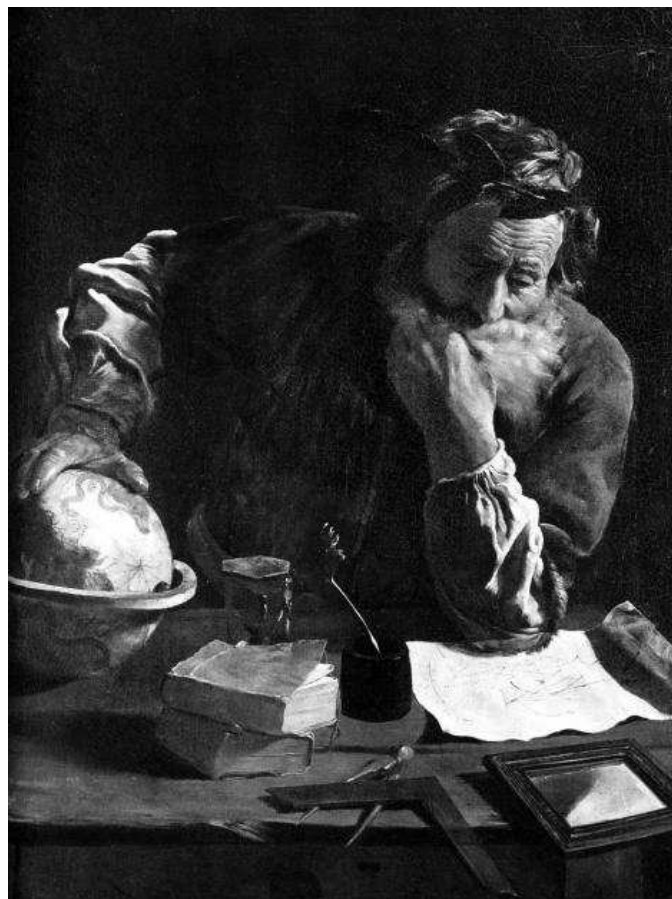


Probabilidad y Estadística (Borradores, Curso 23)
Variables aleatorias (2)
Momentos

Sebastian Grynberg

4 - 6 de abril de 2011



Denme un punto de apoyo y moveré el mundo
(Arquímedes de Siracusa)

Índice

1. Esperanza	2
1.1. Definición	3
1.2. Propiedades y cálculo	6
1.3. Dividir y conquistar	13
2. Varianza	15
2.1. Definición	15
2.2. Propiedades y cálculo	15
3. Algunas desigualdades	18
3.1. Cauchy-Schwartz	18
3.2. Chebyshev	19
4. Covarianza y varianza de sumas	21
4.1. Covarianza	21
4.2. Varianza de sumas	23
5. La ley débil de los grandes números	24
6. Distribuciones particulares	27
7. Bibliografía consultada	34

1. Esperanza

La información relevante sobre el comportamiento de una variable aleatoria está contenida en su función de distribución. Sin embargo, en la práctica, es útil disponer de algunos números representativos de la variable aleatoria que resuman esa información.

Motivación Se gira una rueda de la fortuna varias veces. En cada giro se puede obtener alguno de los siguiente números x_1, x_2, \dots, x_k -que representan la cantidad de dinero que se obtiene en el giro- con probabilidades $p(x_1), p(x_2), \dots, p(x_k)$, respectivamente. ¿Cuánto dinero se “espera” obtener como recompensa “por cada giro”? Los términos “espera” y “por cada giro” son un tanto ambiguos, pero se pueden interpretar de la siguiente manera.

Si la rueda se gira n veces y $n(x_i)$ es la cantidad de veces que se obtiene x_i , la cantidad total de dinero recibida es $\sum_{i=1}^k n(x_i)x_i$ y la cantidad media por giro es $\mu = \frac{1}{n} \sum_{i=1}^k n(x_i)x_i$. Interpretando las probabilidades como frecuencias relativas obtenemos que para n suficientemente grande la cantidad de dinero que se “espera” recibir “por cada giro” es

$$\mu = \frac{1}{n} \sum_{i=1}^k x_i n(x_i) = \sum_{i=1}^k x_i \frac{n(x_i)}{n} \approx \sum_{i=1}^k x_i p(x_i).$$

1.1. Definición

Definición 1.1 (Esperanza de una variable discreta). Sea X una variable aleatoria discreta. La *esperanza de X* , denotada por $\mathbb{E}[X]$, es el promedio ponderado

$$\mathbb{E}[X] := \sum_{x \in \mathbb{A}} x \mathbb{P}(X = x), \quad (1)$$

donde $\mathbb{A} = \{x \in \mathbb{R} : F(x) - F(x-) > 0\}$ es el conjunto de todos los átomos de la función distribución de X .

La esperanza como centro de gravedad. La noción de esperanza es análoga a la noción de centro de gravedad para un sistema de partículas discreto.

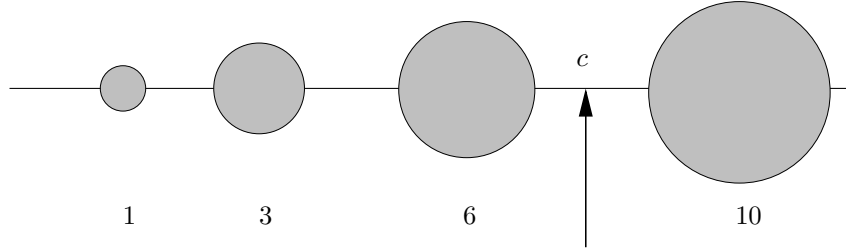


Figura 1: Interpretación de la esperanza como centro de gravedad. Se considera un sistema de cuatro “partículas” de pesos p_i proporcionales a las áreas de los círculos de radio $1/3, 2/3, 3/3, 4/3$ centrados en los puntos $x_i = 1, 3, 6, 10$, respectivamente. No se pierde generalidad si se supone que el peso total del sistema es la unidad. El centro de gravedad del sistema se encuentra en el punto $c = \sum_{i=1}^4 x_i p_i = 227/30 = 7.56 \dots$

Se consideran n partículas ubicadas en los puntos x_1, \dots, x_n cuyos pesos respectivos son $p(x_1), \dots, p(x_n)$. No se pierde generalidad si se supone que $\sum_{i=1}^n p(x_i) = 1$. El centro de gravedad, c , del sistema es el punto respecto de la cual la suma de los momentos causados por los pesos $p(x_i)$ es nula. Observando que

$$\sum_{i=1}^k (x_i - c) p(x_i) = 0 \quad \Leftrightarrow \quad c = \sum_{i=1}^k x_i p(x_i)$$

resulta que el centro de gravedad del sistema coincide con la esperanza de una variable aleatoria X a valores en $\{x_1, \dots, x_n\}$ tal que $\mathbb{P}(X = x_i) = p(x_i)$. \square

La esperanza como promedio. Sea X una variable aleatoria discreta que toma valores x_1, x_2, \dots, x_n con igual probabilidad, $\mathbb{P}(X = x_i) = 1/n$, $i = 1, \dots, n$. Por definición,

$$\mathbb{E}[X] = \sum_{i=1}^n x_i \mathbb{P}(X = x_i) = \frac{1}{n} \sum_{i=1}^n x_i.$$

\square

Ejemplo 1.2 (Dado equilibrado). Sea X el resultado del lanzamiento de un dado equilibrado. X es una variable aleatoria discreta tal que $\mathbb{P}(X = x) = 1/6$ para todo $x \in \{1, 2, 3, 4, 5, 6\}$. De acuerdo con la definición 1.1, la esperanza de X es

$$\mathbb{E}[X] = \sum_{x=1}^6 x\mathbb{P}(X = x) = \frac{1}{6} \sum_{x=1}^6 x = \frac{21}{6} = \frac{7}{2}.$$

□

Ejemplo 1.3 (Uniforme discreta). La variable aleatoria del Ejemplo 1.2 es un caso particular de una variable aleatoria discreta X uniformemente distribuida sobre el “intervalo” de números enteros $\{1, 2, \dots, n\}$. Para ser más precisos, X es una variable aleatoria tal que

$$\mathbb{P}(X = x) = \frac{1}{n} \mathbf{1}_{\{x \in \{1, 2, \dots, n\}\}}.$$

Por definición, la esperanza de X es

$$\mathbb{E}[X] = \sum_{x=1}^n x\mathbb{P}(X = x) = \frac{1}{n} \sum_{x=1}^n x = \frac{1}{n} \left(\frac{n(n+1)}{2} \right) = \frac{1+n}{2}.$$

□

Ejemplo 1.4 (Moneda equilibrada). Sea N la cantidad de veces que debe lanzarse una moneda equilibrada hasta que salga cara. N es una variable aleatoria discreta tal que $\mathbb{P}(N = n) = (1/2)^n$, $n = 1, 2, \dots$. De acuerdo con la definición 1.1, la esperanza de N es

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} n\mathbb{P}(N = n) = \sum_{n=1}^{\infty} n \left(\frac{1}{2} \right)^n.$$

Derivando ambos lados de la igualdad $\sum_{n=0}^{\infty} x^n = (1-x)^{-1}$, que vale para $|x| < 1$, se deduce que $\sum_{n=0}^{\infty} nx^{n-1} = (1-x)^{-2}$ y de allí resulta que $\sum_{n=1}^{\infty} nx^n = x(1-x)^{-2}$. Evaluando en $x = 1/2$ se obtiene que

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} n \left(\frac{1}{2} \right)^n = \left(\frac{1}{2} \right) \left(\frac{1}{2} \right)^{-2} = 2.$$

□

Proposición 1.5 (Esperanza de la función indicadora). Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad. Cualquiera sea el evento $A \in \mathcal{A}$ vale que

$$\mathbb{E}[\mathbf{1}_{\{\omega \in A\}}] = \mathbb{P}(A). \quad (2)$$

Demostración. Por definición, $\mathbb{E}[\mathbf{1}\{\omega \in A\}] = 0 \cdot (1 - \mathbb{P}(A)) + 1 \cdot \mathbb{P}(A)$. \square

La noción de esperanza se puede extender a *variables aleatorias absolutamente continuas* cambiando en (1) la suma por la integral y la función de probabilidades $P(X = x)$, $x \in \mathbb{A}$, por la densidad de probabilidades de la variable X .

Definición 1.6 (Esperanza de una variable absolutamente continua). Sea X una variable aleatoria absolutamente continua con densidad de probabilidades $f_X(x)$. La *esperanza de X* , denotada por $\mathbb{E}[X]$, se define por

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f_X(x) dx. \quad (3)$$

Ejemplo 1.7 (Fiabilidad). Sea T el tiempo de espera hasta la primer falla de un sistema electrónico con función intensidad de fallas de la forma $\lambda(t) = 2t\mathbf{1}\{t > 0\}$. La función de distribución de T es $F_T(t) = (1 - \exp(-t^2))\mathbf{1}\{t > 0\}$. En consecuencia, T es una variable aleatoria absolutamente continua con densidad de probabilidad $f_T(t) = 2t \exp(-t^2)\mathbf{1}\{t > 0\}$. De acuerdo con la definición 1.6, la esperanza de T es

$$\mathbb{E}[T] = \int_{-\infty}^{\infty} t f_T(t) dt = \int_0^{\infty} t 2t \exp(-t^2) dt = \int_0^{\infty} \exp(-t^2) dt = \frac{\sqrt{\pi}}{2}.$$

La tercera igualdad se deduce de la fórmula de integración por partes aplicada a $u = t$ y $v' = 2t \exp(-t^2)$ y la cuarta se deduce de la identidad $\int_0^{\infty} \exp(-x^2/2) dx = \sqrt{2\pi}/2$ mediante el cambio de variables $t = x/\sqrt{2}$. \square

Extendiendo la noción a variables mixtas. La noción de esperanza para variables mixtas se obtiene combinando adecuadamente las nociones anteriores.

Definición 1.8 (Esperanza de una variable mixta). Sea X una variable aleatoria mixta con función de distribución $F_X(x)$. La *esperanza de X* , denotada por $\mathbb{E}[X]$, se define de la siguiente manera:

$$\mathbb{E}[X] = \sum_{x \in \mathbb{A}} x \mathbb{P}(X = x) + \int_{-\infty}^{\infty} x F'_X(x) dx, \quad (4)$$

donde $\mathbb{A} = \{x \in \mathbb{R} : F_X(x) - F_X(x-) > 0\}$ es el conjunto de todos los átomos de $F_X(x)$ y $F'_X(x)$ es una función que coincide con la derivada de $F_X(x)$ salvo en el conjunto de puntos

$$\mathbb{A} \cup \left\{ x \in \mathbb{R} : \nexists \lim_{h \rightarrow 0} \frac{F_X(x+h) - F_X(x)}{h} \right\}.$$

Ejemplo 1.9 (Mixtura). Sea X una variable aleatoria mixta cuya función de distribución es $F_X(x) = \left(\frac{2x+5}{8}\right)\mathbf{1}\{-1 \leq x < 1\} + \mathbf{1}\{x \geq 1\}$. De acuerdo con fórmula (4), la esperanza de X es

$$\mathbb{E}[X] = -1 \cdot \mathbb{P}(X = -1) + 1 \cdot \mathbb{P}(X = 1) + \int_{-1}^1 F'_X(x) dx = -\frac{3}{8} + \frac{1}{8} + \int_{-1}^1 \frac{2}{8} dx = \frac{1}{4}.$$

\square

Nota Bene. En todas las definiciones anteriores, se presupone que las series y/o integrales involucradas son absolutamente convergentes.

Ejemplo 1.10 (Distribución de Cauchy). Sea X una variable aleatoria con *distribución de Cauchy*. Esto es, X es absolutamente continua y admite una densidad de probabilidades de la forma

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Debido a que

$$\int_{-\infty}^{\infty} |x|f(x)dx = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}dx = \infty,$$

X no tiene esperanza. □

1.2. Propiedades y cálculo

Proposición 1.11.

- (a) Si X es una variable positiva, entonces $\mathbb{E}[X] > 0$.
- (b) $\mathbb{E}[1] = 1$;
- (c) $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$, para todo $a, b \in \mathbb{R}$.

Demostración. Para evitar dificultades técnicas supondremos que X es una variable discreta.

(a) Si la variable es positiva, su esperanza no es otra cosa que una suma de cantidades positivas, por lo tanto es positiva.

(b) 1 puede ser visto como una variable discreta que asume solamente el valor 1. Tiene por lo tanto, distribución de probabilidades $\mathbb{P}(X = 1) = 1$ y $\mathbb{P}(X = x) = 0$ para $x \neq 1$, el resultado es consecuencia inmediata de la definición (1).

(c)

$$\begin{aligned} \mathbb{E}[aX + b] &= \sum_{y \in a\mathbb{A} + b} y\mathbb{P}(aX + b = y) = \sum_{x \in \mathbb{A}} (ax + b)\mathbb{P}(X = x) \\ &= a \sum_{x \in \mathbb{A}} x\mathbb{P}(X = x) + b \sum_{x \in \mathbb{A}} \mathbb{P}(X = x) = a\mathbb{E}[X] + b. \end{aligned}$$

□

Ejemplo 1.12 (Uniforme discreta). La variable aleatoria del Ejemplo 1.3 es un caso particular de una variable aleatoria discreta X uniformemente distribuida sobre un “intervalo” de números enteros $\{a, a+1, \dots, b\}$, donde a y b son dos enteros tales que $a < b$. Para ser más precisos, X es una variable aleatoria tal que

$$\mathbb{P}(X = x) = \frac{1}{b-a+1} \mathbf{1}\{x \in \{a, a+1, \dots, b\}\}.$$

Notando que la distribución de X coincide con la de la variable $X^* + a - 1$, donde X^* está uniformemente distribuida sobre $\{1, \dots, b - a + 1\}$, resulta que

$$\mathbb{E}[X] = \mathbb{E}[X^*] + a - 1 = \frac{1 + (b - a + 1)}{2} + a - 1 = \frac{a + b}{2}.$$

□

Teorema 1.13. Sea X una variable aleatoria no negativa (i.e., $F_X(x) = \mathbb{P}(X \leq x) = 0$ para todo $x < 0$). Vale que

$$\mathbb{E}[X] = \int_0^\infty [1 - F_X(x)] dx. \quad (5)$$

Demostración. El argumento principal está contenido en la Figura 2. El caso general se deduce usando técnicas de “paso al límite”.

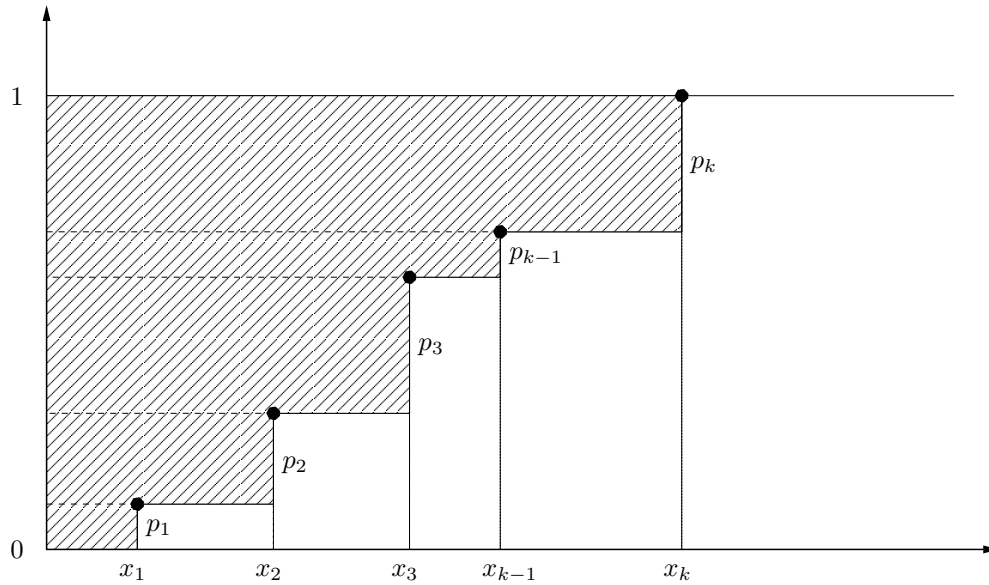


Figura 2: Argumento geométrico que muestra la validez de la identidad (5) en el caso en que X es no negativa, discreta y a valores $0 \leq x_1 < x_2 < \dots < x_k$. Si $p_i = \mathbb{P}(X = x_i)$, el área de la región sombreada es la suma $x_1 p_1 + \dots + x_k p_k = \mathbb{E}[X]$ de las áreas de los rectángulos horizontales y coincide con la integral de la altura $\mathbb{P}(X > x)$.

□

Corolario 1.14. Sea X una variable aleatoria con función de distribución $F_X(x)$. Vale que

$$\mathbb{E}[X] = \int_0^\infty [1 - F_X(x)] dx - \int_{-\infty}^0 F_X(x) dx. \quad (6)$$

Demostración. Ejercicio. □

Nota Bene. Las identidades (5) y (6) son interesantes porque muestran que para calcular la esperanza de una variable aleatoria basta conocer su función de distribución. De hecho, la identidad (6) ofrece una definición alternativa y unificada de la noción de esperanza. □

Ejemplo 1.15. Una máquina fue diseñada para prestar servicios en una instalación productiva. La máquina se enciende al iniciar la jornada laboral y se apaga al finalizar la misma. Si durante ese período la máquina falla, se la repara y en esa tarea se consume el resto de la jornada.

(a) Si se supone que la función intensidad de fallas de la máquina es una constante $\lambda > 0$ (y que el tiempo se mide en jornadas laborales), cuál es el máximo valor de λ que permite asegurar que la máquina prestará servicios durante una jornada laboral completa con probabilidad no inferior a los $2/3$?

(b) Para ese valor de λ , hallar (y graficar) la función de distribución del tiempo, T , de funcionamiento de la máquina durante una jornada laboral y calcular el tiempo medio de funcionamiento, $\mathbb{E}[T]$.

Solución.

(a) Si T_1 es el tiempo que transcurre desde que se enciende la máquina hasta que ocurre la primer falla, el evento “la máquina funciona durante una jornada laboral completa” se describe mediante $\{T_1 > 1\}$. Queremos hallar el máximo $\lambda > 0$ tal que $\mathbb{P}(T_1 > 1) \geq 2/3$. Debido a que la función intensidad de fallas es una constante λ se tiene que $\mathbb{P}(T_1 > t) = e^{-\lambda t}$. En consecuencia, $\mathbb{P}(T_1 > 1) \geq 2/3 \iff e^{-\lambda} \geq 2/3 \iff \lambda \leq -\log(2/3)$. Por lo tanto,

$$\lambda = -\log(2/3).$$

En tal caso, $\mathbb{P}(T > 1) = 2/3$.

(b) El tiempo de funcionamiento de la máquina durante una jornada laboral es

$$T := \min\{T_1, 1\}.$$

Para $t > 0$ vale que

$$\begin{aligned} F_T(t) &= \mathbb{P}(T \leq t) = 1 - \mathbb{P}(T > t) = 1 - \mathbb{P}(\min\{T_1, 1\} > t) \\ &= 1 - \mathbb{P}(T_1 > t) \mathbf{1}\{1 > t\} = 1 - e^{\log(2/3)t} \mathbf{1}\{t < 1\} \\ &= (1 - e^{\log(2/3)t}) \mathbf{1}\{0 \leq t < 1\} + \mathbf{1}\{t \geq 1\}. \end{aligned}$$

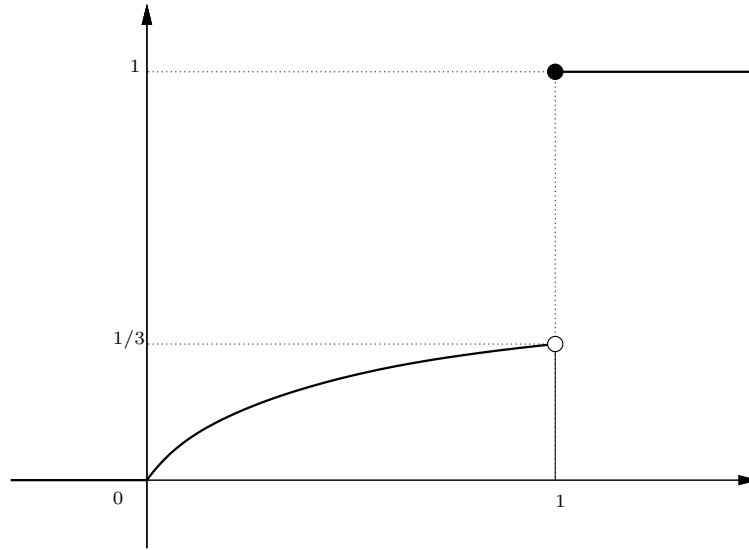


Figura 3: Gráfico de la función de distribución de T .

Como $T > 0$ y conocemos la función $\mathbb{P}(T > t)$ lo más sencillo para calcular la esperanza es usar la fórmula $\mathbb{E}[T] = \int_0^\infty \mathbb{P}(T > t) dt$:

$$\begin{aligned} \mathbb{E}[T] &= \int_0^\infty \mathbb{P}(T > t) dt = \int_0^1 e^{\log(2/3)t} dt = \left. \frac{e^{\log(2/3)t}}{\log(2/3)} \right|_0^1 = \frac{2/3 - 1}{\log(2/3)} \\ &= \frac{-1/3}{\log(2/3)} \approx 0.822... \end{aligned}$$

□

Cálculo

Consideramos una variable aleatoria X de la que conocemos su función de distribución y queremos calcular la esperanza de alguna función de X , digamos, $g(X)$. ¿Cómo podríamos efectuar ese cálculo? Una manera es la siguiente: $Y = g(X)$ es una variable aleatoria y su función de distribución $F_Y(y)$ se puede determinar a partir del conocimiento que tenemos sobre la distribución de X :

$$F_Y(y) := \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \in g^{-1}(-\infty, y]).$$

Una vez que ha sido determinada la distribución de Y , la esperanza $\mathbb{E}[g(X)] = \mathbb{E}[Y]$ se calcula por definición.

Ejemplo 1.16. Queremos calcular $\mathbb{E}[X^2]$ para una variable aleatoria discreta X tal que

$$\mathbb{P}(X = 0) = 0.2, \quad \mathbb{P}(X = 1) = 0.5, \quad \mathbb{P}(X = 2) = 0.3$$

Poniendo $Y = X^2$ obtenemos una variable aleatoria a valores en $\{0^2, 1^2, 2^2\}$ tal que

$$\mathbb{P}(Y = 0) = 0.2, \quad \mathbb{P}(Y = 1) = 0.5, \quad \mathbb{P}(Y = 4) = 0.3$$

En consecuencia,

$$\mathbb{E}[X^2] = \mathbb{E}[Y] = 0(0.2) + 1(0.5) + 4(0.3) = 1.7$$

□

Ejemplo 1.17. Sea X una variable aleatoria con distribución uniforme sobre el intervalo $(0, 1)$. Queremos calcular $\mathbb{E}[X^3]$.

Ponemos $Y = X^3$ y calculamos su función de distribución. Para cada $0 < y < 1$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^3 \leq y) = \mathbb{P}(X \leq y^{1/3}) = y^{1/3}.$$

Derivamos $F_Y(y)$ y obtenemos la densidad de probabilidad de Y

$$f_Y(y) = \frac{1}{3}y^{-2/3}\mathbf{1}_{\{0 < y < 1\}}.$$

En consecuencia,

$$\mathbb{E}[X^3] = \mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \frac{1}{3} y^{-2/3} dy = \frac{1}{3} \int_0^1 y^{1/3} dy = \frac{1}{3} \frac{3}{4} y^{4/3} \Big|_0^1 = \frac{1}{4}.$$

□

Existe una manera mucho más simple para calcular la esperanza de $Y = g(X)$ que no recurre al procedimiento de determinar primero la distribución de Y para luego calcular su esperanza por definición. La siguiente proposición muestra cómo hacerlo.

Teorema 1.18. Sea X una variable aleatoria con función de distribución $F_X(x)$ y sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función tal que $g(X)$ es una variable aleatoria.

(a) Si X es discreta con átomos en el conjunto \mathbb{A} , entonces

$$\mathbb{E}[g(X)] = \sum_{x \in \mathbb{A}} g(x) \mathbb{P}(X = x). \quad (7)$$

(b) Si X es continua con densidad de probabilidad $f_X(x)$ y $g(X)$ es continua, entonces

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \quad (8)$$

(c) Si X es mixta,

$$\mathbb{E}[g(X)] = \sum_{x \in \mathbb{A}} g(x) \mathbb{P}(X = x) + \int_{-\infty}^{\infty} g(x) F'_X(x) dx, \quad (9)$$

donde \mathbb{A} es el conjunto de todos los átomos de $F_X(x)$ y $F'_X(x)$ es una función que coincide con la derivada de $F_X(x)$ en todos los puntos donde ésta derivada existe y vale cero en otro lado.

Ejemplo 1.19. Aplicando la parte (a) de la Proposición 1.18 al Ejemplo 1.16 se obtiene

$$\mathbb{E}[X^2] = 0^2(0.2) + 1^2(0.5) + 2^2(0.3) = 1.7$$

□

Ejemplo 1.20. Aplicando la parte (b) de la Proposición 1.18 al Ejemplo 1.17 se obtiene

$$\mathbb{E}[X^3] = \int_0^1 x^3 dx = \frac{1}{4}$$

□

Demostración del Teorema 1.18.

(a) La demostración se obtiene agrupando todos los términos que tienen el mismo valor $g(x)$. Sea $Y = g(X)$, por definición,

$$\begin{aligned} \mathbb{E}[g(X)] &= \mathbb{E}[Y] = \sum_{y \in g(\mathbb{A})} y \mathbb{P}(Y = y) = \sum_{y \in g(\mathbb{A})} y \mathbb{P}(g(X) = y) \\ &= \sum_{y \in g(\mathbb{A})} y \mathbb{P}(X \in g^{-1}(y)) = \sum_{y \in g(\mathbb{A})} y \sum_{x \in g^{-1}(y)} \mathbb{P}(X = x) \\ &= \sum_{y \in g(\mathbb{A})} \sum_{x \in g^{-1}(y)} y \mathbb{P}(X = x) = \sum_{y \in g(\mathbb{A})} \sum_{x \in g^{-1}(y)} g(x) \mathbb{P}(X = x) \\ &= \sum_{x \in \mathbb{A}} g(x) \mathbb{P}(X = x) \quad \left(\text{pues } \mathbb{A} = \bigcup_{y \in g(\mathbb{A})} g^{-1}(y) \right). \end{aligned}$$

(b) Para simplificar la prueba vamos a suponer que $g \geq 0$. De acuerdo con el Teorema 1.13

$$\mathbb{E}[g(X)] = \int_0^\infty \mathbb{P}(g(X) > y) dy = \int_0^\infty \left(\int_{\{x: g(x) > y\}} f(x) dx \right) dy.$$

Intercambiando el orden de integración obtenemos

$$\mathbb{E}[g(X)] = \int_{-\infty}^\infty \left(\int_0^{g(x)} dy \right) f(x) dx = \int_{-\infty}^\infty g(x) f(x) dx.$$

(c) Se obtiene combinando adecuadamente los resultados (a) y (b). □

Nota Bene. Como consecuencia del Teorema 1.18 se deduce que

$$\mathbb{E} \left[\sum_{k=0}^n a_k X^k \right] = \sum_{k=0}^n a_k \mathbb{E}[X^k]. \quad (10)$$

□

Teorema 1.21 (Cálculo de Esperanzas). Sea \mathbf{X} un vector aleatorio y sea $g : \mathbb{R}^n \rightarrow \mathbb{R}$ una función tal que $g(\mathbf{X})$ es una variable aleatoria. Si la variable aleatoria $g(\mathbf{X})$ tiene esperanza finita, entonces

$$\mathbb{E}[g(\mathbf{X})] = \begin{cases} \sum_{\mathbf{x}} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) & \text{en el caso discreto,} \\ \int_{\mathbb{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} & \text{en el caso continuo,} \end{cases}$$

donde, según sea el caso, $p_{\mathbf{X}}(\mathbf{x})$ y $f_{\mathbf{X}}(\mathbf{x})$ son la función de probabilidad y la densidad conjunta del vector \mathbf{X} , respectivamente.

Demostración. Enteramente análoga a la que hicimos en dimensión 1. \square

Sobre el cálculo de esperanzas. El Teorema 1.21 es una herramienta práctica para calcular esperanzas. Su resultado establece que si queremos calcular la esperanza de una transformación unidimensional del vector \mathbf{X} , $g(\mathbf{X})$, *no necesitamos* calcular la distribución de $g(\mathbf{X})$. La esperanza $\mathbb{E}[g(\mathbf{X})]$ puede calcularse directamente a partir del conocimiento de la distribución conjunta de \mathbf{X} . \square

Corolario 1.22. Si la *media* $\mu_{X_i} := \mathbb{E}[X_i]$ de la variable X_i existe, se puede calcular de la siguiente manera

$$\mu_{X_i} = \sum_{\mathbf{x}} x_i p_{\mathbf{X}}(\mathbf{x}), \quad (11)$$

en el caso discreto; y

$$\mu_{X_i} = \int_{\mathbb{R}^n} x_i f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (12)$$

en el caso continuo.

Corolario 1.23 (Linealidad de la esperanza). Si X_1, X_2, \dots, X_n son variables aleatorias con media finita y a_1, a_2, \dots, a_n son n constantes, entonces

$$\mathbb{E} \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i \mathbb{E}[X_i] \quad (13)$$

Corolario 1.24 (Monotonía). Si las variables X e Y tienen esperanza finita y satisfacen que $X \leq Y$, entonces $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Corolario 1.25 (Regla del producto independiente). Si las variables X_1, \dots, X_n son independientes con esperanza finita, entonces el producto tiene esperanza finita y coincide con el producto de las esperanzas:

$$\mathbb{E} \left[\prod_{i=1}^n X_i \right] = \prod_{i=1}^n \mathbb{E}[X_i]. \quad (14)$$

1.3. Dividir y conquistar

Teorema 1.26 (Esperanza de una variable truncada). Sea X una variable aleatoria definida sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$. Sea $A \subset \mathbb{R}$ un conjunto tal que $\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{A}$ y tal que $\mathbb{P}(X \in A) > 0$. Entonces

$$\mathbb{E}[X|X \in A] = \frac{1}{\mathbb{P}(X \in A)} \mathbb{E}[X \mathbf{1}\{X \in A\}] \quad (15)$$

Demostración. Para simplificar la exposición vamos a suponer que la variable aleatoria X es discreta. En ese caso la función de probabilidad de $X|X \in A$ adopta la forma

$$p_{X|X \in A}(x) = \frac{\mathbb{P}(X = x)}{\mathbb{P}(X \in A)} \mathbf{1}\{x \in A\}.$$

En consecuencia, la esperanza de la variable truncada $X|X \in A$ es

$$\begin{aligned} \mathbb{E}[X|X \in A] &= \sum_{x \in X(\Omega)} x \frac{\mathbb{P}(X = x)}{\mathbb{P}(X \in A)} \mathbf{1}\{x \in A\} \\ &= \frac{1}{\mathbb{P}(X \in A)} \sum_{x \in X(\Omega)} x \mathbf{1}\{x \in A\} \mathbb{P}(X = x) \\ &= \frac{1}{\mathbb{P}(X \in A)} \mathbb{E}[X \mathbf{1}\{X \in A\}]. \end{aligned}$$

La última igualdad es consecuencia del Teorema 1.18. □

Ejemplo 1.27 (Dado equilibrado). Sea X el resultado del tiro de un dado equilibrado y sea $A = \{2, 4, 6\}$. De acuerdo con (15) la esperanza de $X|X \in A$ es

$$\mathbb{E}[X|X \in A] = \frac{1}{\mathbb{P}(X \in A)} \mathbb{E}[X \mathbf{1}\{X \in A\}] = \frac{1}{1/2} \left(\frac{2}{6} + \frac{4}{6} + \frac{6}{6} \right) = 4.$$

Resultado que por otra parte es intuitivamente evidente. □

Motivación. El resultado siguiente es una versión general del resultado que permite calcular el promedio de una muestra de $n + m$ valores $x_1, \dots, x_n, x_{n+1}, \dots, x_{n+m}$ a partir de los promedios de las submuestras, x_1, \dots, x_n y x_{n+1}, \dots, x_{n+m} .

Si

$$\bar{x}_{1,n} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{x}_{n+1,n+m} = \frac{1}{m} \sum_{j=1}^m x_{n+j}$$

son los promedios de las submuestras x_1, \dots, x_n y x_{n+1}, \dots, x_{n+m} , respectivamente y

$$\bar{x}_{1,n+m} = \frac{1}{n+m} \sum_{i=1}^{n+m} x_i$$

es el promedio de la muestra x_1, \dots, x_{n+m} . Entonces, el promedio $\bar{x}_{1,n+m}$ es igual al promedio ponderado de los promedios parciales, $\bar{x}_{1,n}$ y $\bar{x}_{n+1,n+m}$:

$$\bar{x}_{n+m} = \bar{x}_n \left(\frac{n}{n+m} \right) + \bar{x}_m \left(\frac{m}{n+m} \right).$$

Teorema 1.28 (Fórmula de probabilidad total). Sea X una variable aleatoria. Si A_1, A_2, \dots, A_n es una partición medible de \mathbb{R} tal que $\mathbb{P}(X \in A_i) > 0$, $i = 1, \dots, n$. Entonces,

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X|X \in A_i] \mathbb{P}(X \in A_i). \quad (16)$$

Demostración. Descomponer la variable X en una suma de variables (dependientes de la partición) $X = \sum_{i=1}^n X \mathbf{1}\{X \in A_i\}$. Usar la linealidad de la esperanza para obtener

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X \mathbf{1}\{X \in A_i\}]$$

y usar que

$$\mathbb{E}[X|X \in A_i] = \frac{1}{\mathbb{P}(X \in A_i)} \mathbb{E}[X \mathbf{1}\{X \in A_i\}].$$

□

Nota Bene. Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función tal que $g(X)$ es una variable aleatoria. Bajo las hipótesis del Teorema 1.28 también vale que

$$\mathbb{E}[g(X)] = \sum_{i=1}^n \mathbb{E}[g(X)|X \in A_i] \mathbb{P}(X \in A_i). \quad (17)$$

La fórmula (17) se puede extender sin ninguna dificultad al caso multidimensional. □

Ejemplo 1.29 (Dividir y conquistar). Todas las mañanas Lucas llega a la estación del subte entre las 7:10 y las 7:30 (con distribución uniforme en el intervalo). El subte llega a la estación cada quince minutos comenzando a las 6:00. Calcular la media del tiempo que tiene que esperar hasta subirse al subte.

Sea X el tiempo de llegada de Lucas a la estación del subte. El tiempo de espera hasta subirse al subte es

$$T = (7.15 - X) \mathbf{1}\{X \in [7 : 10, 7 : 15]\} + (7 : 30 - X) \mathbf{1}\{X \in (7 : 15, 7 : 30]\}.$$

Ahora bien, dado que $X \in [7 : 10, 7 : 15]$, la distribución de T es uniforme sobre el intervalo $[0, 5]$ minutos y dado que $X \in (7 : 15, 7 : 30]$ la distribución de T es uniforme sobre el intervalo $[0, 15]$ minutos. De acuerdo con (17)

$$\mathbb{E}[T] = \frac{5}{2} \left(\frac{5}{20} \right) + \frac{15}{2} \left(\frac{15}{20} \right) = 6.25.$$

□

2. Varianza

2.1. Definición

La esperanza de una variable aleatoria X , $\mathbb{E}[X]$, también se conoce como *la media* o el primer momento de X . La cantidad $\mathbb{E}[X^n]$, $n \geq 1$, se llama el *n -ésimo momento* de X . Si la esperanza $\mathbb{E}[X]$ es finita, la cantidad $\mathbb{E}[(X - \mathbb{E}[X])^n]$ se llama el *n -ésimo momento central*.

La segunda cantidad en orden de importancia, después de la esperanza, para resumir el comportamiento de una variable aleatoria X es su segundo momento central también llamado la *varianza de X* .

Definición 2.1 (Varianza). Sea X una variable aleatoria con esperanza, $\mathbb{E}[X]$ finita. La *varianza* de X se define por

$$\mathbb{V}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]. \quad (18)$$

En otras palabras, la varianza de X es la esperanza de la variable aleatoria $(X - \mathbb{E}[X])^2$. Puesto que $(X - \mathbb{E}[X])^2$ sólo puede tomar valores no negativos, la varianza es no negativa.

La varianza de X es una de las formas más utilizadas para medir la dispersión de los valores de X respecto de su media. Otra medida de dispersión es el *desvío estándar* de X , que se define como la raíz cuadrada de la varianza y se denota $\sigma(X)$:

$$\sigma(X) := \sqrt{\mathbb{V}(X)}. \quad (19)$$

A diferencia de la varianza, el desvío estándar de una variable aleatoria es más fácil de interpretar porque tiene las mismas unidades de X .

Nota Bene: Grandes valores de $\mathbb{V}(X)$ significan grandes variaciones de los valores de X alrededor de la media. Al contrario, pequeños valores de $\mathbb{V}(X)$ implican una pronunciada concentración de la masa de la distribución de probabilidades en un entorno de la media. En el caso extremo, cuando la varianza es 0, la masa total de la distribución de probabilidades se concentra en la media. Estas afirmaciones pueden hacerse más precisas y serán desarrolladas en la sección 3.

2.2. Propiedades y cálculo

Proposición 2.2. Para todo $a, b \in \mathbb{R}$

$$\mathbb{V}(aX + b) = a^2 \mathbb{V}(X). \quad (20)$$

Demostración. Debido a que la esperanza es un operador lineal, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$. En consecuencia, $aX + b - \mathbb{E}[aX + b] = a(X - \mathbb{E}[X])$. Por lo tanto,

$$\mathbb{V}(aX + b) = \mathbb{E}[(aX + b - \mathbb{E}[aX + b])^2] = \mathbb{E}[a^2(X - \mathbb{E}[X])^2] = a^2 \mathbb{V}(X).$$

□

Cálculo

Una manera “brutal” de calcular $\mathbb{V}(X)$ es calcular la función de distribución de la variable aleatoria $(X - \mathbb{E}[X])^2$ y usar la definición de esperanza. En lo que sigue mostraremos una manera más simple de realizar este tipo cálculo.

Proposición 2.3 (Expresión de la varianza en términos de los momentos). Sea X una variable aleatoria con primer y segundo momentos finitos, entonces

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \quad (21)$$

En palabras, la varianza es la diferencia entre el segundo momento y el cuadrado del primer momento.

Demostración. Desarrollar el cuadrado $(X - \mathbb{E}[X])^2$ y usar las propiedades de la esperanza. Poniendo $(X - \mathbb{E}[X])^2 = X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2$ se obtiene

$$\mathbb{V}(X) = \mathbb{E}[X^2] - 2X\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

□

Ejemplo 2.4 (Dado equilibrado). Sea X el resultado del lanzamiento de un dado equilibrado. Según lo visto en el Ejemplo 1.2, $\mathbb{E}[X] = \frac{7}{2}$. Por otra parte

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=1}^6 x^2 \mathbb{P}(X = x) \\ &= 1^2 \left(\frac{1}{6}\right) + 2^2 \left(\frac{1}{6}\right) + 3^2 \left(\frac{1}{6}\right) + 4^2 \left(\frac{1}{6}\right) + 5^2 \left(\frac{1}{6}\right) + 6^2 \left(\frac{1}{6}\right) = \frac{91}{6}. \end{aligned}$$

Por lo tanto, de acuerdo con la Proposición 2.3, la varianza de X es

$$\mathbb{V}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{32}{12}.$$

□

Ejemplo 2.5 (Uniforme discreta). Sean a y b dos números enteros tales que $a < b$ y sea X una variable aleatoria discreta con distribución uniforme sobre el “intervalo” $[a, b] := \{a, a+1, \dots, b\}$. Para calcular la varianza de X , consideramos primero el caso más simple donde $a = 1$ y $b = n$.

Por inducción en n se puede ver que

$$\mathbb{E}[X^2] = \frac{1}{n} \sum_{k=1}^n k^2 = \frac{(n+1)(2n+1)}{6}.$$

La varianza puede obtenerse en términos de los momentos de orden 1 y 2:

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= \frac{(n+1)[2(2n+1) - 3(n+1)]}{12} = \frac{n^2 - 1}{12}.\end{aligned}$$

Para el caso general, notamos que la variable aleatoria uniformemente distribuida sobre $[a, b]$ tiene la misma varianza que la variable aleatoria uniformemente distribuida sobre $[1, b - a + 1]$, puesto que esas dos variables difieren en la constante $a - 1$. Por lo tanto, la varianza buscada se obtiene de la fórmula anterior sustituyendo $n = b - a + 1$

$$\mathbb{V}(X) = \frac{(b - a + 1)^2 - 1}{12} = \frac{(b - a)(b - a + 2)}{12}.$$

□

Ejemplo 2.6 (Fiabilidad). Sea T el tiempo de espera hasta la primer falla de un sistema electrónico con función intensidad de fallas de la forma $\lambda(t) = 2t\mathbf{1}\{t > 0\}$. Según lo visto en el Ejemplo 1.7, $\mathbb{E}[T] = \sqrt{\pi}/2$. Por otra parte,

$$\mathbb{E}[T^2] = \int_{-\infty}^{\infty} t^2 f(t) dt = \int_0^{\infty} t^2 2t \exp(-t^2) dt = \int_0^{\infty} x e^{-x} dx = 1.$$

La tercera igualdad se obtiene mediante el cambio de variables $t^2 = x$ y la cuarta se deduce usando la fórmula de integración por partes aplicada a $u = x$ y $v' = e^{-x}$.

Por lo tanto, de acuerdo con la Proposición 2.3, la varianza de T es

$$\mathbb{V}(T) = 1 - \left(\frac{\sqrt{\pi}}{2}\right)^2 = 1 - \frac{\pi}{4}.$$

□

Proposición 2.7 (Varianza de la función indicadora). Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad. Cualquiera sea el evento $A \in \mathcal{A}$ vale que

$$\mathbb{V}(\mathbf{1}\{\omega \in A\}) = \mathbb{P}(A)(1 - \mathbb{P}(A)). \quad (22)$$

Demostración. Observar que $\mathbf{1}\{\omega \in A\}^2 = \mathbf{1}\{\omega \in A\}$ y usar la identidad (21) combinada con la identidad (2): $\mathbb{V}(\mathbf{1}\{\omega \in A\}) = \mathbb{P}(A) - \mathbb{P}(A)^2 = \mathbb{P}(A)(1 - \mathbb{P}(A))$. □

Error cuadrático medio. Una manera de “representar” la variable aleatoria X mediante un valor fijo $c \in \mathbb{R}$ es hallar el valor c que minimice el llamado *error cuadrático medio*, $\mathbb{E}[(X - c)^2]$.

Teorema 2.8 (Pitágoras). Sea X una variable aleatoria con esperanza y varianza finitas. Para toda constante $c \in \mathbb{R}$ vale que

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

En particular, el valor de c que minimiza el error cuadrático medio es la esperanza de X , $\mathbb{E}[X]$.

Demostración. Escribiendo $X - c$ en la forma $X - \mathbb{E}[X] + \mathbb{E}[X] - c$ y desarrollando cuadrados se obtiene $(X - c)^2 = (X - \mathbb{E}[X])^2 + (\mathbb{E}[X] - c)^2 + 2(X - \mathbb{E}[X])(\mathbb{E}[X] - c)$. El resultado se obtiene tomando esperanza en ambos lados de la igualdad y observando que $\mathbb{E}[X - \mathbb{E}[X]] = 0$. \square

3. Algunas desigualdades

3.1. Cauchy-Schwartz

Teorema 3.1 (Cauchy-Schwartz).

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[X^2]\mathbb{E}[Y^2])^{1/2} \quad (23)$$

Demostración. Observar que para todo $t \in \mathbb{R}$:

$$0 < \mathbb{E}[(t|X| + |Y|)^2] = t^2\mathbb{E}[X^2] + 2t\mathbb{E}[|XY|] + \mathbb{E}[Y^2].$$

Como la cuadrática en t que aparece en el lado derecho de la igualdad tiene a lo sumo una raíz real se deduce que

$$4\mathbb{E}[|XY|]^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0.$$

Por lo tanto,

$$\mathbb{E}[|XY|]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

\square

Corolario 3.2. Sea X una variable aleatoria tal que $\mathbb{E}[X^2] < \infty$ y $a < \mathbb{E}[X]$. Entonces

$$\mathbb{P}(X > a) \geq \frac{(\mathbb{E}[X] - a)^2}{\mathbb{E}[X^2]}.$$

Demostración. De la desigualdad $X\mathbf{1}\{X > a\} \leq |X\mathbf{1}\{X > a\}|$ y de la propiedad de monotonía de la esperanza se deduce que

$$\mathbb{E}[X\mathbf{1}\{X > a\}] \leq \mathbb{E}[|X\mathbf{1}\{X > a\}|]. \quad (24)$$

Aplicando la desigualdad de Cauchy-Schwartz a $|X\mathbf{1}\{X > a\}|$ se obtiene que

$$\mathbb{E}[|X\mathbf{1}\{X > a\}|] \leq (\mathbb{E}[X^2]\mathbb{E}[\mathbf{1}\{X > a\}^2])^{1/2} = (\mathbb{E}[X^2]\mathbb{P}(X > a))^{1/2} \quad (25)$$

Observando que $X = X\mathbf{1}\{X > a\} + X\mathbf{1}\{X \leq a\}$ y que $X\mathbf{1}\{X \leq a\} \leq a$ se deduce que

$$\mathbb{E}[X] = \mathbb{E}[X\mathbf{1}\{X > a\}] + \mathbb{E}[X\mathbf{1}\{X \leq a\}] \leq \mathbb{E}[X\mathbf{1}\{X > a\}] + a$$

y en consecuencia,

$$\mathbb{E}[X] - a \leq \mathbb{E}[X\mathbf{1}\{X > a\}]. \quad (26)$$

Combinando las desigualdades (26), (24) y (25) se obtiene que

$$\mathbb{E}[X] - a \leq (\mathbb{E}[X^2]\mathbb{P}(X > a))^{1/2}$$

y como $\mathbb{E}[X] - a > 0$, elevando al cuadrado, se concluye que

$$(\mathbb{E}[X] - a)^2 \leq \mathbb{E}[X^2]\mathbb{P}(X > a).$$

El resultado se obtiene despejando. □

3.2. Chebyshev

Teorema 3.3 (Desigualdad de Chebyshev). *Sea $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ tal que $\varphi \geq 0$ y $A \in \mathcal{B}(\mathbb{R})$. Sea $i_A := \inf\{\varphi(x) : x \in A\}$. Entonces,*

$$i_A \mathbb{P}(X \in A) \leq \mathbb{E}[\varphi(X)] \quad (27)$$

Demostración. La definición de i_A y el hecho de que $\varphi \geq 0$ implican que

$$i_A \mathbf{1}\{X \in A\} \leq \varphi(X) \mathbf{1}\{X \in A\} \leq \varphi(X)$$

El resultado se obtiene tomando esperanza. □

En lo que sigue enunciaremos algunos corolarios que se obtienen como casos particulares del Teorema 3.3.

Corolario 3.4 (Desigualdad de Markov). *Sea X una variable aleatoria a valores no negativos. Para cada $a > 0$ vale que*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \quad (28)$$

Demostración. Aplicar la desigualdad de Chebyshev usando la función $\varphi(x) = x$ restringida a la semi-recta no negativa $[0, \infty)$ y el conjunto $A = [a, \infty)$ para obtener

$$a\mathbb{P}(X \geq a) \leq \mathbb{E}[\varphi(X)] = \mathbb{E}[X].$$

y despejar. □

Corolario 3.5. *Sea $a > 0$. Vale que*

$$\mathbb{P}(X > a) \leq \frac{1}{a^2} \mathbb{E}[X^2]. \quad (29)$$

Demostración. Aplicar la desigualdad de Chebyshev usando la función $\varphi(x) = x^2$ y el conjunto $A = (a, \infty)$ para obtener

$$a^2 \mathbb{P}(X > a) \leq \mathbb{E}[X^2]$$

y despejar. □

Corolario 3.6 (Pequeña desigualdad de Chebyshev). Sea X una variable aleatoria de varianza finita. Para cada $a > 0$ vale que

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}. \quad (30)$$

Demostración. Debido a que $(X - \mathbb{E}[X])^2$ es una variable aleatoria no negativa podemos aplicar la desigualdad de Markov (poniendo a^2 en lugar de a) y obtenemos

$$\mathbb{P}((X - \mathbb{E}[X])^2 \geq a^2) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} = \frac{\mathbb{V}(X)}{a^2}.$$

La desigualdad $(X - \mathbb{E}[X])^2 \geq a^2$ es equivalente a la desigualdad $|X - \mathbb{E}[X]| \geq a$. Por lo tanto,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}.$$

Lo que concluye la demostración. □

Nota Bene. *Grosso modo* la pequeña desigualdad de Chebyshev establece que si la varianza es pequeña, los grandes desvíos respecto de la media son improbables.

Corolario 3.7. Sea X una variable aleatoria con media μ y varianza σ^2 , entonces para cada $\lambda > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \alpha\sigma(X)) \leq \frac{1}{\alpha^2}. \quad (31)$$

El resultado se obtiene poniendo $a = \alpha\sigma(X)$ en la pequeña desigualdad de Chebyshev. □

Ejemplo 3.8. La cantidad X de artículos producidos por una fábrica durante una semana es una variable aleatoria de media 500.

(a) ¿Qué puede decirse sobre la probabilidad de que la producción semanal supere los 1000 artículos? Por la desigualdad de Markov,

$$\mathbb{P}(X \geq 1000) \leq \frac{\mathbb{E}[X]}{1000} = \frac{500}{1000} = \frac{1}{2}.$$

(b) Si la varianza de la producción semanal es conocida e igual a 100, ¿qué puede decirse sobre la probabilidad de que la producción semanal se encuentre entre 400 y 600 artículos? Por la desigualdad de Chebyshev,

$$\mathbb{P}(|X - 500| \geq 100) \leq \frac{\sigma^2}{(100)^2} = \frac{1}{100}.$$

Por lo tanto, $\mathbb{P}(|X - 500| < 100) \geq 1 - \frac{1}{100} = \frac{99}{100}$, la probabilidad de que la producción semanal se encuentre entre 400 y 600 artículos es al menos 0.99. \square

Las desigualdades de Markov y Chebyshev son importantes porque nos permiten deducir cotas sobre las probabilidades cuando solo se conocen la media o la media y la varianza de la distribución de probabilidades.

El que mucho abarca poco aprieta. Es muy importante tener en cuenta que las desigualdades de Markov y de Chebyshev producen cotas universales que no dependen de las distribuciones de las variables aleatorias (dependen pura y exclusivamente de los valores de la esperanza y de la varianza). Por este motivo su comportamiento será bastante heterogéneo: en algunos casos producirán cotas extremadamente finas, pero en otros casos solamente cotas groseras. \square

4. Covarianza y varianza de sumas

4.1. Covarianza

Sean X e Y dos variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$. Entonces, $X + Y$ y XY son variables aleatorias. El objetivo de esta sección es calcular, cuando exista, $\mathbb{V}(X + Y)$. Para ello se introduce la noción de *covarianza*.

Observando que $|xy| \leq \frac{1}{2}(x^2 + y^2)$, se deduce que si las esperanzas $\mathbb{E}[X^2]$ y $\mathbb{E}[Y^2]$ son finitas, entonces también lo es la esperanza del producto $\mathbb{E}[XY]$. En tal caso, también existen las esperanzas $\mathbb{E}[X]$, $\mathbb{E}[Y]$ y las variables $X - \mathbb{E}[X]$ y $Y - \mathbb{E}[Y]$ tienen esperanza cero. La esperanza de su producto $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$ se puede calcular distribuyendo términos y usando la linealidad de la esperanza

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] &= \mathbb{E}[XY - \mathbb{E}[Y]X - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

Definición 4.1 (Covarianza). Sean X e Y dos variables aleatorias de varianzas finitas definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$. La *covarianza* de X e Y se define por

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \quad (32)$$

Sobre la esperanza del producto. Si se conoce la covarianza y la esperanza de las marginales, la identidad (32) puede ser útil para calcular la esperanza del producto:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] + \text{Cov}(X, Y).$$

Nota Bene. Si X e Y son independientes, entonces por el Corolario 1.25 resulta que $\text{Cov}(X, Y) = 0$. Pero la *recíproca no es cierta*. \square

Ejemplo 4.2. Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y sean $A \in \mathcal{A}$ y $B \in \mathcal{A}$ dos eventos de probabilidad positiva. Consideremos las variables aleatorias $X = \mathbf{1}\{\omega \in A\}$ e $Y = \mathbf{1}\{\omega \in B\}$. Entonces,

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{P}(XY = 1) - \mathbb{P}(X = 1)\mathbb{P}(Y = 1) \\ &= \mathbb{P}(X = 1, Y = 1) - \mathbb{P}(X = 1)\mathbb{P}(Y = 1). \end{aligned} \quad (33)$$

La segunda y la tercera igualdad se obtienen de (2) observando que XY es una variable a valores 0 o 1 que vale 1 si y solo si X e Y son ambas 1.

De (33) vemos que

$$\begin{aligned} \text{Cov}(X, Y) > 0 &\iff \mathbb{P}(X = 1, Y = 1) > \mathbb{P}(X = 1)\mathbb{P}(Y = 1) \\ &\iff \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(X = 1)} > \mathbb{P}(Y = 1) \\ &\iff \mathbb{P}(Y = 1|X = 1) > \mathbb{P}(Y = 1). \end{aligned} \quad (34)$$

Esto es, la covarianza de X e Y es positiva si y solamente si el resultado $X = 1$ hace que sea más probable que $Y = 1$. \square

Ejemplo 4.3. En una urna hay 6 bolas rojas y 4 bolas negras. Se extraen 2 bolas al azar sin reposición. Consideramos los eventos

$$A_i = \{\text{sale una bola roja en la } i\text{-ésima extracción}\}, \quad i = 1, 2,$$

y definimos las variables aleatorias X_1 y X_2 como las funciones indicadoras de los eventos A_1 y A_2 respectivamente. De acuerdo con el Ejemplo anterior es intuitivamente claro que $\text{Cov}(X_1, X_2) < 0$. (*¿Por qué?*)

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \mathbb{P}(X_1 = 1, X_2 = 1) - \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) \\ &= \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2) \\ &= \frac{6}{10} \times \frac{5}{9} - \frac{6}{10} \left(\frac{5}{9} \times \frac{6}{10} + \frac{6}{9} \times \frac{4}{10} \right) \\ &= -\frac{2}{75} = -0.02666\dots \end{aligned}$$

\square

Nota Bene. En general se puede mostrar que $Cov(X, Y) > 0$ es una indicación de que Y tiende a crecer cuando X lo hace, mientras que $Cov(X, Y) < 0$ es una indicación de que Y decrece cuando X crece. \square

Ejemplo 4.4 (Dos bolas en dos urnas). El experimento aleatorio consiste en ubicar dos bolas en dos urnas. El espacio muestral puede representarse de la siguiente manera

$$\Omega = \left\{ \underline{|b_1 b_2| - |}; \underline{|-| b_1 b_2|}; \underline{|b_1| b_2|}; \underline{|b_2| b_1|} \right\}$$

Sean N la cantidad de urnas ocupadas y X_i la cantidad de bolas en la urna i . La función de probabilidad conjunta de N y X_1 se muestra en el Cuadro 1

$N \setminus X_1$	0	1	2	p_N
1	1/4	0	1/4	1/2
2	0	1/2	0	1/2
p_{X_1}	1/4	1/2	1/4	

Cuadro 1: Función de probabilidad conjunta de (N, X_1) .

La esperanza del producto NX_1 se calcula usando el Teorema 1.21

$$\begin{aligned} \mathbb{E}[NX_1] &= \sum_{n=1}^2 \sum_{x_1=0}^3 nx_1 p_{N, X_1}(n, x_1) = \sum_{n=1}^2 \sum_{x_1=1}^3 nx_1 p_{N, X_1}(n, x_1) \\ &= 1 \cdot 1 \cdot p_{N, X_1}(1, 1) + 1 \cdot 2 \cdot p_{N, X_1}(1, 2) + 2 \cdot 1 \cdot p_{N, X_1}(2, 1) + 2 \cdot 2 \cdot p_{N, X_1}(2, 2) \\ &= 1 \cdot 0 + 2 \cdot 1/4 + 2 \cdot 1/2 + 4 \cdot 0 = 3/2. \end{aligned}$$

Es fácil ver que $\mathbb{E}[N] = 3/2$ y $\mathbb{E}[X_1] = 1$. Por lo tanto, $Cov(N, X_1) = 0$. Sin embargo, es claro que las variables N y X_1 no son independientes. \square

Lema 4.5 (Propiedades). Para variables aleatorias X, Y, Z y constantes a , valen las siguientes propiedades

1. $Cov(X, X) = \mathbb{V}(X)$,
2. $Cov(X, Y) = Cov(Y, X)$,
3. $Cov(aX, Y) = aCov(X, Y)$,
4. $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$.

Demostración. *Ejercicio.* \square

4.2. Varianza de sumas

Usando las propiedades de la covarianza enunciadas en Lema 4.5 se puede demostrar que

$$Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m Cov(X_i, Y_j) \quad (35)$$

En particular, se obtiene que

$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n \mathbb{V}(X_i) + 2 \sum_{i=1}^n \sum_{j<i}^n \text{Cov}(X_i, X_j). \quad (36)$$

Finalmente, si las variables son independientes

$$\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i). \quad (37)$$

5. La ley débil de los grandes números

Teorema 5.1 (Ley débil de los grandes números para variables aleatorias i.i.d.). *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes idénticamente distribuidas, tales que $\mathbb{V}(X_1) < \infty$. Sea S_n , $n \geq 1$, la sucesión de las sumas parciales definida por $S_n := \sum_{i=1}^n X_i$. Entonces, para cualquier $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}[X_1]\right| > \epsilon\right) = 0.$$

Demostración. Se obtiene aplicando la desigualdad de Chebyshev a la variable aleatoria S_n/n . Usando que la esperanza es un operador lineal se obtiene que

$$\mathbb{E}[S_n/n] = \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X_1].$$

Como las variables X_1, X_2, \dots son independientes tenemos que

$$\mathbb{V}(S_n/n) = \frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) = \frac{\mathbb{V}(X_1)}{n}.$$

Entonces, por la desigualdad de Chebyshev, obtenemos la siguiente estimación

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}[X_1]\right| > \epsilon\right) \leq \frac{\mathbb{V}(X_1)}{n\epsilon^2}. \quad (38)$$

Como $\mathbb{V}(X_1) < \infty$ el lado derecho de la última desigualdad tiende a 0 cuando $n \rightarrow \infty$. \square

Nota Bene. La ley débil de los grandes números establecida en el Teorema 5.1 sirve como base para la noción intuitiva de probabilidad como medida de las frecuencias relativas. La proposición “en una larga serie de ensayos idénticos la frecuencia relativa del evento A se aproxima a su probabilidad $\mathbb{P}(A)$ ” se puede hacer teóricamente más precisa de la siguiente manera: el resultado de cada ensayo se representa por una variable aleatoria (independiente

de las demás) que vale 1 cuando se obtiene el evento A y vale cero en caso contrario. La expresión “una larga serie de ensayos” adopta la forma de una sucesión X_1, X_2, \dots de variables aleatorias independientes cada una con la misma distribución que la indicadora del evento A . Notar que $X_i = 1$ significa que “en el i -ésimo ensayo ocurrió el evento A ” y la suma parcial $S_n = \sum_{i=1}^n X_i$ representa la “frecuencia del evento A ” en los primeros n ensayos. Puesto que $\mathbb{E}[X_1] = \mathbb{P}(A)$ y $\mathbb{V}(X_1) = \mathbb{P}(A)(1 - \mathbb{P}(A))$ la estimación (38) adopta la forma

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{P}(A)\right| > \epsilon\right) \leq \frac{\mathbb{P}(A)(1 - \mathbb{P}(A))}{n\epsilon^2}. \quad (39)$$

Por lo tanto, *la probabilidad de que la frecuencia relativa del evento A se desvíe de su probabilidad $\mathbb{P}(A)$ en más de una cantidad prefijada ϵ , puede hacerse todo lo chica que se quiera, siempre que la cantidad de ensayos n sea suficientemente grande.*

Ejemplo 5.2 (Encuesta electoral). Se quiere estimar la proporción del electorado que pretende votar a un cierto candidato. Cuál debe ser el tamaño muestral para garantizar un determinado *error* entre la proporción poblacional, p , y la proporción muestral S_n/n ?

Antes de resolver este problema, debemos reflexionar sobre la definición de *error*. Habitualmente, cuando se habla de error, se trata de un número real que expresa la (in)capacidad de una cierta cantidad de representar a otra. En los problemas de estimación estadística, debido a que una de las cantidades es una variable aleatoria y la otra no lo es, no es posible interpretar de un modo tan sencillo el significado de la palabra error.

Toda medida muestral tiene asociada una incerteza (o un riesgo) expresada por un modelo probabilístico. En este problema consideramos que el voto de cada elector se comporta como una variable aleatoria X tal que $\mathbb{P}(X = 1) = p$ y $\mathbb{P}(X = 0) = 1 - p$, donde $X = 1$ significa que el elector vota por el candidato considerado. Por lo tanto, cuando se habla de que queremos encontrar un tamaño muestral suficiente para un determinado error máximo, por ejemplo 0.02, tenemos que hacerlo con una medida de certeza asociada. Matemáticamente, queremos encontrar n tal que $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq 0.02\right) \geq 0.9999$ o, equivalentemente, queremos encontrar n tal que

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > 0.02\right) \leq 0.0001.$$

Usando la estimación (39) se deduce que

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > 0.02\right) \leq \frac{p(1-p)}{n(0.02)^2}.$$

El numerador de la fracción que aparece en el lado derecho de la estimación depende de p y el valor de p es desconocido. Sin embargo, sabemos que $p(1-p)$ es una parábola convexa con raíces en $p = 0$ y $p = 1$ y por lo tanto su máximo ocurre cuando $p = 1/2$, esto es $p(1-p) \leq 1/4$. En la peor hipótesis tenemos:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > 0.02\right) \leq \frac{1}{4n(0.02)^2}.$$

Como máximo estamos dispuestos a correr un riesgo de 0.0001 y en el peor caso tenemos acotada la máxima incerteza por $(4n(0.02)^2)^{-1}$. El problema se reduce a resolver la desigualdad $(4n(0.02)^2)^{-1} \leq 0.0001$. Por lo tanto,

$$n \geq ((0.0001)4(0.02)^2)^{-1} = 6250000.$$

Una cifra absurdamente grande!! Más adelante, mostraremos que existen métodos más sofisticados que permiten disminuir el tamaño de la muestra. \square

6. Distribuciones particulares

Para facilitar referencias posteriores presentaremos tablas de esperanzas y varianzas de algunas distribuciones importantes de uso frecuente y describiremos el método para obtenerlas.

Discretas

No.	Nombre	Probabilidad	Soporte	Esperanza	Varianza
1.	Bernoulli	$p^x(1-p)^{1-x}$	$x \in \{0, 1\}$	p	$p(1-p)$
2.	Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	$0 \leq x \leq n$	np	$np(1-p)$
3.	Geométrica	$(1-p)^{x-1}p$	$x \in \mathbb{N}$	$1/p$	$(1-p)/p^2$
4.	Poisson	$\frac{\lambda^x}{x!}e^{-\lambda}$	$x \in \mathbb{N}_0$	λ	λ

Cuadro 2: Esperanza y varianza de algunas distribuciones discretas de uso frecuente.

Continuas

No.	Nombre	Densidad	Soporte	Esperanza	Varianza
1.	Uniforme	$\frac{1}{b-a}$	$x \in [a, b]$	$(a+b)/2$	$(b-a)^2/12$
2.	Exponencial	$\lambda e^{-\lambda x}$	$x > 0$	$1/\lambda$	$1/\lambda^2$
3.	Gamma	$\frac{\lambda^\nu}{\Gamma(\nu)}x^{\nu-1}e^{-\lambda x}$	$x > 0$	ν/λ	ν/λ^2
4.	Beta	$\frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)}x^{\nu_1-1}(1-x)^{\nu_2-1}$	$x \in (0, 1)$	$\frac{\nu_1}{\nu_1+\nu_2}$	$\frac{\nu_1\nu_2}{(\nu_1+\nu_2)^2(\nu_1+\nu_2+1)}$
5.	Normal	$\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$	$x \in \mathbb{R}$	μ	σ^2

Cuadro 3: Esperanza y varianza de algunas distribuciones continuas de uso frecuente.

Cuentas con variables discretas

1. Distribución Bernoulli.

Sea $p \in (0, 1)$. Se dice que la variable aleatoria X tiene distribución *Bernoulli* de parámetro p , y se denota $X \sim \text{Bernoulli}(p)$, si X es discreta y tal que

$$\mathbb{P}(X = x) = p^x(1 - p)^{1-x}, \text{ donde } x = 0, 1.$$

Por definición,

$$\mathbb{E}[X] = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Por otra parte,

$$\mathbb{E}[X^2] = 0^2 \cdot \mathbb{P}(X = 0) + 1^2 \cdot \mathbb{P}(X = 1) = p.$$

Por lo tanto,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1 - p).$$

□

2. Distribución Binomial.

Sean $p \in (0, 1)$ y $n \in \mathbb{N}$. Se dice que la variable aleatoria X tiene distribución *Binomial* de parámetros n y p , y se denota $X \sim \text{Binomial}(n, p)$, si X es discreta y tal que

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \text{ donde } x = 0, 1, \dots, n.$$

Por definición,

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=0}^n x \mathbb{P}(X = x) = \sum_{x=0}^n x \binom{n}{x} p^x (1 - p)^{n-x} \\ &= \sum_{x=1}^n \frac{xn!}{(n-x)!x!} p^x (1 - p)^{n-x} = \sum_{x=1}^n \frac{n!}{(n-x)!(x-1)!} p^x (1 - p)^{n-x} \\ &= np \sum_{x=1}^n \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} (1 - p)^{n-x} = np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1 - p)^{n-1-y} \\ &= np(p + (1 - p))^{n-1} = np. \end{aligned}$$

Análogamente puede verse que

$$\mathbb{E}[X^2] = np((n-1)p + 1).$$

Por lo tanto,

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = np((n-1)p + 1) - (np)^2 \\ &= np((n-1)p + 1 - np) = np(1 - p). \end{aligned}$$

□

3. Distribución Geométrica.

Sea $p \in (0, 1)$. Se dice que la variable aleatoria X tiene distribución *Geométrica de parámetro p* , y se denota $X \sim \text{Geométrica}(p)$, si X es discreta y tal que

$$\mathbb{P}(X = x) = (1 - p)^{x-1}p, \text{ donde } x \in \mathbb{N} = \{1, 2, 3, \dots\}.$$

Por definición,

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=1}^{\infty} x \mathbb{P}(X = x) = \sum_{x=1}^{\infty} x(1 - p)^{x-1}p \\ &= p \sum_{x=1}^{\infty} x(1 - p)^{x-1}. \end{aligned}$$

La serie se calcula observando que $x(1 - p)^{x-1} = -\frac{d}{dp}(1 - p)^x$ y recordando que las series de potencias se pueden derivar término a término:

$$\begin{aligned} \sum_{x=1}^{\infty} x(1 - p)^{x-1} &= -\frac{d}{dp} \sum_{x=1}^{\infty} (1 - p)^x = -\frac{d}{dp} (p^{-1} - 1) \\ &= p^{-2}. \end{aligned}$$

Por lo tanto,

$$\mathbb{E}[X] = p \left(\frac{1}{p^2} \right) = \frac{1}{p}.$$

Para calcular $\mathbb{V}(X)$ usaremos la misma técnica: derivamos dos veces ambos lados de la igualdad $\sum_{x=1}^{\infty} (1 - p)^{x-1} = p^{-1}$ y obtenemos

$$\begin{aligned} 2p^{-3} &= \frac{d^2}{dp^2} p^{-1} = \frac{d^2}{dp^2} \sum_{x=1}^{\infty} (1 - p)^{x-1} \\ &= \sum_{x=1}^{\infty} (x-1)(x-2)(1 - p)^{x-3} = \sum_{x=1}^{\infty} (x+1)x(1 - p)^{x-1} \\ &= \sum_{x=1}^{\infty} x^2(1 - p)^{x-1} + \sum_{x=1}^{\infty} x(1 - p)^{x-1}. \end{aligned}$$

Multiplicando por p los miembros de las igualdades obtenemos, $2p^{-2} = \mathbb{E}[X^2] + \mathbb{E}[X] = \mathbb{E}[X^2] + p^{-1}$. En consecuencia,

$$\mathbb{E}[X^2] = \frac{2}{p^2} - \frac{1}{p}.$$

Por lo tanto,

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}. \end{aligned}$$

□

4. Distribución de Poisson.

Sea $\lambda > 0$. Se dice que la variable aleatoria X tiene distribución de *Poisson de intensidad* λ , y se denota $X \sim \text{Poisson}(\lambda)$, si X es discreta y tal que

$$\mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \text{ donde } x \in \mathbb{N}_0 = \{0, 1, 2, \dots\}.$$

Por definición,

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=0}^{\infty} x \mathbb{P}(X = x) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

Derivando término a término, se puede ver que

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=0}^{\infty} x^2 \mathbb{P}(X = x) = \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} \frac{d}{d\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \frac{d}{d\lambda} (\lambda e^{\lambda}) \\ &= \lambda e^{-\lambda} (e^{\lambda} + \lambda e^{\lambda}) = \lambda + \lambda^2. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda + \lambda^2 - \lambda^2 \\ &= \lambda. \end{aligned}$$

□

Cuentas con variables continuas

1. Distribución uniforme.

Sean $a < b$. Se dice que la variable aleatoria X tiene distribución *uniforme sobre el intervalo* $[a, b]$, y se denota $X \sim \mathcal{U}[a, b]$, si X es absolutamente continua con densidad de probabilidades

$$f(x) = \frac{1}{b-a} \mathbf{1}\{x \in [a, b]\}.$$

Por definición,

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \frac{1}{b-a} \mathbf{1}\{x \in [a, b]\} dx \\ &= \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left(\frac{b^2 - a^2}{2} \right) \\ &= \frac{a+b}{2}.\end{aligned}$$

Por otra parte,

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{b-a} \int_a^b x^2 dx \\ &= \frac{1}{b-a} \left(\frac{b^3 - a^3}{3} \right) = \frac{a^2 + ab + b^2}{3}.\end{aligned}$$

Finalmente,

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2} \right)^2 \\ &= \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12}.\end{aligned}$$

□

2. Distribución exponencial.

Sea $\lambda > 0$. Se dice que la variable aleatoria X tiene distribución *exponencial de intensidad* λ , y se denota $X \sim \text{Exp}(\lambda)$, si X es absolutamente continua con función densidad de probabilidades

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}\{x \geq 0\}.$$

El cálculo de $\mathbb{E}[X]$ y $\mathbb{V}(X)$ se reduce al caso $X \sim \text{Exp}(1)$. Basta observar que $Y \sim \text{Exp}(\lambda)$ si y solo si $Y = \lambda^{-1}X$, donde $X \sim \text{Exp}(1)$ y usar las identidades $\mathbb{E}[\lambda^{-1}X] = \lambda^{-1}\mathbb{E}[X]$ y $\mathbb{V}(\lambda^{-1}X) = \lambda^{-2}\mathbb{V}(X)$. En lo que sigue suponemos que $X \sim \text{Exp}(1)$.

Integrando por partes se obtiene,

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x e^{-x} \mathbf{1}\{x \geq 0\} \\ &= \int_0^{\infty} \lambda x e^{-x} dx = -x e^{-x} \Big|_0^{\infty} + \int_0^{\infty} e^{-x} dx \\ &= 1.\end{aligned}$$

Por otra parte,

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^{\infty} x^2 e^{-x} dx \\ &= -x^2 e^{-x} \Big|_0^{\infty} + \int_0^{\infty} 2x e^{-x} dx = 2.\end{aligned}$$

Por lo tanto,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2 - 1 = 1.$$

□

3. Distribución gamma.

La *función gamma* se define por

$$\Gamma(t) := \int_0^\infty x^{t-1} e^{-x} dx \quad t > 0.$$

Integrando por partes puede verse que $\Gamma(t) = (t-1)\Gamma(t-1)$ para todo $t > 0$. De aquí se deduce que la función gamma interpola a los números factoriales en el sentido de que

$$\Gamma(n+1) = n! \quad \text{para} \quad n = 0, 1, \dots$$

Sean $\lambda > 0$ y $\nu > 0$. Se dice que la variable aleatoria X tiene distribución *gamma de parámetros* ν, λ , y se denota $X \sim \Gamma(\nu, \lambda)$, si X es absolutamente continua con función densidad de probabilidades

$$f(x) = \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x} \mathbf{1}\{x > 0\}.$$

El cálculo de $\mathbb{E}[X]$ y $\mathbb{V}(X)$ se reduce al caso $X \sim \Gamma(\nu, 1)$. Para ello, basta observar que $Y \sim \Gamma(\nu, \lambda)$ si y solo si $Y = \lambda^{-1}X$, donde $X \sim \Gamma(\nu, 1)$ y usar las identidades $\mathbb{E}[\lambda^{-1}X] = \lambda^{-1}\mathbb{E}[X]$ y $\mathbb{V}(\lambda^{-1}X) = \lambda^{-2}\mathbb{V}(X)$. En lo que sigue suponemos que $X \sim \Gamma(\nu, 1)$

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty x f(x) dx = \int_0^\infty \frac{1}{\Gamma(\nu)} x^\nu e^{-x} dx \\ &= \frac{1}{\Gamma(\nu)} \Gamma(\nu+1) = \nu. \end{aligned}$$

Del mismo modo se puede ver que

$$\mathbb{E}[X^2] = (\nu+1)\nu = \nu^2 + \nu.$$

Por lo tanto,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \nu.$$

□

4. Distribución beta

Sean $\nu_1 > 0$ y $\nu_2 > 0$. Se dice que la variable aleatoria X tiene distribución *beta* de parámetros ν_1, ν_2 , y se denota $X \sim \beta(\nu_1, \nu_2)$, si X es absolutamente continua con función densidad de probabilidades

$$f(x) = \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} x^{\nu_1-1} (1-x)^{\nu_2-1} \mathbf{1}\{x \in (0, 1)\}.$$

Por definición,

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} x^{\nu_1-1} (1-x)^{\nu_2-1} \mathbf{1}\{x \in (0, 1)\} dx \\ &= \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \int_0^1 x^{\nu_1} (1-x)^{\nu_2-1} dx = \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \frac{\Gamma(\nu_1 + 1)\Gamma(\nu_2)}{\Gamma(\nu_1 + \nu_2 + 1)} \\ &= \frac{\nu_1}{\nu_1 + \nu_2} \end{aligned}$$

Por otra parte,

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \int_0^1 x^{\nu_1+1} (1-x)^{\nu_2-1} dx \\ &= \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \frac{\Gamma(\nu_1 + 2)\Gamma(\nu_2)}{\Gamma(\nu_1 + \nu_2 + 2)} = \frac{\nu_1(\nu_1 + 1)}{(\nu_1 + \nu_2)(\nu_1 + \nu_2 + 1)} \end{aligned}$$

Finalmente,

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{\nu_1(\nu_1 + 1)}{(\nu_1 + \nu_2)(\nu_1 + \nu_2 + 1)} - \left(\frac{\nu_1}{\nu_1 + \nu_2} \right)^2 \\ &= \frac{\nu_1 \nu_2}{(\nu_1 + \nu_2)^2 (\nu_1 + \nu_2 + 1)}. \end{aligned}$$

□

5. Distribución normal.

Sean $\mu \in \mathbb{R}$ y $\sigma > 0$. Se dice que la variable aleatoria X tiene distribución *normal* de parámetros μ, σ^2 , y se denota $X \sim \mathcal{N}(\mu, \sigma^2)$, si X es absolutamente continua con función densidad de probabilidades

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

El cálculo de $\mathbb{E}[X]$ y $\mathbb{V}(X)$ se reduce al caso $X \sim \mathcal{N}(0, 1)$. Para ello, basta observar que $Y \sim \mathcal{N}(\mu, \sigma^2)$ si y solo si $Y = \sigma X + \mu$, donde $X \sim \mathcal{N}(0, 1)$ y usar las identidades $\mathbb{E}[\sigma X + \mu] = \sigma \mathbb{E}[X] + \mu$ y $\mathbb{V}(\sigma X + \mu) = \sigma^2 \mathbb{V}(X)$. En lo que sigue suponemos que $X \sim \mathcal{N}(0, 1)$ y denotamos su densidad mediante

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

Es evidente que $\mathbb{E}[X] = 0$. En consecuencia,

$$\mathbb{V}(X) = \mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \varphi(x) dx$$

Observando que $\varphi'(x) = -x\varphi(x)$ e integrando por partes se obtiene,

$$\mathbb{V}(X) = \int_{-\infty}^{\infty} x(x\varphi(x)) dx = -x\varphi(x) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \varphi(x) dx = 0 + 1.$$

□

7. Bibliografía consultada

Para redactar estas notas se consultaron los siguientes libros:

1. Bertsekas, D. P., Tsitsiklis, J. N.: Introduction to Probability. M.I.T. Lecture Notes. (2000)
2. Billingsley, P.: Probability and Measure. John Wiley & Sons, New York. (1986)
3. Durrett, R. Elementary Probability for Applications. Cambridge University Press, New York. (2009)
4. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 1. John Wiley & Sons, New York. (1957)
5. Kolmogorov, A. N.: The Theory of Probability. Mathematics. Its Content, Methods, and Meaning. Vol 2. The M.I.T. Press, Massachusetts. (1963) pp. 229-264.
6. Ross, S.: Introduction to Probability and Statistics for Engineers and Scientists. Academic Press, San Diego. (2004)
7. Ross, S.: Introduction to Probability Models. Academic Press, San Diego. (2007)
8. Soong, T. T.: Fundamentals of Probability and Statistics for Engineers. John Wiley & Sons Ltd. (2004)