NOTAS DE PROBABILIDAD Y ESTADÍSTICA

JORGE LUIS LÓPEZ LÓPEZ
(BASADO EN EL LIBRO *UNDERSTANDING PROBABILITY,* DE HENK TIJMS)

En el fondo, la teoría de la probabilidad no es más que el sentido común reducido a cálculo; nos permite apreciar con exactitud lo que las mentes precisas sienten con una especie de instinto que muchas veces son incapaces de explicar. . . . Nos enseña a evitar las ilusiones que a menudo nos engañan; . . . No hay ciencia más digna de nuestra contemplación ni más útil para incluirla a nuestro sistema de educación pública.

Pierre Simon Laplace, Théorie Analytique des Probabilités (1812)

1. Introducción

El dado existe desde en la India y en el Oriente Medio desde hace más de catorce siglos. Aunque los juegos de azar existen desde antes: en Egipto se encontró alrededor del año 3500 a.C. un tablero de un juego de azar en que se lanzaban cuatro huesos de borrego.

Los juegos de dados y otros juegos de azar fueron sujetos por primera vez a un estudio matemático formal por el matemático y físico italiano Gerolamo Cardano, quien escribió un libro titulado *El Libro de los Juegos de Azar* (1564, en latín, publicado hasta 1663). La aportación de Cardano consistió en enfocarse en el conjunto de resultados posibles de un experimento, y para los casos en que todos los resultados son igualmente probables definió la probabilidad de que ocurra un suceso como lo grande que es el conjunto que representa a ese suceso con respecto al conjunto que representa a todos los sucesos posibles, es decir, la probabilidad de que suceda el suceso A es igual a

$$p(A) = \frac{\text{tamaño del conjunto que representa al suceso } A}{\text{tamaño del conjunto que representa a todos los sucesos posibles}}. \tag{1}$$

Ejemplo 1. Al lanzar tres dados, hay más posibilidad de que la suma de los números obtenidos sea 10 a que sea 9.

Esta afirmación se puede justificar de la siguiente manera. Los números obtenidos en cada lanzamiento se pueden codificar por un vector con tres entradas, siendo la primera entrada el primer número que sale, la segunda el segundo que sale, y la tercera el tercero en salir. Entonces, todos los vectores que codifican un lanzamiento cuya suma es 10 son los siguientes

$$\begin{array}{cccccc}
(1,3,6) & (2,2,6) & (3,1,6) & (4,1,5) & (5,1,4) & (6,1,3) \\
(1,4,5) & (2,3,5) & (3,2,5) & (4,2,4) & (5,2,3) & (6,2,2) \\
(1,5,4) & (2,4,4) & (3,3,4) & (4,3,3) & (5,3,2) & (6,3,1) \\
(1,6,3) & (2,5,3) & (3,4,3) & (4,4,2) & (5,4,1) & \\
 & (2,6,2) & (3,5,2) & (4,5,1) & & \\
 & & (3,6,1) & & &
\end{array}$$

que son 27 en total. Un análisis análogo se puede hacer para ver que los vectores que codifican un lanzamiento cuya suma es 9 son 25 en total. Esto muestra que hay más chance de que la suma de los números obtenidos sea 10 a que sea 9 (si el dado es honesto).

Si además se quiere calcular la probabilidad de que el lanzamiento arroje una suma igual a 10, notar que hay $6^3 = 216$ vectores que codifican todos los resultados posibles al lanzar los tres dados.

Por lo tanto, aplicando la fórmula (1), la probabilidad de que el lanzamiento arroje una suma igual a 10 es $\frac{27}{216}$. Análogamente, la probabilidad de que el lanzamiento resulte en una suma igual a 9 es $\frac{25}{216}$.

Ejemplo 2. Se escogen al azar los números b y c en el intervalo $(-1,1)$. ¿Cuál es la probabilidad de que la ecuación $x^2+bx+c=0$ tenga raíces reales?

Para contestar la pregunta, es natural representar al par de números b y c mediante el punto (b,c) en el plano cartesiano $\mathbb{R}^2$. Entonces la hipótesis se traduce en que el punto (b,c) se escoge al azar en el cuadrado $\mathcal{C}=(-1,1)\times(-1,1)$, que es de tamaño 2×2, y la condición de tener raíces reales se traduce en la desigualdad $0\leq b^2-4c$, que determina una región de área $2+\int_{-1}^{1}\frac{x^2}{4}dx=2+\frac{1}{6}$ dentro del cuadrado $\mathcal{C}$. Usando la fórmula (1), se concluye que la probabilidad buscada es $\frac{13/6}{4}=\frac{13}{24}$.

1.0.1. Tarea. Verificar las afirmaciones siguiente.

1. Si una familia tiene cuatro hijos, es menos probable tener dos hijos de cada sexo que tener tres hijos de un sexo y otro del otro sexo.
2. Tres niños quieren elegir a uno de ellos. Para esto, cada uno de ellos tira un volado simultáneamente con los demás. Se elige aquel que obtenga un resultado distinto. Entonces la probabilidad de que uno de ellos sea elegido en el primer volado simultáneo es mayor a que probabilidad de que tengan que seguir jugando más volados para elegir a uno.
3. Fulano y Sutano lanzan simultáneamente un dado cada uno y luego calculan la diferencia de los números que resultan (la diferencia es el valor absoluto de la resta). Fulano gana si la diferencia es 0, 1 o 2. En cualquier otro caso gana Sutano. Este juego no es justo: tiene más probabilidad de ganar Fulano.
4. Sea θ un ángulo escogido al azar en el intervalo $[0,2\pi)$. La probabilidad de que $\sin^2\theta<1/2$ resulta ser igual a $1/2$.
5. Fulano tiene un dado con los números 5, 5, 5, 1, 1, 1. Sutano tiene un dado con los números 4, 4, 4, 4, 0, 0. Ambos lanzan sus dados. La probabilidad de que Fulano obtenga un número mayor que Sutano es $2/3$.

   Mengano tiene un dado con los números 3, 3, 3, 3, 3, 3. Perengano tiene un dado con los números 6, 6, 2, 2, 2, 2. Ambos lanzan sus dados. La probabilidad de que Mengano obtenga un número mayor que Perengano es de $2/3$.

   La probabilidad de que Sutano obtenga un número mayor que Mengano también es $2/3$.

   En vista de lo anterior, uno podría pensar que la probabilidad de que Fulano obtenga un número mayor que Perengano sea mayor que $1/2$, pero resulta ser $1/3$.

La definición de probabilidad mediante la fórmula (1) puede ser en estos días algo obvia. Pero a partir de ella uno puede desarrollar la teoría para calcular probabilidades más complejas que incluso nos permiten formalizar situaciones que no son obvias, como en el siguiente ejemplo, que a su vez también pone a prueba nuestra intuición acerca de la probabilidad.

El dilema de Monty Hall. Se acerca el clímax de un concurso. Suena un redoble de tambores. El presentador del concurso te lleva a una pared con tres puertas cerradas. Detrás de una de las puertas está el automóvil de tus sueños, y detrás de cada una de las otras dos hay una lata de comida para perros. Las tres puertas tienen las mismas posibilidades de ocultar el automóvil. El presentador, una persona de confianza que sabe exactamente qué hay detrás de cada una de las tres puertas, te explica cómo funcionará el juego. Primero, elegirás una puerta sin abrirla, sabiendo que después de que lo hayas hecho, el presentador abrirá una de las dos puertas restantes para mostrar una lata de comida para perros. Cuando hayas hecho esto, tendrás la oportunidad de cambiar de puerta; ganarás lo que esté detrás de la puerta que elijas en esta etapa del juego. ¿Aumentas tus posibilidades de ganar el automóvil si cambias de puerta? Probaremos más adelante que sí. Mostraremos después que la probabilidad de ganar si cambiamos de puerta es 2/3.

Este dilema apareció en el el New York Times en 1991. Algunas personas insisten con vehemencia que no importa si un jugador cambia de puerta al final del juego, mientras otras sostienen con confianza que el jugador debe cambiar de puerta. Incluso muchos profesores de matemáticas se equivocan al responder esta cuestión. Este tipo de ejemplos demuestran que, en situaciones de incertidumbre, uno necesita métodos racionales para evitar trampas mentales. La teoría de la probabilidad nos proporciona esos métodos.

1.1. La probabilidad en la práctica. En 1654 el apostador Chevalier De Méré mencionó, en una carta dirigida a Pascal, que había descubierto una contradicción en la aritmética basándose en el siguiente razonamiento: De Méré sabía que era ventajoso apostar que al menos un seis saldría al tirar un dado 4 veces, pero su experiencia como jugador le enseñó que no era ventajoso apostar a que al menos caería un doble seis al lanzar un par de dados 24 veces, y entonces concluyó que no se valía la ley de las proporciones, pues 4 es a 6 (siendo 6 los posibles resultados al lanzar un dado) como 24 es a 36 (siendo 36 los resultados posibles al lanzar dos dados). El mismo Cardano considera el problema en el capítulo 11 de su libro y calcula de manera incorrecta, sin usar su fórmula (1), que con 18 lanzamientos de un par de dados la probabilidad de que aparezca al menos un doble seis es mayor que 1/2. Más allá de los razonamientos incorrectos, al usar la fórmula (1) resulta que la probabilidad de que al menos un seis salga en un lanzamiento de cuatro dados es igual a 0.5177, mientras que la probabilidad de que al menos un seis doble salga al lanzar un par de dados 24 veces es 0.4914. ¿Cómo fue capaz de estimar empíricamente De Méré estas probabilidades sin hacer el cálculo matemático necesario? Lo que sucede es que hay una manera empírica de calcular la probabilidad, además de la teórica que resulta del uso de la fórmula (1).

Suponer que cierto experimento de azar se repite un gran número de veces bajo las mismas condiciones, y de manera que las repeticiones del experimento son independientes una de otra (es decir, el resultado en cada repetición no influye en el resultado en otras repeticiones). Sea A un suceso posible en el experimento. La *frecuencia relativa* del suceso A en las primeras n repeticiones de ese experimento se define como

$$f_n(A) = \frac{\text{número de veces que el suceso } A \text{ ocurre en las primeras } n \text{ repeticiones del experimento}}{n}.$$

Entonces uno podría esperar que al repetir más y más veces el experimento, la frecuencia relativa se va aproximando a la probabilidad de que suceda A, es decir,

$$p(A) = \lim_{n\to\infty} f_n(A), \tag{2}$$

lo cual de paso daría una aplicación práctica de la probabilidad. Los fenómenos que satisfacen la fórmula (2) se denominan *aleatorios* o *estocásticos*. La capacidad de las compañias de seguros, o de los casinos, para generar beneficios se basan en esta fórmula.

Un problema de cumpleaños. Vas con un amigo a ver un partido de fútbol. En el partido participan 22 jugadores de los dos equipos y un árbitro. Tu amigo apuesta a que, entre esas 23 personas que hay en la cancha, al menos dos tendrán la misma fecha de cumpleaños. Si no es así, recibirás cien pesos de tu amigo. ¿Cuánto dinero deberías pagarle a tu amigo para que la apuesta sea justa? Uno puede responder esta pregunta usando la fórmula (2) como sigue.

Vamos a excluir el 29 de febrero de nuestros cálculos y vamos a suponer que los 365 días que quedan son igualmente probables (la cual puede ser una hipótesis cuestionable). Los días de cumpleaños de las 23 personas en la cancha pueden arreglarse en un vector con 23 entradas; en la primer entrada está el cumpleaños de la persona 1, en la segunda el de la persona 2, y así sucesivamente. Para contar la cantidad de vectores que poseen entradas repetidas podemos considerar todos los vectores posibles y quitarle los vectores que no tienen entradas repetidas:

$$\begin{array}{c}\text{cantidad de vectores}\\ \text{con entradas repetidas}\end{array} = \underbrace{365^{23}}_{\text{vectores posibles}} - \underbrace{365 \times 364 \times 363 \times 362 \times \cdots \times 344 \times 343}_{\text{vectores que no tienen entradas repetidas}}.$$

Luego la probabilidad de que el vector tenga entradas repetidas es

$$\frac{365^{23} - 365 \times 364 \times \cdots \times 343}{365^{23}} = 1 - \left(\frac{365}{365}\right) \times \left(\frac{364}{365}\right) \times \cdots \times \left(\frac{343}{365}\right) \approx 0.5073.$$

Sea x la cantidad que recibirá tu amigo si gana la apuesta. Entonces tu amigo ganará x pesos con probabilidad 0.5073 y perderá 100 pesos con probabilidad 0.4927. Aquí es donde entra en juego la fórmula (2). Supongamos que la apuesta se repite una gran cantidad de veces. Denotar por g_n las veces que tu amigo gana en las primeras n apuestas. Por lo tanto tu amigo ganará xg_n pesos y perderá $100(n-g_n)$ pesos en las primeras n apuestas. Decir que la apuesta es justa se traduce en decir que $\lim_{n\to\infty} [xg_n - 100(n-g_n)] = 0$; en otras palabras, al incrementar el número de apuestas la ganancia del amigo tiende a cero (esto significa que tu amigo no se enriquece ni se empobrece repitiendo la apuesta una cantidad indefinida de veces, y tú tampoco). Entonces $\lim_{n\to\infty} \frac{xg_n - 100(n-g_n)}{n} = 0$ por el límite anterior. Puesto que $\frac{xg_n-100(n-g_n)}{n} = x\frac{g_n}{n} - 100(1 - \frac{g_n}{n}) = xf_n - 100(1 - f_n)$ donde f_n es la frecuencia relativa con la que el amigo gana la apuesta en las primeras n apuestas, el límite anterior se traduce en

$$\lim_{n\to\infty} [xf_n - 100(1 - f_n)] = x \lim_{n\to\infty} f_n - 100(1 - \lim_{n\to\infty} f_n) = 0.$$

Usando la fórmula (2), esta última igualdad se convierte en $xp - 100(1 - p) = 0$, donde $p \approx 0.5073$ es la probabilidad de que el amigo gane la apuesta. Sustituyendo el valor de p en esta ecuación con incógnita x se tiene $0.5073x = 49.27$, dando como resultado $x \approx 97.122$.

En el ejemplo anterior se muestra el concepto de *ganancia esperada*:

$$\begin{array}{c}\text{ganancia}\\ \text{esperada}\end{array} = \left(\begin{array}{c}\text{cantidad ganada en}\\ \text{cada experimento}\end{array}\right) \times \left(\begin{array}{c}\text{probabilidad de}\\ \text{ganar tal cantidad}\end{array}\right) - \left(\begin{array}{c}\text{cantidad perdida en}\\ \text{cada experimento}\end{array}\right) \times \left(\begin{array}{c}\text{probabilidad de}\\ \text{perder tal cantidad}\end{array}\right)$$

En particular, la ganancia esperada de tu amigo fue $xp-100(1-p)$. Este concepto, o su generalización que lleva el nombre de *valor esperado*, el cual revisaremos más adelante, aparecen en el libro *Sobre el razonamiento en los juegos de azar* (1660, en alemán) del astrónomo alemán Christiaan Huygens. Entre los problemas analizados en este libro se encontraba el de cómo dos jugadores en un juego de azar deberían dividir las apuestas si el juego termina prematuramente. Veamos otros dos ejemplos del concepto de ganancia esperada, aplicado a seguros y a casinos.

Seguro contra incendios. Un industrial pide a una compañía de seguros que le asegure su fábrica contra un incendio. La compañía de seguros averigua que la probabilidad anual de incendio de ese tipo de fábricas es igual a $\frac{2}{1000}$, y que la fábrica del industrial está valuada en $3,000,000$ pesos. Suponer que la póliza de seguro que ofrece la compañía al industrial tiene un cuesto de x pesos, y que el seguro se renovará cada año una gran cantidad de veces. Denotar por q_n las veces que se incendia la fábrica en los primeros n años. Entonces la ganancia de la compañía de seguros durante los primeros n años será de $xn - 3,000,000q_n$ pesos, que representa una ganancia anual de $x - 3,000,000f_n$ pesos, donde $f_n = q_n/n$ es la frecuencia relativa de incendios. Esta ganancia debe ser positiva con el paso del tiempo para que la compañía de seguros obtenga un beneficio por su servicio, es decir, $0 < \lim_{n\to\infty} (x - 3,000,000f_n) = x - 3,000,000 \lim_{n\to\infty} f_n = x - 3,000,000(\frac{2}{1000})$ (en la última igualdad se usó la fórmula (2)). Se concluye que la póliza contra incendios debe tener un valor x mayor que $6,000$ pesos. Notar que la ganancia esperada por año de la compañia de seguros es precisamente $x(1) - 3,000,000(\frac{2}{1000}) = x - 6,000$.

Ruleta europea. La ruleta europea usa los números del 0 al 36, siendo 0 un número reservado para la casa. El jugador compite contra la casa escogiendo k números, con $k \in \{1, 2, \ldots, 36\}$. El jugador gana si cae cualquiera de sus números escogidos, recibiendo $\frac{36}{k} - 1$ veces la cantidad que apostó, adicional a lo apostado. La ganancia esperada del casino por cada euro apostado es

$$1 \times \left(\frac{37-k}{37}\right) - \left(\frac{36}{k} - 1\right) \times \frac{k}{37} = \frac{1}{37}, \quad \text{i.e. aproximadamente 2.7 centavos.}$$

En decir, se espera que a la larga por cada 37 euros apostados el casino se quede con 1 euro.

Analicemos un poco más la fórmula (2). Por ejemplo, si A es el evento de que caiga ágila en un volado, la fórmula (2) no dice nada de lo que sucederá en cada volado por sí solo, más bien predice el comportamiento a la larga en la situación hipotética de que los volados se repitan una cantidad ilimitada de veces. Al lanzar un volado la probabilidad de obtener águila es $1/2$, pero esto no quiere decir que al lanzar muchos volados la cantidad de águilas y de caras debería parecerse cada vez más, esta es la *falacia del apostador*: Al jugar volados, después de un gran número de veces consecutivas que cae águila el apostador esperaría que caiga una cara, pero esto no necesariamente sucede; es un error pensar que los volados previos influirán en los siguientes, pues la moneda no tiene memoria. Lo que sucede típicamente es que al ir lanzando cada vez más volados la diferencia entre el número de caras y de águilas se hace arbitrariamente grande. Por ejemplo, imaginar que al lanzar $2n$ volados se obtienen $a_n = n + \lfloor\sqrt{n}\rfloor$ águilas, donde $\lfloor x\rfloor$ denota la función *mayor entero menor o igual que* x. Entonces en esos $2n$ volados deben caer $c_n = n - \lfloor\sqrt{n}\rfloor$ caras. En esta situación particular la diferencia entre águilas y caras se hace arbitrariamente grande mientras que cualquiera de los cocientes $a_n/(2n)$ y $c_n/(2n)$ tiende a $1/2$, respetando la fórmula (2).

1.1.1. Ley de los grandes números de Kolmogorov. Hay un teorema conocido como la *Ley de los Grandes Números* que explica la fórmula (2) en un sentido matemático, demostrado alrededor de 1930 por Andrey Nikolayevich Kolmogorov, uno de los matemáticos más sobresalientes del siglo XX.

No hay razón para que la fórmula (2) se cumpla siempre. Por ejemplo, supongamos la situación "rara" (¿en qué sentido?) en la que al aventar una cantidad indefinida de volados, primero cae 1 águila, luego caen 2 caras, luego 4 águilas, luego 8 caras, luego 16 águilas, luego 32 caras, luego 64 águilas, y así sucesivamente. Entonces se puede ver que $f_{2^{2k}-1} = 1/3$ para todo entero positivo k, mientras que $f_{2^{2k+1}-1} = (4 \times 4^k - 1)/(6 \times 4^k - 3)$, la cual tiende a $2/3$ cuando $k \to \infty$, y por lo tanto $f_n(A)$ no converge. También existe la situación "rara" en la que a partir de cierto volado caen solamente caras, por ejemplo, en cuyo caso $\lim_{n\to\infty} f_n(A) = 0$, en la que eventualmente caen solo águilas, en cuyo caso $\lim_{n\to\infty} f_n(A) = 1$. Sin embargo, Kolmogorov lográ formalizar en qué sentido tales situaciones son "raras":

> **Afirmación**. Si se avientan volados una cantidad ilimitada de veces, el resultado es una palabra infinita ω de C's y A's, donde cada letra registra si cayó cara (C) o águila (A) en cada volado. Se define
>
> $$K_n(\omega) = \text{ la cantidad de } C\text{'s en las primeras } n \text{ letras de } \omega.$$
>
> Por ejemplo, si $\omega = CAACCCACC\ldots$, se tiene que $K_5(\omega) = 3$ y $K_8(\omega) = 5$. El cociente $K_n(\omega)/n$ es llamado *frecuencia relativa* del número de caras en la palabra ω. El conjunto de todas las palabras infinitas de C's y A's posee una *medida* (i.e. una forma de medir subconjuntos del conjunto de tales palabras) que es natural y que se puede construir explícitamente, en la cual el subconjunto
>
> $$T = \left\{\omega \ : \ \lim_{n\to\infty} \frac{K_n(\omega)}{n} = \frac{1}{2}\right\}$$
>
> es de medida total (i.e. su complemento es de medida cero).

Naturalmente, lo mismo ocurre para la frecuencia relativa del número de águilas. Esta afirmación, junto con la construcción de dicha medida, sería el enunciado de la Ley de los Grandes Números en el contexto de los volados, pero Kolomogorov demostró este teorema en un contexto más general que el de los volados, incluyendo cualquier experimento de azar que se repite una cantidad ilimitada de veces bajo las mismas condiciones, en el cual cada repetición es independiente de las demás, concluyendo que la fórmula (2) se cumplirá, excepto en situaciones “muy raras” que forman un conjunto de medida cero en el espacio de todas las situaciones posibles.

1.1.2. Tarea.

1. Realizar 15 repeticiones del experimento del Ejemplo 1, indicando la frecuencia con la que se dió la suma 10 y la suma 9.
2. Realizar 20 volados, indicando la frecuencia con la que cayó águila.
3. Resulta que la probabilidad de que en un grupo de 101 personas al menos dos tengan la misma fecha de cumpleaños es aproximadamente igual a 0.9999998. En su libro *Innumeracy: Mathematical Illiteracy and its consecuences* (1988), John Allen Paulos expone ejemplos de interpretación incorrecta de resultados o teorías matemáticas. En uno de los ejemplos del libro, se menciona que Johnny Carson en su programa de televisión The Tonight Show usó esta probabilidad aproximada de 0.9999998 para asegurar que entre los 100 invitados a su show debe haber uno que tenga la misma fecha de cumpleaños que él. De esta manera, Carson confundió dos problemas de probabilidad distintos:
   - La probabilidad de que dos personas de un grupo de 101 personas tengan el mismo cumpleaños.
   - La probabilidad de que al menos uno de los 100 invitados tengan el mismo cumpleños que Carson (mostrar que esta probabilidad es aproximadamente 0.2399).

   Verificar que la audiencia debería estar compuesta por al menos 253 personas para que la probabilidad de que alguno de ellos tuviera la misma fecha de cumpleaños que Carson supere 1/2.

1.2. Definiciones formales. El *espacio muestral* de un experimento es un conjunto $\mathfrak{M}$ que representa todos y cada uno de los resultados posibles del experimento. Cualquier subconjunto $A \subset \mathfrak{M}$ es un *evento*. Una *medida de probabilidad* es una función p que asigna un número real a cada evento, de manera que se cumplen los siguientes axiomas:

- $p(A) \geq 0$ para todo evento A.
- $p(\mathfrak{M}) = 1$, donde $\mathfrak{M}$ es el espacio muestral.
- $p(A \cup B) = p(A) + p(B)$ para eventos ajenos A y B, es decir, $A \cap B = \emptyset$ (lo cual se interpreta como que los eventos A y B no pueden ocurrir simultáneamente).

El número $p(A)$ es llamado *probabilidad* del evento A.

1.2.1. Ejemplos de medidas de probabilidad.

1. Cuando $\mathfrak{M}$ es un conjunto finito con m elementos, se puede considerar
$$p(A) = \frac{\text{número de elementos de } A}{m},$$
que coincide con la definición de Cardano (1). Esta medida de probabilidad se caracteriza por ser aquella en la que todo elemento de $\mathfrak{M}$ es igualmente probable, es decir, $p(\{x\}) = 1/m$ para todo $x \in \mathcal{M}$, donde $\{x\} \subset \mathfrak{M}$ es el evento que consiste solamente del elemento x.
2. Cuando el espacio muestral $\mathfrak{M}$ está representado como una superficie y todos los puntos de la superficie se consideran igualmente probables (ya sea por simetría, o incluso por ignorancia),

la probabilidad asignada a un evento $A \subset \mathfrak{M}$ podría definirse intuitivamente como

$$P(A) = \frac{\text{área de } A}{\text{área de } \mathfrak{M}},$$

que coincide con la definición de Cardano (1) y cumple los tres axiomas.

3. Cuando el espacio muestral $\mathfrak{M}$ está representado como una curva y todos los puntos de la superficie se consideran igualmente probables (ya sea por simetría, o incluso por ignorancia), la probabilidad asignada a un evento A podría definirse intuitivamente como

$$P(A) = \frac{\text{longitud de } A}{\text{longitud de } \mathfrak{M}},$$

que coincide con la definición de Cardano (1) y cumple los tres axiomas.

1.2.2. Ejemplo de la aguja de Buffon, 1777. El siguiente es un ejemplo clásico en el que se usa el área para comparar el tamaño de conjuntos que representan sucesos a fin de obtener una probabilidad. Considerar un piso que tiene rectas paralelas equiespaciadas a distancia d. Una aguja de longitud $\ell \leq d$ es lanzada al piso por una persona con los ojos vendados. La probabilidad p de que la aguja caiga cruzando alguna de las rectas resulta ser $p = \frac{2\ell}{\pi d}$.

En 1901 el matemático Mario Lazzarini reportó haber lanzado una estaca 3408 veces en un piso reglado, de manera las longitudes ℓ y d satisfacían la razón $\ell/d = 5/6$. Lazzarini reportó que la estaca cruzó las rectas paralelas en 1808 ocasiones. Luego $\pi = \frac{2\ell}{d}\frac{1}{p} \approx \frac{5}{3} \times \frac{3408}{1808} = \frac{355}{113} \approx 3.14159292$ y, por otro lado $\pi \approx 3.14159265$. Posteriormente veremos que casi con seguridad Lazzarini mintió.

1.2.3. Variable aleatoria. En ocasiones podemos estar interesados, más que en los resultados de un experimento, en un valor numérico determinado por los resultados de ese experimento, es decir, en una función $X : \mathfrak{M} \to \mathbb{R}$. Esto lleva al concepto de variable aleatoria: una *variable aleatoria* es una función que asigna un valor numérico a cada elemento del espacio muestral. Por ejemplo, al lanzar tres dados, la suma de los números que resultan es una variable aleatoria. Otro ejemplo: cuando uno escoge un ángulo θ al azar, el número $\sin^2\theta$ es una variable aleatoria.

Es común usar letras mayúsculas, como X, Y y Z, para denotar variables aleatorias, y las respectivas minúsculas x, y y z para denotar sus valores numéricos posibles. El evento $\{X = x\}$ consta de aquellos elementos del espacio muestral para los cuales la variable aleatoria X toma el valor x. La probabilidad del evento $\{X = x\}$ se denota por $p(X = x)$. Por ejemplo, si X es la variable aleatoria dada por la suma de los números que resultan al lanzar tres dados, $X = 10$ es el evento que consta de los 27 elementos listados en el Ejemplo 1 de la pág. 1, y $p(X = 10) = 27/216$.

1.2.4. Valor esperado de una variable aleatoria finita. Considerar un juego de casino donde el jugador tiene 0.7 de probabilidad de perder 100 pesos, 0.25 de probabilidad de ganar 200 pesos, y 0.05 de ganar 300 pesos. Un jugador que juega este juego un número n de veces, de manera que pierde a_n de ellos, gana 200 en b_n de ellos, y gana 300 en c_n de ellos, gana en promedio por cada juego $-100\frac{a_n}{n} + 200\frac{b_n}{n} + 300\frac{c_n}{n}$. Al hacer n arbitrariamente grande y usar la fórmula (2), vemos que a la larga la ganancia por juego del jugador es $-100 \times 0.7 + 200 \times 0.25 + 300 \times 0.05 = -5$ pesos. Este es el concepto de valor esperado, que ya hemos usado y puede definirse para cualquier variable aleatoria como sigue.

El *valor esperado* (o *esperanza*, o *media*, o *primer momento*) de una variable aleatoria que toma un número finito de valores $\{x_1, x_2, \ldots, x_M\}$ se define por

$$E(X) = x_1 p(X = x_1) + x_2 p(X = x_2) + \cdots + x_M p(X = x_M).$$

1.3. Ejemplos.

1. En el experimento de lanzar un dado dos veces, el valor esperado de la suma es 7, pues si X es la variable aleatoria que asigna la suma de los números que salen en ambos dados, se tiene

$$E(X) = 2\times\tfrac{1}{36}+3\times\tfrac{2}{36}+4\times\tfrac{3}{36}+5\times\tfrac{4}{36}+6\times\tfrac{5}{36}+7\times\tfrac{6}{36}+8\times\tfrac{5}{36}+9\times\tfrac{4}{36}+10\times\tfrac{3}{36}+11\times\tfrac{2}{36}+12\times\tfrac{1}{36}=7.$$

2. Un barista famoso en Amsterdan ofrece a sus clientes el siguiente juego. El cliente primero escoge un número de $\{1,\ldots,6\}$ y luego avienta un dado 3 veces. Si el número que el cliente escogió no aparece en ningún lanzamiento, entonces debe pagar 100 euros al barista, pero si el número cae una, dos o tres veces, el barista paga 100, 200 o 300 euros al cliente, respectivamente. ¿Qué tanta ganancia presenta el juego para el barista?

   Sea X la variable aleatoria que cuenta las veces que sale el número escogido por el cliente. Entonces $E(X) = 100\times\dfrac{5^3}{6^3}-100\times\dfrac{3\times5^2}{6^3}-200\times\dfrac{3\times5}{6^3}-300\times\dfrac{1}{6^3}$, que es aproximadamente 7.87 euros por juego.
3. Si una familia tiene 4 hijos, denotar por X al número de hijas mujeres de esos 4 hijos. Entonces $E(X) = 2$.
4. Sea X al número que resulta en el experimento de lanzar un dado. Entonces $E(X) = 7/2$.

1.3.1. Tarea.

1. Se elige una sola carta al azar del mazo de 52 cartas. Mostrar que la probabilidad de que la carta escogida sea as o corazón es igual a 4/13.
2. Se reciben 5 cartas tomadas al azar del mazo de 52 cartas. ¿Cuál es la probabilidad de que al menos un as se encuentre entre las cartas que se recibieron?
3. Dos personas quedan en verse en el aeropuerto entre las 10:00 y las 11:00 hrs. Cada persona llega en algún momento al azar en ese lapso de tiempo. Mostrar que la probabilidad de que las dos personas lleguen con menos de 10 minutos de diferencia es 11/36.
4. Probar que para cualquier evento A se tiene $p(A) = 1 - p(A^c)$ donde $A^c = \mathfrak{M} - A$.
5. Probar que $p(\emptyset) = 0$.
6. Probar que $p(A) \leq p(B)$ si $A \subset B$.
7. $p(A_1\cup\cdots\cup A_n) = p(A_1)+\cdots+p(A_n)$ para cualesquiera eventos mutuamente ajenos $A_1,\ldots,A_n$.
8. Probar que $p(A\cup B) = p(A) + p(B) - p(A\cap B)$ para cualesquiera dos eventos A y B.
9. La probabilidad de que suceda el evento $A\cap B$ es 0.3. La probabilidad de que suceda el evento A es 0.7. La probabilidad de que suceda el evento B es 0.5. Mostrar que la probabilidad de que no ocurra $A\cup B$ es 0.1.
10. Probar que $p(A\cup B\cup C) = p(A)+p(B)+p(C)-p(A\cap B)-p(B\cap C)-p(C\cap A)+p(A\cap B\cap C)$ para cualesquiera tres eventos A, B y C.
11. Se escoge un entero al azar entre el 1 y el 1000. ¿Cuál es la probabilidad de que el entero escogido sea divisible entre 3, 5 o 7?
12. Si X es la variable aleatoria que cuenta el número de águilas que resultan al lanzar la moneda $2n$ veces, mostrar que $E(X) = n$.
13. Cierto juego te cobra \$10 por jugada y consiste en lanzar 4 dados. Te dan un premio de \$1000 si los cuatro dados caen mostrando el mismo número, y de \$100 si exactamente dos dados muestran el mismo número. ¿Se espera ganar o perder dinero a la larga jugando este juego?
14. Un tablero para dardos consiste de tres círculos concétricos de radios r, $2r$ y $3r$. El costo por jugar a los dardos es de \$30 por cada dardo, y recibes premios de \$60 por atinar dentro del círculo de radio r, \$40 por atinar al anillo que se encuentra entre los círculos de radios r y $2r$, y \$20 por atinar al anillo que se encuentra entre los círculos de radios $2r$ y $3r$. Suponer que siempre que se tira un dardo se atina a alguno de los círculos. ¿Cuál es la ganancia esperada por juego? ¿Es recomendable jugar este juego de dardos?

2. Primer examen parcial

1. Se lanza un dado y se escoje una carta del mazo de 52 cartas. ¿Cuál es la probabilidad de que coincidan el número en que cae el dado y el número de la carta? (Pensar que el as corresponde al número 1.)
2. Se tiran cuatro volados.
   a) Hallar la probabilidad de obtener exactamente 2 águilas.
   b) Hallar la probabilidad de obtener al menos 2 águilas.
3. Una escuela tiene 100 estudiantes. La escuela ofrece 3 idiomas extracurriculares: inglés, francés y alemán. Hay 28 estudiantes en la clase de inglés, 26 en la de francés y 16 en la de alemán. Hay 12 estudiantes inscritos en inglés y francés simultáneamente, 4 en inglés y alemán simultáneamente, y 6 en francés y alemán simultáneamente. Hay 2 estudiantes que cursan los 2 idiomas.
   a) Si un estudiante es escogido al azar, ¿cuál es la probabilidad de que no tome ningún idioma?
   b) Si un estudiante es escogido al azar, ¿cuál es la probabilidad de que tome exactamente un idioma?
   c) Si dos estudiantes son escogidos al azar, ¿cuál es la probabilidad de que al menos 1 tome un idioma?
4. Una urna contiene 8 bolas blancas, 4 negras y 2 rojas. Se meten ambas manos a la urna para sacar simultáneamente dos bolas. Suponer que se ganan \$20 por cada bola negra que sale y se pierden \$10 por cada bola blanca que sale. Sea X la ganancia en este juego. ¿Cuáles son los valores que puede tomar X? ¿Qué probabilidad está asociada a cada uno de estos valores? ¿Cuál es la ganancia esperada por juego?
5. Una compañía de seguros saca una póliza que paga una cantidad A si un siniestro E ocurre en un año. La compañía estima que E puede ocurrir al año con probabilidad p. ¿Cuánto deberá costar la póliza a fin de que el beneficio esperado sea el 10 % de A?

3. Probabilidad condicional e independencia

Considerar un experimento con espacio muestral $\mathfrak{M}$ y con medida de probabilidad p. Sea A un evento. El número $p(A)$ nos da idea de la posibilidad de que A suceda *antes* de realizar el experimento. Supongamos que *después* de realizar el experimento se nos ha informado que ocurrió cierto evento B y no cualquier resultado en $\mathfrak{M}$. Esta información puede cambiar nuestra percepción de qué tan probable es A; en efecto, nuestro espacio muestral ya no debería ser $\mathfrak{M}$, más bien debería ser B, pues los resultados posibles del experimento ya no están codificados por $\mathfrak{M}$, sino por B. La notación para la probabilidad de que ocurra el evento A si se sabe que ocurrió el evento $B \subset \mathfrak{M}$ como resultado en el experimento es $p(A \mid B)$.

Ejemplos. Se lanza un dado dos veces.

1. Calculemos la probabilidad de que salgan dos 6 si sabemos que en uno de los lanzamientos salió 6. Sea A el evento en que salen dos 6, es decir, $A = \{(6,6)\}$. Sea B el evento en que sale un 6, es decir,
$$B = \{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6),(6,5),(6,4),(6,3),(6,2),(6,1)\}.$$
   Entonces $p(A \mid B) = 1/11$.
2. Calculemos la probabilidad de que salgan dos 6 si se sabe que en el primer lanzamiento salió 6. Sea C el evento $C = \{(6,6),(6,5),(6,4),(6,3),(6,2),(6,1)\}$ y A como antes. Entonces $p(A \mid C) = 1/6$ (esto concuerda con el hecho de que la moneda no tiene memoria, cada lanzamiento no es influenciado por los anteriores).

3. Sea D el evento "cae suma igual a 7", y sea E el evento "no cae el mismo número dos veces". Entonces D tiene seis elementos y $D \subset E$, luego $p(E \mid D) = 6/6 = 1$ y $p(D \mid E) = 6/30 = 1/5$.

En el caso en que $\mathfrak{M}$ tiene m elementos y todos sus elementos son igualmente probables (como en el ejemplo anterior del dado que se lanza dos veces), haciendo que B juegue el papel de espacio muestral en lugar de $\mathfrak{M}$ se obtiene

$$p(A \mid B) = \frac{\text{número de elementos de } A \cap B}{\text{número de elementos de } B} = \frac{\frac{\text{número de elementos de } A \cap B}{n}}{\frac{\text{número de elementos de } B}{n}} = \frac{p(A \cap B)}{p(B)}.$$

También con la frecuencia relativa y la Ley de los Grandes Números se llega a esta misma igualdad. Suponer que en n repeticiones de un experimento el evento B ocurrió r_n veces simultáneamente con el evento A y s_n veces sin el evento A. Entonces $f_n(A \cap B) = \frac{r_n}{n}$ y $f_n(B) = \frac{r_n + s_n}{n}$, son las frecuencias relativas con que ocurren $A \cap B$ y B, y la probabilidad de que suceda A cuando sucede B es

$$p(A \mid B) = \lim_{n\to\infty} \frac{r_n}{r_n + s_n} = \lim_{n\to\infty} \frac{f_n(A \cap B)}{f_n(B)} = \frac{\lim_{n\to\infty} f_n(A \cap B)}{\lim_{n\to\infty} f_n(B)} = \frac{p(A \cap B)}{p(B)}$$

por la Ley de los Grandes Números. Esto motiva el concepto de probabilidad condicional:

Definición. Para cualesquiera dos eventos A y B, con $p(B) > 0$, la *probabilidad condicional de A dado B* (es decir, de que suceda A si sabemos que sucede B) se define como $p(A \mid B) = \dfrac{p(A \cap B)}{p(B)}$.

3.1. Fórmulas útiles donde aparece la probabilidad condicional.

1. $p(A \cap B) = p(A \mid B)p(B)$ es llamada *regla de la multiplicación* de probabilidades, y se deduce inmediatamente a partir de la definición anterior.
2. Aplicando la regla de la multiplicación varias veces se obtiene

$$p(A_1 \cap A_2 \cap \ldots \cap A_n) = p(A_1) \times p(A_2 \mid A_1) \times p(A_3 \mid A_1 \cap A_2) \times \cdots \times p(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

para cualesquiera eventos $A_1, A_2, \ldots, A_n$.

3. Sea A un evento que puede suceder solamente cuando suceden los eventos $B_1, \ldots, B_n$, i.e. $A \subset B_1 \cup \cdots \cup B_n$. Suponer también que los eventos $B_1, \ldots, B_n$ son mutuamente ajenos. Entonces $p(A) = p(A \mid B_1)p(B_1) + p(A \mid B_2)p(B_2) + \cdots + p(A \mid B_n)p(B_n)$. Esta fórmula es conocida como la *fórmula de la probabilidad condicional*, y se deduce aplicando la regla de multiplicación de la probabilidades a cada sumando de $p(A) = p(A \cap B_1) + p(A \cap B_2) + \cdots + p(A \cap B_n)$.
4. La igualdad $p(A \mid B) = \dfrac{p(B \mid A)p(A)}{p(B)}$ es conocida como la *fórmula de Bayes* (1740s), y se deduce sustituyendo la definición de $p(B \mid A)$ en el lado derecho de la fórmula.

3.2. Ejemplos (los abogados deberían saber probabilidad condicional).

1. Se ha cometido un asesinato. Los sospechosos son los sujetos X y Y. Ambos huyeron al comenzar las investigaciones y no han podido ser localizados. Ambos tienen las mismas posibilidades de haber cometido el crimen. Las investigaciones posteriores arrojaron la información de que el criminal tenía sangre tipo A. Sólo el 10 % de la población tiene sangre de este tipo. Al continuar la investigación, se supo que X tiene sangre tipo A, pero no se pudo hallar el tipo de sangre de Y. A raíz de esta nueva información, ¿cuál es la probabilidad de que X sea el criminal? Un minuto de reflexión nos hace pensar que la probabilidad debe ser alta, pues es poca la población que tiene este tipo de sangre.

   Sea C el evento "X es el criminal". Sea B el evento "X tiene sangre tipo A". Entonces

$$p(C \mid B) = \frac{p(B \mid C)p(C)}{p(B)} = \frac{p(B \mid C)p(C)}{p(B \mid C)p(C) + p(B \mid C^c)p(C^c)} = \frac{1(1/2)}{1(1/2) + (1/10)(1/2)} = \frac{10}{11}.$$

Notar que $p(B) = \frac{11}{20} > \frac{1}{10}$, pues X no es cualquier individuo de la población; es uno de los dos sospechosos.

2. El siguiente es un caso famoso de la vida real, e ilustra la importancia de sustituir el espacio muestral por un evento adecuado.

   Nicole Brown fue asesinada en su casa de Los Angeles, California, el 12 de junio de 1994, junto con un amigo. El famoso ex-esposo de Nicole, O.J. Simpson, quien en ese momento era un famoso actor de TV y había sido un exitoso jugador de football americano, era el principal sospechoso de los asesinatos. El hecho de que anteriormente Simpson había violentado físicamente a su ex-esposa salió a relucir por el fiscal del estado durante el juicio. La defensa trató de minimizar este hecho argumentando que "un porcentaje infinitesimal, menor a 1 en 2500, de los hombres que violentan físicamente a sus parejas terminan asesinándolas". El estado no pudo debatir este argumento. Al final Simpson fue declarado inocente.

   La afirmación de la defensa, aunque respaldada con pruebas, fue una aplicación incorrecta de la probabilidad, un truco inteligente para engañar a la audiencia y al jurado. El prolífico estadístico I.J. Good aclaró el asunto en una carta dirigida al editor de la revista *Nature* (When battered turns murdered, Nature, 375, 15 de junio de 1995, pág. 541). Los argumentos más o menos van así. Según datos de 1992, la defensa obtendría aproximadamente 1 entre 2444 al dividir

$$\frac{1432}{3500000} = \frac{\text{número de mujeres violentadas y asesinadas por sus parejas}}{\text{número de mujeres violentadas por sus parejas}} = p(A \mid V),$$

   donde A es el evento "asesinada por su pareja" y V el evento "violentada por su pareja". En efecto, es poco probable que una mujer maltratada haya sido asesinada por su agresor.

   Pero para Good la cuestión es: si una mujer maltratada por su pareja ha sido asesinada, ¿cuál es la probabilidad de que haya sido asesinada por su pareja? Pues es importante usar el hecho crucial de que Nicole fue asesinada (ese evento ocurrió y es la información adicional que debe ser usada al calcular la probabilidad condicional). Entonces hay que calcular $p(V \cap A \mid M)$, donde M es el evento "mujer asesinada", que resulta ser $p(V \cap A \mid M) = \frac{1432}{4936}$, aproximadamente $2/7$. Es decir, la probabilidad de que una mujer asesinada haya sido asesinada por su agresor es considerable, y por lo tanto se podría alegar entonces que el hecho de que Simpson haya violentado anteriormente a Nicole no debe ser minimizado y sí es relevante, ya que 2 de cada 7 mujeres asesinadas fueron asesinadas por sus agresores.

   Jon F. Merz y Jonathan P. Caulkins también dan un argumento similar al de Good (Propensity to Abuse—Propensity to Murder?, Chance Magazine, Primavera 1995, pág, 14).

3.3. El problema de Monty Hall. En un programa popular, el conductor Monty Hall muestra al concursante tres puertas idénticas y le pide escoger una. Una de las puertas esconde un auto nuevo y, con el fin de quedarse con el auto, el concursante debe adivinar cuál es. Después de que el concursante escoge una puerta, Monty abre una de las dos puertas que el concursante no escogió, y lo hace de acuerdo a las siguientes reglas (Monty conoce de antemano cuál es la puerta que esconde al auto):

1. Si la puerta que escogió el concursante es la que esconde al auto, Monty escoge al azar cuál de las dos puertas abrir.
2. Si la puerta que escogió el concursante no esconde al auto, Monty siempre abre la otra puerta que no escode al auto.

Después de que Monty abre la puerta, él ofrece al concursante la opción de cambiar de puerta o quedarse con la puerta escogida inicialmente. ¿Debe el concursante mantener su elección original o cambiar a la otra puerta que permanece cerrada?

Después de reflexionar, puede parecerte que el concursante tiene más oportunidad de ganar el coche cambiando de puerta, ¿verdad? En efecto, después de que Monty abrió una puerta de acuerdo a sus reglas, la probabilidad de que el auto esté detrás de la puerta no escogida inicialmente sube ¡de 1/3 a 2/3!

Existe un reporte de un experimento con computadora simulando la situación. Se ejecutó 20,000 veces el experimento. La computadora mantuvo la elección inicial 9,954 veces, y cambió la elección inical las otras 10,046 veces. Observar que es casi mitad y mitad, pues la decisión de mantenerse o cambiar de puerta se hizo también al azar. Cuando se mantuvo la elección inicial, se tuvo éxito en 3,343 ocasiones (el cociente es aproximadamente 0.3353, cercano a 1/3). Por otro lado, al cambiar la elección inicial, se tuvo éxito en 6,736 ocasiones (el cociente es aproximadamente 0.6705, cercano a 2/3).

Analicemos formalmente lo que sucede. Supongamos que el concursante escoge la puerta I, y que Monty ha abierto la puerta II (esto no ocasiona pérdida de generalidad, simplemente se enumeran así las puertas). Sólo hay dos escenarios posibles:

1. El auto está detrás de la puerta I. Monty tuvo que elegir al azar entre las puertas II y III.
2. El auto está detrás de la puerta III. Monty fué forzado a abrir la puerta II por la situación.

Denotar por C_i el evento en que el auto está detrás de la puerta i, y M el evento en que Monty abre la puerta II. Suponer, como se dijo antes, que el jugador escoge la puerta I y Monty abre la puerta II. Entonces

$$p(C_1 \mid M) = \frac{p(M \mid C_1)p(C_1)}{p(M)} = \frac{p(M \mid C_1)p(C_1)}{p(M \mid C_1)p(C_1) + p(M \mid C_2)p(C_2) + p(M \mid C_3)p(C_3)}$$

$$= \frac{(1/2)(1/3)}{(1/2)(1/3) + 0(1/3) + 1(1/3)} = \frac{1}{3}.$$

Análogamente se muestra que $p(C_3 \mid M) = 2/3$.

Ejemplo. Se lanza un dado y luego se lanzan tantos volados como la cantidad de veces indicada por el número que mostró el dado. Queremos calcular la probabilidad de que no caiga ningún águila. Sea B_k el evento "dado cae en k", para $k \in \{1, \ldots, 6\}$ y A el evento "no cae ningún águila". Entonces $p(B_k) = 1/6$ y $p(A \mid B_k) = 1/2^k$ para todo $k \in \{1, \ldots, 6\}$, y además $B_1, \ldots, B_6$ son ajenos mutuamente y $\mathfrak{M} = B_1 \cup \cdots \cup B_6$. Usando la fórmula de la probabilidad condicional se obtiene que $p(A) = \frac{1}{2}(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^5})\frac{1}{6} \approx 0.1641$.

3.4. Ejemplos (los médicos deberían saber probabilidad condicional).

1. Considerar la siguiente situación. Un examen clínico barato para cierta enfermedad no es 100 % confiable: entre los pacientes que tienen la enfermedad, el examen da resultado positivo el 99 % de las veces (tal porcentaje es conocido como la *sensibilidad* de la prueba), y para personas que no tienen la enfermedad, el examen da positivo el 2 % de las veces (estos casos son llamados *falsos positivos*). En la población general, 1 de cada 1,000 tienen la enfermedad. ¿Será buena idea hacer el examen a toda la población para detectar a los enfermos? Piensa un poco tu respuesta, enfocándote en la proporción de gente sana y gente enferma.

   Después de reflexionar, puede parecer que no es buena idea hacer el examen a toda la población pues habrá muchísimos falsos positivos. En concreto, si se hiciera el examen a 100,000 personas escogidas al azar, sólo 100 tendrían la enfermedad y las otras 99,900 no. Esto significa que, en promedio, el examen daría positivo para 99 personas que sí tienen la enfermedad y para 1998 que no tienen la enfermedad (muchísimos falsos positivos).

   Calculemos ahora la probabilidad de que alguien esté enfermo si su examen dió positivo. Sea E el evento en que el paciente está enfermo, y F el evento en que su examen da positivo.

Entonces

$$p(E \mid F) = \frac{p(F \mid E)p(E)}{p(F)} = \frac{p(F \mid E)p(E)}{p(F \mid E)p(E) + p(F \mid E^c)p(E^c)}$$

$$= \frac{(99/100)(1/1000)}{(99/100)(1/1000) + (2/100)(999/1000)} = 0.0472103,$$

es decir, aproximadamente 5 % de probabilidad, que es muy baja. Esto es compatible con el párrafo anterior: de las $99 + 1998 = 2097$ personas que dieron positivo, el 4.72103 % es 99.

2. Se sabe que aproximadamente el 1 % de mujeres entre 40 y 50 años de edad desarrollan cáncer de seno. La mastografía no es 100 % confiable: el 80 % de las mujeres con cáncer de seno dan positivo en la mastografía, y el 9.69 % de las mujeres sin cáncer de seno dan positivo en la mastografía. ¿Cuál es la probabilidad de que una mujer entre 40 y 50 años tenga cáncer de seno si da positivo en la mastografía?

   Como en el ejemplo anterior, esperamos que haya muchos falsos positivos, pues la proporción de mujeres que desarrolla este cáncer es pequeña. Sin embargo, un reporte (D.M. Eddy, Probabilistic reasoning in clinical medicine: problems and opportunities, Cambridge University Press, 1982) indica que al pedir a 100 médicos con trayectoria que estimen $p(B \mid A)$ partiendo de los datos anteriores, 95 de ellos dieron respuestas entre 70 % y 80 %, lo cual está muy lejos del valor real. Para hacer el cálculo, denotamos por B el evento de tener cáncer de seno, y A el de dar positivo en la mastografía, entonces

$$p(B \mid A) = \frac{p(A \mid B)p(B)}{p(A \mid B)p(B) + p(A \mid B^c)p(B^c)} = \frac{(8/10)(1/100)}{(8/10)(1/100) + (97/1000)(99/100)} \approx 0.0769,$$

   es decir, alrededor de un 7.7 %. Otro reporte que muestra la dificultad que tiene el personal médico de la Escuela de Medicina de Harvard al momento de evaluar la posibilidad de tener cáncer después de dar positivo a la mastografía es el de W. Casscells, A. Schoenberg y T. Grayboys (Interpretation by physicians of clinical laboratory results, New England Journal of Medicine, 299, pág. 999-1000, 1978).

   Algo similar sucede con el VIH y el cáncer de próstata, como se verá en los siguientes ejemplos.

3. Aproximadamente 1 de cada 10,000 personas se enferman de VIH. El 100 % de las personas con VIH dan positivo a la prueba, y el 0.1 % de las personas sin VIH dan positivo a la prueba. Sea A el evento en que una persona da positivo a la prueba de VIH, y B el evento en que una persona tiene VIH. Entonces $p(B \mid A) \approx 0.0909$.
4. El antígeno prostático da positivo el 71 % de las veces que la persona tiene cáncer de próstata, y da negativo el 91 % de las veces que la persona no tiene cáncer de próstata (especificidad). El 0.7 % de la población masculina es diagnosticada con cáncer de próstata cada año. Sea A el evento en que un hombre da positivo con el antígeno prostático, y B el evento en que un hombre tiene cáncer de próstata. Entonces $p(B \mid A) \approx 0.053$.

3.4.1. Tarea.

1. Un náufrago se encuentra en cierta área del océano con probabilidad 0.4. Una búsqueda en dicha área detectará al náufrago con probabilidad 0.9 si efectivamente se encuentra allí. ¿Cuál es la probabilidad de que el naúfrago esté en dicha área si una búsqueda allí resultó infructuosa?
2. Tienes dos cajas idénticas delante de ti. Una de las cajas contiene 10 bolas numeradas del 1 al 10, y la otra contiene 25 bolas numeradas del 1 al 25. Después de escoger una de las cajas

al azar, tomas una bola de la caja escogida. ¿Cuál es la probabilidad de escoger una bola con el número 7?

3. Se tienen dos urnas.
   a) Suponer que una contiene 9 bolas blancas y 1 negra, y que la otra contiene 9 bolas negras y 1 blanca. Se escoge una urna al azar y se saca de ella al azar 1 bola, que resulta ser blanca. ¿Cuál es la probabilidad de que la urna escogida sea la que tiene más bolas blancas?
   b) Suponer ahora que una contiene 9 bolas blancas y 1 negra, y que la otra contiene 20 negras y 5 blancas. Se escoge una urna al azar y se saca de ella al azar 1 bola, que resulta ser blanca. ¿Cuál es la probabilidad de que la urna escogida sea la que tiene más bolas blancas?
4. Se tiene un vuelo programado de París a Sidney, con escalas en Dubai y Singapur. Se tiene documentada una maleta. En cada escala el equipaje documentado cambia de un avión a otro. En el aeropuerto de París hay probabilidad del 5 % de que el equipaje no sea colocado en el vuelo correcto. Esta probabilidad es del 3 % en Dubai y del 2 % en Singapur. ¿Cuál es la probabilidad de que el equipaje no llegue a Sidney con su dueño?

   Si el equipaje no llega a Sidney con el dueño, ¿cuál es la probabilidad de que se haya extraviado en Dubai?
5. Existen dos compañías de taxis en la ciudad, los de una compañía son amarillos y los de la otra blancos. El 85 % de los taxis de la ciudad son amarillos y el 15 % son blancos. Una noche hubo un accidente provocado por un taxi pero el taxi huyó. En la corte solo se tiene el testimonio de un testigo que dice que el taxi que provocó el accidente era blanco. Para analizar la confiabilidad de ese testimonio, se hizo una prueba en una situación similar a la de la noche del accidente con varios participantes. El resultado de la prueba fue que el 80 % de los participantes identificaron correctamente el color del taxi, y el 20 % no identificó el color correctamente. ¿Cuál es la probabilidad de que el taxi que causó el accidente fuera blanco, dado el testimonio del testigo?
6. En el reality show Big Sisters quedan 12 personas. El público escogerá cuatro personas para la gran final. Cada candidata tiene las mismas posibilidades de pasar a la gran final. Fulanita es una de las candidatas, y su familia calcula que tiene una probabilidad aproximada de 38.5 % de pasar a la gran final ya que tiene 1/12 de posibilidades de ser escogida primero, 1/11 de posibilidades de ser escogida en segundo lugar, 1/10 de ser escogida en tercero, y 1/9 de ser escogida en cuarto (y $1/12 + 1/11 + 1/10 + 1/9 \approx 0.385$). Mostrar que la familia está equivocada y calcular la probabilidad correcta.
7. El 70 % de las personas que desaparecen son localizadas posteriormente. De las que son localizadas, el 60 % cuentan con un pariente influyente, mientras que el 90 % de las que no son localizadas no tienen parientes influyentes. Suponer que una persona ha desaparecido. Calcular la probabilidad de que sea localizada en cada uno de los siguientes casos:
   a) Si tiene pariente influyente.
   b) Si no tiene pariente influyente.
8. Con los datos del primer ejemplo de la sección §3.4, usar proporciones para evaluar la probabilidad de tener la enfermedad después de dar positivo a dos exámenes clínicos.

3.5. Eventos independientes. La igualdad $p(A \mid B) = p(A)$ indica que el saber que el evento B sucede no altera la probabilidad de que el evento A suceda. Esto nos motiva a pensar que el evento A no depende del evento B. De la definición de probabilidad condicional, se tiene que $p(A \mid B) = p(A)$ es equivalente a $p(A \cap B) = p(A)p(B)$ y a su vez equivalente a $p(B \mid A) = p(B)$. Es decir, si el evento A no depende del B, estamos también obligados a decir que el evento B no depende del A.

Definición. Dos eventos A y B son *independientes* si $p(A \cap B) = p(A)p(B)$.

Ejemplo. Se lanza un dado dos veces. El evento A de que salga 6 en el primer lanzamiento es independiente del evento B de que salga 6 en el segundo, pues $p(A \cap B) = 1/36$ y, por otro lado, $p(A) = 6/36 = p(B)$.

No hay que confundir ajenos con independientes; en este ejemplo, A y B resultaron ser independientes, pero no son ajenos pues $A \cap B \neq \emptyset$.

Un ejemplo de eventos que no son independientes, pueden ser el evento A y el evento C en que en el primer lanzamiento cae un número mayor que en el segundo lanzamiento, ya que $p(C) = 15/36$ y $p(A \cap C) = 5/36$.

Ejemplo. Considerar los sexos de los hijos de una familia con tres hijos. Suponer que cada una de las posibilidades $\{fff, ffm, fmf, mff, mmf, mfm, fmm, mmm\}$ tiene probabilidad igual a $1/8$. Sea A el evento "la familia tiene hijos de ambos sexos" y B el evento "hay a lo más una hija". Entonces $p(A) = 6/8$, $p(B) = 4/8$ y $p(A \cap B) = 3/8$, por lo tanto A y B son eventos independientes en familias con tres hijos. Notar que esto no es cierto en familias con dos o cuatro hijos. Esto muestra que no siempre es tan obvio si dos eventos son o no independientes.

Ejemplo. Cada mañana, *de manera independiente*, dos estaciones meteorológicas dan sus predicciones para el clima del siguiente día. En promedio, la predicción de la estación I es correcta en el 90 % de los casos, y la predicción de la estación II es correcta en el 80 % de los casos. Cierto día, la estación I predice temperaturas altas para el siguiente día, en cambio la estación II predice temperaturas bajas. ¿Cuál es la probabilidad de que la predicción de la estación I sea correcta?

Antes de comenzar el análisis del problema, notemos que la probabilidad debería ser menor que 90 % (pues el hecho de que la otra estación da el pronóstico contrario baja la probabilidad de que la estación I esté en lo correcto) pero mayor que 50 % (pues es más certera la estación I que la estación II).

Proponemos como espacio muestral para este problema el conjunto $S = \{(c, c), (c, i), (i, c), (i, i)\}$, donde c quiere decir correcto e i incorrecto, y la primera coordenada corresponde a la estación I y la segunda a la II. Sea A el evento en que la predicción de la estación I es correcta, y B el evento en que las estaciones dan predicciones contrarias. El problema nos pide calcular

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{p(\{(c,i)\})}{p(\{(c,i),(i,c)\})} = \frac{p(\{(c,i)\})}{p(\{(c,i)\}) + p(\{(i,c)\})} = \frac{(9/10)(2/10)}{(9/10)(2/10) + (1/10)(8/10)} \approx 0.692.$$

Definición. $A_1, \ldots, A_n$ son eventos *independientes* si $p(A_{j_1} \cap \cdots \cap A_{j_k}) = p(A_{j_1}) \times \cdots \times p(A_{j_k})$ para cualquier subconjunto $\{j_1, \ldots, j_k\} \subset \{1, \ldots, n\}$.

Ejemplo. Suponer que se lanza un dado dos veces. Sea A el evento que consiste en que sale número par en el primer lanzamiento. Sea B el evento consistente en que salen dos números cuya suma es impar. Sea C el evento que consiste en que sale número par en el segundo lanzamiento. Verificar que A, B y C eventos independientes dos a dos, pero $p(A \cap B \cap C) \neq p(A)p(B)p(C)$.

3.5.1. Tarea.

1. Mostrar que si A y B son eventos independientes, entonces A^c y B también son independientes.
2. Se lanza una moneda honesta dos veces. Sea A el evento en que el primer volado cae águila y B el evento en que ambos volados caen en lo mismo. ¿Son A y B independientes?
3. Fultano, Sutano y Perengano son expertos en el juego de los dardos. La probabilidad de que Fulano atine al disco central del tablero es $1/3$, para Sutano es $1/4$ y para Perengano es $1/5$. Los tres jugadores lanzan simultáneamente un dardo. Dos de los dardos atinan al disco central y el otro no. ¿Cuál es la probabilidad de que Fulano haya sido quien no acertó?
4. El color de ojos de una persona está determinado por un solo par de genes. Si ambos genes son de ojos azules, la persona tendrá ojos azules; si ambos genes son de ojos cafés, la persona tendrá ojos cafés; y si un gen es de ojos azules y otro de ojos cafés, la persona tendrá ojos cafés (debido a esto, decimos que el gen de ojos cafés es *dominante* sobre el gen de ojos azules). El

par de genes de cada bebé consiste de un gen heredado de su madre y uno de su padre. La madre puede heredar con igual probabilidad cada gen del par que ella posee, el padre también hereda con igual probabilidad uno de sus dos genes, y lo hacen de manera independiente la madre del padre, es decir, la herencia de uno no afecta la del otro. (El descubrimiento que publicó Gregor Mendel en 1866, siete años después del *Origen de las Especies* de Darwin, fué precisamente este que ya se mencionó: cada gen de un progenitor tiene probabilidad $1/2$ de heredarse a sus descendientes y los progenitores heredan sus genes de manera independiente. Esto permite entender y cuantizar la descendencia, usando la Ley de los grandes Números. Mendel respaldó sus ideas experimentando con alrededor de 30,000 guisantes, que heredan la característica de ser rugosos o lisos.)

Suponer que Fulanito tiene ojos cafés y que su hermana tiene ojos azules.

a) ¿Cuál es la probabilidad de que Fulanito posea el gen de ojos azules?
Respuesta: Sea A el evento "Fulanito tiene gen azul" y B el evento "Fulanito tiene ojos cafés". La pregunta pide calcular $p(A \mid B)$.
Sea C el evento "Fulanito tiene gen café", entonces los eventos B y C son el mismo pues el gen de ojos cafés es dominante, luego $p(A \mid B) = p(A \mid C) = \frac{p(A\cap C)}{p(C)}$.
Además $C = \{(c,c),(c,a),(a,c)\}$ y $A \cap C = \{(c,a),(a,c)\}$, donde la primera coordenada denota el gen que Fulanito hereda de su madre y la segunda coordenada el que hereda de su padre, y c quiere decir *gen café* y a quiere decir *gen azul*.
Cualquier elemento del espacio muestral $\mathcal{M} = \{(c,c),(c,a),(a,c),(a,a)\}$ tiene probabilidad $\frac{1}{2} \times \frac{1}{2}$ de suceder (aquí es donde se usan las leyes de Mendel), y esto puede usarse para calcular $\frac{p(A\cap C)}{p(C)}$ pues las probabilidades en el numerador y denominador no son condicionales. Se concluye que $\frac{p(A\cap C)}{p(C)} = \frac{2/4}{3/4} = 2/3$.

b) Suponer que la esposa de Fulanito tiene ojos azules. ¿Cuál es la probabilidad de que su primogénit@ tenga ojos azules?

c) Si el primer hijo de Fulanito resultó con ojos cafés, ¿cuál es la probabilidad de que su siguiente hij@ también tenga ojos cafés?
Respuesta: El color de ojos de l@s hij@s direct@s de Fulanito están determinados por el gen que heredan de Fulanito, pues su esposa siempre hereda un gen a. Entonces solamente vamos a analizar los genes que hereda Fulanito a sus dos hij@s.
Sea X el evento "Fulanito hereda c a primer hijo" y Y el evento "Fulanito hereda c a segund@ hij@". La pregunta pide calcular $p(Y \mid X) = \frac{p(Y\cap X)}{p(X)}$.
Además $X = \{[c;c],[c;a]\}$ y $Y \cap X = \{[c;c]\}$, donde la primera coordenada entre los corchetes denota el gen que Fulanito hereda a su primer hijo y la segunda coordenada el que hereda a su segund@ hij@.
Cualquier elemento del espacio muestral $\{[c;c],[c;a],[a;c],[a;a]\}$ tiene igual probabilidad de suceder si Fulanito tiene el gen a (ya sabemos que tiene el gen c, pero no sabemos que tenga el gen a), es decir, $p(\{[c;c]\} \mid D) = p(\{[c;a]\} \mid D) = 1/4$ donde D es el evento "Fulanito tiene el gen a" (que ya se trabajó en el primer inciso, sabiendo que su hermana tiene ojos azules, lo cual aplica también es este inciso). Por la fórmula de la probabilidad condicional, usando la respuesta $p(D) = 2/3$ del primer inciso, se tiene

$$p(\{[c;c]\}) = p(\{[c;c]\} \mid D)p(D) + p(\{[c;c]\} \mid D^c)p(D^c) = (1/4)(2/3) + (1)(1/3) = 1/2,$$
$$p(\{[c;a]\}) = p(\{[c;a]\} \mid D)p(D) + p(\{[c;a]\} \mid D^c)p(D^c) = (1/4)(2/3) + (0)(1/3) = 1/6,$$

de donde se concluye que $p(Y \mid X) = \frac{p(Y\cap X)}{p(X)} = \frac{p(\{[c;c]\})}{p(\{[c;c]\})+p(\{[c;a]\})} = \frac{1/2}{1/2+1/6} = 3/4$.

3.6. Segundo examen parcial.

1. El 52 % de los estudiantes del tecnológico son mujeres. El 5 % de los estudiantes del tecnológico estudian sistemas computacionales. El 2 % de los estudiantes del tecnológico son mujeres que estudian sistemas computacionales. Si un estudiante del tecnológico es escogido al azar, hallar la probabilidad de que:
   a) sea mujer, dado que estudia sistemas computacionales;
   b) estudie sistemas computacionales, dado que es mujer.
2. Le pides a tu vecino regar con agua tu planta mientras te vas de vacaciones. Si no la riegan, la probabilidad de que la planta muera es 0.8; y si la riegan, la probabilidad de que muera es 0.15. La probabilidad de que el vecino riegue la planta es 0.9.
   a) ¿Cuál es la probabilidad de que la planta esté viva cuando regreses del viaje?
   b) Si al regresar encuentras la planta muerta, ¿cuál es la probabilidad de que el vecino no la regó?
3. Sean E, F y G tres eventos que cumplen las tres condiciones siguientes:
   - E y F son independientes,
   - F y G son independientes,
   - E y $F \cap G$ son independientes.

   Verificar que G y $E \cap F$ son independientes.
4. Un modelo simplificado para el comportamiento del precio de una acción supone que cada día el precio de la acción sube 1 unidad con probabilidad p o baja 1 unidad con probabilidad $1 - p$; y que el cambio en cada día es independiente de lo que sucede en los otros días.
   a) ¿Cuál es la probabilidad de que después de 2 días la acción regrese a su precio original?
   b) ¿Cuál es la probabilidad de que después de 3 días el precio de la acción se haya incrementado en 1 unidad?
   c) Dado que después de 3 días el precio de la acción se incrementó en 1 unidad, ¿cuál es la probabilidad de que haya bajado el primer día?

4. Media y varianza

Recordar que el *valor esperado* (o *esperanza*, o *media*, o *primer momento*) de una variable aleatoria que toma un número finito de valores $\{x_1, x_2, \dots, x_M\}$ se define por

$$E(X) \;=\; x_1 p(X = x_1) \;+\; x_2 p(X = x_2) \;+\cdots+\; x_M p(X = x_M).$$

Ejemplo. Ya se vió al comenzar la Sección 1.3, que en el experimento de lanzar un dado dos veces el valor esperado de la suma es 7. Por otro lado, el valor esperado del número que aparece en el primer lanzamiento es 7/2.

4.1. Propiedades del valor esperado. Sean X y Y dos variables aleatorias definidas en el mismo espacio muestral con medida de probabilidad p. Suponer que ambas variables aleatorias toman un número finito de valores. Entonces se cumplen las identidades:

$$\begin{aligned} &E(X+Y) = E(X) + E(Y) \\ &E(aX) = aE(X) \quad \text{para cualquier } a \in \mathbb{R} \quad \text{y} \\ &E(X) = x_0 \quad \text{si } X \text{ es constante igual a } x_0. \end{aligned} \tag{3}$$

Demostración. Sea $Z = X + Y$. Sean $\{x_1, \dots, x_M\}$ y $\{y_1, \dots, y_N\}$ los valores que toman las variables aleatorias X y Y, respectivamente. Notar que

$$p(X = x_i) \;=\; \sum_{j=1}^{N} p(X = x_i, Y = y_j) \qquad \text{y} \qquad p(Y = y_j) \;=\; \sum_{i=1}^{M} p(X = x_i, Y = y_j), \tag{4}$$

donde $p(X = x_i, Y = y_j)$ denota la probabilidad de la intersección del evento $X = x_i$ con el evento $Y = y_j$. Puesto que Z toma los valores $z_{ij} = x_i + y_j$,

$$\begin{aligned}E(Z) = \sum_{i,j} z_{ij}p(Z = z_{ij}) &= \sum_{i,j}(x_i + y_j)p(X = x_i, Y = y_j) = \sum_{i,j} [x_ip(X = x_i, Y = y_j) + y_jp(X = x_i, Y = y_j)] \\ = \sum_{i,j} x_ip(X = x_i, Y = y_j) + \sum_{i,j} y_jp(X = x_i, Y = y_j) &= \sum_i x_i \sum_j p(X = x_i, Y = y_j) + \sum_j y_j \sum_i p(X = x_i, Y = y_j) \\ &= \sum_i x_ip(X = x_i) + \sum_j y_jp(Y = y_j) = E(X) + E(Y),\end{aligned}$$

donde la penúltima igualdad se debe a (4). Esto prueba la primer identidad en (3).

Si $a = 0$, la segunda identidad de (3) es inmediata, de otra forma,

$$E(aX) = \sum_i (ax_i)p(aX = ax_i) = \sum_i (ax_i)p(X = x_i) = a\sum_i x_ip(X = x_i) = aE(X).$$

Finalmente, si X toma el valor constante x_0 entonces $E(X) = x_0p(X = x_0) = x_0$. □

Por inducción, la primera igualdad de (3) se puede generalizar a sumas finitas de variables aleatorias finitas: $E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$.

Dos variables aleatorias X y Y que toman un número finito de valores son *independientes* si para cualesquiera valores x y y de X y Y, respectivamente, se tiene $p(X = x, Y = y) = p(X = x)p(Y = y)$.

Observación 1. *$E(XY) = E(X)E(Y)$ si X y Y son variables aleatorias independientes que toman un número finito de valores.*

Demostración. Sean $\{x_1, \ldots, x_M\}$ y $\{y_1, \ldots, y_N\}$ los valores que toman las variables X y Y, respectivamente. Sea Z la variable aleatoria dada por $Z = XY$, que toma los valores $z_{ij} = x_iy_j$. Entonces

$$\begin{aligned}E(Z) = \sum_{i,j} z_{ij}p(Z = z_{ij}) &= \sum_{i,j} x_iy_jp(X = x_i, Y = y_j) = \sum_{i,j} x_iy_jp(X = x_i)p(Y = y_j) \\ &= \sum_i x_ip(X = x_i) \sum_j y_jp(Y = y_j) = E(X)E(Y).\end{aligned}$$

□

Sea X una variable aleatoria finita con valores $\{x_1, \ldots, x_M\}$. Cualquier función de X es también una variable aleatoria finita (por ejemplo, X^2 o $\sin X$).

Regla de Substitución. *Para cualquier función $g(x)$ con valores reales, el valor esperado de la variable aleatoria finita $g(X)$ puede calcularse mediante*

$$E[g(X)] = \sum_{k=1}^{M} g(x_k)p(X = x_k).$$

4.2. Varianza y desviación estándar. Sea X una variable aleatoria que toma un número finito de valores, cuya media es $\mu = E(X)$. Nos gustaría considerar el valor esperado de $|X - \mu|$ como una medida de la manera en que los valores de X se esparcen alrededor de μ. Resulta que es más sencillo trabajar con $E(|X - \mu|^2)$ que con $E(|X - \mu|)$, así definimos la *varianza* de X, denotada por $\sigma^2(X)$ o por $\mathrm{var}(X)$, mediante $\sigma^2(X) = E(|X - \mu|^2)$. La raíz cuadrada $\sigma(X) = \sqrt{E[(X - \mu)^2]}$ es llamada *desviación estándar* de X.

Una consecuencia de las identidades (3) es que $\sigma^2(X) = E(X^2) - \mu^2$.

A diferencia del valor esperado, no es cierto en general que la varianza de una suma es igual a la suma de las varianzas. Para dar una fórmula para la varianza de una suma de dos variables aleatorias se necesita otro concepto: la *covarianza* de dos variables aleatorias X y Y, denotada por $\mathrm{cov}(X; Y)$, y definida mediante $\mathrm{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$.

Propiedad 1. *Para cualesquiera dos variables aleatorias X y Y que toman un número finito de valores se tiene* $\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y) + 2\text{cov}(X,Y)$.

Demostración. La prueba es inmediata usando las identidades (3):

$$E[(X+Y-E(X+Y))^2] = E[(X-E(X))^2] + E[(Y-E(Y))^2] + E[2(X-E(X))(Y-E(Y))].$$

□

La fórmula que generaliza la Propiedad 1 es $\sigma^2(X_1+\cdots+X_n) = \sum_{i=1}^{n} \sigma^2(X_i) + 2 \sum_{1\le i<j\le n} \text{cov}(X_i, X_j)$.

Propiedad 2. *Si X y Y son dos variables aleatorias independientes que toman un número finito de valores, se tiene* $\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y)$.

Demostración. Puesto que $(X-E(X))(Y-E(Y)) = (XY) - XE(Y) - YE(X) + E(X)E(Y)$, otra expresión para la covarianza es $E[(X-E(X))(Y-E(Y))] = E(XY) - E(Y)E(X)$ por las identidades (3). Luego, basta aplicar la Observación 1 de la Sección 4.1 para concluir que $\text{cov}(X,Y) = 0$ cuando X y Y son independientes. □

4.3. Desigualdad de Chebychev. *Sea X una variable aleatoria que toma un número finito de valores. Entonces*

$$p(|X-\mu| \ge a) \le \frac{\sigma^2(X)}{a^2}.$$

Esta relación es sobresaliente pues es fácil de probar y da una manera de estimar en el espacio muestral $\mathfrak{M}$ cómo se esparcen los valores de X alrededor de μ, a partir de la varianza.

Demostración. Sean $\{x_1, \ldots, x_M\}$ los valores de la variable aleatoria finita X. Suponer, después de renumerar si es necesario, que los primeros $N \le M$ valores $\{x_1, \ldots, x_N\}$ cumplen la desigualdad $|x_j - \mu| \ge a$. Entonces $p(|X-\mu| \ge a) = \sum_{j=1}^{N} p(X = x_j)$. Puesto que $|x_j - \mu| \ge a$ para $j = 1, \ldots, N$, se tiene $1 \le |x_j - \mu|^2/a^2$ para tales j, de donde

$$\sum_{j=1}^{N} p(X = x_j) \le \sum_{j=1}^{N} \frac{|x_j-\mu|^2}{a^2} p(X = x_j) = \frac{1}{a^2} \sum_{j=1}^{N} |x_j - \mu|^2 p(X = x_j) \le \frac{1}{a^2} E(|X-\mu|^2).$$

□

4.3.1. Tarea.

1. Sea X la variable aleatoria que indica los puntos que muestra un dado al caer después de ser lanzado. Sea Y la variable aleatoria que indica la suma de los puntos obtenidos al lanzar dos veces el dado. Suponiendo que el dado es honesto, calcular $E(X)$, $\sigma^2(X)$, $E(Y)$ y $\sigma^2(Y)$. Suponiendo que $P(X=1) = P(X=6) = 1/2$ y $P(X=2) = P(X=3) = P(X=4) = P(X=5) = 0$, hallar $E(X)$, $\sigma^2(X)$, $E(Y)$ y $\sigma^2(Y)$.
2. Sea X una variable aleatoria que toma los valores $\{1,2,3\}$. Hallar $P(X=1)$, $P(X=2)$ y $P(X=3)$ si $E(X) = 23/10$ y $E(X^2) = 59/10$.

5. Distribución binomial

Considerar un experimento en que sólo nos interesa si ocurre o no cierto evento, es decir, solamente dos resultados son posibles, y los llamaremos “acierto” y “error”. Este experimento tan básico es llamado *experimento de Bernoulli*. Vamos a suponer que la probabilidad de obtener un acierto es p y, por lo tanto, un error se obtiene con probabilidad $1-p$.

Suponer ahora que un experimento de Bernoulli se repite n veces bajo las mismas condiciones. Es natural suponer que cada repetición es *independiente* de las demás. Sea X la variable aleatoria que indica el número de aciertos en esas n repeticiones. Entonces se tiene que

$$p(X=k) \;=\; \binom{n}{k} p^k(1-p)^{n-k}, \qquad k=0,1,\ldots,n, \tag{5}$$

donde p es la probabilidad de acierto en cada repetición, como se mencionó anteriormente.

Cuando se tiene una medida de probabilidad y una variable aleatoria X que solamente toma valores $\{0,1,\ldots,n\}$ de manera que se cumple (5), decimos que la variable X representa una *distribución binomial*. (El término *distribución* se debe a que la fórmula (5) indica cómo se distribuyen las probabilidades en función de los valores de la variable aleatoria.)

A partir de (5) notamos que

$$\frac{p(X=k)}{p(X=k-1)} = \frac{(n-k+1)p}{k(1-p)} = 1 + \frac{(n+1)p-k}{k(1-p)}.$$

Entonces $p(X=k) > p(X=k-1) \Leftrightarrow k < (n+1)p$ y $p(X=k) < p(X=k-1) \Leftrightarrow k > (n+1)p$. Sea m aquel entero que cumple $(n+1)p - 1 < m \leq (n+1)p$, entonces se concluye lo siguiente.

Observación. *Cuando k recorre valores enteros entre 0 y n, la probabilidad $p(X=k)$ crece estrictamente y luego decrece estrictamente, alcanzando su valor máximo cuando $k=m$. Este valor máximo es único, excepto cuando $m=(n+1)p$, en cuyo caso $p(X=m-1) \;=\; p(X=m)$.*

El valor $p(X=m)$ es conocido como el término *central*, y m se puede pensar como la "cantidad de aciertos más probable".

5.1. Ejemplos.

1. El número de águilas obtenidos al lanzar una moneda n veces es un ejemplo de una distribución binomial, con $p=1/2$. Por ejemplo, en $n=100$ volados, el número de águilas más probable es 50, aunque $p(X=50) < 8/100$ (este número se puede calcular con WolframAlpha, por ejemplo, mediante la instrucción

   ```
   P(X=50) binomial distribution with n=100, p=0.5
   ```

   que arroja 0.0795892 como resultado aproximado y muestra una gráfica de la distribución binomial.
2. Diariamente hay un vuelo de Guadalajara a Monterrey a las 8:30 am que es extremadamente solicitado. La aeronave que realiza este vuelo tiene capacidad para 150 pasajeros. La aerolínea tiene la política de vender 160 boletos pues algunos pasajeros no se presentan. Por experiencia, se sabe que la probabilidad de que un pasajero no se presente es igual a $1/10$. Suponer que el que un pasajero no se presente es independientemente de los demás. ¿Cuál es la probabilidad de tener que dejar fuera a algún(os) pasajero(s) debido a que el vuelo ya va lleno?

   Este problema puede plantearse como un experimento de Bernoulli que se repite $n=160$ veces (una vez por cada boleto vendido) de manera independiente, con probabilidad de acierto $p=9/10$ en cada ocasión. De esta manera, la variable aleatoria X que cuenta el número total de aciertos corresponde al total de pasajeros que se presentan a tomar su vuelo, y resulta ser una distribución binomial. La probabilidad que se busca es

   $$p(X>150) = \; p(X=151) + \cdots + p(X=160) \;=\; \sum_{k=1}^{10} \binom{160}{150+k} \frac{9^{150+k}}{10^{160}}.$$

   Este cálculo se puede realizar con WolframAlpha, bien sea pidiéndole que calcule la última sumatoria, tecleando por ejemplo

   ```
   sum_{k=1}^10 binom(160,150+k)9^{150+k}/10^160
   ```

o también se le puede dar la instrucción

```
P(X>150) binomial distribution with n=160, p=0.9
```

y de cualquiera de las dos maneras WolframAlpha arroja el resultado $P(X > 150) \approx 0.0359$.

3. Dado un grupo de m personas escogidas al azar, recordar que la probabilidad de que ninguna pareja de personas escogidas en ese grupo celebre su cumpleaños el mismo día es igual a
$$\frac{365 \times 364 \times \cdots \times (365 - m + 1)}{365^m}.$$
A partir de esto se puede obtener inmediatamente la probabilidad de que al menos dos personas en ese grupo celebren su cumpleaños el mismo día.

   Este problema también se puede aproximar como un experimento de Bernoulli que se repite $n = \binom{m}{2}$ veces de manera independiente con probabilidad de acierto $p = 1/365$. Esto solamente brinda una aproximación, pues en realidad las repeticiones del experimento no son independientes. La fórmula obtenida de esta manera es
$$P(X \geq 1) = 1 - P(X = 0) \; = 1 - \left(\frac{364}{365}\right)^{\frac{1}{2}m(m-1)}.$$
Por ejemplo, cuando $m = 23$ se tiene:
$$1 - \frac{365 \times 364 \times \cdots \times (365 - m + 1)}{365^m} \approx 0.5073,$$
$$1 - \left(\tfrac{364}{365}\right)^{\frac{1}{2}m(m-1)} \approx 0.500477.$$
   Una pregunta que resulta mucho más difícil de calcular con exactitud es la siguiente: ¿Cuál es la probabilidad de que al menos dos personas en un grupo de m personas escogidas al azar celebren su cumpleaños con a lo más un día de diferencia?[1] Sin embargo, es posible aproximar el resultado mediante una distribución binomial considerando $n = \binom{m}{2}$ y $p = 3/365$. Por ejemplo, para $m = 14$ la probabilidad exacta es alrededor de 0.537493, mientras que la aproximación binomial es $1 - (1 - p)^n \approx 0.5281225$.
4. Sea X una variable aleatoria que es una distribución binomial con valores $0, 1, \ldots, n$ y con probabilidad de acierto p. Entonces $E(X) = np$ y $\sigma^2(X) = np(1 - p)$.

5.1.1. Tarea.

1. El tesorero de un reino recibe de parte del rey la orden de llenar 100 sacos con 100 monedas de oro cada uno. Mientras realiza el encargo, el tesorero sustituye una moneda de oro por una moneda de plomo en cada saco. El rey sospecha del engaño de su súbdito y le aconsejan dos métodos para verificar los contenidos de los sacos. El primer método consiste en sacar al azar una moneda de cada uno de los 100 sacos. El segundo método consiste en escoger 25 sacos y sacar al azar 4 monedas de cada uno de ellos. ¿Qué método da la mayor probabilidad de descubrir al impostor?
2. Sea X una variable aleatoria que es una distribución binomial con $p = 1/2$. Usar el teorema de Chebychev para hallar una cota para $p(X \leq 10)$ si $n = 36$. Comparar con el valor exacto de $p(X \leq 10)$.

[1]J.I. Nauss (An extension of the birthday problem, *The American Statistician*, 1968) encontró la respuesta para la situación más general en la que el grupo consta de m personas y se pregunta por la probabilidad de que dos o más de ellos celebren su cumpleaños con a lo más d días de diferencia, siendo esta igual a $1 - \frac{(364-md)!}{365^{m-1}(365-m(d+1))!}$.

3. En una campaña publicitaria durante el próximo campeonato europeo de fútbol, coca cola ha impreso la imagen de un balón en la parte inferior de la tapa de una de cada mil latas de coca cola. Cualquiera que entregue una de esas tapas antes de una fecha determinada recibirá una entrada gratuita para la final del torneo de fútbol. Se han comprado 1500 latas de coca cola para una fiesta escolar y se consumirán en dicha fiesta. ¿Cuál es la probabilidad de que coca cola entregue una o mś entradas gratuitas entre los asistentes a la fiesta escolar?
4. El daltonismo aparece en el 1 % de las personas de cierta población. ¿Qué tan grande debe ser una muestra de personas elegidas al azar en esa población para que la probabilidad de que al menos una persona daltónica esté en dicha muestra sea mayor que 0.95?
5. Considerar un grupo de 25 personas seleccionadas al azar. ¿Cuál es aproximadamente la probabilidad de encontrar en este grupo tres o más personas que celebren su cumpleaños el mismo día? ¿Y cuál es aproximadamente la probabilidad de encontrar tres o más personas que celebren su cumpleaños con a lo más un día de diferencia?

5.2. Como la estadística ayudó en un descubrimiento merecedor de un premio Nobel. Las bacterias han sido estudiadas intensamente desde que han sido reconocidas como causas de enfermedad. La reprodución bacteriana parece ser completamente asexual y parece que las bacterias transmiten sus características a su descendencia. Por ejemplo, los experimentos muestran que si una población de bacterias se somete a una toxina particular, casi todas las bacterias mueren; sin embargo, unas pocas sobreviven y formarán colonias que son resistentes permanente y específicamente a esa toxina en particular. Debido a esto, muchos investigadores argumentaron (ofreciendo incluso modelos matemáticos) que el contacto con esa toxina en particular inducía a algunas bacterias a adquirir resistencia y heredarla a su descendencia.

Por otro lado, había investigadores que afirmaban que las nuevas características adquiridas por las bacterias pueden explicarse mediante mutaciones genéticas accidentales y las bacterias que las sufren se convierten en dominantes en ese entorno. En particular, una mutación puede ser reversible y las nuevas características adquiridas por las bacterias pueden perderse. Estos investigadores argumentaban que las aparentes características adquiridas por las bacterias bajo el estrés provocado por el entorno son producto de métodos experimentales inadecuados y modelos que fallan en reconocer el mecanismo de mutación en poblaciones tan vastas con individuos que se reproducen asexualmente con rapidez. En 1943 Salvador Luria y Max Delbrück mostraron que la hipótesis de la selección natural mediante mutación accidental era la correcta, haciendo un uso ingenioso y crucial de la media y la varianza. Por este trabajo se les otorgó el Premio Nobel de Fisiología y Medicina en 1969.

En el experimento se usó cierto tipo de bacteria que es atacada por cierto virus, de manera que al reproducirse el virus en ella ocasiona su muerte. Si la población de bacterias es pequeña, digamos con alrededor de 10^5 individuos, la población muere por completo, pero en una problación de alrededor de 10^9 individuos hay buenos chances de que sobrevivan bacterias y formen colonias. Se quieren comparar entonces dos teorías: la teoría de la mutación (que afirma que la resistencia de la bacteria se debe a una mutación accidental y que, por lo tanto, la bacteria era resistente desde antes del ataque con el virus) y la teoría de la adaptación (que afirma que la resistencia de la bacteria se crea a partir del ataque con el virus). No se puede saber desde antes del ataque si una bacteria es resistente al virus, solamente es posible identificar la resistencia mediante el ataque del virus.

En el experimento de Luria y Delbrück, una cantidad grande K de bacterias iniciaron cada una su colonia. Suponer que la unidad de tiempo es igual al promedio que dura una bacteria en reproducirse por división celular, de manera que después de un tiempo T cada colonia tendría aproximadamente 2^T bacterias. Después de un tiempo en que se dan las condiciones idóneas para el crecimiento de las colonias hasta alcanzar una población con N individuos cada una, las colonias son expuestas al virus, y poco después se cuentan las bacterias sobrevivientes de cada colonia, dando una lista $(x_1, x_2, \ldots, x_K)$.

Si la teoría de la adaptación es correcta, cada bacteria tiene una probabilidad pequeña p de desarrollar inmunidad a la infección. La probabilidad de que la k-ésima colonia tenga $X_k = n$ sobrevivientes está dada por una distribución binomial $p(X_k = n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$ con $E(X_k) = Np$ y $\sigma^2(X_k) = Np(p-1)$. Luego, si $X = \sum_{k=1}^K X_k$ se tiene

$$\frac{\sigma^2(X)}{E(X)} = 1 - p.$$

Por lo tanto, aunque no se conozca el valor exacto de p, podemos afirmar que si la hipótesis de la adaptación es correcta entonces $\sigma^2_{\mathrm{exp}}/\mu_{\mathrm{exp}}$ debería ser independiente de N y cercano a 1.

Si la teoría de la mutación es correcta, hay una probabilidad pequeña q de encontrar individuos que hayan mutado antes de la exposición al virus para ser inmunes. Para simplificar vamos a concentrarnos en los K fundadores de colonias. La idea intuitiva es que cuando un fundador de una colonia es inmune, prácticamente toda su colonia sobrevivirá, y cuando un fundador no es inmute, prácticamente toda su colonia desaparecerá, y esta diferencia debe notarse en que hay una mayor varianza en los datos. La hipótesis de fijarse básicamente en el fundador no está tan alejada de la realidad, pues así como pueden aparecer eventualmente bacterias inmunes en una colonia con probabilidad pequeña, puede suceder también que la inmunidad se pierda eventualmente en ciertos individuos de una colonia con probabilidad pequeña, compensándose. La distribución binomial $p(Y = k) = \frac{K!}{k!(K-k)!} q^k (1-q)^{K-k}$ nos da la probabilidad de tener k fundadores de colonias inmunes. Entonces el valor esperado de la población total de bacterias es $E(X) = E(NY) = NE(Y) = NKq$ y la varianza es $\sigma^2(X) = \sigma^2(NY) = N^2\sigma(Y) = N^2Kq(1-q)$, por lo tanto

$$\frac{\sigma^2(X)}{E(X)} = N(1 - q).$$

En conclusión, aunque no se conozca el valor exacto de q, podemos afirmar que si la hipótesis de la mutación es correcta entonces $\sigma^2_{\mathrm{exp}}/\mu_{\mathrm{exp}}$ debería ser aproximadamente N.

Lo que resta es comparar con el resultado del experimento, codificado por el listado $(x_1, x_2, \ldots, x_K)$ de las cantidades de bacterias sobrevivientes en cada colonia. En nuestra descripción simplificada del experimento será suficiente considerar la media experimental y la varianza experimental de la población sobreviviente

$$\mu_{\mathrm{exp}} = \frac{1}{K}\sum_{k=1}^K x_k, \qquad \sigma^2_{\mathrm{exp}} = \frac{1}{K}\sum_{k=1}^K (x_k - \mu_{\mathrm{exp}})^2.$$

Los datos estadísticos que se obtuvieron a partir del experimento favorecieron esta hipótesis de manera concluyente. En conclusión, la habilidad aparente de las bacterias para responder rápidamente y adaptarse a los cambios en su medio no es sino consecuencia de mutaciones al azar en poblaciones enormes que se reproducen rápidamente.

6. Tercer examen parcial

1. Un examen consiste de 5 preguntas de opción múltiple. Cada pregunta tiene 3 opciones de respuesta, de las cuales solamente una es correcta. Un estudiante responde al azar cada pregunta del examen. ¿Cuál es la probabilidad de contestar correctamente 4 o más preguntas?
2. Sea X una variable aleatoria que toma un número finito de valores. Sean μ y σ^2 la media y la varianza de X, respectivamente. Hallar la media y la varianza de $Y = (X - \mu)/\sigma$.

3. La probabilidad de que en un parto nazca una niña o un niño no es exactamente igual a $1/2$. Típicamente, en cada 205 partos nacen 100 niñas y 105 niños. Determinar el número esperado de niñas en una familia con 6 hij@s.
4. Sean X y Y variables aleatorias que toman un número finito de valores. Si $E(X) = 75 = E(Y)$, $\sigma^2(X) = 10$, $\sigma^2(Y) = 12$, y $\mathrm{cov}(X, Y) = -3$, dar una cota superior para $p(|X - Y| \geq 15)$ aplicando el teorema de Chebychev a la variable aleatoria $X - Y$.
5. Sea X la cantidad de 1's y Y la cantidad de 2's que caen al lanzar n veces un dado honesto. Verificar que $\mathrm{cov}(X, Y) = -n/36$.
   Sugerencia: definir para cada $i = 1, \dots, n$, las variables
$$X_i = \begin{cases} 1 & \text{si en el } i\text{-ésimo lanzamiento cae 1,} \\ 0 & \text{si en el } i\text{-ésimo lanzamiento no cae 1.} \end{cases}$$
$$Y_i = \begin{cases} 1 & \text{si en el } i\text{-ésimo lanzamiento cae 2,} \\ 0 & \text{si en el } i\text{-ésimo lanzamiento no cae 2.} \end{cases}$$
6. N personas llegan por separado a la cena de una compañía. Al llegar, cada persona da un vistazo para localizar a sus amigos presentes; luego esa persona se sienta en la mesa de un amigo presente o se sienta en una mesa desocupada si no hay amigos presentes. Suponiendo que cada uno de los $N(N-1)/2$ pares de personas son amigos con probabilidad p, y que cada par de personas son amigos o no independientemente de los demás pares, mostrar que la cantidad esperada de mesas ocupadas es $\frac{1-(1-p)^N}{p}$.
   Sugerencia: definir para cada $i = 1, \dots, N$,
$$X_i = \begin{cases} 1 & \text{si la } i\text{-ésima persona en llegar se sienta en una mesa desocupada,} \\ 0 & \text{si la } i\text{-ésima persona en llegar se sienta en una mesa ocupada.} \end{cases}$$

7. Variables aleatoria continuas

En la Sección 4 trabajamos los conceptos de media y varianza para variables aleatorias $X : \mathcal{M} \to \mathbb{R}$ que tomaban un número finito de valores, pero también son interesantes otras variables aleatorias, como veremos en esta sección.

Más precisamente, vamos a considerar una variable aleatoria $x : \mathcal{M} \to [a, b]$ con una distribución de probabilidad de la forma
$$p(c \leq x \leq d) = \int_c^d f(x)dx \quad \text{para todo } a \leq c < d \leq b,$$
donde $f : [a, b] \to \mathbb{R}$ es una función continua. Tal función f será llamada *densidad de probabilidad* de la variable aleatoria x.

Ejemplo 1. Se escoge al azar un número x en el intervalo $[a, b]$. De acuerdo con la fórmula de Cardano, podemos admitir que la probabilidad de escoger un número en un subintervalo $[c, d] \subset [a, b]$ está dada por $p(c \leq x \leq d) = (d - c)/(b - a)$. Entonces la densidad de probabilidad debe ser una función f que cumple $\int_c^d f(x)dx = (d - c)/(b - a)$, para todo $a \leq c < d \leq b$. Luego $f : [a, b] \to \mathbb{R}$ es la función constante $f(x) = 1/(b - a)$.

Ejemplo 2. Se escoge un punto al azar en el interior de un disco de radio r. Sea ρ la variable aleatoria que denota la distancia de este punto al centro del disco. De acuerdo con la fórmula de Cardano, admitiremos que $p(c \leq \rho \leq d) = (d^2 - c^2)/r^2$ para todo $0 \leq c < d \leq r$. Por lo tanto, la densidad de probabilidad debe cumplir $\int_c^d f(\rho)d\rho = (d^2 - c^2)/r^2$, luego $f(\rho) = \dfrac{d}{d\rho}\dfrac{\rho^2}{r^2} = \dfrac{2\rho}{r^2}$ por el Teorema Fundamental del Cálculo.

Ejemplo 3. Una vara de longitud uno es quebrada al azar en dos partes. La variable aleatoria q denota la razón de la longitud de la pieza más corta entre la longitud de la mayor. En este caso se cumple $p(c \leq q \leq d) = 2d/(1+d) - 2c/(1+c)$ para todo $0 < c < d < 1$. El Teorema Fundamental del Cálculo nos dice que $f(q) = \frac{d}{dq}\frac{2q}{1+q} = \frac{2}{(1+q)^2}$.

7.1. Tarea.

1. Se escoge al azar un punto dentro de una esfera de radio r. ¿Cuál es la densidad de la variable aleatoria dada por la distancia del punto escogido al centro de la esfera? (Puede ser útil recordar que el volumen de una esfera de radio r es $4\pi r^3/3$.)
2. Se escoge un punto (x, y) al azar dentro del círculo unitario. La variable aleatoria v denota el valor absoluto de la coordenada x del punto escogido. ¿Cuál es la densidad de la variable v?
3. Se escoge un punto al azar dentro de un triángulo cuyas longitudes de altura y base son h y b, respectivamente. ¿Cuál es la densidad de la variable que indica la distancia del punto escogido a la base del triángulo?
4. Considerar una variable aleatoria x con valores en el intervalo $[-\pi/2, \pi/2]$ y con densidad de probabilidad $f(x) = \frac{1}{2}\cos x$. Verificar las hipótesis de una medida de probabilidad para concluir que efectivamente $f(x)$ se puede considerar como una densidad de probabilidad. En particular, si x toma valores en el intervalo $[-\pi/2, \pi/2]$, la función $f(x) = \alpha \cos x$ solamente puede ser densidad de probabilidad cuando $\alpha = 1/2$.

Definición. Sea $x : \mathcal{M} \to [a, b]$ una variable aleatoria con densidad de probabilidad $f : [a, b] \to \mathbb{R}$. El *valor esperado de* x se define mediante $E(x) = \int_a^b xf(x)dx$.

Ejemplo 1′. Se escoge al azar un número x en el intervalo $[a, b]$. Entonces $E(x) = (a + b)/2$.

Ejemplo 2′. Se escoge un punto al azar en el interior de un disco de radio r. Sea ρ la variable aleatoria que denota la distancia de este punto al centro del disco. Entonces $E(\rho) = 2r/3$.

Ejemplo 3′. Una vara de longitud uno es quebrada al azar en dos partes. La variable aleatoria q denota la razón de la longitud de la pieza más corta entre la longitud de la mayor. Entonces

$$E(q) = \int_0^1 \frac{2q\,dq}{(1+q)^2} = \int_0^1 \left(\frac{2}{1+q} - \frac{2}{(1+q)^2}\right) dq = \left[2\ln(1+q) + \frac{2}{1+q}\right]_0^1 = 2\ln 2 - 1 \approx 0.3863.$$

7.2. Tarea.

1. Se escoge al azar un punto dentro de una esfera de radio r. ¿Cuál es el valor esperado de la variable aleatoria dada por la distancia del punto escogido al centro de la esfera?
2. Se escoge un punto (x, y) al azar dentro del círculo unitario. La variable aleatoria v denota el valor absoluto de la coordenada x del punto escogido. ¿A qué es igual $E(v)$?
3. Se escoge un punto al azar dentro de un triángulo cuyas longitudes de altura y base son h y b, respectivamente. ¿Cuál es el valor esperado de la variable que indica la distancia del punto escogido a la base del triángulo?

Sea x una variable aleatoria con densidad de probabilidad $f(x)$ y sea g es una función con valores reales. No es necesariamente cierto que la variable aleatoria $g(x)$ tiene a $f(x)$ como densidad de probabilidad. Por ejemplo, si x es una variable aleatoria con valores positivos que tiene densidad de probabilidad $f(x)$, entonces la densidad de probabilidad de la variable aleatoria $y = x^2$ es $\frac{f(\sqrt{y})}{2\sqrt{y}}$. (En general, si g es diferenciable y estrictamente creciente o estrictamente decreciente, la densidad de probabilidad de la variable $y = g(x)$ es $f(h(y))|h'(y)|$, donde $h(y)$ es la función inversa de $g(x)$.) Por lo tanto puede ser útil usar la Regla de Substitución, que sigue siendo válida para variables continuas.

Regla de Substitución. Si x es una variable aleatoria con densidad $f(x)$, entonces para cualquier función diferenciable biyectiva g, el valor esperado de la variable aleatoria $g(x)$ puede calcularse mediante $E[g(x)] = \int_a^b g(x)f(x)dx$.

Ejemplo 3′′. Una vara de longitud uno es quebrada al azar en dos partes. La variable aleatoria q denota la razón de la longitud de la pieza más corta entre la longitud de la mayor. Si $x \in (0, 1/2]$ es la variable que indica que la vara se partió en pedazos de longitudes x y $1-x$, entonces $q = \frac{x}{1-x}$. Aceptamos que $p(c \leq x \leq d) = 2(d-c)$ para cualquier subintervalo $[c, d] \subset (0, 1/2]$, y por lo tanto la densidad es $f(x) = 2$. Aplicando la Regla de Substitución,

$$E(q) = \int_0^{1/2} \frac{2x\ dx}{1-x} = \int_0^{1/2} \left(\frac{2}{1-x} - 2\right) dx = [-2\ln(1-x) - 2x]_0^{1/2} = 2\ln 2 - 1 \approx 0.3863.$$

Ejemplo 4. El precio de un diamante es proporcional al cuadrado de su peso. Luego, cuando un diamante se parte en dos pedazos accidentalmente, pierde valor. Queremos calcular la precio esperado por ambos pedazos.

Para esto, vamos a suponer que el diamante completo tiene peso ω y precio a. Sea $x \in (0, 1/2]$ la variable que indica que el diamante se partió en dos pedazos con pesos $x\omega$ y $(1-x)\omega$. Aceptaremos que $p(c \leq x \leq d) = 2(d-c)$ para cualquier subintervalo $[c, d] \subset (0, 1/2]$. El costo total por los dos pedazos de diamante es $y = ax^2 + a(1-x)^2$, y el valor esperado del costo total es $E(y) = 2a/3$.

Definición. Sea $x : \mathcal{M} \to [a, b]$ una variable aleatoria con densidad de probabilidad $f : [a, b] \to \mathbb{R}$. La varianza de x se define mediante $\sigma^2(x) = E(|x-\mu|^2)$, donde $\mu = E(x)$. Se verifica fácilmente que $\sigma^2(x) = E(x^2) - \mu^2$.

Ejemplo 1′′. Se escoge al azar un número x en el intervalo $[a, b]$. Entonces $\sigma^2(x) = (a-b)^2/12$.

Ejemplo 2′′. Se escoge un punto al azar en el interior de un disco de radio r. Sea x la variable aleatoria que denota la distancia de este punto al centro del disco. Entonces $\sigma^2(x) = r^2/18$.

Nota: *La desigualdad de Chebychev sigue siendo válida.*

7.3. Tarea.

1. Se escoge al azar un punto dentro de una esfera de radio r. ¿Cuál es la desviación estándar de la distancia del punto escogido al centro de la esfera?
2. Una vara de longitud uno es quebrada al azar en dos partes. Sea q la variable aleatoria que denota la razón de las longitudes de la pieza más larga entre la pieza más corta. Calcular el valor esperado $E(q)$ e interpretar el resultado.
3. Considerar una variable aleatoria x con valores en el intervalo $[-\pi/2, \pi/2]$ y con densidad de probabilidad $f(x) = \frac{1}{2}\cos x$.
   a) Mostrar que $p(0 \leq \sin^2 x \leq 1/5) = 1/\sqrt{5}$.
   b) Mostrar que la media y la desviación estándar para la variable aleatoria $\sin^2 x$ son $\mu = 1/3$ y $\sigma = \sqrt{4/45}$.

8. Distribución normal

La Figura 1 muestra curvas en forma de campana que son gráficas de funciones de la forma

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}, \tag{6}$$

en donde $\mu, \sigma \in \mathbb{R}$ son parámetros que pueden variar, con la única restricción $\sigma > 0$. Una curva de este estilo es llamada *curva normal*, o *curva de Gauss*, o *campana de Gauss*[2].

Es posible mostrar que la curva normal cumple lo siguiente:

[2]Es una curiosidad que vale la pena señalar el hecho de que tres números importantes aparecen en la fórmula de la curva normal: $\sqrt{2}$, π y e.

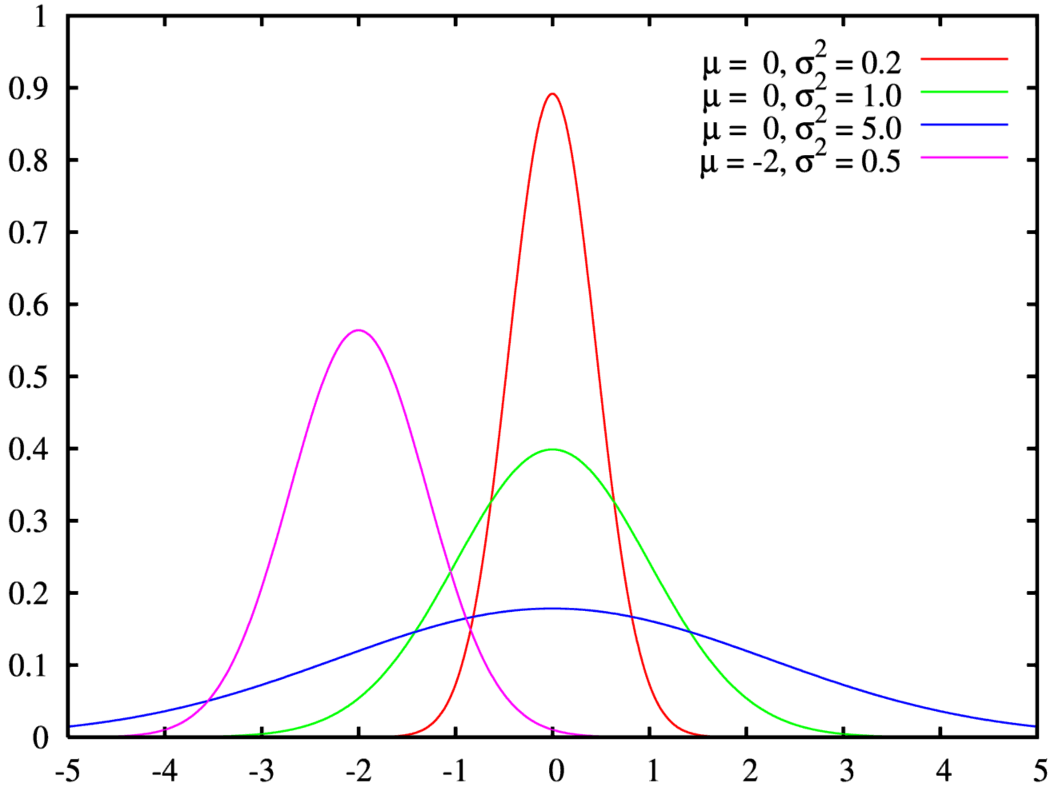

FIGURA 1. Ilustración tomada de https://es.wikipedia.org/wiki/Distribución_normal

1. Tiene un máximo en $x = \mu$.
2. Es simétrica con respecto a la recta $x = \mu$.
3. El área total bajo la curva es igual a 1. Esto es un poco más difícil de probar debido a que $f(x)$ no posee primitiva (y por lo tanto no es posible la integración usando el Teorema Fundamental del Cálculo).

Una *distribución normal* con parámetros μ y σ consiste de una variable aleatoria x que toma valores en todo $\mathbb{R}$ tal que

$$p(c \leq x \leq d) = \frac{1}{\sigma\sqrt{2\pi}} \int_c^d e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx.$$

La distribución normal con parámetros $\mu = 0$ y $\sigma = 1$, cuya ϕ tiene la gráfica ilustrada de color verde en la Figura 1, es llamada *distribución normal estándar*. Un cálculo sencillo muestra que, como la notación sugiere, $\mu = E(x)$ y $\sigma^2 = \mathrm{var}(x)$.

La función $\Phi(\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tau} e^{-\frac{1}{2}x^2} dx$, que indica la probabilidad $p(X \leq \tau)$ para la distribución normal estándar, se encuentra en tablas matemáticas para muchos valores de τ, y también es una función rutinaria en varios programas. Usando Φ se puede hallar la medida de probabilidad para cualquier distribución normal: si x es una distribución normal con parámetros μ y σ, entonces
(7)

$$p(c \leq x \leq d) = \frac{1}{\sigma\sqrt{2\pi}} \int_c^d e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx = \frac{1}{\sqrt{2\pi}} \int_{(c-\mu)/\sigma}^{(d-\mu)/\sigma} e^{-\frac{1}{2}y^2} dy = \Phi\left(\frac{d-\mu}{\sigma}\right) - \Phi\left(\frac{c-\mu}{\sigma}\right)$$

(la segunda igualdad se obtiene a partir del cambio de variable $y = (x - \mu)/\sigma$).

Ejemplo. El periodo de gestación humano satisface una distribución normal con $\mu = 266$ días y $\sigma = 16$ días. Entonces el porcentaje de nacimientos que se retrasan más de 20 días del valor esperado es aproximadamente 10.6 % pues $p(286 < x) \approx 0.10565$.

8.1. Utilidad de la distribución normal: teorema de Límite Central. Sean $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes (no importa si son finitas o continuas, aunque para nuestras aplicaciones basta que sean finitas) definidas en el mismo espacio muestral con una medida de probabilidad p. Suponer que $E(X_k) = \mu_k$ y $\operatorname{var}(X_k) = \sigma_k^2$ para cada $k \in \mathbb{N}$. Entonces $E(X_1 + \cdots + X_n) = \mu_1 + \cdots + \mu_n$ y $\operatorname{var}(X_1 + \cdots + X_n) = \sigma_1^2 + \cdots + \sigma_n^2$ para cualquier $n \in \mathbb{N}$. Resulta que la variable aleatoria $X = X_1 + \cdots + X_n$ se comporta como una distribución normal con parámetros $\mu_1 + \cdots + \mu_n$ y $\sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$ cuando n es suficientemente grande:

Teorema del Límite Central. *Para cualesquiera $c, d \in \mathbb{R}$ con $c < d$, se tiene*

$$\lim_{n\to\infty} p(c \le X_1 + \cdots + X_n \le d) = \Phi\left(\frac{b - (\mu_1 + \cdots + \mu_n)}{\sqrt{\sigma_1^2 + \cdots + \sigma_n^2}}\right) - \Phi\left(\frac{a - (\mu_1 + \cdots + \mu_n)}{\sqrt{\sigma_1^2 + \cdots + \sigma_n^2}}\right).$$

No revisaremos la demostración de este profundo resultado, solamente veremos ejemplos de su aplicación.

En particular, cuando $\mu_k = \mu$ y $\sigma_k = \sigma$ para cada $k \in \mathbb{N}$ (sería el caso en que el mismo experimento se repite una infinidad de veces, por ejemplo) entonces la fórmula anterior se expresa como

$$\lim_{n\to\infty} p(c \le X_1 + \cdots + X_n \le d) = \Phi\left(\frac{d - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right).$$

De esta forma, el Teorema del Límite Central puede servir para aproximar el resultado de un experimento de Bernoulli que se repite n veces con probabilidad de acierto p (en cuyo caso la distribución binomial da el resultado exacto, y la distribución normal sería solamente una aproximación). Más precisamente, sea

$$X_k = \begin{cases} 1 & \text{si se obtiene un acierto en la repetición } k \text{ del experimento,} \\ 0 & \text{si no se obtiene acierto en la repetición } k \text{ del experimento.} \end{cases}$$

Puesto que $\mu_k = p$ y $\sigma_k^2 = p(1-p)$ para cada k, el Teorema del Límite Central establece que

$$\lim_{n\to\infty} p(c \le X_1 + \cdots + X_n \le d) = \Phi\left(\frac{d - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{c - np}{\sqrt{np(1-p)}}\right).$$

Esta aproximación se considera adecuada por la teoría cuando $np > 5$ y $n(1-p) > 5$.

Ejemplo del vuelo Guadalajara-Monterrey revisado. Recordar que este ejemplo consistía en un experimento de Bernoulli que se repetía $n = 160$ veces con $p = 9/10$. Si X indica el número de pasajeros que se presentan al vuelo, se obtuvo $p(X \ge 151) \approx 0.0359$ usando distribución binomial. Se hará ahora el cálculo usando la aproximación normal. Sea X_j la variable que indica si el pasajero j se presenta, entonces $\mu = 9/10$ y $\sigma = 3/10$. Se desea conocer $p(151 \le X_1 + \cdots + X_{160})$ sabiendo que $X = X_1 + \cdots + X_{160}$ se comporta aproximadamente como una distribución normal con parámetros $n\mu = 144$ y $\sigma\sqrt{n} \approx 3.7947$, obteniendo $p(151 \le X_1 + \cdots + X_{160}) \approx 0.0325$.

Aguja de Buffon y experimento de Lazzarini. Recordar que el matemático Mario Lazzarini reportó haber lanzado una estaca 3408 veces en un piso reglado, agregando que la estaca cruzó las rectas paralelas del piso reglado en 1808 ocasiones (ver pág. 7). Considerar la variable

$$X_k = \begin{cases} 1 & \text{si la estaca cruza una recta paralela en la repetición } k \text{ del experimento,} \\ 0 & \text{si la estaca cruza una recta paralela en la repetición } k \text{ del experimento,} \end{cases}$$

para $k = 1, \ldots, 3408$. Analizaremos el reporte de Lazzarini usando la probabilidad $p = \frac{5}{3\pi}$ de que la aguja caiga cruzando el reglado del piso. Entonces $np \approx 1808.00015$ y $\sqrt{np(1-p)} \approx 29.13462$, de donde $p(1807 < X < 1809) \approx 0.02738$ y, por lo tanto, hay razón para sospechar que Lazzarini mintió, o que al menos realizó el experimento sin los cuidados necesarios para permitir que el azar jugara su papel[3].

Vacuna contra la polio. Este es un caso con datos reales. La vacuna de Salk contra la polio fue probada en 1954 mediante un experimento cuidadosamente diseñado. Participaron en él 400,000 niños, que se separaron al azar en dos grupos de igual tamaño. La vacuna se suministró a un grupo, el de tratamiento, y se le inyectó placebo al otro grupo, el de control. Ningún niño sabía en cuál de los dos grupos fue colocado. En el grupo de tratamiento 57 niños contrajeron polio, mientras que en el grupo de control 142 niños la contrajeron. Con estos datos, ¿qué tan confiable es la afirmación de que la vacuna funciona?

Supongamos que la vacuna no funciona. Esto se traduce en que cada individuo enfermo tiene la misma probabilidad de pertenecer al grupo de tratamiento o al de control. Bajo esta hipótesis, podemos considerar un experimento de Bernoulli con probabilidad de acierto $p = 1/2$ (un "acierto" consiste en pertenecer al pertenecer al grupo de tratamiento, mientras que un "error" consiste en pertenecer al grupo de control) que se repite 199 veces. ¿Cuál es la probabilidad de que a lo más 57 niños del grupo de tratamiento se enfermen? Se tiene $p(X < 57.5) \approx \Phi(\frac{57.5-99.5}{7.053}) \approx 1.3 \times 10^{-9}$. Esto indica que es extremadamente difícil el que, por azar, solamente 57 (o menos) de los 199 niños enfermos pertenezcan al grupo de tratamiento. Podemos concluir entonces que la hipótesis es incorrecta y entonces la vacuna influye en el resultado; por lo tanto funciona.

El experimento de los volados. Se lanzan n volados. Sea

$$X_k = \begin{cases} 1 & \text{si cae águila en el volado número } k, \\ 0 & \text{si no cae águila.} \end{cases}$$

¿Cuánto es lo máximo que $X = X_1 + \cdots + X_n$ puede diferir del valor esperado $n/2$ para poder considerar el resultado como algo dentro de lo "usual"? Para cualquier constante $c > 0$ se tiene

(8)
$$p\left(\frac{n}{2} - c \leq X \leq \frac{n}{2} + c\right) = \Phi\left(\frac{2c}{\sqrt{n}}\right) - \Phi\left(\frac{-2c}{\sqrt{n}}\right) = \Phi\left(\frac{2c}{\sqrt{n}}\right) - \left(1 - \Phi\left(\frac{2c}{\sqrt{n}}\right)\right) = 2\Phi\left(\frac{2c}{\sqrt{n}}\right) - 1.$$

Nos gustaría hallar c que satisface $2\Phi\left(\frac{2c}{\sqrt{n}}\right) - 1 \leq \frac{95}{100}$, i.e. $\Phi\left(\frac{2c}{\sqrt{n}}\right) \leq \frac{195}{200}$, luego $2c \leq 1.96\sqrt{n}$.

Por lo tanto, en el experimento de los volados, cualquier resultado que se aleje menos de $0.96\sqrt{n}$ del valor esperado $n/2$ es "usual". Por ejemplo, si $n = 400$, se tiene que $0.96\sqrt{400} \approx 19.2$, por lo tanto, si se reportan $200 + 20 = 220$ águilas o más, existiría razón para desconfiar de tal resultado.

8.2. Tarea.

1. Un estudiante presentó su examen final, el cual consistió de 50 preguntas de opción múltiple. Cada pregunta tenía 3 opciones de respuesta y solamente una de las opciones era correcta. El estudiante contestó correctamente 26 preguntas, y afirmó que contestó el examen totalmente al azar pues no estudió para el examen ni entendió nada durante el curso. ¿Deberíamos creerle?
2. En 1986 apareció un artículo en la primera página del New York Times acerca de los resultados de un proyecto de investigación sobre el efecto de una ligera dosis de aspirina en la

[3]Un cálculo de L. Badger en *Mathematics Magazine* (1994) estima que hay que repetir alrededor de $n = 134 \times 10^{12}$ veces el experimento para obtener una aproximación de seis cifras decimales de π, como obtuvo Lazzarini. Recordar que $\pi = \frac{2\ell}{dp} \approx \frac{5}{3f_{3408}}$ pues $p \approx f_{3408}$, la frecuencia relativa de la cantidad de aciertos. Cuando hay 1807 y 1809 aciertos, las aproximaciones de π resultan ser 3.14333 y 3.13986 respectivamente; es decir, a los más dos cifras decimales.

incidencia de paros cardiacos. Mediante un cuidadoso método se seleccionó al azar un grupo de 22000 hombres saludables de mediana edad que fueron divididos al azar en dos grupos del mismo tamaño: el "grupo aspirina" y el grupo placebo. En el grupo aspirina ocurrieron 104 paros cardiacos, mientras que en el grupo placebo ocurrieron 209. Analizar, en base a estos resultados, si es razonable pensar que la aspirina contribuye a prevenir paros cardiacos.

3. Considerar la cantidad de lluvia en Amsterdam durante cada uno de los meses del año como doce variables aleatorias independientes con valores esperados 62.1, 43.4, 58.9, 41.0, 48.3, 67.5, 65.8, 61.4, 82.1, 85.1, 89.0 y 74.9 mm. y desviaciones estándar de 33.9, 27.8, 31.1, 24.1, 29.3, 33.8, 36.8, 32.1, 46.6, 42.4, 40.0 y 36.2 mm. Calcular un valor aproximado para la probabilidad de que la cantidad de lluvia en Amsterdam sea mayor que 1000 mm. el próximo año.
4. Considerar el experimento de escoger un dígito al azar, es decir, escoger al azar un elemento del conjunto $\{0, \ldots, 9\}$. Designar por X al dígito escogido. Suponer que la probabilidad de que cada dígito sea escogido es $1/10$. Calcular $E(X)$ y $\sigma(X)$. ¿Según el Teorema del Límite Central, qué sumas de 100 dígitos generados al azar son consideradas demasiado grandes o demasiado chicas como para ser consideradas extraordinarias (o inusuales)?
5. Una compañía fabrica ligas y las empaca en paquetes de 200 ligas cada uno. El peso de una liga individual es una variable aleatoria con una distribución desconocida, pero se sabe que la media es aproximadamente de medio gramo, con una desviación estándar aproximada de un centécimo de gramo. Hallar la probabilidad de que el paquete pese más de 101 gramos.

9. Cuarto examen parcial

Instrucciones. Se calificará el procedimiento únicamente; un resultado sin procedimiento, o con procedimiento equivocado, o mal redactado, no tendrá NINGUNA validez (aunque el resultado sea correcto).

Promedio final. Si a un alumno le conviene, el cuarto examen parcial le contará como el 50 % de la calificación final del curso. Pero si a un alumno no le conviene lo anterior, el cuarto examen parcial le contará igual que los otros exámenes parciales.

También se eliminará de los tres primeros parciales el de menor calificación de cada alumno. Así que en total solamente se tomarán en cuenta tres parciales para la calificación final del curso: dos parciales de los anteriores y el cuarto parcial (pudiendo tener mayor peso este último, como se enunció en el párrafo anterior).

Finalmente, en el promedio que dan estos tres parciales, los decimales se redondean al siguiente entero si son mayores o iguales que .5 y se redondean al entero anterior si son inferiores a .5, quedando como calificación final del curso dicho entero redondeado. No se tomará en cuenta ningún otro parámetro en la calificación (dedicación, esfuerzo, asistencia, situaciones personales, etc.)