugadores entre 100, habrán perdido dinero después de 900 apuestas. El juego de azar en cuestión es, por lo tanto, ¡muy lucrativo para un casino!

4.7.1 Cotas de error para el Teorema Central Límite.

Hasta ahora hemos aplicado el Teorema Central Límite para aproximar probabilidades asociadas al promedio de varias variables aleatorias independientes e idénticamente distribuidas. Sin embargo, para un valor finito y dado de n cabe preguntarse: *¿cuán precisa es la aproximación*

$$\mathbb{P}\left[a \leq \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma} \leq b\right] \approx \Phi(b) - \Phi(a)?$$

La respuesta viene dada por lo que se llama la *desigualdad de Berry-Esseen* y que enunciamos a continuación [**1**].

Teorema 4.18. *Si $X_1, \ldots, X_n$ son variables aleatorias i.i.d. con esperanza μ y varianza $\sigma^2 > 0$ finitas, entonces, para todo $t \in \mathbb{R}$, aplica que*

$$\left|\mathbb{P}\left[\sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma} \leq t\right] - \Phi(t)\right| \leq \frac{0{,}8}{\sqrt{n}} \cdot \mathbb{E}\left(\left|\frac{X_1 - \mu}{\sigma}\right|^3\right). \tag{4.33}$$

En particular, para todo par de constantes $-\infty \leq a < b \leq +\infty$ se tiene que

$$\left|\mathbb{P}\left[a \leq \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma} \leq b\right] - (\Phi(b) - \Phi(a))\right| \leq \frac{1{,}6}{\sqrt{n}} \cdot \mathbb{E}\left(\left|\frac{X_1 - \mu}{\sigma}\right|^3\right). \tag{4.34}$$

Las desigualdades anteriores cuantifican el error máximo en la aproximación del Teorema Central Límite. Como esta cantidad es inversamente proporcional a $\sqrt{n}$, para reducirla a la mitad es necesario cuadruplicar el número de variables aleatorias que participan del promedio. Dos comentarios importantes a este respecto son los siguientes: (1) el error exacto es en general desconocido, aunque en la mayoría de los casos puede acotarse por debajo, por una cantidad de la forma $c/\sqrt{n}$, con $c > 0$ constante; en particular, las cotas superiores son del orden correcto, y (2) las constantes que no dependen de n al lado derecho de las desigualdades en (4.33) y (4.34) son típicamente mucho más grandes que lo realmente necesario; en particular, para valores pequeños de n, la cuantificación del error basada en la desigualdad de Berry-Esseen es generalmente pesimista. Los siguientes ejemplos aclararán estas observaciones.

Ejemplo 4.31. En el Ejemplo 4.30 vimos que si $X_1, ..., X_n$ son variables aleatorias independientes y del tipo Bernoulli con parámetro común $0 < p < 1$, entonces

$$\sqrt{n} \cdot \frac{\bar{X} - p}{\sqrt{p(1-p)}}$$

tiene una distribución aproximadamente Normal estándar, cuando n es suficientemente grande. Como $\mu = p$ y $\sigma^2 = p(1-p)$, para cuantificar el error en la aproximación usando la Desigualdad de Berry-Esseen, necesitamos determinar el valor esperado

$$\begin{aligned} \mathbb{E}\left(\left|\frac{X_1 - p}{\sqrt{p(1-p)}}\right|^3\right) &= \left|\frac{0-p}{\sqrt{p(1-p)}}\right|^3 \cdot (1-p) + \left|\frac{1-p}{\sqrt{p(1-p)}}\right|^3 \cdot p, \\ &= \sqrt{\frac{p^3}{1-p}} + \sqrt{\frac{(1-p)^3}{p}}. \end{aligned} \tag{4.35}$$

Por otro lado, en el Ejemplo 4.30, aproximamos la probabilidad $\mathbb{P}[\bar{X} \leq 1/2]$ usando el Teorema Central Límite como sigue:

$$\begin{aligned} \mathbb{P}[\bar{X} \leq 1/2] &= \mathbb{P}\left[\sqrt{n} \cdot \frac{\bar{X} - p}{\sqrt{p(1-p)}} \leq \sqrt{n} \cdot \frac{\frac{1}{2} - p}{\sqrt{p(1-p)}}\right], \\ &\approx \Phi\left(\sqrt{n} \cdot \frac{\frac{1}{2} - p}{\sqrt{p(1-p)}}\right). \end{aligned}$$

Para cuantificar el error en esta aproximación, podemos ahora usar la desigualdad en (4.33) y el valor calculado en (4.35), para obtener que

$$\begin{aligned} \left|\mathbb{P}[\bar{X} \leq 1/2] - \Phi\left(\sqrt{n} \cdot \frac{\frac{1}{2} - p}{\sqrt{p(1-p)}}\right)\right| &\leq \frac{0{,}8}{\sqrt{n}} \cdot \mathbb{E}\left(\left|\frac{X_1 - p}{\sqrt{p(1-p)}}\right|^3\right), \\ &= \frac{0{,}8}{\sqrt{n}} \cdot \left\{\sqrt{\frac{p^3}{1-p}} + \sqrt{\frac{(1-p)^3}{p}}\right\}. \end{aligned}$$

Por ejemplo, si $n = 900$ y $p = 49/100$ el lado derecho de lo anterior es 0,026...; en particular, el error incurrido en la aproximación dada en (4.32) no puede ser mayor que un 3 %. Por lo tanto, el suceso $[\bar{X} \leq 1/2]$ tiene una probabilidad exacta entre 69,6 % y 75,6 %.

Ejemplo 4.32. Imagine que los alumnos de un curso grande invierten, en promedio, 5 minutos por cada pregunta asignada en un examen. Si el examen final consiste de 16 preguntas, *¿cuál es la duración mínima del examen para asegurar que alrededor de la mitad de los alumnos terminará a tiempo?*

Para simplificar la notación definiremos $n = 16$. En lo que sigue asumiremos que los tiempos $X_1, \ldots, X_n$ que un alumno invierte en cada problema del examen final son variables aleatorias i.i.d. del tipo exponencial con parámetro $\lambda = 1/5$. (Esto asume que medimos el tiempo en minutos.) En particular, $\mathbb{E}(X_i) = 5$ y $\mathbb{V}(X_i) = 25$. Como el curso en cuestión es grande, para responder a la pregunta anterior, determinaremos un tiempo t tal que el suceso $[\sum_{i=1}^{n} X_i \leq t]$ tiene una probabilidad aproximada de un 50 %. Note, sin embargo, que esto es equivalente a tener que

$$\mathbb{P}\left[\sqrt{n}(\lambda\bar{X} - 1) \leq \frac{t}{20} - 4\right] \approx 0{,}5.$$

Por lo tanto, debido al Teorema Central Límite (ver Identidad (4.30)), necesitamos determinar t tal que $\Phi(t/20 - 4) \approx 0{,}5$. Para esto es suficiente que $(t/20 - 4) = 0$ i.e. que $t = 80$ minutos (*¿por qué?*).

De acuerdo a la Figura 4.6, el supuesto del párrafo anterior i.e. que $\sqrt{n}(\lambda\bar{X} - 1)$ tiene una función de densidad casi idéntica a la de una v.a. Normal estándar, parece razonable cuando $n = 16$. Por lo tanto, $t = 80$ debiera ser una buena aproximación del valor exacto de t para el cual $\mathbb{P}[\sum_{i=1}^{n} X_i \leq t] = 0{,}5$. De hecho, se tiene que (ver Ejercicio 42):

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \leq 80\right] = \int_0^{80} \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}\,dt = 1 - \frac{2648261961071387}{638512875} e^{-16} \approx 0{,}533.$$

Por otro lado, y según hemos explicado, la cota de error dada por la desigualdad de Berry-Esseen es generalmente pesimista cuando n no es suficientemente grande. De hecho, encontramos para el caso bajo estudio que

$$\begin{aligned}
\mathbb{E}\left(\left|\frac{X_1 - 1/\lambda}{1/\lambda}\right|^3\right) &= \mathbb{E}\left(|\lambda X_1 - 1|^3\right), \\
&= \int_0^{\infty} |\lambda x - 1|^3 \cdot \lambda e^{-\lambda x} dx = (12e^{-1} - 2).
\end{aligned}$$

Por lo tanto, la cota máxima para el error, dada por la desigualdad de Berry-Esseen, para la aproximación $\mathbb{P}[\sqrt{n}(\lambda\bar{X} - 1) \leq t/20 - 4] \approx \Phi(t/20 - 4)$, es

$$\frac{0{,}8}{\sqrt{n}} \cdot \mathbb{E}\left(\left|\frac{X_1 - 1/\lambda}{1/\lambda}\right|^3\right) = \frac{2(6e^{-1} - 1)}{5} \approx 48{,}3\,\%.$$

¡Esta cota es bastante exagerada ya que el error exacto es menos de un 3,5 %!

4.8 Ejercicios propuestos

Sección 4.1.

Ejercicio 1. ¿Cuál es la probabilidad del suceso $[X = 0]$, si X es una v.a. del tipo Bernoulli, tal que $\mathbb{E}((X - 1/2)^3) = 1/10$?

Ejercicio 2. Si $X \sim Bernoulli(1/3)$ e $Y \sim Bernoulli(2/3)$, ¿cuál es la probabilidad del suceso $[X \geq Y]$ cuando X e Y son independientes?

Ejercicio 3. Considere el experimento donde se escogen dos personas al azar y se les pregunta el día de la semana en que nacieron i.e. si fue un lunes, martes, miércoles, jueves, viernes, sábado o domingo. Muestre que la función indicadora del suceso "ambas personas nacieron un mismo día de la semana" tiene una distribución del tipo Bernoulli y determine su parámetro.

Sección 4.2.

Ejercicio 4. Explique por qué el número de embarazos, hasta que una cierta madre conciba una niña, es una v.a. geométrica con parámetro $p = 1/2$.

Ejercicio 5. ¿Cuál es la probabilidad del suceso $[X = 2]$, si X es una v.a. geométrica tal que $\mathbb{P}[X \geq 3] = 1/125$?

Ejercicio 6. ¿Cuántos lanzamientos simultáneos de 10 dados equilibrados son necesarios, en promedio, hasta observar todas las caras iguales?

Sección 4.3.

Ejercicio 7. ¿Para qué número entero $n \geq 1$ es posible lo siguiente:
si $X \sim Binomial(n, 1/3)$, entonces $\mathbb{P}[X = 2] = 2/9$?

Ejercicio 8. Si el 20 % de los niños que llegan de urgencia a un cierto hospital sufren de una quebradura de hueso, determine la esperanza y varianza de la v.a. X que cuenta el número de niños con una quebradura de hueso entre 10 admitidos a la sala de Urgencia, ¿cuál es la probabilidad del suceso $[X = 5]$?

Ejercicio 9. Considere una v.a. $X \sim Binomial(n, p)$, tal que $\mathbb{E}(X) = 10$ y $\mathbb{V}(X) = 6$. Determine los parámetros n y p.

Ejercicio 10. Imagine un juego que admite a lo más cinco participantes y que para jugarlo se requieren al menos tres de ellos. Si cinco amigos planean jugarlo y cada amigo se presentará en el lugar y hora acordada con una probabilidad de un 75 %, independientemente del resto, responda: ¿cuál es la probabilidad que al menos tres de los amigos se hagan presentes para el juego?

Sección 4.4.

Ejercicio 11. Considere constantes $\alpha, \beta > 0$. Muestre que si $X \sim Poisson(\alpha)$ e $Y \sim Poisson(\beta)$ son independientes, entonces $(X + Y)$ es también del tipo Poisson y determine su parámetro. Concluya a partir de esto que si $\lambda_1, \ldots, \lambda_n > 0$ son constantes dadas y $X_1, \ldots, X_n$ son variables aleatorias independientes, tales que $X_i \sim Poisson(\lambda_i)$, entonces $\sum_{i=1}^{n} X_i$ es también del tipo Poisson y determine su parámetro.
Hint: Para adivinar el parámetro de la distribución de $(X + Y)$ determine primero su valor esperado.

Ejercicio 12. Si el número de ampolletas que necesitan ser reemplazadas por semana, en un cierto colegio, tiene una distribución de Poisson con parámetro $\lambda = 0{,}1$, responda: ¿cuál es la probablidad que ninguna ampolleta tenga que ser reemplazada en una semana típica?, ¿cuál es la probablidad que exactamente una ampolleta tenga que ser reemplazada en una semana típica?, ¿cuál es la probablidad que dos o más ampolletas tengan que ser reemplazadas en una semana típica? Use el Ejercicio 11 para explicar por qué el número de ampolletas que necesitarán ser reemplazadas en un mes típico es una v.a. del tipo Poisson y determine su parámetro.

Ejercicio 13. Imite el razonamiento utilizado en el Ejemplo 4.17 para justificar que, si en un cierto hospital, se despachan en promedio 2 ambulancias durante cada hora de la madrugada, entonces el número N de despachos entre la 1:00am y 6:00am satisface que $N \sim Poisson(10)$. ¿Cuál es la probabilidad de que ninguna ambulancia sea despachada entre la 1:00am y 6:00am en una noche típica?, ¿cuál es la probabilidad de que se despachen a lo más 6 ambulancias entre la 1:00am y 6:00am en una noche típica?

Sección 4.5.

Ejercicio 14. Si $X \sim Exponencial(\lambda)$, ¿cuál es el valor de λ si $\mathbb{E}(X^3) = 2$?
Hint: Use el Método de Integración por Partes, para determinar el valor esperado de X^3.

Ejercicio 15. Muestre que si $X \sim Exponencial(\lambda)$ entonces $\mathbb{P}[a + s < X \leq b + s \mid X > s] = e^{-\lambda a} - e^{-\lambda b} = \mathbb{P}[a < X \leq b]$, para todo $0 \leq a < b$ y $s > 0$. Explique en palabras lo que esta identidad significa.

Ejercicio 16. Considere dos secretarias que comparten una oficina: Marta recibe en promedio 6 llamadas telefónicas por hora, mientras que Carmen recibe 10. (a) Explique por qué la separación temporal entre dos llamadas consecutivas que recibe Marta es una v.a. exponencial, ¿cuál es el parámetro de la distribución? (b) ¿Qué hay de la otra secretaria? (c) Si en un momento dado ambas secretarias están libres del teléfono, ¿cuál es la probabilidad que ninguna reciba una llamada en las próximas t horas? Explique los supuesto hechos y el rol de la propiedad desmemoriada en su cálculo.

Ejercicio 17. Use el Teorema 2.19 para mostrar que si $X \sim Exponencial(\lambda)$ e $Y \sim Exponencial(\lambda)$ son variables aleatorias independientes, entonces la v.a. $Z = (X+Y)$ tiene una función de densidad dada por la fórmula

$$f_Z(z) = \begin{cases} \lambda^2 z e^{-\lambda z}, & z \geq 0; \\ 0, & z < 0. \end{cases}$$

Sección 4.6.

Ejercicio 18. Muestre que la función de densidad de una v.a. Normal con media μ y varianza σ^2 es simétrica alrededor de μ, alcanza su máximo absoluto en μ, y $(\mu - \sigma)$ y $(\mu + \sigma)$ son sus únicos puntos de inflección.

Ejercicio 19. Considere una v.a. $X \sim Normal(\mu, \sigma^2)$, tal que los sucesos $[X \geq 1]$ y $[X \leq -3]$ tienen la misma probabilidad. Utilice esto para determinar μ, y explique por qué no hay información suficiente para determinar σ.

Ejercicio 20. (a) Si $X \sim Normal(0, \sigma^2)$ es tal que $\mathbb{P}[X \leq 4{,}52] = 0{,}8707$, ¿cuál es aproximadamente el valor de σ? (b) Si $Y \sim Normal(\mu, 9)$ es tal que $\mathbb{P}[Y \geq 14{,}13] = 0{,}0034$, ¿cuál es aproximadamente el valor de μ?
Hint: Utilice la Tabla 4.2.

Ejercicio 21. Si X e Y son variables aleatorias independientes tales que $X \sim Normal(0, 1/4)$ e $Y \sim Normal(-1, 16)$, ¿cuál es la distribución de $(8X - Y)$?

Ejercicio 22. El peso de una barra de chocolate sin envoltorio es una v.a. Normal con media 300 gramos y desviación estándar 1 gramo. Por otro lado, el peso del envoltorio es una v.a. Normal con media 10 gramos y desviación estándar 0,1 gramos. Asumiendo que el peso de la barra y del envoltorio son variables aleatorias independientes, responda: ¿cuál es la distristribución del peso de una barra de chocolate envuelta?, ¿cuál es la distristribución del peso de 25 barras de chocolate con sus respectivos envoltorios? Explique los supuestos hechos en su cálculo.
Hint: Utilice el Teorema 4.15.

Ejercicio 23. En este ejercicio indicaremos cómo demostrar que la densidad de una v.a. del tipo Normal, es efectivamente una función de densidad. El ejercicio usa integrales dobles y coordenadas polares, y puede ser omitido por aquellos lectores no familiarizados con estos conceptos.

(a) Muestre que

$$\int_{-\infty}^{+\infty} \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}\,dx = \int_{-\infty}^{+\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = \sqrt{\frac{2}{\pi}} \cdot \int_0^{+\infty} e^{-x^2/2}\,dx.$$

(b) Muestre que

$$\left(\int_0^{\infty} e^{-x^2/2}\,dx\right)^2 = \int_0^{\infty}\int_0^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy = \frac{\pi}{2}.$$

Hint: Para la primera identidad consulte la Sección 2.6.2. Para la segunda identidad use que si $f(x,y)$ es una función continua y no negativa, para todo $x, y \geq 0$, entonces

$$\int_0^\infty \int_0^\infty f(x,y)\,dx\,dy = \int_0^\infty \int_0^{\pi/2} r \cdot f\big(r\cos(\theta), r\sin(\theta)\big)\,d\theta\,dr.$$

(c) Concluya que la densidad de una v.a. del tipo Normal es efectivamente una función de densidad.

Sección 4.7.

Ejercicio 24. Considere una v.a. $X \sim Binomial(100, 1/2)$. Use el Teorema Central Límite para aproximar la probabilidad del suceso $[X = 51]$.
Hint: Observe que $\mathbb{P}[X = 51] = \mathbb{P}[50{,}5 \leq X \leq 51{,}5]$ (*¿por qué?*).

Ejercicio 25. Imagine que se lanza un dado equilibrado 10.000 veces, ¿qué suceso tiene una probabilidad aproximada de un 50 %?, ¿qué suceso tiene una probabilidad aproximada de un 95,4 %? Justifique sus respuestas usando el Teorema Central Límite.
Hint: Para la segunda pregunta, consulte la Figura 4.5.

Ejercicio 26. Considere variables aleatorias $X_1, \ldots, X_{900}$ como las dadas en el Ejemplo 4.30. (a) Muestre que

$$\mathbb{P}[\bar{X} \leq 1/2] = \sum_{n=0}^{450} \binom{900}{n} \left(\frac{49}{100}\right)^n \left(\frac{51}{100}\right)^{900-n}.$$

Observe que la expresión al lado derecho de lo anterior es exacta, sin embargo, es poco claro cuán grande o pequeña es esta probabilidad a partir de ella. (b) Explique por qué la aproximación de la distribución binomial, usando la distribución de Poisson (Teorema 4.7), no es apropiada en este caso. (El Teorema Central Límite es, por lo tanto, la única herramienta útil que hemos discutido para aproximar probabilidades como la anterior.)

Ejercicio 27. Suponga que un niño consume en promedio 1 litro de agua cada tres horas. (a) En un paseo de curso de tres horas y con 40 niños, ¿cuál es aproximadamente la probabilidad de quedarse sin agua si el profesor de curso sólo llevará 38 litros de agua para el consumo de los niños? Asuma que la desviación estándar del consumo de agua de cada niño es 0,4 litros. (b) ¿Cómo cambia su respuesta si la desviación estándar del consumo de agua es 0,1 litros? (c) En cada uno de los dos casos presentados, ¿cuál es aproximadamente el mínimo número de litros de agua que el profesor tendrá que llevar al paseo para que la probabilidad de quedarse sin agua después de tres horas es menos de un 1 %?

Problemas misceláneos Capítulo 4.

Ejercicio 28. Considere parámetros $0 < p_1, p_2 < 1$ y variables aleatorias independientes X_1, X_2, tales que $X_i \sim Bernoulli(p_i)$. Muestre que $X_1 \cdot X_2$ es también una v.a. del tipo Bernoulli y determine su parámetro.

Ejercicio 29. Considere una urna con 5 bolitas azules y 4 rojas. Imagine que tres bolitas se sacan de la urna simultáneamente y un cierto número X de ellas son azules, ¿es X una v.a. del tipo binomial? Corrobore su respuesta determinando la probabilidad del suceso $[X = k]$, para cada $k \in \{0, 1, 2, 3\}$.
Hint: Para corroborar su respuesta considere el espacio muestral Ω de todos los subconjuntos de $\{a_1, a_2, a_3, a_4, a_5, r_1, r_2, r_3, r_4\}$ de cardinalidad tres, donde a_i representa la i-ésima bolita azul y r_j la j-ésima bolita roja.

Ejercicio 30. Muestre que si $X \sim Normal(0, 1)$, entonces la v.a. X^2 tiene una función de densidad dada por la fórmula

$$f(x) = \begin{cases} 0, & x \leq 0; \\ \frac{e^{-x/2}}{\sqrt{2\pi x}}, & x > 0. \end{cases}$$

Ejercicio 31. Muestre usando la serie de Taylor de la función exponencial que si $X \sim Poisson(\lambda)$, entonces, $\mathbb{E}(X \cdot (X-1) \cdots (X-k)) = \lambda^{k+1}$, para todo número entero $k \geq 0$.

Ejercicio 32. Muestre que si $X \sim Normal(\mu, \sigma^2)$, y $z > 0$ y $0 < \alpha < 1$ son constantes, tales que $\Phi(z) = (1-\alpha)$, entonces $\mathbb{P}[X \leq \sigma \cdot z + \mu] = (1-\alpha)$. Más generalmente, dados $a < b$, tales que $\Phi(b) - \Phi(a) = (1-\alpha)$, determine la probabilidad del suceso $[\mu + a \cdot \sigma \leq X \leq \mu + b \cdot \sigma]$.

Ejercicio 33. Clasifique cada una de las siguientes variables aleatorias como una v.a. del tipo geométrica, binomial o Poisson. En cada caso, especifique los parámetros de la distribución:

X_1 = el número de genios por nacer en los próximos 300 años con un coeficiente intelectual mayor o igual al de Einstein, si en promedio, sólo un ser humano cada 150 años nace con un coeficiente intelectual igual o superior al de éste;

X_2 = el número de personas que debe abordar un mendigo, antes de recibir su primera limosna del día, si la probabilidad de que una persona al azar entrege una limosna es 0,03;

X_3 = el número de computadores defectuosos, devueltos a una multitienda, entre 15 vendidos, si la probabilidad de que un computador salga defectuoso es de 1 %, y la probabilidad de que una persona devuelva un computador defectuoso es de un 99 %.

Ejercicio 34. Muestre que si $X \sim Geométrica(p)$ e $Y \sim Geométrica(p)$ son variables aleatorias independientes, entonces $\mathbb{P}[X = k \mid X + Y = n] = 1/(n-1)$, para cada

$n \geq 2$ y $k \in \{1, \ldots, n-1\}$. Equivalentemente, X es igualmente probable de tomar cualquier valor entre 1 y $(n-1)$ cuando condicionamos sobre el suceso $X + Y = n$. (Para fijar ideas considere el experimento que consiste en lanzar un dado equilibrado hasta que la cara 6 haya salido dos veces. Si esto requirió en total 10 lanzamientos, entonces la probabilidad que el primer 6 haya salido en el primer lanzamiento es 1/9.)

Ejercicio 35. Imagine que el promedio de notas de un alumno es 5,905. Si las n notas del alumno son variables aleatorias i.i.d. Normales con media μ y desviación estándar 0,25, ¿cuál es él número n más pequeño para el cual un profesor tendría un 99 % de confianza que $\mu < 6$?, ¿cómo cambia la respuesta anterior si el profesor requiere una certeza de un 95 %?

Ejercicio 36. Use la desigualdad de Berry-Esseen para demostrar el Teorema Central Límite cuando existe una constante $c > 0$, tal que $|X| \leq c$, y X tiene una varianza estrictamente positiva.

Ejercicio 37. Use el Teorema 4.5 para determinar la distribución de la variable $(X + Y)$, donde $X \sim Binomial(n, p)$ e $Y \sim Binomial(m, p)$ son independientes.

Ejercicio 38. Considere una encuesta donde a cada niño, de un total de 75, se le pide decir su número entero favorito entre 1 y 5. Defina X como el número de niños que nombran "tres" como su número favorito. Asumiendo que cada niño nombra su número favorito sin dejarse influenciar por la respuesta de los otros niños y que cada número es igualmente probable que cualquier otro, responda: (a) ¿cuál es el valor esperado y varianza de X?; (b) ¿cuál es la probabilidad del suceso $[X = 0]$?; (c) ¿cómo se compara la respuesta anterior con la entregada por la aproximación de Poisson?, explique si el error en la aproximación es del todo sorprendente; (d) ¿cómo se compara la probabilidad del suceso $[X = 0]$ con la dada usando el Teorema Central Límite?, explique si el error en la aproximación es del todo sorprendente.

Ejercicio 39. (a) Muestre que si X es una v.a. del tipo exponencial con parámetro λ, entonces $\lambda X \sim Exponencial(1)$. (b) Use lo anterior para justificar que si $X_1, \ldots, X_n$ son i.i.d. del tipo exponencial con parámetro λ, entonces la distribución de $\sqrt{n}(\lambda\bar{X} - 1)$ no depende de λ.
Hint: Use el Teorema 2.11 para la parte (a).

Ejercicio 40. Muestre que si $X \sim Geométrica(p)$, entonces, para todo par de constantes $k, l \geq 1$, se tiene que $\mathbb{P}[X - k = l \mid X > k] = (1-p)^{l-1} \cdot p = \mathbb{P}[X = l]$. En palabras, la v.a. $(X - k)$, cuando condicionamos sobre el suceso $[X > k]$, tiene la misma distribución que la variable X. (Para fijar ideas considere el número de lanzamientos X de un dado equilibrado, hasta observar por primera vez el número 3. Claramente, $X \sim Geométrica(1/6)$. Sin embargo, si en los primeros $k = 7$ lanzamientos no hemos observado ninguna vez el número 3 i.e. si sabemos que $X > 7$, entonces el número de lanzamientos adicionales hasta observar por primera vez este número i.e. la cantidad $(X - 7)$, también debiera tener una distribución geométrica con parámetro $p = 1/6$.)

Ejercicio 41. La Sociedad Mundial de la Salud estima que el número promedio de casos de Influenza Porcina es 1 cada 28 días, por cada 10,000 habitantes que residen en una zona urbana de 10 km^2. En particular, el número de casos de Influenza Porcina en la ciudad de Santiago (población=5 millones, área urbana=642 km^2) en un periodo de una semana tiene una distribución de Poisson. ¿Cuál es el parámetro de esta distribución? Aplique esto para determinar la probabilidad de al menos un caso de Influenza Porcina en la ciudad de Santiago, en un periodo de una semana.

Ejercicio 42. Muestre usando Inducción Matemática que, para todo $n \geq 1$, si $X_1, \ldots, X_n$ son variables aleatorias i.i.d. del tipo exponencial con parámetro λ, entonces $\sum_{i=1}^{n} X_i$ tiene función de densidad:

$$f(t) = \begin{cases} \lambda^n t^{n-1} e^{-\lambda t}/(n-1)!, & \text{si } t \geq 0; \\ 0, & \text{si } t < 0. \end{cases}$$

Una v.a. continua con función de densidad como la anterior, se dice tener una *distribución gama con parámetros* (n, λ).

Ejercicio 43. Considere variables aleatorias $X_1, \ldots, X_n$ i.i.d. del tipo exponencial con parámetro λ. Use la desigualdad de Berry-Esseen para acotar el error incurrido al aproximar probabilidades relacionadas con la v.a. $\sum_{i=1}^{n} X_i$ usando el Teorema Central Límite.

Ejercicio 44. Muestre que si $Y \sim Binomial(n, p)$, donde n es extremadamente grande y p extremadamente cercano a uno entonces, para todo $k \in \{0, \ldots, n\}$ la probabilidad del suceso $[Y = n - k]$ es aproximadamente la probabilidad del suceso $[Z = k]$, con $Z \sim Poisson(n \cdot (1 - p))$. Determine una cota superior para el error en la aproximación.

Comentarios de final de capítulo

La *función característica* de una v.a. X se define como la función $\varphi_X(\theta) = \mathbb{E}(e^{i\theta X})$, para todo $\theta \in \mathbb{R}$. Esta función, al igual que la función de distribución, caracteriza completamente la distribución de la variable aleatoria: *X e Y tienen la misma distribución si y solo si* $\varphi_X = \varphi_Y$. Por ejemplo, si $X \sim Normal(0, 1)$ entonces $\varphi_X(\theta) = \exp(-\theta^2/2)$. En particular, si Y es otra v.a. tal que $\varphi_Y(\theta) = \exp(-\theta^2/2)$, entonces Y es también una variable aleatoria Normal estándar. Otras propiedades importantes acerca de las funciones características, son las siguientes:

(a) $\varphi_{\alpha X + \beta}(\theta) = e^{i\beta\theta} \cdot \varphi_X(\alpha\theta)$, para todo par de constantes $\alpha, \beta \in \mathbb{R}$.

(b) si $X_1, \ldots, X_n$ son variables aleatorias independientes, entonces

$$\varphi_{\sum\limits_{i=1}^{n} X_i} = \prod_{i=1}^{n} \varphi_{X_i}.$$

(c) si $X_1, X_2, \ldots$ son variables aleatorias tales que $\lim_{n\to\infty} \varphi_{X_n}(\theta) = \varphi_X(\theta)$, para todo $\theta \in \mathbb{R}$, entonces $\lim_{n\to\infty} \mathbb{P}[a \leq X_n \leq b] = \mathbb{P}[a \leq X \leq b]$, para todo par de constantes $-\infty < a \leq b < +\infty$ tales que F_X es continua sobre a y b.

El Teorema Central Límite es en gran parte consecuencia de las tres propiedades anteriores. De hecho, si $X_1, X_2, \ldots$ son variables aleatorias i.i.d. con esperanza μ y varianza $\sigma^2 > 0$ finitas, entonces existe una secuencia de números reales $\epsilon_1, \epsilon_2, \ldots$ tales que $\lim_{n\to\infty} n\epsilon_n = 0$, y

$$\lim_{n\to\infty} \varphi_{\sqrt{n}\cdot\frac{\bar{X}-\mu}{\sigma}}(\theta) = \lim_{n\to\infty} \left(1 - \frac{\theta^2}{2n} + \epsilon_n\right)^n = \exp(-\theta^2/2).$$

En particular, como la función de distribución de una v.a. Normal estándar es continua, la propiedad (c) anterior implica que:

$$\lim_{n\to\infty} \mathbb{P}\left[a \leq \sqrt{n}\cdot\frac{\bar{X}-\mu}{\sigma} \leq b\right] = \Phi(b) - \Phi(a),$$

para todo $-\infty < a \leq b < +\infty$, estableciendo así el Teorema Central Límite.

Capítulo 5: Simulación Estocástica

En este último capítulo presentaremos dos métodos para simular variables aleatorias. Ambos métodos requieren, y de hecho asumen, que es posible simular una variable aleatoria uniforme en el intervalo $(0, 1)$. ¡Esto último es un requerimiento de toda simulación!

En la Sección 5.1 presentaremos el Método de la Transformada Inversa. Este método puede utilizarse para simular en teoría cualquier tipo de variable aleatoria, ya sea discreta, continua, o incluso una mezcla de ambos tipos. En la práctica, sin embargo, este método presenta ciertas limitaciones para simular variables aleatorias continuas. Debido a esto presentaremos un segundo método, el Método de Rechazo, que puede implementarse para simular una gran diversidad de variables aleatorias continuas. El lector interesado en profundizar en este tema podrá consultar la referencia [**2**].

A lo largo del capítulo presentaremos una multiplicidad de algoritmos para simular variables aleatorias en específico. Se invita al lector a implementar estos algoritmos en una computadora usando paquetes matemáticos como EXCEL®, MAPLE®, MATHEMATICA® o MATLAB®, o su lenguaje computacional favorito. Las herramientas que presentaremos permitirán al lector osado simular los experimentos de esencialmente cualquier ejemplo o ejercicio propuesto en esta monografía: el lanzamiento de dados y monedas, extracción de bolitas de una urna o cartas de mazos, e incluso los tiempos de llamadas de teléfono a una oficina o de ocurrencia de sismos.

5.1 Método de la Transformada Inversa

El Método de la Transformada Inversa consiste en simular una v.a. X utilizando su función de distribución y una variable aleatoria uniforme en el intervalo abierto $(0, 1)$. Para implementar el método necesitamos encontrar primero una función $G : (0, 1) \to \mathbb{R}$ que satisfaga la siguiente propiedad:

$$G(u) \leq v \text{ si y solo si } u \leq F_X(v), \tag{5.1}$$

para todo $u \in (0, 1)$ y $v \in \mathbb{R}$. Previsto lo anterior, el siguiente resultado nos permitirá simular la v.a. X.

Teorema 5.1. *Si X es una v.a. con función de distribución F_X y G satisface la condición dada en (5.1), entonces la v.a. $Y = G(U)$, con $U \sim Uniforme(0, 1)$, tiene la misma distribución que X.*

Demostración. Basta con demostrar que $F_X = F_Y$ (Teorema 2.11). En efecto, dado $x \in \mathbb{R}$, y debido a la propiedad en (5.1), obtenemos que

$$\begin{aligned} F_Y(x) = \mathbb{P}[Y \leq x] &= \mathbb{P}[G(U) \leq x], \\ &= \mathbb{P}[U \leq F_X(x)] = \mathbb{P}[U \in [0, F_X(x)]] = F_X(x), \end{aligned}$$

donde para la última identidad hemos usado que $F_X(x) \in [0,1]$. Esto demuestra el teorema. $\square$

La existencia de una función G que satisface la condición dada en (5.1) es directa cuando $F_X : \mathbb{R} \to [0,1]$ es estrictamente creciente. En este caso, basta con considerar $G = F_X^{-1}$ i.e. la función inversa de F_X. En efecto, si F_X es estrictamente creciente entonces F_X^{-1} está bien definida y es también estrictamente creciente (*¿por qué?*). Por lo tanto, para todo $u \in (0,1)$ y $v \in \mathbb{R}$ se tienen las siguientes equivalencias lógicas:

$$G(u) \leq v \Longleftrightarrow F_X^{-1}(u) \leq v \Longleftrightarrow F_X(F_X^{-1}(u)) \leq F_X(v) \Longleftrightarrow u \leq F_X(v),$$

lo que demuestra (5.1).

Ejemplo 5.1. Para fijar ideas considere una v.a. X con la función de distribución dada en la Figura 5.1. Como F_X es estrictamente creciente, entonces $F_X^{-1}(0{,}2) = -2{,}37$, $F_X^{-1}(0{,}4) = 0$, $F_X^{-1}(0{,}6) = 2{,}01$, y $F_X^{-1}(0{,}8) = 4{,}37$. Para simular X basta con simular una v.a. $U \sim Uniforme(0,1)$ y, por ejemplo, asignar a X el valor $-2{,}37$ si $U = 0{,}2$, o asignar a X el valor $4{,}37$ si $U = 0{,}8$. Por otro lado, si $U = 0{,}4$, entonces $X = 0$; y si $U = 0{,}6$, entonces $X = 2{,}01$. Más generalmente, si $U = u$, entonces $X = x$, donde x es la única solución de la ecuación $F_X(x) = u$.

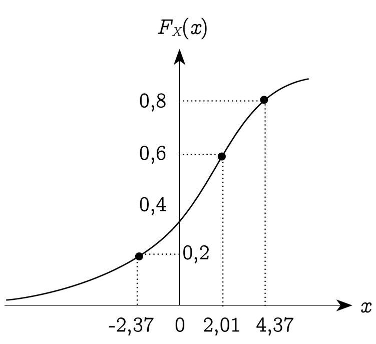


Figura 5.1

Ejemplo 5.2. Considere el experimento donde se escoge un número U uniformemente distribuido en el intervalo $(0,1]$ e independientemente se lanza una moneda que tiene probabilidad p de salir cara y $(1-p)$ de salir sello. Considere la variable aleatoria

$$X = \begin{cases} 1 - \frac{1}{U}, & \text{si la moneda sale cara;} \\ \frac{1}{U} - 1, & \text{si la moneda sale sello.} \end{cases} \tag{5.2}$$

Observe que X puede tomar en principio cualquier valor real. Esto se debe a que el rango de las funciones $u \to (1-1/u)$ y $u \to (1/u-1)$, con $0 < u \leq 1$, son los intervalos $(-\infty, 0]$ y $[0, +\infty)$, respectivamente (*¿por qué?*).

Para determinar F_X distinguiremos dos casos. Si $v < 0$, entonces la Fórmula de Probabilidades Totales implica que

$$\begin{aligned} F_X(v) = \mathbb{P}[X \leq v] &= \mathbb{P}\left[1 - \frac{1}{U} \leq v \,\middle|\, \text{moneda=cara}\right] \cdot \mathbb{P}[\text{moneda=cara}] \\ &\qquad + \mathbb{P}\left[\frac{1}{U} - 1 \leq v \,\middle|\, \text{moneda=sello}\right] \cdot \mathbb{P}[\text{moneda=sello}], \\ &= \mathbb{P}\left[U \leq \frac{1}{1-v} \,\middle|\, \text{moneda=cara}\right] \cdot p + 0 \cdot (1-p), \\ &= \mathbb{P}\left[U \leq \frac{1}{1-v}\right] \cdot p = \frac{p}{1-v}, \end{aligned}$$

donde, para la penúltima igualdad, hemos usado que el lanzamiento de la moneda es independiente del valor que U tomará. Por otro lado, si $v \geq 0$, entonces se encuentra similarmente que

$$F_X(v) = \frac{p+v}{1+v}.$$

En resumen,

$$F_X(v) = \begin{cases} \frac{p}{1-v}, & v < 0; \\ \frac{p+v}{1+v}, & v \geq 0. \end{cases}$$

Observe que F_X es estrictamente creciente (ver Figura 5.2) y, por lo tanto, F_X admite una función inversa. Para determinar esta función distinguiremos nuevamente dos casos. Si $0 < u < p$, entonces las siguientes equivalencias lógicas aplican:

$$F_X(v) = u \iff \frac{p}{1-v} = u \iff v = \frac{u-p}{u},$$

y si $p \leq u < 1$, entonces

$$F_X(v) = u \iff \frac{p+v}{1+v} = u \iff v = \frac{u-p}{1-u}.$$

La función inversa de F_X está, por lo tanto, dada por la fórmula

$$G(u) = \begin{cases} \frac{u-p}{u}, & 0 < u < p; \\ \frac{u-p}{1-u}, & p \leq u < 1. \end{cases}$$

De acuerdo al Método de la Transformada Inversa, si $U \sim Uniforme(0,1)$, entonces

$$G(U) = \begin{cases} \frac{U-p}{U}, & 0 < U < p; \\ \frac{U-p}{1-U}, & p \leq U < 1; \end{cases} \tag{5.3}$$

tiene la misma distribución que la v.a. X.

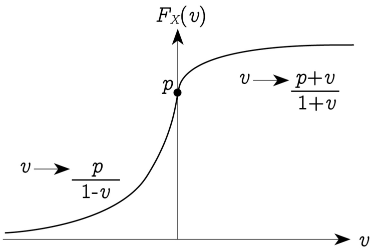


FIGURA 5.2

Lamentablemente, en muchos casos de interés la función de distribución F_X es creciente, pero no estrictamente creciente y, por lo tanto, su función inversa no está definida (*¿por qué?*). En estos casos, la implementación del Método de la Transformada Inversa depende del siguiente resultado. Recuerde que F_X es siempre una función creciente y continua por la derecha.

Teorema 5.2. *Si $F : \mathbb{R} \to [0,1]$ es una función creciente y continua por la derecha, entonces $G : (0,1) \to \mathbb{R}$ definida como $G(u) = \text{ínf}\{v \in \mathbb{R} : F(v) \geq u\}$ satisface la condición dada en (5.1).*

Demostración. Primero demostraremos que

$$F(G(u)) \geq u, \tag{5.4}$$

para todo $u \in \mathbb{R}$. En efecto, como $G(u) = \text{ínf}\{t \in \mathbb{R} : F(t) \geq u\}$, entonces existe una secuencia de números reales $(t_n)_{n\geq 1}$, tal que $t_n \downarrow G(u)$ y $F(t_n) \geq u$, para todo n. Pero observe que $\text{lím}_{n\to\infty} F(t_n) = F(G(u))$ debido a que F es continua por la derecha. Como $F(t_n) \geq u$ obtenemos que $F(G(u)) = \text{lím}_{n\to\infty} F(t_n) \geq u$, lo que demuestra (5.4).

Demostraremos ahora que G satisface la condición dada en (5.1). Para esto considere $u \in (0,1)$ y $v \in \mathbb{R}$. Si $G(u) \leq v$ y como F es creciente, obtenemos que $F(G(u)) \leq F(v)$. En particular, y de acuerdo a la desigualdad en (5.4), $u \leq F(v)$. Recíprocamente, si $u \leq F(v)$, entonces $v \in \{t \in \mathbb{R} : F(t) \geq u\}$ y, por lo tanto, $v \geq \text{ínf}\{t \in \mathbb{R} : F(t) \geq u\} = G(u)$. Esto demuestra la condición en (5.1) y concluye la demostración del teorema. □

Ejemplo 5.3. Para fijar ideas considere la función F dada en la Figura 5.3. La función G toma los siguientes valores debido a las siguientes razones: (a) $G(1) = 3$, porque $F(v) \geq 1$ si y solo si $v \geq 3$, (b) $G(0{,}75) = 2{,}5$, porque $F(v) \geq 0{,}75$ si y solo si $v \geq 2{,}5$, (c) $G(0{,}5) = 2{,}1$, porque $F(v) \geq 0{,}5$ si y solo si $v \geq 2{,}1$, y (d) $G(0{,}25) = 1$, porque $F(v) \geq 0{,}25$ si y solo si $v \geq 1$.

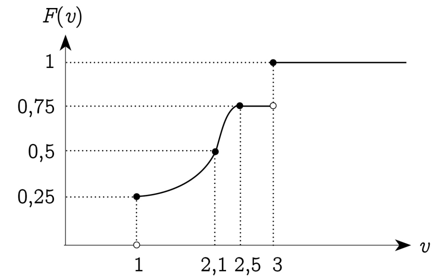


FIGURA 5.3

Una consecuencia importante de la discusión anterior, y que adelantamos en la Sección 2.2, es la siguiente: *si F es una función de distribución i.e. F es creciente y continua a la derecha con límite a la izquierda y tal que* $\lim_{t\to-\infty} F(t) = 0$ *y* $\lim_{t\to+\infty} F(t) = 1$ *(Teorema 2.2), entonces existe una v.a. X tal que $F = F_X$.* La demostración de este hecho es una consecuencia directa de los teoremas 5.1 y 5.2: si $U \sim Uniforme(0,1)$, entonces $G(U) = \inf\{v \in \mathbb{R} : F(v) \geq U\}$ es una v.a. con función de distribución F.

Ejemplo 5.4. Recordemos que si (X,Y) es un punto equidistribuido en el círculo unitario, entonces $R = \sqrt{X^2+Y^2}$ tiene una función de distribución dada por la fórmula (Ejemplo 2.12):

$$F_R(r) = \begin{cases} 0, & r < 0; \\ r^2, & 0 \leq r < 1; \\ 1, & r \geq 1. \end{cases}$$

Para simular R bastaría con simular (X,Y) y luego determinar la distancia entre este punto y el origen. Sin embargo, podemos simular R directamente usando el Método de la Transformada Inversa. Como F_R no es estrictamente creciente (e.g. $(-1/2) < 0$, pero $F_R(-1/2) = F_R(0) = 0$), para implementar el método sólo necesitamos determinar explícitamente para $u \in (0,1)$ la función

$$G(u) = \inf\{v \in \mathbb{R} : F_R(v) \geq u\}.$$

Observe que si $0 < u < 1$, entonces $0 < \sqrt{u} < 1$ y $F_R(\sqrt{u}) = (\sqrt{u})^2 = u$. Como F_R es estrictamente creciente en el intervalo abierto $(0,1)$ (*¿por qué?*), concluimos que $F_R(v) \geq u$ si y solo si $v \geq \sqrt{u}$, y la igualdad $F_R(v) = u$ es posible solamente cuando $v = \sqrt{u}$ (ver Figura 5.4). Por lo tanto, $G : (0,1) \to \mathbb{R}$ está alternativamente dada por la fórmula

$$G(u) = \sqrt{u}.$$

En particular, si $U \sim Uniforme(0,1)$, entonces $\sqrt{U}$ y R tienen la misma distribución. Una consecuencia inesperada de esto es que $R^2 \sim Uniforme(0,1)$. Lo anterior se debe a que la función $x \longrightarrow x^2$ es continua y, por lo tanto, R^2 y $(\sqrt{U})^2$ tienen la misma distribución (Teorema 2.12). En otras palabras: *si escogemos un punto* (X, Y) *equidistribuido en el círculo unitario, entonces* $(X^2 + Y^2)$ *i.e. el cuadrado de su distancia al origen tiene una distribución uniforme en el intervalo* $[0, 1]$.

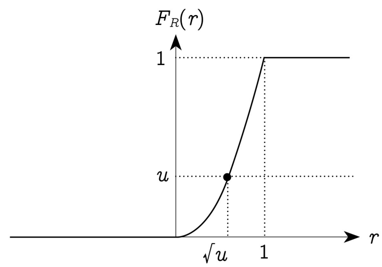


Figura 5.4

5.1.1 Simulación de variables aleatorias exponenciales.

Usando el Método de la Transformada Inversa podemos simular fácilmente variables aleatorias exponenciales. Para esto, observe que si $X \sim \mathit{Exponencial}(\lambda)$, entonces $F_X(t) = 1 - e^{-\lambda t}$, para todo $t \geq 0$; en particular, si $u \in (0, 1)$, entonces existe un único $v_0 > 0$ tal que $F_X(v_0) = u$ (ver Figura 5.5). De hecho, resolviendo la ecuación $1 - e^{-\lambda v_0} = u$ se encuentra que $v_0 = -\ln(1 - u)/\lambda$. Más aún, según se aprecia en la Figura 5.5, $F_X(v) \geq u$ si y solo si $v \geq v_0$. Por lo tanto, la función G en el Teorema 5.2 viene en este caso dada por la fórmula

$$G(u) = \text{ínf}\{v \in \mathbb{R} : \ 1 - e^{-\lambda v} \geq u\} = -\frac{\ln(1 - u)}{\lambda}.$$

De acuerdo al Teorema 5.1, si $U \sim Uniforme(0, 1)$, entonces la variable aleatoria $Y = -\ln(1 - U)/\lambda$ es del tipo exponencial con parámetro λ. Note, sin embargo, que U tiene la misma distribución que $(1 - U)$ (*¿por qué?*). En particular, y debido al Teorema 2.12, tenemos que

$$Z = -\frac{\ln U}{\lambda} \sim \mathit{Exponencial}(\lambda).$$

Para apreciar una posible aplicación de la discusión anterior, imagine que el bibliotecario de un colegio determinado quiere simular los tiempos en que los próximos n estudiantes entrarán a la biblioteca (e.g. para ver cómo mejorar el pedido de libros). Si λ denota el número promedio de estudiantes que entran a la biblioteca por unidad de tiempo, entonces el tiempo de llegada X_k del k-ésimo estudiante viene dado por la variable aleatoria

$$X_k = T_1 + \ldots + T_k = X_{k-1} + T_k,$$

donde $T_1, T_2, \ldots$ son i.i.d. con distribución exponencial con parámetro λ (ver Sección 4.5). Usando el Método de la Transformada Inversa podemos simular cada una

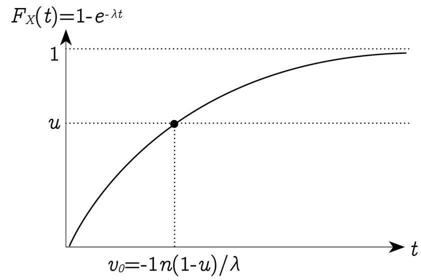


FIGURA 5.5

de las variables $T_1, \ldots, T_n$ y, por lo tanto, también los tiempos aleatorios de llegada $X_1, \ldots, X_n$ de los próximos n estudiantes. El algoritmo que resulta se describe en la Tabla 5.1.

En el algoritmo, la variable k mantiene cuenta de las variables $X_1, \ldots, X_n$ que ya han sido simuladas; en particular, el algoritmo comienza con $k = 0$ y termina cuando $k = n$. La variable X_k es asignada el valor $X_{k-1} + T_k$. Una observación importante (y que aplica a todos los algoritmos que presentaremos) es que las simulaciones en cada línea son independientes de las simulaciones ejecutadas previamente en cualquiera de ellas.

Línea 1:	Asigne $k = 0$ y $X_0 = 0$.
Línea 2:	Simule $U \sim Uniforme(0, 1)$.
Línea 3:	Si $k < n$, entonces asigne $k = k + 1$, $X_k = X_{k-1} - \ln(U)/\lambda$, y vuelva a la línea 2. Si no, termine.

TABLA 5.1. Algoritmo basado en el Método de la Transformada Inversa para simular los tiempos de llegada de los próximos n estudiantes a una biblioteca donde λ estudiantes llegan en promedio por unidad de tiempo.

5.1.2 Simulación de variables aleatorias discretas.

En esta sección especializaremos el Método de la Transformada Inversa para simular variables aleatorias discretas. Para fijar la idea principal consideraremos primero un ejemplo ilustrativo.

Ejemplo 5.5. Considere el problema de simular el lanzamiento de una moneda con probabilidad $3/4$ para cara y $1/4$ para sello. Identificando "cara" con el valor 0 y "sello" con 1 obtenemos una v.a. X que sólo puede tomar valores en el conjunto finito $\{0, 1\}$. La función de distribución de X está entonces dada por la fórmula (ver

Figura 5.6):

$$F_X(x) = \begin{cases} 0, & x < 0; \\ 3/4, & 0 \leq x < 1; \\ 1, & x \geq 1. \end{cases}$$

Esta función no es estrictamente creciente ya que, por ejemplo, $1/10 < 9/10$ y, sin embargo, $F_X(1/10) = 3/4 = F_X(9/10)$. Para simular X usando el Método de la Transformada Inversa tenemos que hacer explícita la función G dada en el Teorema 5.2. En efecto, usando el grafo de F_X en la Figura 5.6, fácilmente encontramos que

$$G(u) = \begin{cases} 0, & 0 < u \leq 3/4; \\ 1, & 3/4 < u < 1. \end{cases}$$

Por lo tanto, si $U \sim Uniforme[0,1]$, entonces

$$Y = \begin{cases} \text{cara}, & \text{si } 0 \leq U \leq 3/4; \\ \text{sello}, & \text{si } 3/4 < U \leq 1; \end{cases}$$

simulará el lanzamiento de la moneda.

Retrospectivamente la simulación anterior es bastante obvia: como la probabilidad del suceso $[U \leq 3/4]$ es $3/4$ (la probabilidad que queremos para cara), mientras que la del suceso $[U > 3/4]$ es $1/4$ (la probabilidad que queremos para sello), la simulación producirá los valores “cara” y “sello” con las probabilidades requeridas. Esta idea es la esencia del Método de la Transformada Inversa para simular variables aleatorias discretas en general.

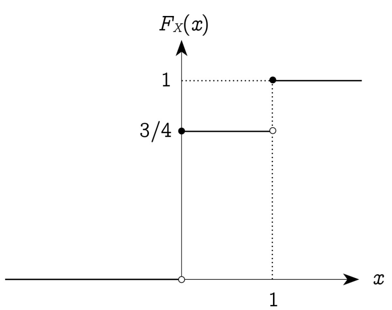


Figura 5.6

De manera más general, considere una v.a. discreta X y un conjunto contable $A \subset \mathbb{R}$ tal que $\mathbb{P}[X \in A] = 1$ y

$$\mathbb{P}[X = a] > 0, \tag{5.5}$$

para cada $a \in A$. (Si $\mathbb{P}[X = a] = 0$ para algún $a \in A$, entonces podemos eliminar a de A a modo de satisfacer esta última restricción.)

Al igual que como lo hicimos para simular el lanzamiento de la moneda en el Ejemplo 5.5, simularemos la v.a. X construyendo, para cada $a \in A$, un intervalo $I(a) \subset [0,1]$ del mismo largo que la probabilidad del suceso $[X = a]$. La única restricción importante sobre estos intervalos es que tienen que ser disjuntos. En particular, y debido a que

$$\sum_{a \in A} largo(I(a)) = \sum_{a \in A} \mathbb{P}[X = a] = 1,$$

todo punto entre 0 y 1 pertenecerá exactamente a un único intervalo. Esto permite definir una v.a. Y como sigue: *primero simulamos $U \sim Uniforme[0,1]$ y vemos a cuál intervalo pertenece U. Si $U \in I(a)$, entonces asignamos a Y el valor a.* El siguiente resultado garantiza que Y es una realización de X.

Teorema 5.3. *Si Y es la v.a. definida como recién vimos, entonces X e Y tienen la misma distribución.*

Para entender la validez del teorema basta con observar que si $a \in A$, entonces

$$\mathbb{P}[Y = a] = \mathbb{P}[U \in I(a)] = largo(I(a)) = \mathbb{P}[X = a],$$

donde la primera igualdad es consecuencia de la definición de Y, la segunda del hecho que $U \sim Uniforme[0,1]$, y la tercera a que el largo de $I(a)$ fue precisamente escogido como la probabilidad del suceso $[X = a]$.

Ejemplo 5.6. Para fijar ideas consideremos el problema de simular el lanzamiento de un dado sesgado donde la cara lanzada X satisface que $\mathbb{P}[X = k] = k/21$, para $k = 1, \ldots, 6$.

El problema de diseñar un dado con esta propiedad, por ejemplo, para un casino en la ciudad de Las Vegas, Estados Unidos, puede ser un problema ingenieril bastante complicado. Sin embargo, usando un computador es un problema bastante simple: ¡todo se reduce a la simulación de una v.a. uniforme! Para esto dividimos el intervalo $[0,1]$ en los intervalos $[0,1/21]$, $(1/21,3/21]$, $(3/21,6/21]$, $(6/21,10/21]$, $(10/21,15/21]$, $(15/21,1]$, respectivamente de largo $1/21$, $2/21$, $3/21$, $4/21$, $5/21$ y $6/21$. Si $U \sim Uniforme[0,1]$, entonces, y de acuerdo al Teorema 5.3, la siguiente v.a. se comporta como lo haría el dado que queremos simular

$$Y = \begin{cases} 1, & 0 \leq U \leq 1/21; \\ 2, & 1/21 < U \leq 3/21; \\ 3, & 3/21 < U \leq 6/21; \\ 4, & 6/21 < U \leq 10/21; \\ 5, & 10/21 < U \leq 15/21; \\ 6, & 15/21 < U \leq 1. \end{cases}$$

Los algoritmos en las Tablas 5.2 y 5.3 pueden implementarse en un computador para simular el lanzamiento del dado. Destacamos que para el mismo valor de U ambos algoritmos asignarán a Y el mismo valor. Sin embargo, el Algoritmo No. 2 es más eficiente que el Algoritmo No. 1. Esto se debe a que el Algoritmo No. 2 chequea los sucesos $U \in (15/21, 1], \ldots, U \in [0, 1/21]$ desde el más probable hasta el menos probable; sin embargo, el Algoritmo No. 1 chequea lo mismo pero en el orden inverso.

Línea 1:	Asigne $i = 0$, $j = 1$.
Línea 2:	Simule $U \sim Uniforme[0, 1]$.
Línea 3:	Asigne $i = (i + j)$.
Línea 4:	Si $U \leq i/21$, entonces asigne $Y = j$ y termine. Si no, asigne $j = (j + 1)$ y vuelva a la línea 3.

Tabla 5.2. Algoritmo No. 1 para la simulación de un dado sesgado donde la cara k tiene probabilidad $k/21$ de ocurrir. El algoritmo chequea iteradamente si $U \leq 1/21$, $U \leq 3/21$, etc., hasta asignar un valor a Y.

Línea 1:	Asigne $i = 21$, $j = 6$.
Línea 2:	Simule $U \sim Uniforme[0, 1]$.
Línea 3:	Asigne $i = i - j$.
Línea 4:	Si $U > i/21$ entonces asigne $Y = j$ y termine. Si no, asigne $j = (j - 1)$ y vuelva a la línea 3.

Tabla 5.3. Algoritmo No. 2 para la simulación de un dado sesgado donde la cara k tiene probabilidad $k/21$ de ocurrir. El algoritmo chequea iteradamente si $15/21 < U$, $10/21 < U$, etc., hasta asignar un valor a Y.

Ejemplo 5.7. Considere el problema de simular la extracción al azar de una carta de una baraja inglesa. Para esto, y sin pérdida de generalidad, identificaremos cada carta con un par ordenado de la forma (i, j), donde $i \in \{1, \ldots, 4\}$ representa la pinta de la carta y $j \in \{1, \ldots, 13\}$ su número.

Si X denota la pinta de la carta extraída e Y su número, entonces X e Y son variables aleatorias independientes (*¿por qué?*). Note que para cada i, el suceso $[X = i]$ tiene probabilidad $1/4$. En particular, para simular X usando el Método de la Transformada Inversa, basta con simular una v.a. $U_1 \sim Uniforme[0, 1]$ y asignar a X el valor i si $(i - 1)/4 < U_1 \leq i/4$. Similarmente, para simular Y basta con simular independientemente de U_1 otra v.a. $U_2 \sim Uniforme[0, 1]$ y asignar a Y el valor j si $(j - 1)/13 < U_2 \leq j/13$. El algoritmo en la Tabla 5.4 implementa la extracción al azar de una carta de una baraja inglesa usando estos hechos.

Imagine ahora que se quiere simular la extracción de n cartas al azar del mazo, donde $1 \leq n \leq 52$ es un entero fijo. Para fijar ideas supondremos que $n = 2$ y que en una primera iteración del algoritmo en la Tabla 5.4 obtenemos el valor $(X, Y) = (3, 11)$. Si aplicaramos nuevamente el algoritmo para simular la extracción de una segunda carta, entonces no estaremos simulando lo que queremos ya que existe una probabilidad positiva de que el algoritmo entregue nuevamente el valor $(3, 11)$. Sin embargo, si el resultado de la segunda simulación es diferente de $(3, 11)$, entonces el par ordenado es igualmente probable de ser cualquiera de los asociados a las otras cartas. Para simular la extracción de dos cartas al azar basta, por lo tanto, con aplicar el algoritmo en la Tabla 5.4 hasta observar un par ordenado diferente del primero. Esta idea es incorporada en la Tabla 5.5 para simular la extracción de n cartas del mazo. En el algoritmo, el conjunto $\mathcal{C}$ mantiene cuenta de las diferentes cartas que ya han sido extraídas y el índice i de la cardinalidad de $\mathcal{C}$. La condición de término del algoritmo es, por lo tanto, cuando $i = n$.

Línea 1:	Asigne $i = 0$, $j = 0$.
Línea 2:	Simule $U_1, U_2 \sim Uniforme[0, 1]$.
Línea 3:	Asigne $i = i + 1$.
Línea 4:	Si $U_1 \leq i/4$ entonces asigne $X = i$ y continue a la línea 5. Si no, vuelva a la línea 3.
Línea 5:	Asigne $j = j + 1$.
Línea 6:	Si $U_2 \leq j/13$ entonces asigne $Y = j$ y termine. Si no vuelva a la línea 5.

Tabla 5.4. Algoritmo basado en el Método de la Transformada Inversa para simular la extracción al azar de una carta de una baraja inglesa. X denota el número de la carta e Y su pinta. Aquí las pintas $1, 2, 3$ y 4 representan e.g. corazón, diamante, trébol y espada, y los números $1, 11, 12$ y 13 representan As, Jota, Reina y Rey, respectivamente.

Línea 1:	Asigne $i = 0$ y $\mathcal{C} = \emptyset$.
Línea 2:	Genere (X, Y) usando el algoritmo de la Tabla 5.4.
Línea 3:	Si $(X, Y) \in \mathcal{C}$ entonces vuelva a la línea 2. Si no asigne $i = (i + 1)$ y $\mathcal{C} = \mathcal{C} \cup \{(X, Y)\}$.
Línea 4:	Si $i < n$ entonces vuelva a la línea 2. Si no, termine.

Tabla 5.5. Algoritmo basado en la Tabla 5.4 para simular la extracción al azar de n cartas de una baraja inglesa.

5.1.3 Simulación de variables aleatorias del tipo Bernoulli.

El Método de la Transformada Inversa nos permite simular directamente una v.a. del tipo Bernoulli con parámetro p. Todo se reduce a dividir el intervalo $[0, 1]$ en los intervalos $[0,p)$ y $[p,1]$, de largo p y $(1-p)$, respectivamente. De acuerdo al Teorema 5.3, si $U \sim Uniforme[0,1]$, entonces la v.a. definida como

$$Y = \begin{cases} 1, & \text{si } U < p; \\ 0, & \text{si } U \geq p; \end{cases} .$$

es del tipo Bernoulli con parámetro p. Observe que Y corresponde a la función indicadora del suceso $[U \leq p]$.

5.1.4 Simulación de variables aleatorias geométricas.

Considere una v.a. $X \sim$ *Geométrica*(p). Para simular X, usando el Método de la Transformada Inversa, dividiremos primero—de izquierda a derecha—el intervalo $[0,1]$ en intervalos más pequeños y de largo p, $(1-p)p$, $(1-p)^2p$, etc. Si $U \sim Uniforme[0,1]$, entonces la v.a. definida como

$$Y = k \Longleftrightarrow U \text{ toma un cierto valor en el } k\text{-ésimo intervalo,}$$

tiene una distribución geométrica con parámetro p.

Una manera eficiente de determinar el valor de Y es verificar secuencialmente si $U \leq \sum_{i=1}^{k}(1-p)^{i-1}p = 1-(1-p)^k$, para $k = 1, 2, \ldots$ (*¿por qué?*). Si definimos $V = (1-U)$, lo anterior es equivalente a verificar secuencialmente si $(1-p)^k \leq V$, para $k = 1, 2, \ldots$ Usando que $V \sim Uniforme[0,1]$ (*¿por qué?*), el algoritmo en la Tabla 5.6 es suficiente para simular la v.a. X.

Línea 1:	Asigne $k = 0$.
Línea 2:	Simule $V \sim Uniforme[0,1]$.
Línea 3:	Asigne $k = k+1$.
Línea 4:	Si $(1-p)^k \leq V$, entonces asigne $Y = k$ y termine. Si no, vuelva a la línea 3.

Tabla 5.6. Algoritmo No. 1 basado en el Método de la Transformada Inversa para la simulación de una v.a. geométrica con parámetro p.

Ejemplo 5.8. Para simular el lanzamiento de una moneda equilibrada hasta observar una cara por primera vez, basta con aplicar el algoritmo en la Tabla 5.6 con $p = 1/2$. Por ejemplo, si V en la línea 2 fuera asignado el valor $V = 0{,}2755...$, entonces Y será asignado el valor $Y = 2$ debido a que $(1/2)^1 = 0{,}5 > V$, pero $(1/2)^2 = 0{,}25 \leq V$. Note, sin embargo, $\log_2(V) = -1{,}8...$; en particular, $(1/2)^2 < V < (1/2)^1$. Por lo tanto, usando el logaritmo en base-2 de V, podemos determinar directamente el valor más pequeño de k en la línea 4 para el cual la condición $(1/2)^k \leq V$ será satisfecha.

Para fijar ideas, imagine que a V es asignado el valor $V = 0{,}0135...$ en la línea 2. Entonces $\log_2(V) = -6{,}2...$ e Y será asignada el valor $Y = 7$. Para corroborar esto note que $(1/2)^6 = 0{,}015625 > V$, sin embargo, $(1/2)^7 = 0{,}0078125 \leq V$. Esta idea es implementada en la Tabla 5.12 en el Ejercicio 11 para simular más fácilmente una v.a. geométrica con un parámetro arbitrario p.

5.1.5 Simulación de variables aleatorias binomiales.

Para simular una v.a. $X \sim Binomial(n,p)$ podemos dividir el intervalo $[0,1]$ en intervalos de largo $\binom{n}{k}p^k(1-p)^{n-k}$ con $k=0,\ldots,n$, y luego simular una v.a. uniforme. Sin embargo, esta tarea resulta un poco engorrosa y puede simplificarse tremendamente usando el Teorema 4.5: para simular X es suficiente sumar n variables aleatorias independientes y del tipo Bernoulli con parámetro común p. Este procedimiento ha sido implementado en el algoritmo de la Tabla 5.7.

Línea 1:	Asigne $j=0$ y $X=0$.
Línea 2:	Simule $U \sim Uniforme[0,1]$.
Línea 3:	Asigne $j=j+1$.
Línea 4:	Si $j>n$, termine. Si no, si $0 \leq U < p$ entonces asigne $X=(X+1)$, y vuelva a la línea 2.

TABLA 5.7. Algoritmo basado en el Método de la Transformada Inversa para la simulación de una v.a. binomial con parámetros n y p.

(Recuerde que cada ejecución de la Línea 2 en el algoritmo es independiente de todas las ejecuciones previas y futuras.) En el algoritmo propuesto la variable j mantiene cuenta de las variables binomiales $X_1,\ldots,X_n$ que ya han sido simuladas; en particular, el algoritmo comienza con $j=0$. La variable X corresponde a la suma parcial $\sum_{i=1}^{j} X_i$. La última iteración del algoritmo ocurre cuando $j=n$ razón por la cual $j>n$ es la condición de término.

5.1.6 Simulación de variables aleatorias del tipo Poisson.

Considere una v.a. $X \sim Poisson(\lambda)$. Ciertamente, podemos simular X usando el Método de la Transformada Inversa. Dejaremos como ejercicio para el lector diseñar un algoritmo basado en este método. En esta sección, sin embargo, veremos una manera alternativa para simular variables aleatorias de tipo Poisson, basada en la conexión entre este tipo de variable y las del tipo exponencial.

Recordemos que si en promedio ocurren λ eventos por unidad de tiempo (e.g. sismos, llamadas de teléfono, o visitas a una biblioteca), entonces el número de eventos que ocurren en un intervalo de duración t tiene una distribución del tipo Poisson con parámetro λt (Sección 4.4). Usando esto, vimos que la separación temporal entre eventos consecutivos tiene una distribución del tipo exponencial con parámetro λ (Sección 4.5). En particular, para simular X basta con registrar el número de eventos que ocurren en el intervalo $[0,1]$. (Esto se relaciona de alguna manera con la simulación de los tiempos de llegada de alumnos a una biblioteca presentada en la Sección 5.1.1.

Entonces fijamos el número n de alumnos, mientras que ahora fijaremos el tiempo t en que permitiremos a los alumnos entrar a la biblioteca.)

Debido a la discusión anterior y si $T_1, T_2, \ldots$ son los entretiempos de eventos consecutivos (ver Figura 5.7), para simular X basta con determinar el número de eventos en el intervalo $[0,1]$. Este número corresponde al índice $n \geq 0$ más grande, tal que

$$\sum_{i=1}^{n} T_i \leq 1.$$

Se entiende que $n = 0$, cuando $T_1 > 1$. Como $-\ln(U_1)/\lambda, -\ln(U_2)/\lambda, \ldots$ son variables aleatorias independientes y del tipo exponencial con parámetro λ, cuando $U_1, U_2, \ldots$ son i.i.d. $Uniformes[0,1]$ (*¿por qué?*), el algoritmo descrito en la Tabla 5.8 simulará la v.a. X.

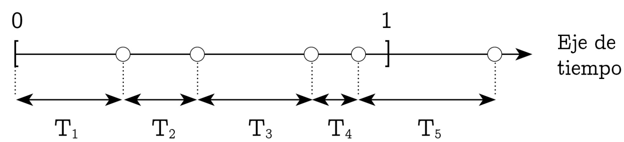


FIGURA 5.7. Simulación de una v.a. $X \sim Poisson(\lambda)$. Como $T_1, T_2, T_3, \ldots$ son variables aleatorias i.i.d. del tipo exponencial con parámetro λ, y $\sum_{i=1}^{4} T_i \leq 1$, pero $\sum_{i=1}^{5} T_i > 1$, X es asignado el valor 4 en este caso.

Línea 1:	Asigne $X = 0$ y $T = 0$.
Línea 2:	Simule $U \sim Uniforme[0,1]$.
Línea 3:	Si $(T - \ln(U)/\lambda) > 1$, termine. Si no, asigne $T = (T - \ln(U)/\lambda)$ y $X = X + 1$. Vuelva a la línea 2.

TABLA 5.8. Algoritmo para la simulación de una v.a. del tipo Poisson con parámetro λ, basado en la relación entre variables aleatorias del tipo Poisson y exponenciales.

5.2 Método de Rechazo

El Método de la Transformada Inversa puede, en teoría, utilizarse para simular cualquier tipo de variable aleatoria, particularmente variables aleatorias del tipo discreto. En la práctica, sin embargo, este método presenta complicaciones para simular variables aleatorias continuas en general. El problema radica en determinar la función G dada en el Teorema 5.2 de forma relativamente explícita.

En esta sección discutiremos un método alternativo para simular variables aleatorias continuas llamado *Método de Rechazo*. Este método tiene una variante para simular variables aleatorias discretas que, por razones de espacio, hemos omitido de la presente monografía. El lector interesado podrá usar la referencia [**2**].

Considere una v.a. continua X que se desea simular. *El Método de Rechazo consiste en simular X usando otra v.a. continua Z que ya sabemos como simular (e.g. usando el Método de la Transformada Inversa).* Para implementar el método lo único que necesitamos es una constante finita $c > 0$ tal que

$$f_X(z) \leq c \cdot f_Z(z), \tag{5.6}$$

para todo $z \in \mathbb{R}$. Observe que la condición anterior requiere que $f_Z(z) > 0$ si $f_X(z) > 0$ i.e. todo valor posible para X tiene que ser un valor posible para Z.

Línea 1:	Simule Z y asigne $z = Z$.
Línea 2:	Simule $U \sim Uniforme[0,1]$.
Línea 3:	Si $U \leq \frac{f_X(z)}{c \cdot f_Z(z)}$ entonces asigne $Y = z$ y termine. Si no, vuelva a la línea 1.

TABLA 5.9. Algoritmo general para la simulación de una v.a. continua X a través de otra v.a. continua Z usando el Método de Rechazo. La v.a. Y producida por el algoritmo tiene la misma distribución que X.

La parte más intrincada del Método de Rechazo es la elección de la constante c. Observe que la condición en (5.6) implica que $c \geq 1$ (*¿por qué?*). Por lo tanto, el meollo del problema no es tener $c > 0$, sino $c < +\infty$. Esta última condición es equivalente a tener que

$$\sup_{z \in \mathbb{R}: f_Z(z) > 0} \frac{f_X(z)}{f_Z(z)} < \infty.$$

En este caso, cualquier valor c mayor o igual al valor supremo anterior, puede usarse para la aplicación del método (*¿por qué?*). Sin embargo, y debido al siguiente resultado, es recomendable escoger el valor más pequeño para c a modo de minimizar el número promedio de iteraciones del algoritmo. La selección óptima de c es precisamente el valor supremo anterior.

Teorema 5.4. *La v.a. N que cuenta el número de iteraciones de la Línea 3 en el algoritmo del Método de Rechazo, tiene una distribución geométrica con valor esperado c. Más aún, X e Y tienen la misma distribución.*

Presentaremos la demostración de este resultado al final del capítulo.

Ejemplo 5.9. Para fijar ideas considere dos v.a.'s continuas X y Z, tales que

$$\frac{f_X(t)}{f_Z(t)} = \frac{25t^2}{2(1+t^4)},$$

para todo $t \in \mathbb{R}$, *¿es posible simular X a partir de Z usando el Método de Rechazo?, ¿qué hay de Z a partir de X?*

Para responder la primera pregunta observe que

$$\frac{\partial}{\partial t}\left[\frac{f_X(t)}{f_Z(t)}\right] = \frac{25t(1-t^4)}{(1+t^4)^2}.$$

Como $\lim_{t\to\pm\infty} f_X(t)/f_Z(t) = 0$, podemos concluir a partir de la identidad anterior que $f_X(t)/f_Z(t)$ alcanza su máximo absoluto cuando $t = \pm 1$. En particular, para implementar el Método de Rechazo, podemos seleccionar $c = f_X(1)/f_Z(1) = 25/4$ para simular X a partir de Z.

Para responder la segunda pregunta note que

$$\sup_{t\in\mathbb{R}} \frac{f_Z(t)}{f_X(t)} = +\infty.$$

En particular, no es posible simular la v.a. Z a partir de la v.a. X usando el Método de Rechazo.

Ejemplo 5.10. Considere una v.a. X con función de densidad $f(x) = k \cdot x^{2/5} \cdot (1-x)^{4/5}$, para $x \in [0,1]$, donde k es una constante tal que $\int_0^1 f(x)\,dx = 1$ i.e. $k = (\int_0^1 x^{2/5} \cdot (1-x)^{4/5}\,dx)^{-1}$.

La función de distribución de X viene dada por

$$F_X(x) = \int_0^x k \cdot t^{2/5} \cdot (1-t)^{4/5}\,dt,$$

para todo $x \in [0,1]$. Observe que si quisiéramos simular X usando el Método de la Transformada Inversa necesitaríamos para cada $u \in (0,1)$ resolver la ecuación $F_X(v) = u$. Aunque esta tarea no es fácil, podemos simular X fácilmente usando el Método de Rechazo. Algo interesante del cálculo es que la simulación no requiere conocer explícitamente la constante k en la densidad f.

Como $f_X > 0$ en el intervalo $(0,1)$, para implementar el Método de Rechazo resulta natural escoger $Z \sim Uniforme[0,1]$. En particular, tenemos que

$$c = \sup_{z\in[0,1]} \frac{f_X(z)}{f_Z(z)} = \sup_{z\in[0,1]} f_X(z) = k \cdot \sup_{z\in[0,1]} z^{2/5}(1-z)^{4/5}.$$

Note, sin embargo, que

$$\frac{\partial}{\partial z}\left[z^{2/5}(1-z)^{4/5}\right] = \frac{2(1-3z)}{5z^{3/5}(1-z)^{1/5}}.$$

Esto implica que la función $z \longrightarrow z^{2/5}(1-z)^{4/5}$, con $z \in [0,1]$, alcanza su único máximo en el punto $z = 1/3$. Por lo tanto, $c = k \cdot (1/3)^{2/5} \cdot (2/3)^{4/5}$ y en particular:

$$\frac{f_X(z)}{c \cdot f_Z(z)} = \frac{z^{2/5}(1-z)^{4/5}}{(1/3)^{2/5} \cdot (2/3)^{4/5}} = (3z)^{2/5}(3/2 - 3z/2)^{4/5}.$$

El algoritmo en la Tabla 5.10 implementa el Método de Rechazo para simular la v.a. X.

Línea 1:	Simule $Z \sim Uniforme[0,1]$ y asigne $z = Z$.
Línea 2:	Simule $U \sim Uniforme[0,1]$.
Línea 3:	Si $U \leq (3z)^{2/5}(3/2 - 3z/2)^{4/5}$ entonces asigne $X = z$ y termine. Si no, vuelva a la línea 1.

Tabla 5.10

5.2.1 Simulación de variables aleatorias normales.

Considere el problema de simular una v.a. $X \sim Normal(\mu, \sigma^2)$. Como la v.a. $Y = (X - \mu)/\sigma$ tiene una distribución Normal estándar (Teorema 4.13) y $X = \mu + \sigma \cdot Y$, para simular X basta con simular Y. Veremos cómo hacer esto último usando el Método de Rechazo.

Como $f_Y > 0$ sobre $\mathbb{R}$, para satisfacer la condición dada en (5.6) necesitamos una v.a. Z que sabemos como simular y tal que $f_Z > 0$ sobre todo $\mathbb{R}$. Lamentablemente, una v.a. del tipo exponencial no sirve para este efecto, ya que su función de densidad se anula sobre el intervalo abierto $(-\infty, 0)$. Para corregir este problema, sin embargo, podemos multiplicar una v.a. exponencial por ± 1 con probabilidad $1/2$. El lector podrá verificar por su cuenta que la v.a. Z que resulta es continua y tiene una función de densidad dada por la fórmula $f_Z(z) = e^{-|z|}/2$, para todo $z \in \mathbb{R}$. En particular, se tiene que

$$F_Z(z) = \begin{cases} e^z/2, & z < 0; \\ (2 - e^{-z})/2, & z \geq 0. \end{cases}$$

Como esta función es estrictamente creciente y su función inversa viene dada por

$$F_Z^{-1}(u) = \begin{cases} \ln(2u), & 0 < u \leq 1/2; \\ -\ln(2 - 2u), & 1/2 < u < 1; \end{cases}$$

el Teorema 5.1 implica que si $U \sim Uniforme(0,1)$, entonces $F_Z^{-1}(U)$ tiene la misma distribución que Z.

Finalmente, para implementar el Método de Rechazo, sólo necesitamos acotar por arriba la función f_Y/f_Z. Para esto observe que

$$c = \sup_{y \in \mathbb{R}} \frac{f_Y(y)}{f_Z(y)} = \sup_{y \in \mathbb{R}} \sqrt{\frac{2}{\pi}} \cdot e^{|y| - y^2/2} = \sup_{y \geq 0} \sqrt{\frac{2}{\pi}} \cdot e^{y - y^2/2} = \sqrt{\frac{2e}{\pi}} \approx 1{,}3.$$

Para la tercera igualdad, hemos usado que la función $y \to (|y| - y^2/2)$ es simétrica alrededor de $y = 0$, y para la última, hemos usado que esta función alcanza su máximo en el punto $y = 1$ (*¿por qué?*). En particular, para simular Y a partir de Z (y, por lo tanto, X a partir de Z) necesitaremos en promedio 1,3 iteraciones del Método de Rechazo. Como

$$\frac{f_Y(y)}{c \cdot f_Z(y)} = e^{|y|-y^2/2-1/2},$$

en algoritmo en la Tabla 5.11 simulará una v.a. del tipo Normal.

Línea 1:	Simule $U_1 \sim Uniforme[0,1]$ y $U_2 \sim Uniforme[0,1]$ independientemente
Línea 2:	Si $U_1 \leq 1/2$ entonces asigne $z = \ln(2U_1)$, sino asigne $z = -\ln(2-2U_1)$
Línea 3:	Si $U_2 \leq e^{|z|-z^2/2-1/2}$ entonces asigne $X = (\mu + \sigma \cdot z)$ y termine. Si no, vuelva a la línea 1.

Tabla 5.11. Algoritmo basado en el Método de Rechazo para simular una v.a. Normal con media μ y varianza σ^2.

Demostración del Teorema 5.4. La demostración requiere el uso de integrales dobles y el lector no familiarizado con este concepto puede omitirla (ver Sección 2.6.2). Para simplificar la demostración supondremos que $f_Z(z) > 0$, para todo $z \in \mathbb{R}$. Denotemos como p la probabilidad de que Y es asignado un valor en la primera iteración del algoritmo. Como cada iteración del algoritmo es independiente de las iteraciones previas, $\mathbb{P}[N = n] = (1-p)^{n-1} \cdot p$, para cada $n \geq 1$. Esto demuestra que $N \sim Geométrica(p)$.

Para determinar el valor de p observe lo siguiente: para que Y sea asignado un valor en la primera iteración, necesitamos que (Z, U) pertenezca a la región de puntos en el plano:

$$R_1 = \left\{ (z,u) : z \in \mathbb{R} \text{ y } 0 \leq u \leq \frac{f_X(z)}{c \cdot f_Z(z)} \right\}.$$

Como U y Z son variables aleatorias independientes, tenemos que (Teorema 2.19)

$$\begin{aligned} p = \mathbb{P}[(Z,U) \in R_1] &= \iint_{R_1} f_Z(z) \cdot f_U(u)\, du\, dz, \\ &= \int_{\mathbb{R}} \int_0^{\frac{f_X(z)}{c \cdot f_Z(z)}} f_Z(z) \cdot 1\, du\, dz, \\ &= \int_{\mathbb{R}} \frac{f_X(z)}{c \cdot f_Z(z)} \cdot f_Z(z)\, dz = \frac{1}{c} \int_{\mathbb{R}} f_X(z)\, dz = \frac{1}{c}. \end{aligned}$$

En particular, $\mathbb{E}(N) = 1/p = c$ debido a la identidad en (4.3). Esto demuestra la primera parte del teorema.

Finalmente demostraremos que X e Y tienen la misma distribución i.e. que $F_Y(t) = F_X(t)$, para todo $t \in \mathbb{R}$ (Teorema 2.11). En efecto, dado $t \in \mathbb{R}$ y $n \geq 1$, para que Y sea asignado un valor menor a igual a t en la n-ésima iteración i.e. para que el suceso $[Y \leq t, N = n]$ ocurra, necesitamos que Y no sea asignado ningún valor en las primeras $(n-1)$-iteraciones, pero que en la n-ésima iteración el punto (Z, U) pertenezca a la región

$$R_2 = \left\{(z,u) :\, z \leq t \text{ y } 0 \leq u \leq \frac{f_X(z)}{c \cdot f_Z(z)}\right\}.$$

En particular, y debido a la Fórmula de Probabilidades Totales, encontramos que

$$\begin{aligned}
F_Y(t) = \mathbb{P}[Y \leq t] &= \sum_{n=1}^{\infty} \mathbb{P}[Y \leq t, N = n], \\
&= \sum_{n=1}^{\infty} (1-p)^{n-1} \cdot \mathbb{P}[(Z,U) \in R_2], \\
&= \iint_{R_2} f_Z(z) \cdot f_U(u)\, du\, dz \cdot \sum_{n=1}^{\infty} (1-p)^{n-1}, \\
&= \int_{-\infty}^{t} \int_{0}^{\frac{f_X(z)}{c \cdot f_Z(z)}} f_Z(z)\, du\, dz \cdot \frac{1}{1-(1-p)}, \\
&= \frac{1}{c} \int_{-\infty}^{t} f_X(z)\, dz \cdot \frac{1}{p} = \frac{1}{cp} \cdot \mathbb{P}[X \leq t] = F_X(t),
\end{aligned}$$

donde hemos usado que $cp = 1$ en la penúltima identidad. Esto concluye la demostración del teorema.

5.3 Ejercicios propuestos

Sección 5.1.

Ejercicio 1. Considere una v.a. X con función de densidad $f(x) = 3/x^4$ si $x \geq 1$, y $f(x) = 0$ si $x < 1$. (a) Determine la función de distribución de X. (b) Muestre cómo simular esta v.a., usando el Método de la Transformada Inversa.

Ejercicio 2. Determine la función G dada en el Teorema 5.2 si F es la función

$$F(x) = \begin{cases} 0, & x < -1; \\ (3x+5)/6, & -1 \leq x < 0; \\ 5/6, & 0 \leq x < 2/3; \\ (x+1)/2, & 2/3 \leq x < 1; \\ 1, & x \geq 1. \end{cases}$$

Ejercicio 3. Muestre cómo simular el lanzamiento de un dado equilibrado usando el Método de la Transformada Inversa.

Ejercicio 4. Considere el experimento que consiste en lanzar dos veces un dado con probabilidad p de cara y $(1-p)$ de sello, y la v.a. X definida a continuación. Diseñe un algoritmo para simular X, a partir del Método de la Transformada Inversa.

$$X = \begin{cases} 1, & \text{primer lanzamiento=cara, segundo lanzamiento=cara;} \\ 2, & \text{primer lanzamiento=cara, segundo lanzamiento=sello;} \\ 3, & \text{primer lanzamiento=sello, segundo lanzamiento=cara;} \\ 4, & \text{primer lanzamiento=sello, segundo lanzamiento=sello.} \end{cases}$$

Ejercicio 5. Considere las variables aleatorias $V \sim Uniforme(-\pi/2, +\pi/2)$ y $X = \tan(V)$. Determine cómo simular la v.a. X, usando el Método de la Transformada Inversa. Retrospectivamente, ¿es la fórmula encontrada del todo sorprendente? Explique.

Sección 5.2.

Ejercicio 6. Explique por qué el Método de Rechazo no puede utilizarse para simular la v.a. X, a partir de la v.a. Z, en cada uno de los siguientes casos:

(a) $X \sim \mathit{Exponencial}(1)$, $Z \sim Uniforme[0, 1]$.
(b) $X \sim \mathit{Exponencial}(2)$, $Z \sim \mathit{Exponencial}(4)$.
(c) $X \sim \mathit{Exponencial}(1)$, $Z \sim Normal(0, 1)$.

Ejercicio 7. Implemente el Método de Rechazo para simular una v.a. uniforme en el intervalo $[0, 1]$ a partir de una v.a. exponencial con parámetro $\lambda = 1$, ¿cuántas iteraciones requiere en promedio el algoritmo para simular la v.a. uniforme?

Ejercicio 8. Sean $\alpha, \beta > 0$ dos constantes dadas. Implemente el Método de Rechazo para simular la v.a. X con función de densidad dada por la fórmula $f(x) = kx^\alpha(1-x)^\beta$, para $x \in [0, 1]$, donde k es una constante tal que $\int_0^1 f(x)\,dx = 1$.

Ejercicio 9. Considere un punto aleatorio (X, Y) uniformemente distribuido en el círculo unitario. (a) Muestre que X es una v.a. continua con función de densidad dada por la fórmula $f_X(x) = 2\sqrt{1-x^2}/\pi$, si $-1 \leq x \leq 1$, y $f_X(x) = 0$, en el caso contrario. (b) Dise
ne un algoritmo basado en el Método de Rechazo para simular la v.a. X. (c) Intuitivamente, si $X = x$, entonces $Y \sim Uniforme[-\sqrt{1-x^2}, \sqrt{1-x^2}]$. Use esto para dise nar un algoritmo que simule el punto aleatorio (X, Y).

Problemas misceláneos Capítulo 5.

Ejercicio 10. Considere el experimento donde se lanza una moneda equilibrada y dos variables aleatorias X e Y (independientes de la moneda) tales que $f_X(t) = e^{-t}$ y $f_Y(t) = 2e^{-2t}$ si $t \geq 0$, y $f_X(t) = f_Y(t) = 0$ si $t < 0$. Muestre cómo simular la siguiente v.a. usando el Método de la Transformada Inversa:

$$Z = \begin{cases} X, & \text{si la moneda sale cara;} \\ Y, & \text{si la moneda sale sello.} \end{cases}$$

Ejercicio 11. Use el algoritmo descrito en la Tabla 5.6 para demostrar directamente que el algoritmo descrito en la Tabla 5.12 también puede utilizarse para simular una v.a. geométrica con parámetro p.

Línea 1:	Simule $V \sim Uniforme[0,1]$.
Línea 2:	Asigne $X = \lceil \ln(V)/\ln(1-p) \rceil$ y termine.

TABLA 5.12. Algoritmo No. 2 basado en el Método de la Transformada Inversa para la simulación de una v.a. geométrica con parámetro p. Para $x \in \mathbb{R}$, $\lceil x \rceil$ denota el número entero más pequeño que es mayor o igual a x.

Ejercicio 12. Describa dos algoritmos diferentes para simular el dado sesgado del Ejemplo 5.6 usando el Método de la Transformada Inversa con los intervalos $[0, 6/21]$, $(6/21, 11/21]$, $(11/21, 15/21]$, $(15/21, 18/21]$, $(18/21, 20/21]$ y $(20/21, 1]$.

Ejercicio 13. Muestre que si U_1, U_2 son variables aleatorias i.i.d. y uniformes en el intervalo abierto $(0,1)$, entonces

$$Y = \begin{cases} 1 - \frac{1}{U_2}, & 0 < U_1 < p; \\ \frac{1}{U_2} - 1, & p \leq U_1 < 1; \end{cases}$$

tiene la misma distribución que la v.a. X en el Ejemplo 5.2. (Note que la simulación de X, basada en el Método de la Transformada Inversa en el Ejemplo 5.2, es más eficiente ya que sólo requiere la simulación de una única v.a. uniforme.)

Ejercicio 14. Considere una sección S del plano de área A y contenida en un cierto rectángulo R con vértices (a_1, b_1), (a_2, b_1), (a_2, b_2) y (a_1, b_2), donde $a_1 < a_2$ y $b_1 < b_2$ (ver Figura 5.8). (a) Use el Teorema 2.19 para mostrar que si $U_1 \sim Uniforme[a_1, a_2]$ y $U_2 \sim Uniforme[b_1, b_2]$ son independientes, entonces (U_1, U_2) es un punto equidistribuido en el rectángulo R.
Hint: Para esto basta con demostrar que $\mathbb{P}[(U_1, U_2) \in R_1] = \mathbb{P}[(U_1, U_2) \in R_2]$, para todo par de rectángulos $R_1, R_2 \subset R$ con lados paralelos a los ejes y de la misma área. (b) Considere el algoritmo en la Tabla 5.13. Muestre que si N cuenta el número de iteraciones de la Línea 3, entonces

$$N \sim Geométrica\left(\frac{A}{(a_2 - a_1) \cdot (b_2 - b_1)}\right).$$

(c) Muestre que el punto aleatorio (X, Y), producido por el algoritmo, es equidistribuido en la sección S. (d) Si S es el círculo unitario: ¿cuál es la selección óptima de a_1, a_2, b_1 y b_2? Usando estos valores reescriba el algoritmo en la Tabla 5.13 a modo de verificar la condición $(U_1, U_2) \in S$ directamente.

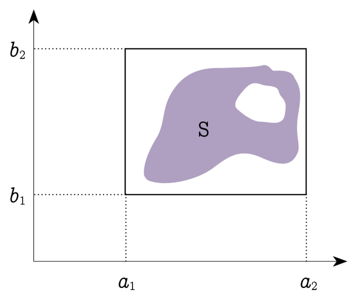


FIGURA 5.8

Línea 1:	Simule $U_1 \sim Uniforme[a_1, a_2]$.
Línea 2:	Simule $U_2 \sim Uniforme[b_1, b_2]$.
Línea 3:	Si $(U_1, U_2) \in S$ entonces asigne $(X, Y) = (U_1, U_2)$. Si no, vuelva a la línea 1.

TABLA 5.13. Algoritmo para simular un punto (X, Y) equidistribuido en una sección S del plano, contenida en el rectángulo con vértices (a_1, b_1), (a_2, b_1), (a_2, b_2) y (a_1, b_2).

Ejercicio 15. Implemente el Método de la Transformada Inversa para simular la extracción—sin reposición—de dos bolitas al azar de una urna que contiene tres bolitas de color rojo, tres de color blanco, y cuatro de color azul.

Ejercicio 16. Use el Teorema 4.3 para desarrollar un algoritmo basado en la simulación de variables aleatorias del tipo Bernoulli, para simular una v.a. geométrica con parámetro p.

Ejercicio 17. Considere el problema de simular un punto (X, Y) equidistribuido en la región elíptica $x^2/9 + y^2/16 \leq 1$.

(a) Muestre que X tiene una función de densidad dada por la fórmula

$$f_X(x) = \begin{cases} 0 & , \quad x \notin [-3, 3]; \\ \frac{2}{3\pi}\sqrt{1 - \frac{x^2}{9}} & , \quad x \in [-3, 3]. \end{cases}$$

(b) Observe que si $X = x$, entonces necesitamos $|Y| \leq 4\sqrt{1 - x^2/9}$, para que (X, Y) pertenezca a la región elíptica. Más aún, como (X, Y) es equidistribuido en la región elíptica resulta intuitivo que

$$Y \sim Uniforme\left[-4\sqrt{1 - \frac{x^2}{9}}, 4\sqrt{1 - \frac{x^2}{9}}\right],$$

cuando $X = x$. Use esto para diseñar un algoritmo para simular (X, Y).

Bibliografía

[1] Barbour, A. D. and Chen, L. H. Y., editors. *An Introduction to Stein's Method*. Singapore University Press, 2005.
[2] Ross, S. M. *Simulation*. Academic Press, fourth edition, 2006.
[3] Lacourly, N. *Introducción a la Estadística*. J.C. Sáez Editor, Santiago, 2011.
[4] Romagnoli, P. P. *Probabilidades doctas con discos, árboles, bolitas y urnas*. J.C. Sáez Editor, Santiago, 2011.
[5] Durrett, R. *Essentials of Stochastic Processes*. Springer, corrected edition, 2001.
[6] Durrett, R. *Probability: Theory and Examples*. Duxbury Press, third edition, 2004.
[7] Républica de Chile, Ministerio de Educación: *Matemática: Programa de Estudio Segundo Año Medio*. 2004.
[8] Républica de Chile, Ministerio de Educación: *Matemática: Programa de Estudio Tercer Año Medio*. 2004.
[9] Républica de Chile, Ministerio de Educación: *Matemática: Programa de Estudio Cuarto Año Medio*. 2004.
[10] Pitman, J. *Probability*. Springer, 1999.
[11] Ross, S. M. *A First Course in Probability*. Prentice Hall, Séptima Edición, 2006.
[12] Ross, S. M. *Introduction to Probability Models*. Academic Press, Novena Edición, 2006.
[13] Bremaud, P. *Markov Chains*. Springer, Edición Corregida, 2008.

Índice de Términos

círculo unitario, 75
cardinalidad de un cojunto, 28
coeficiente binomial, 139
complemento de un suceso, 24
conjunto contable, 27
conjunto enumerable, 27
Corolario 4.16, 162
covarianza entre dos variables aleatorias, 128

desigualdad de Berry-Esseen, 170
desigualdad de Chebyshev, 129
desviación estándar de una v.a., 117
diferencia de sucesos, 30
distribución gama, 179
distribución Normal, 156
distribución Normal estándar, 156

espacio muestral, 23
esperanza condicional, 126
esperanza de una v.a., 100
esperanza de una v.a. continua, 105
esperanza de una v.a. dado un suceso, 126
esperanza de una v.a. discreta, 102
esperanza v.a. del tipo Bernoulli, 134
esperanza v.a. del tipo Poisson, 145
esperanza v.a. del tipo uniforme, 106
esperanza v.a. exponencial, 152
esperanza v.a. geométrica, 137
esperanza v.a. Normal, 156
esperanza variables aleatorias con la misma distribución, 110

Fórmula Binomial, 139
Fórmula de Inclusión-Exclusión, 37, 49
Fórmula de Integración por Partes, 152
Fórmula de Probabilidades Totales, 31–33
factorial de un número, 139
fenómenos modelados con la distribución de Poisson, 150
función de densidad, 62
función de distribución de una v.a. Normal estándar, 158
función de masa de probabilidad, 89
función de variables aleatorias con la misma distribución, 76
función generadora de momento, 128
función indicadora de un suceso, 98
función indicadora de una proposición, 98
función con límite a la izquierda, 58
función continua a la derecha, 58
función de distribución, 57, 58
funciones de variables aleatorias independientes, 82

igualdad de sucesos, 24
igualdad de variables aleatorias, 73
igualdad en distribución de variables aleatorias continuas, 78
igualdad en distribución de variables aleatorias discretas, 77
igualdad en distribución, 73
inclusión de sucesos, 24
independencia de dos sucesos, 34
independencia de variables aleatorias, 79
independencia de variables aleatorias continuas, 85
independencia de variables aleatorias discretas, 83
independencia de varios sucesos, 36
integración por partes, 152

intensidad de un evento, 153
interpretación del parámetro λ en la distribución exponencial, 153
interpretación de la varianza, 121
interpretación frecuentista de probabilidades, 26, 97
intersección contable de sucesos, 41
intersección de sucesos, 25, 41
intervalo de confianza para μ, 163

Ley Débil de los Grandes Números, 130
Leyes de Morgan, 41
linealidad de la esperanza, 111

Método de Integración por Partes, 152
Método de la Transformada Inversa, 181
Método de Rechazo, 194
Método del Jacobiano, 69, 72
medida de probabilidad, 26

ocurrencia de sucesos, 23

paréntesis de Iverson, 98
Principio de Inclusión-Exclusión, 49
probabilidad condicional, 31
probabilidad de un suceso, 26
probabilidad de un suceso dado otro, 31
propiedad desmemoriada de la distribución exponencial, 155
punto equidistribuido en un cuadrado, 81
punto equidistribuido en un intervalo, 57
punto equidistribuido en una región, 65, 75
punto fijo de una función, 45

representación de sucesos como conjuntos, 23

secuencia contable de sucesos, 27
secuencia creciente de sucesos, 38
secuencia decreciente de sucesos, 38
secuencia monótona de sucesos, 38
serie de Taylor función exponencial, 145
serie geométrica, 29
sigma aditividad, 27
simulación de variables aleatorias
    binomiales, 193
    del tipo Bernoulli, 191
    del tipo Poisson, 193
    exponenciales, 186
    geométricas, 192
    normales, 197
substracción de sucesos, 30
suceso, 23
suceso asociado a una v.a., 53
suceso asociado a varias variables aleatorias, 54
suceso vacío, 24
sucesos disjuntos, 25
sucesos independientes, 34, 36

Teorema Central Límite, 165
Teorema de Fubini, 107
transformación de variables aleatorias con la misma distribución, 76
transformaciones de variables aleatorias independientes, 82

unión contable de sucesos, 26
unión de sucesos, 24, 26

v.a., 53
v.a. binomial, 139
v.a. con esperanza bien definida, 100
v.a. con esperanza finita, 100
v.a. constante, 110
v.a. continua, 62
v.a. del tipo Bernoulli, 133
v.a. del tipo gama, 179
v.a. del tipo Poisson, 144, 150
v.a. discreta, 61
v.a. exponencial, 78, 151
v.a. geométrica, 135
v.a. Normal, 78, 156
v.a. uniforme, 66
v.a. uniforme en un intervalo general, 69
valor esperado de una v.a., 100
variable aleatoria, 53
variables aleatorias continuas con la misma distribución, 78

variables aleatorias discretas con la misma distribución, 77
variables aleatorias i.i.d., 108
varianza, 116
varianza de una v.a., 116
varianza muestral, 121
varianza v.a. del tipo Bernoulli, 134
varianza v.a. del tipo Poisson, 145
varianza v.a. exponencial, 152
varianza v.a. geométrica, 137
varianza v.a. Normal, 156