Capítulo 5

Esperanza Matemática

5.1. Esperanza Matemática de Variables Aleatorias Discretas.

Recordemos que una variable aleatoria X es discreta, si existe una sucesión $(x_n)_{n\geq 1}$ de números reales tales que

$$\sum_{n=1}^{\infty} P(X = x_n) = 1.$$

Comenzamos por definir la noción de valor esperado para variables aleatorias discretas.

Definición 5.1 Sea X una variable aleatoria discreta con la notación anterior, y llamemos $p_n = P(X = x_n),\ n = 1, 2, \ldots$. Diremos que existe el *valor esperado*, la *media* o la *esperanza matemática* de X si la serie

$$\sum_{n=1}^{\infty} |x_n| p_n \tag{5.1}$$

es convergente. En este caso, el valor esperado se denota $\mathrm{E}(X)$ y se define mediante la serie

$$\mathrm{E}(X) = \sum_{n=1}^{\infty} x_n p_n. \tag{5.2}$$

Ejemplo 5.1
Sea X el resultado de lanzar un dado, entonces X toma valores $\{1, 2, 3, 4, 5, 6\}$ con probabilidad uniforme en este conjunto. Por lo tanto

$$\mathrm{E}(X) = \sum_{i=1}^{6} i\frac{1}{6} = \frac{21}{6} = 3.5$$

Observamos que en este caso el valor esperado no es un valor posible de la variable aleatoria.

5.1.1. Propiedades

La esperanza matemática de una variable discreta tiene las siguientes propiedades.

Propiedad 1. Si $X \geq 0$ y existe $\mathrm{E}(X)$, entonces $\mathrm{E}(X) \geq 0$.

Es obvio que si $X \geq 0$, los valores x_n que figuran en la suma (5.1) son no-negativos, y si dicha serie es convergente, la suma también será no-negativa.

Propiedad 2. Si X es una variable aleatoria acotada entonces existe $\mathrm{E}(X)$.

Decir que la variable aleatoria X es acotada es decir que existe una constante C tal que

$$|x_n| \leq C \quad \text{para todo } n,$$

y, por lo tanto,

$$\sum_{n=1}^{N} |x_n| p_n \leq C \sum_{n=1}^{N} p_n \leq C.$$

Es decir que las sumas parciales de la serie (5.1) resultan estar acotadas por la constante C. Ahora bien, recordemos que para una serie de términos no-negativos – es el caso de la serie (5.1) – es necesario y suficiente para que converja que sus sumas parciales estén acotadas. Como las de (5.1) lo están, esta serie es convergente y el valor esperado de X existe.

Observación 5.1 Una primera observación es que si una variable aleatoria toma sólamente un número finito de valores, entonces su esperanza matemática está bien definida, ya que (5.1) se reduce a una suma finita.

Por otra parte, no siempre existe el valor esperado de una variable aleatoria. Por ejemplo, consideremos la variable aleatoria X tal que

$$x_n = 2^n \qquad \text{y} \qquad p_n = P(X = 2^n) = \frac{1}{2^n}, \quad n \geq 1.$$

Es claro que

$$\sum_{n=1}^{\infty} p_n = 1.$$

Se tiene que $x_n p_n = 1$, para todo n y por lo tanto, la serie (5.1) es divergente, de modo que esta variable aleatoria no tiene esperanza matemática.

En algunos textos el lector encontrará que, en casos como el de este ejemplo, en el cual la variable aleatoria X es no-negativa y la serie (5.1) diverge, se conviene en definir $\mathrm{E}(X) = +\infty$ en lugar de hacer como hemos optado, esto es, decir que no existe la esperanza.

Se debe hacer especial atención, sin embargo, en que si la variable no es no-negativa (o, en general, de signo constante), no hay esa opción convencional. Por ejemplo, si modificamos el ejemplo considerado y tomamos la variable aleatoria Y tal que

$$P(Y = 2^n) = \frac{1}{2^{n+1}} \qquad P(Y = -2^n) = \frac{1}{2^{n+1}} \quad n \geq 1,$$

entonces, nuevamente (5.1) diverge y no existe $\mathrm{E}(Y)$.

Propiedad 3. Sea A un evento y $\mathbf{1}_A$ la variable aleatoria definida por:

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Entonces
$$\mathrm{E}(\mathbf{1}_A) = P(A).$$

Es claro que $\mathbf{1}_A$ es discreta, ya que toma solamente dos valores y

$$\mathrm{E}(\mathbf{1}_A) = 1 \times P(\mathbf{1}_A = 1) + 0 \times P(\mathbf{1}_A = 0) = P(A).$$

Propiedad 4. Sea $g : \mathbb{R} \to \mathbb{R}$ una función y consideremos la nueva variable aleatoria (discreta)

$$Y = g(X).$$

Entonces, si existe la esperanza matemática de Y, ésta es

$$\mathrm{E}(Y) = \sum_{n=1}^{\infty} g(x_n) p_n. \tag{5.3}$$

Demostración. Llamemos $\{y_m\}$ a los valores de Y, entonces

$$\mathrm{E}(Y) = \sum_m y_m P(Y = y_m). \tag{5.4}$$

Por otra parte el evento $\{Y = y_m\}$ es igual a

$$\{X = x_n \text{ para algún } x_n \text{ tal que } g(x_n) = y_m\},$$

puesto que decir que Y toma el valor y_m equivale a decir que X toma un valor cuya imagen por la función g es y_m. Por lo tanto

$$P(Y = y_m) = \sum_{\{n : g(x_n) = y_m\}} p_n$$

donde el conjunto de los valores de n sobre los que se suma es el conjunto de los valores tales que $g(x_n) = y_m$. Sustituyendo en (5.4)

$$\mathrm{E}(Y) = \sum_m y_m \sum_{\{n : g(x_n) = y_m\}} p_n = \sum_m \sum_{\{n : g(x_n) = y_m\}} g(x_n) p_n$$

y ahora, un instante de reflexión mostrará al lector que la suma doble que aparece en la última igualdad es, simplemente, la suma sobre todos los valores de n, ya que cada n aparece una y sólo una vez allí. En resumen

$$\mathrm{E}(Y) = \sum_n g(x_n) p_n$$

como afirmamos en el enunciado de la propiedad 4. ∎

Observación 5.2 La ventaja que tiene la fórmula (5.3), es que nos permite calcular $\mathrm{E}(Y)$ sin necesidad de calcular previamente la función de probabilidad de la variable aleatoria Y, bastándonos con la función de probabilidad de X.

Propiedad 5. Si existe $\mathrm{E}(X)$ entonces $|\,\mathrm{E}(X)| \leq \mathrm{E}(|X|)$.

Demostración. De acuerdo a la propiedad 4 (tomando $g(x) = |x|$), se tiene

$$\mathrm{E}(|X|) = \sum_n |x_n| p_n \geq |\sum_n x_n p_n| = |\,\mathrm{E}(X)|.$$

Observe que la desigualdad entre las sumas de las series, se deduce simplemente de que

$$\sum_{n=1}^{N} |x_n| p_n \geq |\sum_{n=1}^{N} x_n p_n|$$

en virtud de la desigualdad triangular entre números reales. Pasando al límite cuando $N \to \infty$, se obtiene la desigualdad análoga entre las sumas de las series.

Propiedad 5. Si λ es una constante y $\mathrm{E}(X)$ existe, entonces también existe $\mathrm{E}(\lambda X)$ y

$$\mathrm{E}(\lambda X) = \lambda\, \mathrm{E}(X).$$

La demostración es sencilla y queda a cargo del lector.

Propiedad 7. Si X e Y son variables aleatorias que tienen valor esperado, entonces también existe el valor esperado de $X+Y$ y se tiene

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y).$$

Demostración. Para demostrar esta propiedad utilizamos la notación anteriormente introducida:

$$r_{nm} = P(X = x_n, Y = y_m), \quad p_n = P(X = x_n), \quad q_m = P(Y = y_m),$$

para $n, m = 1, 2, \dots$ ($\{r_{nm}\}$ es la función de probabilidad conjunta de X e Y y $\{p_n\}, \{q_m\}$ las respectivas funciones de probabilidad marginales).

Sea $Z = X + Y$ y $\{z_k\}$ el conjunto de valores posibles de Z. Para ver que la variable aleatoria Z tiene valor esperado, tenemos que probar que

$$\sum_k |z_k| P(Z = z_k) < \infty.$$

Procediendo de manera parecida a lo que hicimos para demostrar la propiedad 4

$$\begin{aligned}\sum_k |z_k| P(Z = z_k) &= \sum_k |z_k| P(X + Y = z_k) \\ &= \sum_k |z_k| \sum_{\{(n,m): x_n + y_m = z_k\}} r_{nm}\end{aligned}$$

donde en la última suma hay que entender que la suma interior se efectúa sobre todas las parejas de índices (n, m) tales que $x_n + y_m = z_k$. La última expresión es:

$$\begin{aligned}\sum_k \sum_{\{(n,m): x_n + y_m = z_k\}} |x_n + y_m| r_{nm} &= \sum_{n,m} |x_n + y_m| r_{nm} \leq \sum_{n,m} (|x_n| + |y_m|) r_{nm} \\ &= \sum_n |x_n| \sum_m r_{nm} + \sum_m |y_m| \sum_n r_{nm} \\ &= \sum_n |x_n| p_n + \sum_m |y_m| q_m < \infty,\end{aligned}$$

ya que, por un lado,

$$\sum_m r_{nm} = p_n, \quad n = 1, 2, \dots; \qquad \sum_n r_{nm} = q_m, \quad m = 1, 2, \dots$$

y por otro, las dos series

$$\sum_n |x_n| p_n; \qquad \sum_m |y_m| q_m$$

son convergentes, dado que existen $\mathrm{E}(X)$ y $\mathrm{E}(Y)$.

Esto prueba que existe $\mathrm{E}(Z)$. Si ahora repetimos el cálculo anterior sin los valores absolutos, resulta

$$\mathrm{E}(Z) = \sum_k z_k P(Z = z_k) = \sum_n x_n p_n + \sum_m y_m q_m = \mathrm{E}(X) + \mathrm{E}(Y),$$

que es la propiedad que queríamos probar. ■

Propiedad 8. Si $X \leq Y$ y existen $\mathrm{E}(X)$ y $\mathrm{E}(Y)$, entonces $\mathrm{E}(X) \leq \mathrm{E}(Y)$.

Demostración. Para probar esta propiedad, recurrimos a las propiedades 1, 6 y 7. Tenemos

$$\mathrm{E}(Y) - \mathrm{E}(X) = \mathrm{E}(Y - X) \geq 0, \quad \text{ya que } Y - X \geq 0.$$

Propiedad 9. Si X e Y son variables aleatorias independientes con valor esperado, entonces existe $\mathrm{E}(XY)$ y

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y). \tag{5.5}$$

Demostración. Procedemos de manera análoga, nuevamente, a lo que hicimos para la propiedad 7, teniendo en cuenta además que la independencia de X e Y implica que

$$r_{nm} = P(X = x_n, Y = y_m) = P(X = x_n)P(Y = y_m) = p_n q_m,$$

y por lo tanto, si llamamos ahora $\{w_k\}$ a los valores de XY

$$\begin{aligned}\sum_k |w_k| P(XY = w_k) &= \sum_k \sum_{\{(n,m):x_n y_m = w_k\}} |w_k| r_{nm} \\ &= \sum_k \sum_{\{(n,m):x_n y_m = w_k\}} |x_n y_m| p_n q_m \\ &= \sum_{n,m} |x_n||y_m| p_n q_m = \sum_n |x_n| p_n \sum_m |y_m| q_m < \infty\end{aligned}$$

lo cual muestra que existe $\mathrm{E}(XY)$.

El mismo cálculo, quitando los valores absolutos, permite obtener la relación (5.5). ■

5.2. Momentos de Orden Superior. Momentos Centrados.

Definición 5.2 Sea X una variable aleatoria discreta,

$$P(X = x_n) = p_n, \qquad \sum_n p_n = 1$$

y m un número positivo. Si

$$\sum_n |x_n|^m p_n < \infty,$$

se define el *momento de orden m de X* mediante

$$\mathrm{E}(X^m) = \sum_n x_n^m p_n.$$

(Ver la propiedad 4, con $g(x) = x^m$).

Se define también el *momento centrado de orden m de X* para $m \geq 2$ mediante

$$\mu_m = \mathrm{E}((X - \mathrm{E}(X))^m) = \sum_n (x_n - \mathrm{E}(X))^m p_n.$$

(Ver la propiedad 4, con $g(x) = (x - \mathrm{E}(X))^m$. En la definición se sobrentiende que existe la esperanza de la variable aleatoria X).

Se acostumbra denominar *varianza* ($\mathrm{Var}(X)$) al momento centrado de orden 2 – cuando existe – es decir

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2) \tag{5.6}$$

y *desviación típica o estándar* de X a $(\mathrm{Var}(X))^{1/2}$.

También se define la *covarianza* de dos variables aleatorias mediante

$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))], \tag{5.7}$$

siempre que la esperanza que figura en el segundo miembro esté bien definida, y el *coeficiente de correlación* de X e Y por

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}, \tag{5.8}$$

siempre que el segundo miembro tenga sentido.

Finalmente, se dice que X e Y no están correlacionadas, son *no-correlacionadas* o *incorreladas* si $\mathrm{Cov}(X, Y) = 0$.

5.2.1. Propiedades

Veamos a continuación algunas propiedades de estos momentos.

Propiedad 1. Si a es una constante y X una variable aleatoria que tiene varianza, entonces

$$\mathrm{Var}(X) = \mathrm{Var}(X + a)$$

y

$$\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X).$$

La demostración resulta de aplicar la definición y queda como ejercicio para el lector.

Propiedad 2. $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$.

Demostración. Observar simplemente que

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

Propiedad 3. (Desigualdad de Cauchy-Schwarz)

$$|\,\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}.$$

Demostración. Para demostrar esta desigualdad, consideremos la función $f : \mathbb{R} \to \mathbb{R}$ definida por

$$f(x) = \mathrm{E}[(U - xV)^2]\ ,$$

donde $U = X - \mathrm{E}(X)$ y $V = Y - \mathrm{E}(Y)$. Usando las propiedades de la esperanza matemática

$$f(x) = \mathrm{E}[U^2 - 2xUV + x^2V^2] = \mathrm{E}(U^2) - 2x\,\mathrm{E}(UV) + x^2\,\mathrm{E}(V^2)$$

de modo que f es un polinomio de segundo grado, y como

$$f(x) \geq 0 \quad \text{para todo } x \in \mathbb{R}$$

porque $(U - xV)^2 \geq 0$ para todo $x \in \mathbb{R}$, se deduce que el discriminante de este polinomio es menor o igual que cero (ya que, de no ser así, habría valores de $x \in \mathbb{R}$ donde $f(x)$ sería negativa). Es decir que

$$(-2\,\mathrm{E}(UV))^2 - 4\,\mathrm{E}(V^2)\,\mathrm{E}(U^2) \leq 0 \Rightarrow (\mathrm{E}(UV))^2 \leq \mathrm{E}(U^2)\,\mathrm{E}(V^2).$$

Tomando en cuenta quienes son U y V y las definiciones de varianza y covarianza, resulta la desigualdad que queríamos probar. ■

Propiedad 4. Sean X e Y dos variables aleatorias tales que está definido el coeficiente de correlación. Entonces

$$-1 \leq \rho(X, Y) \leq 1 \tag{5.9}$$

Demostración. Es una consecuencia sencilla de la propiedad anterior.

Propiedad 5. Si $\mathrm{Var}(X) = 0$ entonces existe una constante c tal que

$$P(X = c) = 1.$$

(Esta propiedad de que $P(X = c) = 1$ se enuncia a veces diciendo que X es una variable aleatoria "*casi seguramente*" igual a la constante c).

Demostración. Para probar esta propiedad, recordemos que X es una variable aleatoria discreta. (El resultado también será cierto en general, como ocurre con casi todas las propiedades que hemos estudiado, pero eso lo veremos más adelante).

Entonces la variable aleatoria $Y = (X - \mathrm{E}(X))^2$ también es discreta y toma valores mayores o iguales a cero, digamos $\{y_m\}$. Nuestra hipótesis es que

$$\mathrm{E}(Y) = 0.$$

Sea $p_m = P(Y = y_m)$. Si hubiera algún $y_{m_0} > 0$ tal que $p_{m_0} > 0$, se tendría que

$$\mathrm{E}(Y) = \sum_m y_m p_m \geq y_{m_0} p_{m_0} > 0$$

contradiciendo la hipótesis. Por lo tanto no hay un tal y_{m_0}, y si $y_m > 0$ necesariamente $p_m = 0$. En consecuencia

$$P(Y > 0) = \sum_{\{m: y_m > 0\}} p_m = 0,$$

de donde se deduce que

$$P(Y = 0) = P(Y \geq 0) - P(Y > 0) = 1 - 0 = 1.$$

Pero $\{Y = 0\}$ es el mismo evento que $\{X = \mathrm{E}(X)\}$, de modo que

$$P(X = \mathrm{E}(X)) = 1,$$

y la propiedad se cumple, tomando $c = \mathrm{E}(X)$. ■

Observación 5.3 Es obvio que el recíproco de esta propiedad es cierto, ya que si

$$P(X = c) = 1 \quad \text{entonces} \quad \mathrm{E}(X) = c$$

y

$$\mathrm{Var}(X) = \mathrm{E}((X - c)^2) = 0,$$

puesto que

$$P(X - c = 0) = P(X = c) = 1.$$

Propiedad 5. Sean $X_1, X_2, \dots, X_n$ variables aleatorias independientes, cada una de las cuales tiene varianza, entonces la suma también tiene varianza y

$$\operatorname{Var}(X_1 + X_2 + \cdots + X_n) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \cdots + \operatorname{Var}(X_n). \tag{5.10}$$

Demostración. Basta probar la igualdad (5.10) para $n = 2$. Para $n > 2$ se sigue por inducción completa, y queda como ejercicio.

Sea entonces $n = 2$ y pongamos $Y_1 = X_1 - \mathrm{E}(X_1)$, $Y_2 = X_2 - \mathrm{E}(X_2)$. Como X_1, X_2 son variables aleatorias independientes, también lo son Y_1, Y_2 y, en consecuencia, por la propiedad 9 de la sección 5.1.1

$$\mathrm{E}(Y_1 Y_2) = \mathrm{E}(Y_1)\,\mathrm{E}(Y_2) = 0$$

Se deduce que

$$\begin{aligned}
\operatorname{Var}(X_1 + X_2) &= \mathrm{E}[(X_1 + X_2 - \mathrm{E}(X_1 + X_2))^2] = \mathrm{E}[(Y_1 + Y_2)^2] \\
&= \mathrm{E}[Y_1^2 + Y_1^2 + 2Y_1Y_2] = \mathrm{E}(Y_1^2) + \mathrm{E}(Y_2^2) + 2\,\mathrm{E}(Y_1Y_2) \\
&= \mathrm{E}(Y_1^2) + \mathrm{E}(Y_2^2) = \mathrm{E}[(X_1 - \mathrm{E}(X_1))^2] + \mathrm{E}[(X_2 - \mathrm{E}(X_2))^2] \\
&= \operatorname{Var}(X_1) + \operatorname{Var}(X_2)
\end{aligned}$$

que es lo que queríamos probar. ■

Propiedad 7. Sea X una variable aleatoria y supongamos que existe su momento de orden m. Entonces también existe su momento de orden k, para cualquier k, que satisfaga $0 \le k < m$.

Demostración. Para probar que existe el momento de orden k hay que ver que

$$\sum_n |x_n|^k p_n < \infty, \quad \text{donde } p_n = P(X = x_n),\ \sum_n p_n = 1.$$

Pero tenemos la desigualdad, válida para todo número $x \in \mathbb{R}$,

$$|x|^k \le 1 + |x|^m. \tag{5.11}$$

En efecto, si $|x| \le 1$ la desigualdad es cierta ya que, en este caso, el primer miembro es menor o igual que 1 y el segundo es mayor o igual que 1. Por otra parte, si $|x| > 1$, es claro que

$$|x|^k < |x|^m < 1 + |x|^m$$

y también se verifica (5.11). En resumen,

$$\sum_n |x_n|^k p_n \le \sum_n (1 + |x_n|^m) p_n = 1 + \sum_n |x_n|^m p_n < \infty.$$

■

Propiedad 8. Sean X e Y variables aleatorias con covarianza $\operatorname{Cov}(X, Y)$, entonces

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

Propiedad 9. Sean X e Y variables aleatorias con covarianza $\operatorname{Cov}(X, Y)$, entonces

$$\operatorname{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y).$$

Las demostraciones quedan como ejercicio.

5.3. Ejemplos y Aplicaciones.

1. *Distribución Binomial.* Recordemos que la distribución binomial es la distribución de la variable S_n que representa el número de veces que ocurre un cierto evento A en n observaciones independientes del mismo. Por ejemplo, el número de veces que extraemos un objeto defectuoso de una cierta población de objetos fabricados en una línea de producción, cuando la extracción se hace al azar y con reposición.

   Definimos las variables aleatorias

   $$X_i = \begin{cases} 1 & \text{si en la } i\text{-ésima observación ocurre } A, \\ 0 & \text{si en la } i\text{-ésima observación no ocurre } A. \end{cases}$$

   que admitimos que son independientes en nuestro modelo. Entonces

   $$S_n = X_1 + X_2 + \cdots + X_n.$$

   Si denotamos $p = P(A)$, entonces

   $$P(X_i = 1) = p \qquad i = 1, 2, \dots$$

   y

   $$\mathrm{E}(X_i) = 1 \times P(X_i = 1) + 0 \times P(X_i = 0) = p$$

   Se deduce que

   $$\mathrm{E}(S_n) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = np.$$

   Además, en virtud de que $X_1, X_2, \dots, X_n$ son variables aleatorias independientes

   $$\mathrm{Var}(S_n) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

   y

   $$\mathrm{Var}(X_i) = \mathrm{E}(X_i^2) - (\mathrm{E}(X_i))^2 = p - p^2 = p(1-p),$$

   ya que $X_i^2 = X_i$ puesto que $X_i = 0$ ó 1. Reemplazando en la igualdad anterior, resulta

   $$\mathrm{Var}(S_n) = np(1-p)$$

   También podemos calcular $\mathrm{E}(S_n)$ y $\mathrm{Var}(S_n)$ directamente, ya que conocemos la función de probabilidad de S_n, que es binomial $b(n, p)$, es decir que

   $$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n$$

   Por lo tanto, aplicando la definición de esperanza matemática en forma directa,

   $$\begin{aligned} \mathrm{E}(S_n) &= \sum_{k=0}^{n} k P(S_n = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np. \end{aligned}$$

Observamos que la última suma es igual a 1 ya que no es otra cosa que la suma de todas las probabilidades ($j = 0, 1, \dots, n-1$) correspondientes a una distribución binomial con $(n-1)$ observaciones. O, también, es el desarrollo de Newton del binomio $(p + 1 - p)^{n-1} = 1$.

Con un cálculo parecido también reencontramos el resultado que hemos obtenido para $\mathrm{Var}(S_n)$; observemos primero que

$$\mathrm{Var}(S_n) = \mathrm{E}(S_n^2) - (\mathrm{E}(S_n))^2 = \mathrm{E}(S_n(S_n-1)) + \mathrm{E}(S_n) - (\mathrm{E}(S_n))^2$$

y entonces

$$\begin{aligned}
\mathrm{E}(S_n(S_n-1)) &= \sum_{k=0}^{n} k(k-1)P(S_n = k) \\
&= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(n-k)!(k-2)!} p^{k-2}(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} = n(n-1)p^2.
\end{aligned}$$

La última suma nuevamente vale 1, por un argumento análogo al ya empleado. Reemplazando en la igualdad anterior

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

2. *Distribución de Poisson.* Recordemos que la variable aleatoria X tiene distribución de Poisson con parámetro $\lambda > 0$ si

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots$$

Entonces

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} = \lambda,
\end{aligned}$$

donde hemos empleado el desarrollo de Taylor de la función exponencial $e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$. Por lo tanto

$$\mathrm{E}(X) = \lambda.$$

Además

$$\begin{aligned}
\mathrm{E}(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\
&= \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2
\end{aligned}$$

y en consecuencia

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \mathrm{E}(X(X-1)) + \mathrm{E}(X) - (\mathrm{E}(X))^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

Es decir
$$\mathrm{Var}(X) = \lambda.$$

3. *Distribución Geométrica.* Sea X una variable aleatoria con función de probabilidad

$$P(X = n) = p(1-p)^{n-1} \qquad n = 1, 2, \dots$$

con $0 < p < 1$. Entonces

$$\mathrm{E}(X) = \sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} np(1-p)^{n-1}. \tag{5.12}$$

Para calcular esta suma, recordemos que si

$$\sum_{n=0}^{\infty} c_n z^n$$

es una serie de potencias y R es su radio de convergencia (lo cual significa que si $|z| < R$ la serie converge y que, cuando $R < \infty$, si $|z| > R$ la serie no converge), si denotamos por $f(z)$ la suma de la serie para $|z| < R$:

$$f(z) = \sum_{n=0}^{\infty} c_n z^n \qquad (|z| < R),$$

entonces f es derivable y

$$f'(z) = \sum_{n=0}^{\infty} nc_n z^{n-1} \qquad (|z| < R).$$

Lo cual significa que, para $|z| < R$, podemos calcular la derivada de la función $f(z)$ como si fuera una suma finita, es decir, sumando las derivadas de cada uno de los términos.

Apliquemos este resultado a la serie de potencias

$$\sum_{n=0}^{\infty} z^n.$$

Primero, se verifica que el radio de convergencia de esta serie es igual a 1. Esto es directo: si $|z| > 1$ entonces $|z|^n \to \infty$, es decir que el término general no tiende a cero y la serie no converge, y si $|z| < 1$, la suma parcial de orden N es

$$\sum_{n=0}^{N} z^n = 1 + z + z^2 + \cdots + z^N = \frac{1 - z^{N+1}}{1 - z} \overset{N\to\infty}{\longrightarrow} \frac{1}{1-z}$$

es decir que la serie converge cuando $|z| < 1$ y su suma es $\frac{1}{1-z}$, o sea

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z} \qquad |z| < 1.$$

Derivando término a término como hemos indicado anteriormente, se tiene

$$\frac{1}{(1-z)^2} = \sum_{n=0}^{\infty} nz^{n-1} = \sum_{n=1}^{\infty} nz^{n-1} \qquad |z| < 1.$$

Volviendo ahora al problema que estábamos considerando, reemplazando z por $1-p$ en la ecuación (5.12) resulta

$$\mathrm{E}(X) = p\sum_{n=1}^{\infty} n(1-p)^{n-1} = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

Para calcular la varianza de esta distribución comenzamos por la relación

$$\operatorname{Var}(X) = \mathrm{E}(X(X-1)) + \mathrm{E}(X) - (\mathrm{E}(X))^2, \tag{5.13}$$

y calculamos

$$\begin{aligned}\mathrm{E}(X(X-1)) &= \sum_{n=1}^{\infty} n(n-1)P(X=n) = \sum_{n=1}^{\infty} n(n-1)p(1-p)^{n-1}\\ &= p(1-p)\sum_{n=2}^{\infty} n(n-1)(1-p)^{n-2}.\end{aligned}$$

Ahora bien, para calcular $f''(z)$ procedemos como antes, derivando término a término la serie que define $f'(z)$. Es decir, que para $|z|<1$,

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n \Rightarrow \frac{1}{(1-z)^2} = \sum_{n=1}^{\infty} nz^{n-1} \Rightarrow \frac{2}{(1-z)^3} = \sum_{n=2}^{\infty} n(n-1)z^{n-2}.$$

En consecuencia, reemplazando z por $1-p$, resulta

$$\mathrm{E}(X(X-1)) = p(1-p)\frac{2}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}$$

y volviendo a (5.13)

$$\operatorname{Var}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

4. **Muestreo de poblaciones finitas.** Se considera una población dividida en r grupos dos a dos disjuntos (llamados “estratos”) cuyas proporciones respectivas con relación al total de los miembros de la población son $q_1, q_2, \dots, q_r$. Se extrae una muestra de tamaño n de la población, al azar y con reposición.

   a. Calcular el valor esperado del número de estratos no representados en la muestra.

   b. Hacer el cálculo efectivo de dicho valor esperado, en los siguientes casos numéricos, suponiendo que todos los estratos son del mismo tamaño:

   i) $r = n = 5$.

   ii) $r = n = 100$.

► a. Consideremos las variables aleatorias $X_1, X_2, \dots, X_r$ definidas por

$$X_i = \begin{cases} 1 & \text{si el estrato } i \text{ no está representado en la muestra,}\\ 0 & \text{si el estrato } i \text{ está representado en la muestra.}\end{cases}$$

Entonces, el número de estratos no representados en la muestra es, evidentemente,

$$X_1 + X_2 + \cdots + X_r$$

y su esperanza matemática

$$\mathrm{E}(X_1 + X_2 + \cdots + X_r) = \sum_{i=1}^{r} \mathrm{E}(X_i) = \sum_{i=1}^{r} P(X_i = 1)$$

ya que

$$\mathrm{E}(X_i) = 1 \times P(X_i = 1) + 0 \times P(X = 0).$$

Ahora bien, ¿cuánto vale $P(X_i = 1)$? El evento $\{X_i = 1\}$ sucede cuando el i-ésimo estrato no está representado en la muestra, es decir, que ninguno de los n individuos seleccionados pertenece al estrato i. En otros términos

$$\begin{aligned}\{X_i = 1\} = &\{1^{er} \text{ individuo no pertenece al estrato } i\} \\ &\cap \{2^{o} \text{ individuo no pertenece al estrato } i\} \\ &\cdots \cap \{n\text{-ésimo individuo no pertenece al estrato } i\}.\end{aligned}$$

Como el muestreo es con reposición, los n eventos que figuran en esta intersección son independientes y, en consecuencia, la probabilidad de su intersección es igual al producto de sus probabilidades:

$$P(X_i = 1) = \prod_{j=1}^{n} P(j\text{-ésimo individuo no pertenece al estrato } i).$$

Además,

$$P(j\text{-ésimo individuo no pertenece al estrato } i) = 1 - P(j\text{-ésimo individuo pertenece al estrato } i)$$

y esta última probabilidad es q_i, es decir, la proporción de individuos que contiene el estrato i con relación al total. Reemplazando obtenemos

$$P(X_i = 1) = (1 - q_i)^n$$

y

$$\mathrm{E}(X_1 + X_2 + \cdots + X_r) = \sum_{i=1}^{r} (1 - q_i)^n.$$

b. Que todos los estratos son de igual tamaño implica que

$$q_1 = q_2 = \cdots = q_r$$

y como

$$\sum_{i=1}^{r} q_i = 1,$$

se deduce que

$$q_i = \frac{1}{r} \qquad (i = 1, 2, \ldots, r).$$

La esperanza matemática del número de estratos no representados en la muestra resulta ser

$$r \left(1 - \frac{1}{r}\right)^n.$$

Si $r = n = 5$, esto es $5\left(\frac{4}{5}\right)^5 \approx 1.64$, mientras que si $r = n = 100$ el resultado es $100\left(1 - \frac{1}{100}\right)^{100} \approx 35.6$.

Es útil ver cómo encontramos un valor aproximado para esta expresión. Usando la fórmula de Taylor para $\log(1-x)$

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots \qquad (|x| < 1)$$

y entonces, si $|x| < 1$

$$(1-x)^{\frac{1}{x}} = e^{\frac{1}{x}\log(1-x)} = e^{\frac{1}{x}(-x-\frac{x^2}{2}-\frac{x^3}{3}-\dots)} = e^{-1-\frac{x}{2}-\frac{x^2}{3}-\dots}$$

que es aproximadamente igual a e^{-1} cuando x es pequeño. Con $x = 1/100$ se tiene que la esperanza matemática es aproximadamente $100e^{-1}$, o sea que esperamos encontrar una proporción $1/e$ de estratos *no* representados en una muestra de $n = r$ individuos. Observe que esto es cierto si $n = r$ es suficientemente grande. ◀

5. Se lanzan dos dados simétricos independientes. Sean X_1, X_2 los resultados obtenidos en cada uno e $Y = \text{máx}\{X_1, X_2\}$.

   a. Hallar la probabilidad conjunta de X_1 e Y.

   b. Calcular $\mathrm{E}(X_1)$, $\mathrm{E}(Y)$, $\mathrm{Var}(X_1)$, $\mathrm{Var}(Y)$ y $\mathrm{Cov}(X_1, Y)$.

► a. Usemos la notación $r_{ij} = P(X_1 = i, Y = j)$ para $i, j = 1, 2, 3, 4, 5, 6$. Es claro que si $i > j$ entonces $r_{ij} = 0$ porque $Y \geq X_1$. Si $i < j$, entonces

$$r_{ij} = P(X_1 = i, X_2 = j) = P(X_1 = i)P(X_2 = j) = \left(\frac{1}{6}\right)^2,$$

mientras que para $i = j$ tenemos

$$r_{ii} = P(X_1 = i, X_2 \leq i) = P(X_1 = i)P(X_2 \leq i) = \frac{1}{6}\frac{i}{6}.$$

Resumiendo

$$r_{ij} = \begin{cases} \frac{1}{36}, & i < j, \\ \frac{i}{36}, & i = j, \\ 0, & i > j. \end{cases}$$

b.

$$\begin{aligned}
\mathrm{E}(X_1) &= \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6}, \\
\mathrm{Var}(X_1) &= \mathrm{E}(X_1^2) - (\mathrm{E}(X_1))^2 \\
&= \frac{1}{6}(1+2^2+3^2+4^2+5^2+6^2) - \left(\frac{21}{6}\right)^2 = \frac{35}{12}.
\end{aligned}$$

Para calcular $\mathrm{E}(Y)$ y $\mathrm{Var}(Y)$, observemos que

$$P(Y = i) = \frac{2i-1}{36} \qquad (i = 1, 2, 3, 4, 5, 6)$$

lo cual puede calcularse directamente o usando la parte (a) de este ejercicio. Por lo tanto

$$\begin{aligned}
\mathrm{E}(Y) &= 1\frac{1}{36} + 2\frac{3}{36} + 3\frac{5}{36} + 4\frac{7}{36} + 5\frac{9}{36} + 6\frac{11}{36} = \frac{161}{36} \\
\mathrm{Var}(Y) &= \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 \\
&= 1\frac{1}{36} + 2^2\frac{3}{36} + 3^2\frac{5}{36} + 4^2\frac{7}{36} + 5^2\frac{9}{36} + 6^2\frac{11}{36} - \left(\frac{161}{36}\right)^2 \\
&= \frac{2555}{(36)^2}.
\end{aligned}$$

Finalmente

$$\operatorname{Cov}(X_1, Y) = \operatorname{E}[(X_1 - \operatorname{E}(X_1))(Y - \operatorname{E}(Y))] = \operatorname{E}(X_1 Y) - \operatorname{E}(X_1)\operatorname{E}(Y).$$

$\operatorname{E}(X_1 Y)$ se calcula utilizando la probabilidad conjunta que encontramos en la parte a.

$$\begin{aligned}
\operatorname{E}(X_1 Y) &= \sum_{i,j=1}^{6} ij r_{ij} = \sum_{i<j} ij \frac{1}{36} + \sum_{i=1}^{6} i^2 \frac{i}{36} \\
&= \frac{1}{36}\left[\sum_{i=1}^{5} i \sum_{j=i+1}^{6} j + \sum_{i=1}^{6} i^3\right] \\
&= \frac{1}{36}\left[\sum_{i=1}^{5} \frac{(6-i)(6+i+1)}{2} i + \sum_{i=1}^{6} i^3\right] = \frac{154}{9}
\end{aligned}$$

lo que implica que $\operatorname{Cov}(X_1, Y) = \frac{35}{24}$. ◀

6. Sean $X_1, X_2, \dots, X_n$ variables aleatorias independientes y supongamos que

$$\operatorname{E}(X_i) = \mu, \qquad \operatorname{Var}(X_i) = \sigma^2, \quad (i = 1, \dots, n).$$

Definimos la media muestral correspondiente a esta muestra por

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

y la varianza muestral por

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Es importante no confundirlas con la media μ y varianza σ^2 de cada una de las observaciones. Estas son constantes, mientras que las primeras son variables aleatorias.

Mostrar que

$$\operatorname{E}(\overline{X}) = \mu; \qquad \operatorname{E}(s^2) = \sigma^2.$$

▶ Comenzamos con

$$\operatorname{E}(\overline{X}) = \operatorname{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\operatorname{E}(X_i) = \mu.$$

Por otro lado

$$\operatorname{E}(s^2) = \operatorname{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\operatorname{E}((X_i - \mu + \mu - \overline{X})^2). \tag{5.14}$$

Calculemos el valor de cada sumando:

$$\operatorname{E}((X_i - \mu)^2) - 2\operatorname{E}((X_i - \mu)(\overline{X} - \mu)) + \operatorname{E}((\overline{X} - \mu)^2) \tag{5.15}$$

y ahora cada uno de estos términos:

$\mathrm{E}((X_i-\mu)^2)$ es la varianza de X_i, es decir, σ^2.

$$\begin{aligned}\mathrm{E}((X_i-\mu)(\overline{X}-\mu)) &= \mathrm{E}((X_i-\mu)(\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)))\\ &= \mathrm{E}((X_i-\mu)\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu))\\ &= \frac{1}{n}\,\mathrm{E}(\sum_{j=1}^{n}(X_i-\mu)(X_j-\mu))\\ &= \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}((X_i-\mu)(X_j-\mu)).\end{aligned}$$

En esta suma, si $i\neq j$, como $X_i,\ X_j$ son variables aleatorias independientes, tenemos

$$\mathrm{E}((X_i-\mu)(X_j-\mu)) = \mathrm{E}(X_i-\mu)\,\mathrm{E}(X_j-\mu)=0,$$

y por lo tanto

$$\mathrm{E}((X_i-\mu)(\overline{X}-\mu)) = \frac{1}{n}\,\mathrm{E}((X_i-\mu)^2)=\frac{\sigma^2}{n}.$$

Veamos a continuación el tercer término de (5.15):
$\mathrm{E}((\overline{X}-\mu)^2)$ es la varianza de $\overline{X}$, ya que

$$\mathrm{E}(\overline{X}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n}X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{n\mu}{n}=\mu.$$

Pero, por otra parte,

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n}X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n},$$

ya que $X_1,\dots,X_n$ son variables aleatorias independientes.
En consecuencia, resulta que la suma en (5.15) es igual a

$$\sigma^2-\frac{2\sigma^2}{n}+\frac{\sigma^2}{n}=\sigma^2\frac{n-1}{n},$$

y reemplazando en la igualdad (5.14)

$$\mathrm{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right) = \frac{1}{n-1}n\sigma^2\frac{n-1}{n}=\sigma^2.$$

◀

Observación 5.4 Es fácil verificar que todo el cálculo realizado en este ejercicio, y por lo tanto la conclusión, permanece válido si las variables aleatorias $X_1,\dots,X_n$ no están correlacionadas, en lugar de exigirse que sean independientes.

7. Dar un ejemplo de dos variables aleatorias X e Y que no sean independientes, pero que, sin embargo,

$$\mathrm{E}(XY)=\mathrm{E}(X)\,\mathrm{E}(Y),$$

lo cual implica que no están correlacionadas.

► Sea X una variable aleatoria que toma los valores $1, 0$ y -1, cada uno con probabilidad $1/3$, y definamos $Y = X^2$.

Es obvio que X e Y no son variables aleatorias independientes, ya que si conocemos el valor de X, también conocemos el valor de Y. Más precisamente, por ejemplo,

$$P(X = 1, Y = 1) = P(X = 1),$$

debido a que el evento $\{X = 1\}$ está contenido en el evento $\{Y = 1\}$. En consecuencia

$$P(X = 1, Y = 1) = \frac{1}{3} \neq P(X = 1)P(Y = 1) = \frac{1}{3} \times \frac{2}{3}$$

ya que

$$P(Y = 1) = P(X = 1) + P(X = -1) = \frac{2}{3}.$$

Sin embargo, X e Y cumplen que

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y).$$

Calculemos ambos miembros y veamos que dan lo mismo. El segundo es evidentemente igual a cero, ya que

$$\mathrm{E}(X) = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + (-1) \times \frac{1}{3} = 0.$$

En cuanto al primero, obervamos que $XY = X^3 = X$, ya que como X vale $1, 0$ ó -1, X^3 es siempre igual a X. En conclusión

$$\mathrm{E}(XY) = \mathrm{E}(X^3) = \mathrm{E}(X) = 0.$$

◄

Observación 5.5 En el enunciado se señala que si $E(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$, entonces X e Y son variables aleatorias no-correlacionadas. Esto se deduce de la Propiedad 9 de la sección 5.1.1.

5.4. Esperanza Matemática de Variables Aleatorias con Densidad.

Definición 5.3 Sea X una variable aleatoria, F_X su función de distribución y supongamos que ésta posee una densidad f_X. El *valor esperado* o *esperanza matemática* de X existe y es finito si y sólo si la integral

$$\int_{-\infty}^{\infty} |x| f_X(x)dx \tag{5.16}$$

es finita, y en este caso $\mathrm{E}(X)$ se define mediante la fórmula

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx. \tag{5.17}$$

Veamos a continuación algunos ejemplos de aplicación de la fórmula (5.17).

Ejemplos.

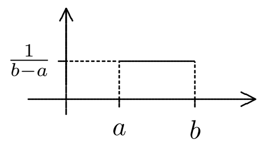


Figura 5.1:

1. *Distribución Uniforme*. Supongamos que X tiene distribución uniforme en el intervalo (a, b), $a < b$. Sabemos que ello implica que X tiene densidad

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

Tenemos

$$\begin{aligned} \mathrm{E}(X) &= \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \int_a^b x \frac{1}{b-a}\, dx \\ &= \frac{1}{b-a}\frac{1}{2}(b^2 - a^2) = \frac{b+a}{2} \end{aligned}$$

1. *Distribución Exponencial.* Si X tiene distribución exponencial con parámetro λ, $(\lambda > 0)$, su densidad es

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

Entonces

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \int_0^{+\infty} x\, \lambda e^{-\lambda x}\, dx.$$

Integrando por partes

$$\begin{aligned} \mathrm{E}(X) &= \left[-e^{-\lambda x} x\right|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\, dx \\ &= \int_0^{+\infty} e^{-\lambda x}\, dx = \frac{1}{\lambda} \end{aligned}$$

es decir $\mathrm{E}(X) = 1/\lambda$.

2. *Distribución Normal.* Si X tiene distribución normal con parámetros (μ, σ^2), su densidad es

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\} \qquad x \in \mathbb{R}.$$

Se verifica fácilmente que

$$\int_{-\infty}^{+\infty} |x|\, f_X(x)\, dx < \infty,$$

y efectuando el cambio de variables $u = (x-\mu)/\sigma$, resulta

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\} dx \\
&= \int_{-\infty}^{+\infty} (\sigma u + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2} du \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u e^{-u^2/2} du + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.
\end{aligned}$$

La integral en el primer término vale cero, dado que la función $u\, e^{-u^2/2}$ es impar y tiene integral finita. En cuanto a la integral en el segundo término, vale 1, ya que no es sino la integral sobre toda la recta de la densidad normal con parámetros $(0, 1)$. Resumiendo $\mathrm{E}(X) = \mu$.

5.5. Cambio de Variables. Momentos de Orden Superior.

Sea $f(x)$ la densidad de la variable aleatoria X y

$$g : \mathbb{R} \to \mathbb{R}$$

una función con alguna regularidad. Consideremos la variable aleatoria

$$Y = g(X).$$

Entonces, el valor esperado de Y se puede calcular mediante la fórmula

$$\mathrm{E}(Y) = \int_{\mathbb{R}} g(x) f(x) dx \tag{5.18}$$

siempre que la integral de la derecha sea absolutamente convergente, es decir que

$$\int_{\mathbb{R}} |g(x)| f(x) dx < \infty$$

Un caso particular merece una mención aparte. Si $g : \mathbb{R} \to \mathbb{R}$ es la función

$$g(x) = x^k,$$

entonces

$$\mathrm{E}(g(X)) = \mathrm{E}(X^k)$$

se denomina el *momento de orden k de la variable aleatoria X*, que es finito si y sólo si

$$\mathrm{E}(|X|^k) < \infty.$$

Si X tiene densidad f_X y si

$$\int_{-\infty}^{+\infty} |x|^k f_X(x) dx < \infty,$$

entonces la fórmula (5.18) se traduce en

$$\mathrm{E}(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x)\, dx.$$

Al igual que en el caso discreto definimos el *momento centrado de orden m* de X para $m \geq 2$ mediante

$$\mu_m = \mathrm{E}((X - \mathrm{E}(X))^m) = \int_{-\infty}^{+\infty} (x - \mathrm{E}(X))^m f_X(x)\, dx,$$

suponiendo que la integral exista.

Las definiciones de *varianza, desviación típica, covarianza* y *correlación* son iguales a las del caso discreto.

Observación 5.6 Las propiedades que demostramos para los momentos de orden superior en la sección 5.2.1 para el caso discreto también son ciertas para el caso continuo y aún con mayor generalidad. Para algunas propiedades las demostraciones no requieren modificación, como en el caso de las propiedades 1, 2 y 6, pero en otros es necesario dar demostraciones distintas, que no incluiremos.

Ejemplos.

1. *Distribución Uniforme.* Ya vimos que si X tiene distribución uniforme en (a, b), entonces

$$\mathrm{E}(X) = \frac{a+b}{2}.$$

Calculemos

$$\begin{aligned}\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}(X))^2] &= \int_{-\infty}^{+\infty} \left(x - \frac{a+b}{2}\right)^2 f_X(x) dx \\ &= \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{1}{b-a} \frac{1}{3} \left(x - \frac{a+b}{2}\right)^3 \Bigg|_{x=a}^{x=b} \\ &= \frac{1}{3(b-a)} \left[\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3\right] = \frac{(b-a)^2}{12}\end{aligned}$$

2. *Distribución Exponencial.* Calculemos los momentos de orden k ($k \geq 1$ entero) de una variable aleatoria X con distribución exponencial con parámetro λ ($\lambda > 0$).

Recurriendo a la forma de la densidad respectiva, e integrando por partes

$$\mathrm{E}(X^k) = \int_0^{+\infty} x^k \lambda e^{-\lambda x} dx = -e^{-\lambda x} x^k \Big|_0^{\infty} + k \int_0^{\infty} x^{k-1} e^{-\lambda x} dx$$

de donde obtenemos

$$\mathrm{E}(X^k) = \frac{k}{\lambda}\, \mathrm{E}(X^{k-1}).$$

Procediendo de manera inductiva se deduce de aquí que

$$\mathrm{E}(X^k) = \frac{k}{\lambda} \frac{k-1}{\lambda} \cdots \frac{1}{\lambda}\, \mathrm{E}(X^0) = \frac{k!}{\lambda^k}$$

3. *Distribución Normal.* Sea X una variable aleatoria con distribución normal de parámetros $(0, 1)$, es decir que su densidad es

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \qquad \text{para } x \in \mathbb{R}.$$

Calculemos el momento de orden k ($k \geq 1$ entero) de X. Es claro que estos momentos son finitos, ya que una acotación sencilla (que dejamos como ejercicio) muestra que

$$\int_{-\infty}^{+\infty} |x|^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx < \infty.$$

Sabemos que

$$\mathrm{E}(X^k) = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \tag{5.19}$$

Si k es impar, como el integrando resulta ser una función impar, la integral vale cero y $\mathrm{E}(X^k) = 0$.

Si k es un número par, digamos $k = 2p$, p entero positivo, integrando por partes en el segundo miembro de (5.19) se obtiene

$$\begin{aligned}\mathrm{E}(X^{2p}) &= \frac{1}{\sqrt{2\pi}} \left[-x^{2p-1} e^{-x^2/2} \Big|_{-\infty}^{+\infty} + (2p-1) \int_{-\infty}^{+\infty} x^{2p-2} e^{-x^2/2} dx \right] \\ &= (2p-1) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^{2p-2} e^{-x^2/2} dx,\end{aligned}$$

y por lo tanto

$$\mathrm{E}(X^{2p}) = (2p-1)\,\mathrm{E}(X^{2p-2}).$$

Aplicando ahora un procedimiento inductivo, se tiene

$$\mathrm{E}(X^{2p}) = (2p-1)(2p-3)\cdots 3\,\mathrm{E}(X) = \frac{(2p)!}{2^p\, p!} \tag{5.20}$$

Una variante sencilla es el cálculo del momento de orden k de la variable aleatoria $X - \mu$, donde X tiene distribución normal con parámetros (μ, σ^2). En este caso

$$\mathrm{E}((X-\mu)^k) = \int_{-\infty}^{+\infty} (x-\mu)^k \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} dx.$$

Haciendo el cambio de variables $t = (x-\mu)/\sigma$, resulta $dx = \sigma dt$ y

$$\mathrm{E}((X-\mu)^k) = \int_{-\infty}^{+\infty} (\sigma t)^k \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2} \sigma dt = \sigma^k \int_{-\infty}^{+\infty} t^k \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \sigma dt$$

y la integral de la derecha es la que venimos de calcular. Por lo tanto

$$\begin{aligned} k \text{ impar } &\Rightarrow \mathrm{E}((X-\mu)^k) = 0 \\ k = 2p \text{ par } &\Rightarrow \mathrm{E}((X-\mu)^{2p}) = \sigma^{2p} \frac{(2p)!}{2^p\, p!} \end{aligned}$$

En especial, si $k = 2$ tenemos $\mathrm{Var}(X) = \mathrm{E}((X-\mu)^2) = \sigma^2$.

5.6. Dos Fórmulas para el Cálculo de Valores Esperados.

En primer lugar, sea X una variable aleatoria discreta, cuyos valores son enteros positivos:

$$P(X = n) = p_n \quad n = 0, 1, \dots, \qquad \sum_{n=0}^{\infty} p_n = 1.$$

Denotamos

$$q_n = P(X > n) = \sum_{k=n+1}^{\infty} p_k, \qquad n = 0, 1, \dots$$

entonces

$$\mathrm{E}(X) = \sum_{n=0}^{\infty} q_n, \tag{5.21}$$

donde se debe entender que si $\mathrm{E}(X) = +\infty$ la serie diverge. Para probar (5.21) observamos que

$$\begin{aligned} \sum_{n=1}^{N} np_n &= (p_1 + \cdots + p_N) + (p_2 + \cdots + p_N) + \cdots + (p_{N-1} + p_N) + p_N \\ &= \sum_{n=0}^{N-1} q_n - Nq_N \end{aligned} \tag{5.22}$$

Esto ya muestra que si $\mathrm{E}(X) = +\infty$, caso en el cual el primer miembro tiende a $+\infty$ cuando $N \to \infty$, entonces

$$\sum_{n=0}^{\infty} q_n = +\infty,$$

ya que

$$\sum_{n=1}^{N} np_n \le \sum_{n=0}^{N-1} q_n.$$

Si, por el contrario, $\mathrm{E}(X) < \infty$, es decir, si la serie

$$\sum_n np_n$$

es convergente, debe cumplirse que

$$Nq_N \to 0 \quad \text{cuando } N \to \infty,$$

porque

$$0 \le Nq_N = N \sum_{n=N+1}^{\infty} p_n \le \sum_{n=N+1}^{\infty} np_n \to 0 \quad \text{cuando } N \to \infty,$$

dado que el último miembro de esta cadena de desigualdades es la “cola”de una serie convergente. Pero entonces, pasando al límite en (5.22) cuando $N \to \infty$, se obtiene (5.21).

La segunda fórmula que veremos, de la cual en realidad (5.21) es un caso particular, es la siguiente: Sea X una variable aleatoria no-negativa y F_X su función de distribución. Entonces

$$\mathrm{E}(X) = \int_0^{+\infty} (1 - F_X(x))\, dx. \tag{5.23}$$

Hay que entender que en la fórmula (5.23),

$$\int_0^{+\infty} (1 - F_X(x))\, dx$$

se define como integral impropia, es decir,

$$\int_0^{+\infty} (1 - F_X(x))\, dx = \lim_{A \to \infty} \int_0^{A} (1 - F_X(x))\, dx,$$

y también que si $\mathrm{E}(X) = +\infty$, la integral en el segundo miembro de (5.36) es divergente, y en este sentido, también vale la igualdad. El lector interesado puede encontrar la demostración de (5.23) en el apéndice de este capítulo.

5.7. Ejemplos.

1. *Distribución Geométrica.* Sea X una variable aleatoria con distribución geométrica de parámetro p, $0 < p < 1$, es decir que

$$P(X = n) = p(1-p)^{n-1}, \qquad n = 1, 2, \dots$$

Calcular $E(X)$ mediante aplicación de la fórmula (5.21).

▶ Sabemos que X es el tiempo que tarda en ocurrir el primer éxito en una sucesión de pruebas de Bernoulli con parámetro p. Por lo tanto, el suceso $\{X > n\}$ corresponde a que en las primeras n experiencias no ocurre ningún éxito. Esto implica que

$$q_n = P(X > n) = (1-p)^n$$

y entonces

$$\mathrm{E}(X) = \sum_{n=0}^{\infty}(1-p)^n = \frac{1}{1-(1-p)} = \frac{1}{p}$$

donde hemos usado la suma de una serie geométrica de razon $1 - p$. ◀

2. *Distribución Normal.* Sea X una variable aleatoria con distribución normal de parámetros $(0, \sigma^2)$. Calcular $\mathrm{E}(|X|)$ y $\mathrm{Var}(|X|)$.

▶ Hagamos la transformación $Y = X/\sigma$, entonces Y es normal típica y

$$\mathrm{E}(|X|) = \sigma\,\mathrm{E}(|Y|), \qquad \mathrm{Var}(|X|) = \sigma^2\,\mathrm{Var}(|Y|),$$

y nos basta considerar el caso de Y. Tenemos

$$\begin{aligned}\mathrm{E}(|Y|) &= \int_{-\infty}^{+\infty} |x|\, f_Y(x)\, dx = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x\, e^{-x^2/2} dx \\ &= \sqrt{\frac{2}{\pi}} \left(-e^{-x^2/2}\right|_0^{+\infty} = \sqrt{\frac{2}{\pi}}.\end{aligned}$$

Por otro lado

$$\mathrm{Var}(|Y|) = \mathrm{E}(|Y|^2) - (\mathrm{E}(|Y|))^2, \qquad \mathrm{E}(|Y|^2) = \mathrm{Var}(Y) = 1.$$

(ya hemos visto que si una variable aleatoria tiene distribución normal de parámetros (μ, σ^2), entonces su varianza es σ^2. Por lo tanto la varianza de Y es igual a 1). Reemplazando

$$\mathrm{Var}(|Y|) = 1 - \frac{2}{\pi},$$

y volviendo al principio

$$\mathrm{E}(|X|) = \sigma\sqrt{\frac{2}{\pi}}, \qquad \mathrm{Var}(|Y|) = \sigma^2(1 - \frac{2}{\pi}).$$

◀

3. Calcular el valor esperado y la varianza de la variable aleatoria

$$Z = X_1^2 + X_2^2 + \cdots + X_n^2,$$

donde $X_1, \dots, X_n$ son variables aleatorias independientes cada una de las cuales tiene distribución normal de parámetros $(0, 1)$.

Nota. La distribución de Z se denomina distribución ji-cuadrado con n grados de libertad *y se denota* χ_n^2.

▶ El cálculo del valor esperado es sencillo:

$$\mathrm{E}(Z) = \mathrm{E}(X_1^2 + X_2^2 + \cdots + X_n^2) = \mathrm{E}(X_1^2) + \mathrm{E}(X_2^2) + \cdots + \mathrm{E}(X_n^2)$$

y como $\mathrm{E}(X_i^2)$ es la varianza de X_i, es claro que $\mathrm{E}(X_i) = 1$ para todo $i = 1, \dots, n$. Por lo tanto

$$\mathrm{E}(Z) = n.$$

En cuanto a la varianza, como $X_1, \dots, X_n$ son variables aleatorias independientes, se deduce que también $X_1^2, \dots, X_n^2$ son variables independientes, y por lo tanto

$$\mathrm{Var}(Z) = \sum_{i=1}^{n} \mathrm{Var}(X_i^2).$$

Como todas las X_i tienen igual distribución de probabilidad, todos los términos de esta suma son iguales. Además

$$\mathrm{Var}(X_1^2) = \mathrm{E}(X_1^4) - (\mathrm{E}(X_1^2))^2 = \frac{4!}{2^2 2!} - 1 = 2$$

donde hemos aplicado la fórmula (5.20) para calcular $\mathrm{E}(X_1^4)$. Resumiendo

$$\mathrm{Var}(Z) = 2n.$$

◀

4. Sean $X_1, X_2, \dots, X_n$ n observaciones independientes con distribución uniforme en el intervalo $(0, 1)$. Ordenando estas variables en forma creciente, obtenemos los llamados *estadísticos de orden* de las observaciones $X_1, X_2, \dots, X_n$:

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

a. Calcular $\mathrm{E}(X_{(n)})$ y $\mathrm{Var}(X_{(n)})$.

b. Calcular, en general, $\mathrm{E}(X_{(k)})$ y $\mathrm{Var}(X_{(k)})$.

▶ a. La distribución de $X_{(n)}$ está dada por

$$P(X_{(n)} \le x) = x^n \quad \text{si } 0 \le x \le 1,$$

ya que

$$\{X_{(n)} \in [0, x]\} = \cap_{i=1}^{n} \{X_i \le x\}$$

y los eventos que aparecen en esta intersección son independientes. Por lo tanto,

$$F_{X_{(n)}}(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^n & \text{si } 0 \le x \le 1 \\ 1 & \text{si } x > 1 \end{cases}$$

y $X_{(n)}$ tiene densidad

$$f_{X_{(n)}}(x) = \begin{cases} 0 & \text{si } x \notin (0,1) \\ nx^{n-1} & \text{si } 0 < x < 1 \end{cases}$$

y calculamos

$$\begin{aligned} \mathrm{E}(X_{(n)}) &= \int_{-\infty}^{+\infty} x f_{X_{(n)}}(x) dx = \int_0^1 xnx^{n-1}\, dx = \frac{n}{n+1} \\ \mathrm{E}(X_{(n)}^2) &= \int_0^1 x^2 nx^{n-1}\, dx = \frac{n}{n+2} \\ \mathrm{Var}(X_{(n)}) &= \mathrm{E}(X_{(n)}^2) - (\mathrm{E}(X_{(n)}))^2 \\ &= \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n}{(n+2)(n+1)^2} \end{aligned}$$

b. En general, la densidad de $X_{(k)}$ está dada por

$$f_{X_{(k)}}(x) = \begin{cases} n\binom{n-1}{k-1} x^{k-1}(1-x)^{n-k}, & \text{si } 0 < x < 1, \\ 0, & \text{si } x \notin (0,1). \end{cases}$$

Entonces

$$\begin{aligned} \mathrm{E}(X_{(k)}) &= \int_0^1 xn\binom{n-1}{k-1} x^{k-1}(1-x)^{n-k}\, dx, \\ \mathrm{E}(X_{(k)}^2) &= \int_0^1 x^2 n\binom{n-1}{k-1} x^{k-1}(1-x)^{n-k}\, dx. \end{aligned}$$

Si introducimos la notación

$$B(k,l) = \int_0^1 x^k (1-x)^l dx \quad \text{donde } k,l = 0,1,2\dots$$

$\mathrm{E}(X_{(k)})$ y $\mathrm{E}(X_{(n)}^2)$ se escriben como

$$\begin{aligned} \mathrm{E}(X_{(k)}) &= n\binom{n-1}{k-1} B(k, n-k), \\ \mathrm{E}(X_{(k)}^2) &= n\binom{n-1}{k-1} B(k+1, n-k), \end{aligned}$$

y nuestro problema se reduce a calcular $B(k,l)$, lo cual hacemos a continuación. Primero observemos que

$$B(k,l) = B(l,k)$$

haciendo el cambio de variables $y = 1-x$ en la integral (5.39). Calculamos $B(k,l)$ integrando por partes. Sea $l \geq 1$:

$$\begin{aligned} B(k,l) &= \int_0^1 x^k (1-x)^l dx \\ &= \frac{1}{k+1} x^{k+1}(1-x)^l \Big|_0^1 + \frac{l}{k+1}\int_0^1 x^{k+1}(1-x)^{l-1} dx \\ &= \frac{l}{k+1} B(k+1, l-1). \end{aligned}$$

Esta fórmula de recurrencia, aplicada reiteradamente nos da

$$\begin{aligned}
B(k,l) = \frac{l}{k+1}B(k+1,l-1) &= \frac{l(l-1)}{(k+1)(k+2)}B(k+2,l-2) \\
&= \frac{l(l-1)\cdots 1}{(k+1)\cdots(k+l)}B(k+l,0) \\
&= \frac{l!\,k!}{(k+l)!}\frac{1}{k+l+1}
\end{aligned}$$

ya que $B(k+l,0) = 1/(k+l+1)$. Resumiendo,

$$B(k,l) = \frac{l!\,k!}{(k+l+1)!}$$

y reemplazando en las fórmulas para $\mathrm{E}(X_{(k)})$ y $\mathrm{E}(X_{(n)}^2)$ obtenemos

$$\begin{aligned}
\mathrm{E}(X_{(k)}) &= \frac{n!}{(k-1)!(n-k)!}\frac{k!(n-k)!}{(n+1)!} = \frac{k}{n+1}, \\
\mathrm{E}(X_{(k)}^2) &= \frac{n!}{(k-1)!(n-k)!}\frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k(k+1)}{(n+1)(n+2)}, \\
\mathrm{Var}((X_{(k)}) &= \frac{k(k+1)}{(n+1)(n+2)} - \frac{k^2}{(n+1)^2} = \frac{k(n-k+1)}{(n+2)(n+1)^2}.
\end{aligned}$$

◀

5. Sea X una variable aleatoria con distribución de Cauchy de densidad:

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

Demostrar que no existe $\mathrm{E}(X)$ pero que, sin embargo

$$\lim_{a\to\infty}\int_{-a}^{a} x\, f(x)\, dx = 0.$$

▶ Para ver que la esperanza matemática de X no existe, mostraremos que la integral

$$\int_{-\infty}^{\infty} |x|\, f(x)\, dx$$

es divergente. Efectivamente

$$\begin{aligned}
\frac{1}{\pi}\int_{-a}^{a}\frac{|x|}{1+x^2}\,dx &= \frac{2}{\pi}\int_{0}^{a}\frac{x}{1+x^2}\,dx \\
&= \frac{2}{\pi}\left(\frac{1}{2}\log(1+x^2)\right|_0^a = \frac{1}{\pi}\log(1+a^2) \\
&\to +\infty \quad \text{cuando } a\to\infty.
\end{aligned}$$

Sin embargo, es obvio que

$$\frac{1}{\pi}\int_{-a}^{a}\frac{x}{1+x^2}\,dx = 0$$

ya que el integrando es una función impar, y por lo tanto,

$$\lim_{a\to\infty}\frac{1}{\pi}\int_{-a}^{a}\frac{x}{1+x^2}\,dx = 0.$$

◀

6. Sean X e Y dos variables aleatorias, $\mathrm{Var}(X) \neq 0,\ \mathrm{Var}(Y) \neq 0$ y sea ρ su coeficiente de correlación

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Probar que si $\rho = 1$ ó $\rho = -1$, entonces existen dos constantes a y b tales que

$$P(Y = aX + b) = 1.$$

► Recordemos cómo probamos que $-1 \leq \rho \leq 1$ (Propiedad 4, sección 5.1.1). Consideramos la función $f(x) = \mathrm{E}((V - xU)^2)$, donde $U = X - \mathrm{E}(X)$ y $V = Y - \mathrm{E}(Y)$. Desarrollando el cuadrado

$$\begin{aligned} f(x) &= \mathrm{E}(V^2) - 2x\,\mathrm{E}(UV) + x^2\,\mathrm{E}(U^2) \\ &= \mathrm{Var}(Y) - 2x\,\mathrm{Cov}(X,Y) + x^2\,\mathrm{Var}(X). \end{aligned}$$

El discriminante de este polinomio de segundo grado es positivo (observar que el grado del polinomio es 2 ya que hemos supuesto que $\mathrm{Var}(X) \neq 0$). Por otra parte, este polinomio tiene discriminante

$$4(\mathrm{Cov}(X,Y))^2 - 4\,\mathrm{Var}(X)\,\mathrm{Var}(Y) = 0$$

ya que $\rho = 1$ ó $\rho = -1$. Esto quiere decir que ese polinomio tiene una raíz real doble, que llamaremos a. O sea que $f(a) = 0$, es decir

$$\mathrm{E}((V - aU)^2) = 0.$$

Pero $(V - aU)^2$ es una variable aleatoria no-negativa, cuya esperanza es nula y por lo tanto (ver propiedad 5, sección 5.2.1 y la observación 5.6 de la sección 5.5)

$$P((V - aU)^2 = 0) = 1 \Rightarrow P(V - aU = 0) = 1$$

ya que $\{V - aU = 0\}$ y $\{(V - aU)^2 = 0\}$ son el mismo evento. Reemplazando V por $Y - \mathrm{E}(Y)$ y U por $X - \mathrm{E}(X)$, resulta

$$P(Y - \mathrm{E}(Y) - a(X - \mathrm{E}(X)) = 0) = 1$$

y eligiendo $b = \mathrm{E}(Y) - a\,\mathrm{E}(X)$, se obtiene el resultado. ◄

Observación 5.7 El número ρ, que sabemos que está comprendido entre -1 y 1, es una medida de la dependencia lineal entre las variables aleatorias X e Y. Lo que dice este ejercicio es que si $\rho = 1$ ó $\rho = -1$, entonces la probabilidad de que el par aleatorio (X,Y) caiga en la recta de ecuación $y = mx + n$ es igual a 1, o, lo que es lo mismo, que la probabilidad de que caiga fuera de dicha recta es igual a cero.

Cuando $\rho = 1$ ó $\rho = -1$, la pendiente m de la recta en cuestión es la raíz doble de la ecuación $f(x) = 0$, es decir que

$$m = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}.$$

De modo que, si $\rho = 1,\ \mathrm{Cov}(X,Y) > 0$ y la pendiente es positiva y si $\rho = -1$, $\mathrm{Cov}(X,Y) < 0$ y m resulta negativo.

7. Un pasajero llega al terminal de autobuses en el instante T, que es una variable aleatoria con distribución uniforme entre las 11 y las 12 horas. De acuerdo con el horario, está previsto que del terminal partan un autobús a las 11 y otro a las 12, pero éstos salen con retardos aleatorios X e Y respectivamente, teniendo cada una de estas dos variables aleatorias distribución uniforme en el intervalo $(0, 1/2$ hora$)$.

Si ambos autobuses le sirven al pasajero y las variables aleatorias $T,\ X,\ Y$ son independientes, ¿cuál es el valor esperado del tiempo que el pasajero permanecerá en el terminal?

► Llamemos Z al tiempo que debe esperar el pasajero. Sea A el evento $\{11 \leq T \leq 11:30\}$, es decir "el pasajero llega entre las 11 y las 11 y 30" y B el evento $\{T - 11 \leq X\}$, es decir, "llega antes de que el primer autobus salga". Entonces

– Si ocurren A y B, el tiempo de espera es $Z = X - (T - 11)$ (ver figura 5.2).

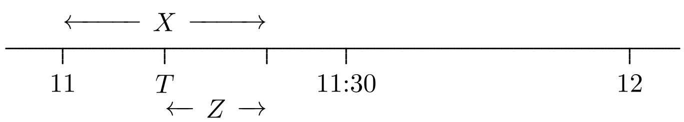


Figura 5.2:

– Si ocurre A y no ocurre B, $Z = 12 - T + Y$ (ver figura 5.3).

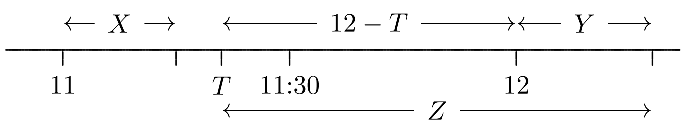


Figura 5.3:

– Si no ocurre A entonces $Z = 12 - T + Y$(ver figura 5.4).

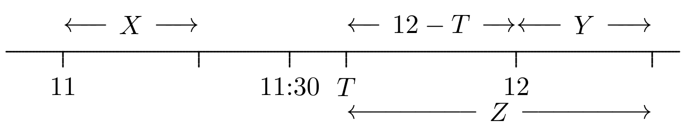


Figura 5.4:

En resumen,

$$\begin{aligned} &\text{en } A \cap B, && Z = X - (T - 11), \\ &\text{en } (A \cap B)^c, && Z = 12 - T + Y, \end{aligned}$$

lo que es lo mismo que

$$Z = (X - (T - 11))\mathbf{1}_{A\cap B} + (12 - T + Y)\mathbf{1}_{(A\cap B)^c}$$

y si observamos que

$$\mathbf{1}_{(A\cap B)^c} = 1 - \mathbf{1}_{A\cap B}$$

obtenemos que

$$\begin{aligned} Z &= (X - (T - 11))\mathbf{1}_{A\cap B} + (12 - T + Y)(1 - \mathbf{1}_{(A\cap B)}) \\ &= (12 - T + Y) + (X - T + 11 - 12 + T - Y)\mathbf{1}_{A\cap B} \\ &= (12 - T + Y) + (X - Y - 1)\mathbf{1}_{A\cap B}. \end{aligned}$$

Por lo tanto

$$\mathrm{E}(Z) = 12 - \mathrm{E}(T) + \mathrm{E}(Y) + \mathrm{E}((X - Y - 1)\mathbf{1}_{A\cap B}).$$

Es claro que

$$\mathrm{E}(T) = 11.5 \qquad \mathrm{E}(Y) = 0.25.$$

Para calcular el último término, observamos que la densidad conjunta de las variables aleatorias T, X e Y es

$$f_{T,X,Y}(t,x,y) = \begin{cases} 4 & \text{si } 11 \le t \le 12;\ 0 \le x \le 0.5;\ 0 \le y \le 0.5 \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Esto resulta inmediatamente de que T, X e Y son independientes y de que tienen densidad uniforme, la primera en $(11,\ 12)$ y en $(0,\ 0.5)$ las otras dos.

Por lo tanto, teniendo en cuenta la definición de las eventos A y B,

$$\mathrm{E}((X - Y - 1)\mathbf{1}_{A\cap B}) = \int_{11}^{11,5} dt \int_{t-11}^{0,5} dx \int_{0}^{0,5} (x - y - 1)4\, dy = -\frac{11}{48}$$

Reemplazando se obtiene

$$\mathrm{E}(Z) = \frac{25}{48} \text{ horas.}$$

◄

5.8. Sumas Aleatorias

Con frecuencia encontramos sumas de la forma $T = X_1 + \cdots + X_N$, donde el número de sumandos es una variable aleatoria. Consideremos una sucesión $X_1, X_2, \dots$ de v.a.i.i.d. y sea N una v.a. discreta, independiente de $X_1, X_2, \dots$ con densidad $p_N(n) = P(N = n)$, $n = 0, 1, \dots$. Definimos la suma aleatoria T como

$$T = \begin{cases} 0 & \text{si } N = 0, \\ X_1 + \cdots + X_N & \text{si } N > 0. \end{cases}$$

Ejemplos 5.2

a) Colas: N representa el número de clientes, X_i es el tiempo de atención de cada cliente, T es el tiempo total de atención.

b) Seguros: N representa el número de reclamos en un período de tiempo dado, X_i es el monto de cada reclamo y T es el monto total de los reclamos en el período.

c) Población: N representa el número de plantas, X_i es el número de semillas de cada planta, T es el total de semillas.

d) Biometría: N es el tamaño de la población, X_i es el peso de cada ejemplar y T representa el peso total de la muestra.

Momentos de una Suma Aleatoria

Supongamos que X_k y N tienen momentos finitos

$$\mathrm{E}[X_k] = \mu, \qquad \mathrm{Var}[X_k] = \sigma^2,$$
$$\mathrm{E}[N] = \nu, \qquad \mathrm{Var}[N] = \tau^2.$$

y queremos determinar media y varianza de $T = X_1 + \cdots + X_N$. Veamos que

$$\mathrm{E}[T] = \mu\nu, \qquad \mathrm{Var}[T] = \nu\sigma^2 + \mu^2\tau^2.$$

Tenemos

$$\begin{aligned}
\mathrm{E}[T] &= \sum_{n=0}^{\infty} \mathrm{E}[T|N=n] p_N(n) = \sum_{n=0}^{\infty} \mathrm{E}[X_1 + \cdots + X_N | N=n] p_N(n) \\
&= \sum_{n=0}^{\infty} \mathrm{E}[X_1 + \cdots + X_n | N=n] p_N(n) = \sum_{n=0}^{\infty} \mathrm{E}[X_1 + \cdots + X_n] p_N(n) \\
&= \sum_{n=0}^{\infty} n\mu p_N(n) = \mu\nu.
\end{aligned}$$

Para determinar la varianza comenzamos por

$$\begin{aligned}
\mathrm{Var}[T] &= \mathrm{E}[(T-\mu\nu)^2] = \mathrm{E}[(T - N\mu + N\mu - \nu\mu)^2] \\
&= \mathrm{E}[(T-N\mu)^2] + \mathrm{E}[\mu^2(N-\nu)^2] + 2\,\mathrm{E}[\mu(T-N\mu)(N-\nu)].
\end{aligned}$$

Calculemos cada uno de estos sumandos por separado, el primero es

$$\begin{aligned}
\mathrm{E}[(T-N\mu)^2] &= \sum_{n=0}^{\infty} \mathrm{E}[(T-N\mu)^2 | N=n] p_N(n) \\
&= \sum_{n=1}^{\infty} \mathrm{E}[(X_1 + \cdots + X_n - n\mu)^2 | N=n] p_N(n) \\
&= \sum_{n=1}^{\infty} \mathrm{E}[(X_1 + \cdots + X_n - n\mu)^2] p_N(n) \\
&= \sigma^2 \sum_{n=1}^{\infty} n p_N(n) = \nu\sigma^2.
\end{aligned}$$

Para el segundo tenemos

$$\mathrm{E}[\mu^2(N-\nu)^2] = \mu^2\,\mathrm{E}[(N-\nu)^2] = \mu^2\tau^2$$

y finalmente el tercero es

$$\begin{aligned}
\mathrm{E}[\mu(T-N\mu)(N-\mu)] &= \mu \sum_{n=0}^{\infty} \mathrm{E}[(T-n\mu)(n-\nu) | N=n] p_N(n) \\
&= \mu \sum_{n=0}^{\infty} (n-\nu)\,\mathrm{E}[(T-n\mu) | N=n] p_N(n) \\
&= 0.
\end{aligned}$$

La suma de estos tres términos demuestra el resultado.

Distribución de una Suma Aleatoria

Supongamos que los sumandos $X_1, X_2, \dots$ son v.a.i. continuas con densidad de probabilidad $f(x)$. Para $n \geq 1$ fijo, la densidad de la suma $X_1 + \cdots + X_n$ es la n-ésima convolución de la densidad $f(x)$, que denotaremos por $f^{(n)}(x)$ y definimos recursivamente por

$$\begin{aligned}
f^{(1)}(x) &= f(x), \\
f^{(n)}(x) &= \int f^{(n-1)}(x-u) f(u)\,du \quad \text{para } n > 1.
\end{aligned}$$

Como N y $X_1, X_2, \dots$ son independientes, $f^{(n)}(x)$ es también la densidad condicional de $T = X_1 + \cdots + X_N$ dado que $N = n \geq 1$.

Supongamos que $P(N = 0) = 0$, es decir, que la suma aleatoria siempre tiene al menos un sumando. Por la ley de la probabilidad total, T es continua y tiene densidad marginal

$$f_T(x) = \sum_{n=1}^{\infty} f^{(n)}(x) p_N(n).$$

Observación 5.8 Si N puede valer 0 con probabilidad positiva entonces $T = X_1 + \cdots + X_N$ es una v.a. mixta, es decir, tiene componentes discreta y continua. Si suponemos que $X_1, X_2, \dots$ son continuas con densidad $f(x)$, entonces

$$P(T = 0) = P(N = 0) = p_N(0)$$

mientras que para $0 < a < b$ ó $a < b < 0$,

$$P(a < T < b) = \int_a^b \big(\sum_{n=1}^{\infty} f^{(n)}(x) p_N(n)\big) dx$$

◀

Ejemplo 5.3 (Suma Geométrica de Variables Exponenciales)
Supongamos que

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0, \\ 0 & \text{para } x < 0. \end{cases}$$

$$p_N(n) = \beta(1-\beta)^{n-1} \quad n = 1, 2, \dots$$

Comenzamos por hallar la convolución de las densidades exponenciales

$$\begin{aligned} f^{(2)}(x) &= \int f(x-u) f(u)\, du = \int \mathbf{1}_{\{x-u \geq 0\}}(u) \lambda e^{-\lambda(x-u)} \mathbf{1}_{\{u \geq 0\}}(u) \lambda e^{-u} \\ &= \lambda^2 e^{-\lambda x} \int_0^x du = x \lambda^2 e^{-\lambda x} \end{aligned}$$

para $x \geq 0$. La siguiente convolución es

$$\begin{aligned} f^{(3)}(x) &= \int f^{(2)}(x-u) f(u)\, du = \int \mathbf{1}_{\{x-u \geq 0\}}(u) \lambda^2 (x-u) e^{-\lambda(x-u)} \mathbf{1}_{\{u \geq 0\}}(u) \lambda e^{-u}\, du \\ &= \lambda^3 e^{-\lambda x} \int_0^x (x-u)\, du = \frac{x^2}{2} \lambda^3 e^{-\lambda x} \end{aligned}$$

para $x \geq 0$. Procediendo inductivamente obtenemos que

$$f^{(n)}(x) = \frac{x^{n-1}}{(n-1)!} \lambda^n e^{-\lambda x}$$

La densidad de $T = X_1 + \cdots + X_N$ es

$$\begin{aligned} f_T(t) &= \sum_{n=1}^{\infty} f^{(n)}(t) p_N(n) = \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} \beta (1-\beta)^{n-1} \\ &= \lambda \beta e^{-\lambda t} \sum_{n=1}^{\infty} \frac{(\lambda(1-\beta)t)^{n-1}}{(n-1)!} = \lambda \beta e^{-\lambda t} e^{\lambda(1-\beta)t} \\ &= \lambda \beta e^{-\lambda \beta t} \end{aligned}$$

para $t \geq 0$, y por lo tanto $T \sim \mathcal{E}xp(\lambda\beta)$. ◀

5.9. Funciones Generadoras de Probabilidad

Consideremos una v.a. ξ con valores enteros positivos y distribución de probabilidad

$$P(\xi = k) = p_k, \quad k = 0, 1, \dots$$

La función generadora de probabilidad (f.g.p.) $\phi(s)$ asociada a la v.a. ξ (o equivalentemente a su distribución (p_k)) se define por

$$\phi(s) = \mathrm{E}[s^\xi] = \sum_{k=0}^{\infty} s^k p_k, \qquad 0 \le s \le 1. \tag{5.24}$$

A partir de la definición es inmediato que si ϕ es una f.g.p. entonces

$$\phi(1) = \sum_{k=0}^{\infty} p_k = 1.$$

Resultados Fundamentales:

1. La relación entre funciones de probabilidad y funciones generadoras es 1-1. Es posible obtener las probabilidades (p_k) a partir de ϕ usando la siguiente fórmula

$$p_k = \frac{1}{k!} \left. \frac{d^k \phi(s)}{ds^k} \right|_{s=0}. \tag{5.25}$$

   Por ejemplo,

$$\phi(s) = p_0 + p_1 s + p_2 s^2 + \cdots \ \Rightarrow \ p_0 = \phi(0)$$

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \cdots \ \Rightarrow \ p_1 = \left. \frac{d\phi(s)}{ds} \right|_{s=0}$$

2. Si $\xi_1, \dots, \xi_n$ son v.a.i. con funciones generadoras $\phi_1(s), \phi_2(s), \dots, \phi_n(s)$ respectivamente, la f. g. p. de su suma $X = \xi_1 + \xi_2 + \cdots + \xi_n$ es el producto de las funciones generadoras respectivas

$$\phi_X(s) = \phi_1(s)\phi_2(s) \cdots \phi_n(s). \tag{5.26}$$

3. Los momentos de una variable que toma valores en los enteros no-negativos se pueden obtener derivando la función generadora:

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \cdots,$$

   por lo tanto

$$\left. \frac{d\phi(s)}{ds} \right|_{s=1} = p_1 + 2p_2 + 3p_3 + \cdots = \mathrm{E}[\xi]. \tag{5.27}$$

   Para la segunda derivada tenemos

$$\frac{d^2\phi(s)}{ds^2} = 2p_2 + 3 \cdot 2p_3 s + 4 \cdot 3p_4 s^2 + \cdots,$$

   evaluando en $s = 1$,

$$\begin{aligned} \left. \frac{d^2\phi(s)}{ds^2} \right|_{s=1} &= 2p_2 + 3 \cdot 2p_3 + 4 \cdot 3p_4 \cdots \\ &= \sum_{k=2}^{\infty} k(k-1)p_k \\ &= \mathrm{E}[\xi(\xi-1)] = \mathrm{E}[\xi^2] - \mathrm{E}[\xi] \end{aligned} \tag{5.28}$$

de modo que

$$\mathrm{E}[\xi^2] = \frac{d^2\phi(s)}{ds^2}\Big|_{s=1} + \mathrm{E}[\xi] = \frac{d^2\phi(s)}{ds^2}\Big|_{s=1} + \frac{d\phi(s)}{ds}\Big|_{s=1},$$

y en consecuencia

$$\mathrm{Var}[\xi] = \mathrm{E}[\xi^2] - (\mathrm{E}[\xi])^2 = \frac{d^2\phi(s)}{ds^2}\Big|_{s=1} + \frac{d\phi(s)}{ds}\Big|_{s=1} - \Big(\frac{d^2\phi(s)}{ds^2}\Big|_{s=1}\Big)^2.$$

Ejemplo 5.4
Supongamos que $\xi \sim \mathcal{P}ois(\lambda)$:

$$p_k = P(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, \dots$$

Su función generadora de probabilidad es

$$\begin{aligned}\phi(s) = \mathrm{E}[s^\xi] &= \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(s\lambda)^k}{k!} = e^{-\lambda}e^{\lambda s} \\ &= e^{-\lambda(1-s)}\end{aligned}$$

Entonces,

$$\frac{d\phi(s)}{ds} = \lambda e^{-\lambda(1-s)}, \qquad \frac{d\phi(s)}{ds}\Big|_{s=1} = \lambda \tag{5.29}$$

$$\frac{d^2\phi(s)}{ds^2} = \lambda^2 e^{-\lambda(1-s)}, \qquad \frac{d^2\phi(s)}{ds^2}\Big|_{s=1} = \lambda^2 \tag{5.30}$$

y obtenemos

$$\mathrm{E}[\xi] = \lambda, \qquad \mathrm{Var}(\xi) = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

◀

5.9.1. Funciones Generadoras de Probabilidad y Sumas de V. A. I.

Sean ξ, η v.a.i. con valores $0, 1, 2, \dots$ y con funciones generadoras de probabilidad

$$\phi_\xi(s) = \mathrm{E}[s^\xi], \qquad \phi_\eta(s) = \mathrm{E}[s^\eta], \qquad |s| < 1,$$

entonces la f.g.p. de la suma $\xi + \eta$ es

$$\phi_{\xi+\eta}(s) = \mathrm{E}[s^{\xi+\eta}] = \mathrm{E}[s^\xi s^\eta] = \mathrm{E}[s^\xi]\,\mathrm{E}[s^\eta] = \phi_\xi(s)\phi_\eta(s) \tag{5.31}$$

El recíproco tambien es cierto, si $\phi_{\xi+\eta}(s) = \phi_\xi(s)\phi_\eta(s)$ entonces las variables ξ y η son independientes.

Como consecuencia, si $\xi_1, \xi_2, \dots, \xi_m$ son v.a.i.i.d. con valores en $\{0, 1, 2, \dots\}$ y f.g.p. $\phi(s) = \mathrm{E}[s^\xi]$ entonces

$$\mathrm{E}[s^{\xi_1+\cdots+\xi_m}] = \phi^m(s) \tag{5.32}$$

¿Qué ocurre si el número de sumandos es aleatorio?

Proposición 5.1 *Sea N una v.a. con valores enteros no-negativos e independiente de $\xi_1, \xi_2, \dots$ con f.g.p. $g_N(s) = \mathrm{E}[s^N]$ y consideremos la suma*

$$X = \xi_1 + \dots + \xi_N.$$

Sea $h_X(s) = \mathrm{E}[s^X]$ la f.g.p. de X. Entonces

$$h_X(s) = g_N(\phi(s)). \tag{5.33}$$

Demostración.

$$\begin{aligned}
h_X(s) &= \sum_{k=0}^{\infty} P(X = k)s^k \\
&= \sum_{k=0}^{\infty} \Big(\sum_{n=0}^{\infty} P(X = k|N = n)P(N = n) \Big) s^k \\
&= \sum_{k=0}^{\infty} \Big(\sum_{n=0}^{\infty} P(\xi_1 + \dots + \xi_n = k|N = n)P(N = n) \Big) s^k \\
&= \sum_{k=0}^{\infty} \Big(\sum_{n=0}^{\infty} P(\xi_1 + \dots + \xi_n = k)P(N = n) \Big) s^k \\
&= \sum_{n=0}^{\infty} \Big(\sum_{k=0}^{\infty} P(\xi_1 + \dots + \xi_n = k)s^k \Big) P(N = n) \\
&= \sum_{n=0}^{\infty} \phi^n(s)P(N = n) = g_N(\phi(s))
\end{aligned}$$

■

Ejemplo 5.5
Sea N una variable aleatoria con distribución de Poisson de parámetro λ. Dado el valor de N, realizamos N experimentos de Bernoulli con probabilidad de éxito p y llamamos X al número de éxitos. En este caso ξ_i tiene distribución de Bernoulli y su f.g.p. es

$$\phi_\xi(s) = \mathrm{E}[s^\xi] = sp + q$$

mientras que $N \sim \mathcal{P}ois(\lambda)$ con f.g.p.

$$g_N(s) = \mathrm{E}[s^N] = e^{-\lambda(1-s)}$$

según vimos en el ejemplo 5.4. Por la proposición anterior obtenemos que la f.g.p. de X es

$$h_X(s) = g_N(\phi_\xi(s)) = g_N(q + sp) = \exp\Big\{ -\lambda(1 - q - sp)\Big\} = \exp\Big\{ -\lambda p(1 - s)\Big\}$$

que es la f.g.p. de una distribución de Poisson de parámetro λp. ◀

5.10. Funciones Generadoras de Momentos.

Dada una variable aleatoria X, o su función de distribución F, vamos a definir otra función generadora, como

$$M_X(t) = \mathrm{E}(e^{tX}).$$

siempre que este valor esperado exista.

Notemos que cuando el recorrido de X son los enteros no-negativos, $M_X(t) = \phi_X(e^t)$. Si X está acotada, M_X está bien definida para todo t real; en cambio, si X no está acotada, es posible que el dominio de M no sea el conjunto de todos los reales. En todo caso, p siempre está definida en cero, y $M(0) = 1$.

Si la función M está definida en un entorno de $t = 0$, entonces las series

$$M_X(t) = \mathrm{E}(e^{tX}) = \mathrm{E}\left(1 + \sum_{n=1}^{\infty} \frac{t^n X^n}{n!}\right) = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathrm{E}(X^n)$$

son convergentes y en consecuencia se puede derivar término a término. Obtenemos

$$M'_X(0) = \mathrm{E}(X); \quad M''_X(0) = \mathrm{E}(X^2) \quad \text{y en general } M_X^{(n)}(0) = \mathrm{E}(X^n).$$

Es por esta última propiedad que esta función se conoce como *función generadora de momentos* (f.g.m.).

Ejemplos 5.6

1. Si $X \sim Bin(n,p)$ veamos que $M(t) = (pe^t + 1 - p)^n$): Un cálculo directo muestra que

$$M(t) = \sum_{j=0}^{n} e^{jt} \binom{n}{j} p^j (1-p)^{n-j} = (pe^t + 1 - p)^n,$$

.

2. Si $X \sim \mathcal{E}xp(\lambda)$, es decir, si $P(X \le x) = 1 - e^{-\lambda x}$, para $x \ge 0$, entonces $M(t) = \lambda/(\lambda - t)$ para $t \le \lambda$.

   El resultado se obtiene a partir del cálculo

$$M(t) = \int_0^{\infty} \lambda e^{-\lambda x}\, e^{tx} dx = \lambda \left. \frac{e^{(t-\lambda)x}}{t - \lambda} \right|_0^{\infty} = \frac{\lambda}{\lambda - t}.$$

   Observamos que en este caso, $M(t)$ no está definida si $t \ge \lambda$.

3. Si $X \sim \mathcal{N}(0,1)$, es decir, si $P(X \le x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-x^2/2} dx$, entonces $M(t) = e^{t^2/2}$.

   Calculemos

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx}\, e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2}\, e^{t^2/2} \mathrm{d}x = e^{t^2/2}$$

   ya que $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} \mathrm{d}x = 1$ puesto que el integrando es la densidad de una variable aleatoria con distribución $\mathcal{N}(t,1)$

Observación 5.9 Por la forma en la cual hemos definido la función generadora de momentos, cuando las f.g.m. de dos variables aleatorias X_1, X_2 coinciden para todos los valores de t en un entorno de $t = 0$, entonces las distribuciones de probabilidad de X_1 y X_2 deben ser idénticas. Este resultado lo enunciamos en el próximo teorema, sin demostración

Teorema 5.1 *Si X tiene función generadora de momentos $M(t)$ que está definida en un entorno $(-a,a)$ de 0, entonces $M(t)$ caracteriza a la distribución de X, es decir, si otra variable Y tiene la misma función generadora de momentos, las distribuciones de X e Y coinciden.*

La función generadora de momentos resulta particularmente útil cuando consideramos sucesiones de variables aleatorias, como lo muestra el siguiente teorema que enunciamos sin demostración.

Teorema 5.2 (de Continuidad) *Sea $F_n(x)$, $n \geq 1$ una sucesión de f.d. con funciones generadoras de momento respectivas $M_n(t)$, $n \geq 1$, que están definidas para $|t| < b$. Supongamos que cuando $n \to \infty$, $M_n(t) \to M(t)$ para $|t| \leq a < b$, donde $M(t)$ es la función generadora de momentos de la distribución $F(x)$. Entonces $F_n(x) \to F(x)$ cuando $n \to \infty$ para todo punto x en el cual F es continua.*

Veamos una aplicación del teorema anterior para demostrar el Teorema de de Moivre y Laplace.

Teorema 5.3 (de Moivre-Laplace) *Sea $S_n \sim Bin(n,p)$ para $n \geq 1$ y $q = 1-p$. Definimos*

$$T_n = \frac{S_n - np}{(npq)^{1/2}}$$

Entonces para todo $x \in \mathbb{R}$,

$$P(T_n \leq x) \to \Phi(x) = \int_{-\infty}^{x} \frac{1}{2\pi} e^{-x^2/2} dx.$$

Demostración. Recordemos que S_n es la suma de n v.a.i. con distribución de Bernoulli de parámetro p: $S_n = \sum_1^n X_i$. Usamos esto para calcular la función generadora de momentos de T_n.

$$\begin{aligned}
\mathrm{E}(e^{tT_n}) &= \mathrm{E}\Big[\exp\Big(\frac{t(S_n - np)}{(npq)^{1/2}}\Big)\Big] = \mathrm{E}\Big[\exp\Big(\frac{t(\sum_1^n (X_i - p))}{(npq)^{1/2}}\Big)\Big] \\
&= \mathrm{E}\Big[\prod_{i=1}^n \exp\Big(\frac{t(X_i - p)}{(npq)^{1/2}}\Big)\Big] = \prod_{i=1}^n \mathrm{E}\Big[\exp\Big(\frac{t(X_i - p)}{(npq)^{1/2}}\Big)\Big] \\
&= \Big(E\Big[\exp\Big(\frac{t(X_1 - p)}{(npq)^{1/2}}\Big)\Big]\Big)^n \\
&= \Big(p\,\exp\Big(\frac{t(1-p)}{(npq)^{1/2}}\Big) + q\exp\Big(\frac{-pt}{(npq)^{1/2}}\Big)\Big)^n. \qquad (5.34)
\end{aligned}$$

Ahora hacemos un desarrollo de Taylor para las dos exponenciales que aparecen en esta última expresión para obtener

$$p\,\exp\Big(\frac{t(1-p)}{(npq)^{1/2}}\Big) = p\Big(1 + \frac{qt}{(npq)^{1/2}} + \frac{q^2t^2}{2npq} + \frac{C_1 q^3 t^3}{3!(npq)^{3/2}}\Big) \qquad (5.35)$$

$$q\,\exp\Big(\frac{-pt}{(npq)^{1/2}}\Big) = q\Big(1 - \frac{pt}{(npq)^{1/2}} + \frac{p^2t^2}{2npq} + \frac{C_2 p^3 t^3}{3!(npq)^{3/2}}\Big). \qquad (5.36)$$

La suma de estas dos expresiones nos da $1 + \frac{t^2}{2n} + O(n^{-3/2})$ y sustituyendo en (5.34) obtenemos

$$\mathrm{E}(e^{tT_n}) = \big(1 + \frac{t^2}{2n} + O(n^{-3/2})\big)^n \to e^{t^2/2}$$

que es la f.g.m. de la distribución normal típica. ∎

Ejercicios

1. Un jugador lanza dos monedas. Si ambas caen águila, gana \$10, si śolo hay un águila, gana \$2, mientras que si no hay ninguń águila pierde \$12. Determine el valor esperado de la ganancia para un lanzamiento de las dos monedas.

2. En una lotería se venden 1,000,000 de boletos de \$10 cada uno. Hay un primer premio de \$3,000,000, 10 premios de \$200,000, 100 premios de \$2,000, 1,000 premios de \$100 y 10,000 boletos reciben un reembolso del costo del boleto. Halle el valor esperado de la ganancia neta por boleto.

3. Extraemos al azar cartas de un juego de barajas con reposición hasta obtener un as y sea X el número de cartas extraídas. Halle $\mathrm{E}(X)$ y $\mathrm{Var}(X)$.

4. Lanzamos una moneda repetidamente hasta obtener dos águilas o dos soles, lo que ocurra primero. Sea X el número de lanzamientos de la moneda. Halle el valor esperado y la varianza de X.

5. Sea X una variable aleatoria con distribución de Poisson de parámetro λ. Halle el valor esperado de la variable $Y = e^X$.

6. Sean X, Y variables aleatorias cada una de las cuales toma únicamente dos valores distintos. Demuestre que X e Y son independientes si y sólo si $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.

7. Lanzamos dos dados, si el lanzamiento es un *doble* (dos caras iguales) los dados se vuelven a lanzar y así hasta que las dos caras que se obtienen sean distintas. Sea X el número de lanzamientos necesarios para que esto ocurra. Halle el valor esperado y la varianza de X. Halle también el valor esperado y la varianza de 2^X

8. En una bolsa hay cinco boletos que corresponden a un premio de \$1,000 y cuatro premios de \$20. Cinco personas sacan, sucesivamente, un boleto al azar y sea X_i el premio que le corresponde a la i-ésima persona en sacar un boleto. Calcule $\mathrm{E}(X_1)$ y $\mathrm{Var}(X_1)$, (b) $\mathrm{E}(X_3)$ y $\mathrm{Var}(X_5)$, (c) $\mathrm{E}(X_1 + \cdots + X_5)$ y $\mathrm{Var}(X_1 + \cdots + X_5)$.

9. Una caja contiene 5 bolas numeradas del 1 al 5. Se extraen 200 bolas con reposición y sea X_i el número en la i-ésima bola extraída, Y la suma de los 200 números obtenidos y Z el promedio de los 200 números. Halle (a) $\mathrm{E}(X)$ y $\mathrm{Var}(X)$, (b) $\mathrm{E}(Y)$ y $\mathrm{Var}(Y)$, (c) $\mathrm{E}(Z)$ y $\mathrm{Var}(Z)$.

10. Lanzamos un par de dados, sea X_1 el número que aparece en el primero y X_2 el del segundo. Definimos $Y = X_1 + X_2$, $Z = XY$. Halle (a) $\mathrm{E}(X_1)$, (b) $\mathrm{E}(X_2^2)$, (c) $\mathrm{E}(Y)$, (d) $\mathrm{E}(Z)$, (e) $\mathrm{E}(YZ)$, (f) $\mathrm{E}(Y^2)$, (g) $\mathrm{E}(Z^2)$.

11. En un concurso hay cinco cajas idénticas cerradas. En una de ellas hay \$100,000, otra contiene \$10,000, una tercera \$1,000, la cuarta \$100 y la última tiene un signo de ALTO. El concursante escoge una caja y gana el contenido. Este proceso se repite hasta que salga el signo de ALTO, momento en el cual el concurso termina. Halle el valor esperado de la cantidad que gana el concursante.

12. Una máquina produce objetos que son defectuosos con probabilidad 0.01 y cuando esto ocurre, la máquina se detiene y es ajustada. Halle el valor promedio del número de objetos buenos producidos entre dos objetos defectuosos.

13. Calcule media y varianza para las distribuciones de los ejercicios 1, 2, 4, 5, 7, 8, 9, 10, 11, 18 y 19 del Capítulo 4.

14. Sea X una variable aleatoria con $\mathrm{E}(X) = 2$, $\mathrm{Var}(X) = 1$, $\mathrm{E}(X^4) = 34$. Calcular media y varianza para las siguientes variables aleatorias: $U = 2X + 1$, $V = X^2$, $Z = -X^2 + 2$.

15. Las variables X, Y son independientes y $\mathrm{E}(X) = -3$, $\mathrm{E}(X^4) = 100$, $\mathrm{E}(Y) = 4$, $\mathrm{E}(Y^4) = 500$, $\mathrm{Var}(X) = 0.5$, $\mathrm{Var}(Y) = 2$. Calcular la media y varianza para las variables $U = 3X - 2Y$ y $V = X^2 - Y^2$.

16. Suponga que X, Y tienen igual media e igual varianza y además son independientes. Demuestre que
$$\mathrm{E}((X - Y)^2) = 2\,\mathrm{Var}(X).$$

17. Suponga que X, Y tienen igual varianza. Demuestre que
$$\mathrm{E}((X + Y)(X - Y)) = \mathrm{E}(X + Y)\,\mathrm{E}(X - Y).$$

18. Suponga que X, Y son independientes y ambas tienen media 3 y varianza 1. Halle la media y varianza de $X+Y$ y XY.

19. Demuestre que
$$\mathrm{Var}(X+Y) + \mathrm{Var}(X-Y) = 2\,\mathrm{Var}(X) + 2\,\mathrm{Var}(Y).$$

20. Sea X una variable con valor esperado finito $\mathrm{E}(X)$. Demuestre que
$$(\mathrm{E}\,X)^2 \leq \mathrm{E}(X^2)$$

21. Sean $X_1, \dots, X_n$ variables aleatorias independientes idénticamente distribuidas con media μ y varianza σ^2. Definimos la media muestral por $\overline{X} = n^{-1}\sum_{k=1}^n X_k$. Demuestre que
$$\mathrm{Cov}(\overline{X}, X_k - \overline{X}) = 0, \quad 1 \leq k \leq n.$$

22. Se considera el siguiente juego de azar entre dos jugadores. El primero elige un punto X al azar en el intervalo $(0,2)$ mientras que el segundo elige un punto Y en el intervalo $(1,3)$, también con distribución uniforme. Suponemos que X e Y son variables aleatorias independientes. Entonces

   – Si $X < Y$, el primer jugador paga $a(Y-X)$ unidades al segundo.

   – Si $X \geq Y$, el segundo jugador paga $b(X-Y)$ unidades al primero,

   donde a y b son constantes positivas.

   a. Hallar la relación b/a para que el juego sea equitativo (esto significa que la esperanza matemática de la ganancia de cada jugador es igual a cero).

   b. Con la relación b/a calculada en la parte anterior, calcular la varianza de la ganancia del primer jugador.

23. Sean X, Y variables aleatorias independientes, ambas con distribución de Bernoulli con probabilidad de éxito $1/2$. Demuestre que $X+Y$ y $|X+Y|$ son dependientes pero no están correlacionadas.

24. Sea X una variable aleatoria con distribución geométrica de parámetro p y M un número entero positivo. Calcular la esperanza matemática de la variable aleatoria $Y = \min\{X, M\}$.

25. (a) Se lanza una moneda balanceada repetidamente. Sea ν la variable aleatoria que indica el número de veces seguidas que ocurre lo mismo que en el primer lanzamiento. (Por ejemplo, si A es águila y S es sol, con la sucesión de resultados $AAASSAS\dots$ se tiene $\nu = 3$ mientras que con la sucesión $SSSSSASAS\dots$ se tiene $\nu = 5$). Calcular $\mathrm{E}(\nu)$ y $\mathrm{Var}(\nu)$.

   (b) Rehacer el cálculo de la parte (a) suponiendo que, en lugar de una moneda perfecta, se dispone de una moneda tal que la probabilidad de águila en un lanzamiento es igual a p. ¿Qué ocurre si $p = 0$ ó $p = 1$?

26. (a) Una ciudad está dividida en cuatro zonas, con número respectivo de habitantes H_1, H_2, H_3 y H_4. Supongamos que el promedio $m = (H_1 + H_2 + H_3 + H_4)/4$ de habitantes de las zonas, es de 1000 personas y sea
$$\sigma^2 = \frac{1}{4}\sum_{i=1}^{4}(x_i - m)^2.$$

   Se eligen al azar en un muestreo sin reposición dos de las cuatro zonas y se cuenta el número de habitantes obtenidos en cada una de ellas, que denominamos X_1 y X_2. Mostrar que
$$\mathrm{E}(X_1 + X_2) = 2000$$
$$\mathrm{Var}(X_1 + X_2) = \frac{4}{3}\sigma^2$$

(b) Generalizar al caso en el cual, en lugar de cuatro, la ciudad está dividida en n zonas, y en lugar de seleccionar al azar dos de ellas, se seleccionan r. Se obtiene

$$\mathrm{E}(X_1 + X_2 + \cdots + X_r) = mr,$$

donde m es el promedio del número de habitantes de las zonas, y

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_r) = \frac{r(n-r)}{n-1}\sigma^2.$$

27. Se escriben n cartas y sus respectivos sobres, y se ensobran las cartas al azar, de modo que la probabilidad de cualquiera de las posibles permutaciones de las cartas, es la misma. Calcular la esperanza y la varianza del número de cartas que se ensobran correctamente.

Sugerencia. Escribir Y como $Y = \sum_{i=1}^{n} X_i$, donde

$$X_i = \begin{cases} 1 & \textit{si la i-ésima carta va en el i-ésimo sobre,} \\ 0 & \textit{en caso contrario.} \end{cases}$$

28. En una bolsa hay n tarjetas numeradas de 1 a n. Se extraen las tarjetas sucesivamente con reposición. (a) ¿Cuál es el valor esperado del número de extracciones hasta repetir el primer número? (b) ¿Cuál es el valor esperado del número de extracciones hasta que ocurra la primera repetición?

29. Se considera un conjunto de n personas. Calcular el valor esperado del número de días del año en que cumplen años exactamente k de ellos. Suponga que el año tiene 365 días, que todas las distribuciones posibles son igualmente probables y que $n \geq k$.

30. Una persona con n llaves trata de abrir una puerta probando las llaves sucesiva e independientemente. Calcule la esperanza y la varianza del número ν de intentos requeridos hasta encontrar la llave correcta, suponiendo que ésta es una sola, en las dos situaciones siguientes.

    (a) Si la selección de las llaves es con reposición, es decir, que una llave inservible no es quitada del lote, una vez probada. (b) Si la selección es sin reposición.

31. Un sistema permite establecer comunicaciones de acuerdo al diagrama 5.1. Cada bloque rectangular demora una unidad de tiempo y T es la variable aleatoria que indica el tiempo que se demora en establecer una comunicación buena.

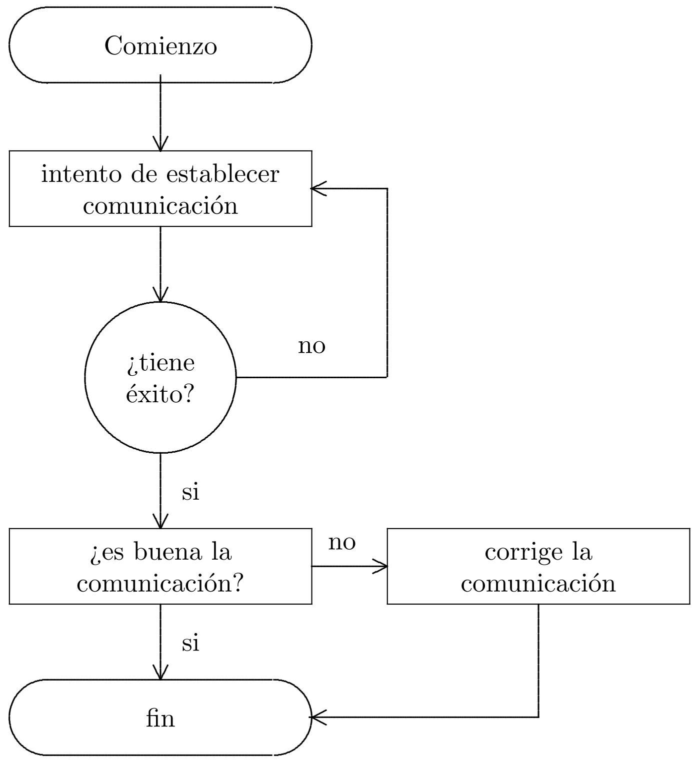


Diagrama 5.1

Se supone que los intentos para establecer comunicación son independientes y que la probabilidad de éxito en cada uno es p. Una vez establecida, la comunicación puede ser buena o mala, y la probabilidad de que sea buena es b. Si es mala, en una operación más se corrige y deviene buena. Calcular la función de probabilidad de T y su esperanza matemática.

32. Se lanza un dado n veces. Sea X_i el número de veces que se obtiene el resultado i. Calcular la covarianza de X_1 y X_6.

33. En un estanque en el que originalmente hay n peces, se pesca sucesivamente por medio de redes, que retiran $n_1,\ n_2,\dots,$ donde n_k denota el número (aleatorio) de peces extraídos la k-ésima vez. Suponiendo que la probabilidad de que cada pez sea capturado es p, calcular la esperanza matemática del número de peces extraídos en la n-ésima ocasión.

34. Una bolsa contiene bolas numeradas de 1 a N. Sea X el mayor número obtenido en n extracciones al azar y con reposición, donde n es un número fijo.
    (a) Hallar la distribución de probabilidad de X.
    (b) Calcular $\mathrm{E}(X)$ y probar que si N es grande, $\mathrm{E}(X)$ es aproximadamente igual a $N/(n+1)$.
    (c) Hallar $\mathrm{E}(X^2)$ y deducir una expresión asintótica para $\mathrm{Var}(X)$ cuando $N\to\infty$.

35. Calcular $\mathrm{E}(X^3)$ donde X tiene distribución binomial con parámetros (n,p).

36. Sea X, Y variables aleatorias discretas con función de probabilidad
$$p_{ij}=\frac{1}{1+2n(n+1)},\quad \text{para } |i-j|\le n, |i+j|\le n.$$
Demuestre que X e Y son independientes y que $\mathrm{Cov}(X,Y)=0$.

37. Sea X, Y variables independientes e idénticamente distribuidas con distribución normal $\mathcal{N}(\mu, \sigma^2)$. Halle media y varianza de la variable $Z = \max(X, Y)$. (Ayuda: si x, y son números reales demuestre que $2\max(x, y) = |x - y| + x + y$).

38. Sea X una variable con distribución $\mathcal{N}(\mu, \sigma^2)$. Halle media y varianza de $|X - c|$ cuando (a) c es una constante dada, (b) $\sigma = \mu = c = 1$, (c) $\sigma = \mu = 1, c = 2$.

39. Calcule media y varianza para las distribuciones de los ejercicios 21, 22, 24, 25, 26, 29 y 30 del Capítulo 4.

40. En el ejercicio 31 del capítulo 4 vimos que de acuerdo a la ley de Maxwell, la velocidad v de una molécula de gas de masa m a temperatura absoluta T es una variable aleatoria con densidad

$$f_v(x) = \frac{4}{\sqrt{\pi}} \frac{1}{\alpha^3} x^2 e^{-x^2/k^2}$$

para $x > 0$ con $\alpha = (2kT/m)^{1/2}$ y k es la constante de Boltzman. Halle media y varianza de (a) la velocidad v de la molécula, (b) la energía cinética $E = mv^2/2$ de la molécula.

41. Sea X una variable aleatoria con distribución uniforme en $(-\frac{\pi}{2}, \frac{\pi}{2})$. Calcular $\mathrm{E}(Y)$ cuando (a) $Y = \operatorname{sen} X$. (b) $Y = \cos X$ (c) $Y = 3X^2 + 2$ (d) $Y = 1/|X|^\alpha$ En el último caso, ¿para cuáles valores de α se tiene que $\mathrm{E}(Y) < \infty$?

42. Calcular la esperanza y la varianza de las distribuciones cuyas densidades se indican a continuación:

$$(a)\ f(x) = \begin{cases} 2x & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases} \qquad (b)\ f(x) = \begin{cases} |x| & \text{si } x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

43. Sea X una variable con distribución exponencial. (a) Halle $P(\operatorname{sen} X > 1/2)$, (b) $\mathrm{E}(X^n)$ para $n \geq 1$.

44. La duración T de cierto tipo de llamadas telefónicas satisface la relación

$$P(T > t) = ae^{-\lambda t} + (1 - a)e^{-\mu t}, \quad t \geq 0,$$

donde $0 \leq a \leq 1, \lambda > 0, \mu > 0$ son constantes. halle media y varianza para T.

45. El tiempo de vida de cierto componente electrónico tiene distribución exponencial con parámetro $\lambda = 0.01$. Calcule el tiempo de vida promedio. Si el aparato ha durado 50 horas, ¿cuál es el valor esperado del tiempo de vida que le queda?

46. Escogemos n puntos al azar en $[0, 1]$ con distribución uniforme. Sea $m_n = \min\{X_1, \dots, X_n\}$, $M_n = \max\{X_1, \dots, X_n\}$, $R_n = M_n - m_n$. Halle el valor esperado de estas tres variables aleatorias.

47. La proporción de tiempo que una máquina trabaja durante una semana de 40 horas es una variable aleatoria con densidad $f(x) = 2x$ para $0 \leq x \leq 1$. Halle $\mathrm{E}(X)$ y $\mathrm{Var}(X)$. La ganancia semanal Y para esta máquina está dada por $Y = 200X - 60$; determine $\mathrm{E}(Y)$ y $\mathrm{Var}(Y)$.

48. Sea X una variable aleatoria con densidad $f(x) = Cx^{a-1}e^{-x^a}$, $x \geq 0$. Halle el valor de C y calcule $\mathrm{E}(X)$ y $\mathrm{Var}(X)$.

49. Sea X una variable aleatoria con densidad $f(x) = 6x(1 - x)$ para $0 < x < 1$. Compruebe que f es una densidad y obtenga media y varianza para esta distribución.

50. Para ciertas muestras de minerales la proporción de impurezas por muestra X es una variable aleatoria con densidad dada por $f(x) = 1.5x^2 + x$ para $0 \leq x \leq 1$. El valor en pesos de cada muestra es $Y = 5 - 0.5X$. Encuentre el valor esperado y la varianza de Y

51. Sea X una variable aleatoria con densidad $f(x) = |\operatorname{sen}(x)|/4$ para $0 \le x \le 2\pi$. Calcule $\mathrm{E}(X)$.

52. La proporción de tiempo por día en que todas las cajas registradoras están ocupadas a la salida de cierto supermercado es una variable aleatoria X con una densidad $f(x) = Cx^2(1-x)^4$ para $0 \le x \le 1$. Determine el valor de C y $\mathrm{E}(X)$.

53. **Distribución de Pareto** La distribución de Pareto con parámetros r y A, $r > 0,\ A > 0$, es aquella que tiene la densidad
$$f(x) = \begin{cases} \frac{rA^r}{x^{r+1}} & \text{si } x \ge A, \\ 0 & \text{si } x < A. \end{cases}$$
(a) Calcular la esperanza y la varianza de esta distribución de probabilidad.

(b) Calcular y graficar la función
$$Q(y) = F(\mu + y\sigma) - F(\mu - y\sigma)$$
para $y \ge 0$ donde μ es la esperanza y σ^2 la varianza halladas en a, y F es la función de distribución de probabilidad de Pareto, en el caso $A = 1,\ r = 3$.

54. **Distribución Beta** Sea X una variable aleatoria con densidad
$$f(x) = Cx^{a-1}(1-x)^{b-1}, \quad 0 \le x \le 1.$$
Halle el valor de C y calcule $\mathrm{E}(X)$ y $\mathrm{Var}(X)$.

55. En un proceso de manufactura se ensamblan cuantro componente sucesivamente de longitudes X_1, X_2, X_3 y X_4, respectivamente con $\mathrm{E}(X_1) = 20,\ \mathrm{E}(X_2) = 30,\ \mathrm{E}(X_3) = 40,\ \mathrm{E}(X_4) = 60$ y $\mathrm{Var}(X_j) = 4$ para $j = 1, 2, 3, 4$. Halle la media y varianza para la longitud total $L = X_1 + X_2 + X_3 + X_4$ (a) si las longitudes de los componentes no están correlacionadas, (b) si $\rho(X_j, X_k) = 0.2$ para $1 \le j < k \le 4$.

56. **Distribución de Rayleigh** Una variable tiene distribución de Rayleigh si su densidad es
$$f(x) = \frac{2x}{\theta} e^{-x^2/\theta}$$
para $x > 0$. Calcule $\mathrm{E}(X)$, $\mathrm{Var}(X)$ y obtenga la densidad de $Y = X^2$.

57. **Distribución de Laplace** Una variable X tiene distribución de Laplace si su densidad está dada por
$$f(x) = \frac{1}{2\beta} \exp\Big(-\frac{|x-\alpha|}{\beta}\Big), \quad x \in \mathbb{R},$$
para ciertas constantes α y $\beta > 0$. Halle media y varianza para una variable con esta distribución.

58. El tiempo de vida de ciertos focos tiene distribución de Rayleigh con función de distribución
$$F(x) = 1 - \exp\Big(\frac{x^2}{300}\Big), \quad \text{para } x > 0.$$
Hallar la densidad de la variable aleatoria $Y = X^2$ y su valor esperado.

59. Sea X una variable aleatoria con tercer momento finito $\mathrm{E}(|X|^3) < \infty$). Definimos el coeficiente de asimetría de X, $\alpha(X)$ por
$$\alpha(X) = \frac{\mathrm{E}(X-\mu)^3}{\sigma^3}$$
donde $\mu = \mathrm{E}(X)$ y $\sigma^2 = \mathrm{Var}(X)$.

(a) Demuestre que para cualquier variable X se tiene que

$$\alpha(X) = \frac{\mathrm{E}(X^3) - 3\mu\,\mathrm{E}(X^2) + 2\mu^3}{\sigma^3}$$

Halle $\alpha(X)$ si X tiene distribución

(b) Bernoulli con parámetro p.

(c) Poisson con parámetro λ.

(d) Geométrica con parámetro p.

(e) Demuestre que si $X_1, \dots, X_n$ son variables aleatorias independientes e idénticamente distribuidas entonces

$$\alpha(X_1 + \cdots + X_n) = \frac{\alpha(X_1)}{\sqrt{n}}.$$

(f) Halle $\alpha(X)$ si X tiene distribución binomial.

60. Sea $\{p_n\}$ la distribución de probabilidad de la variable aleatoria X, concentrada en los enteros no-negativos, es decir que

$$P(X = n) = p_n, \qquad n = 0, 1, 2 \dots \qquad \sum_{n=0}^{\infty} p_n = 1$$

Se define la función generatriz de la distribución definida por $\{p_n\}$ mediante

$$g(z) = \sum_{n=0}^{\infty} p_n z^n = \mathrm{E}(z^X). \tag{5.37}$$

(a) Probar que la serie (5.40) converge, por lo menos, cuando $|z| \le 1$, es decir que su radio de convergencia es mayor o igual que 1.

(b) Observar que la función $g(z)$ determina unívocamente las probabilidades $\{p_n\}$.

(c) Calcular la función generatriz de la distribución binomial con parámetros n y p.

(d) Calcular la función generatriz de la distribución geométrica.

(e) Calcular la función generatriz de la distribución de Poisson de parámetro λ.

(f) Probar que

$$\mathrm{E}(X) = g'(1), \qquad \mathrm{E}(X(X-1)) = g''(1), \dots$$
$$\mathrm{E}(X(X-1)\dots(X-k+1)) = g^{(k)}(1).$$

bajo la hipótesis de que las esperanzas que allí figuran son finitas. ¿Qué ocurre si las esperanzas son infinitas?

(g) Sea g la función generatriz de la variable aleatoria X y h la función generatriz de la variable aleatoria Y. Probar que si X e Y son independientes, entonces $g(z)h(z)$ es la función generatriz de $X + Y$.

(h) Utilizando funciones generatrices, mostrar que $X_1, X_2, \dots, X_n$ son variables aleatorias independientes, cada una con distribución de Poisson, con parámetros respectivos $\lambda_1, \lambda_2, \dots, \lambda_n$, entonces $X_1 + X_2 + \cdots + X_n$ también tiene distribución de Poisson y su parámetro es $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

(i) Hallar la distribución de probabilidad de $X_1 + X_2 + \cdots + X_n$ donde las X_i, $i = 1, \dots, m$ son variables aleatorias independientes, cada una con distribución binomial, y los parámetros respectivos son (n_i, p), $i = 1, \dots, m$.

Apéndice

5.10.1. Definición del Valor Esperado en el Caso General.

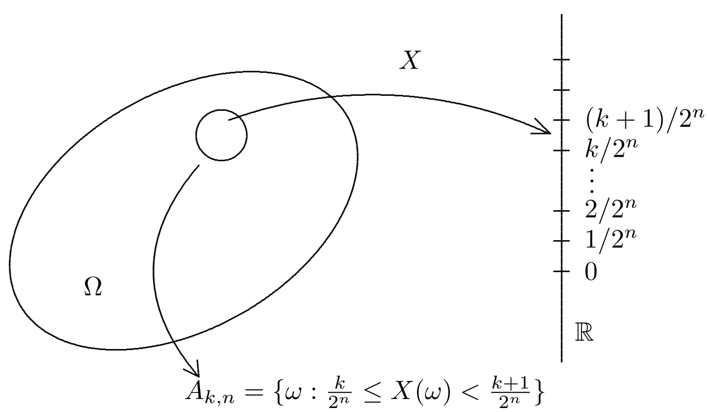
Hasta el momento hemos definido y calculado el *valor esperado* o *esperanza matemática* de una variable aleatoria, solamente en el caso en el que la distribución de probabilidad de la misma es discreta. A continuación nos proponemos hacer lo mismo en el caso general, es decir, para variables no necesariamente discretas. El estudio del tema requiere una extensión bastante mayor de la que estamos en condiciones de asignarle aquí. Aún así, el lector que no esté especialmente interesado, en una primera instancia, en contar con una fundamentación sobre el tema, puede pasar directamente a la sección 5.5 en la cual se estudia el caso de las variables aleatorias que tienen densidad, y adoptar la fórmula (5.21) como una definición. Así mismo, en una primera instancia, puede saltar las demostraciones de las propiedades básicas de la esperanza matemática en el caso general, que están marcadas con una estrella *.

Como criterio general, la manera de definir la esperanza matemática y sus propiedades, consiste en aproximar una variable aleatoria cualquiera por variables discretas, y utilizar la definición y propiedades de la esperanza para estas últimas, que ya conocemos de nuestro estudio anterior. Este proceso de aproximación tiene algunas dificultades, que aparecen en las secciones mencionadas, marcadas también con una estrella *.

Sea X una variable aleatoria no negativa. Para cada número natural n, $n = 1, 2, \dots$ definimos la variable aleatoria aproximante X_n mediante

$$X_n(\omega) = \frac{k}{2^n} \quad \text{si} \quad \frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n} \quad (k = 0, 1, 2, \dots) \tag{5.38}$$

Para entender esta definición procedemos de la siguiente forma: dividimos la semirecta $[0, \infty)$ de los números reales no negativos, en intervalos de logitud $1/2^n$, a partir de 0.



Para cada intervalo

$$\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right),$$

consideramos la preimagen del intervalo por la variable aleatoria X, recordando que X es una función de Ω en $\mathbb{R}$. Dicha preimagen es

$$A_{k,n} = \left\{\omega : \frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}\right\},$$

es decir, el conjunto de puntos de Ω cuya imagen por X está en

$$\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right).$$

Entonces, definimos X_n en $A_{k,n}$ igual a $k/2^n$. Por lo tanto, una manera de escribir X_n es

$$X_n(\omega) = \sum_{k=0}^{\infty} \frac{k}{2^n} \mathbf{1}_{A_{k,n}}(\omega) = \sum_{k=0}^{\infty} \frac{k}{2^n} \mathbf{1}_{\{\omega: \frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}\}}(\omega)$$

donde, como es habitual, la función indicatriz $\mathbf{1}_{A_{k,n}}(\omega)$ está definida por

$$\mathbf{1}_{A_{k,n}}(\omega) = \begin{cases} 1 & \text{si } \omega \in A_{k,n}, \\ 0 & \text{si } \omega \notin A_{k,n}. \end{cases}$$

Las siguientes observaciones son consecuencia de la definición de X_n:

1. Para cada n, X_n es una variable aleatoria discreta y no negativa. Esto es claro ya que X_n sólo puede tomar los valores
$$0, \frac{1}{2^n}, \frac{2}{2^n}, \frac{3}{2^n}, \dots, \frac{k}{2^n}, \dots$$
que es un conjunto numerable de números mayores o iguales que cero.

2. Para cada $\omega \in \Omega$, $X_n(\omega)$ es una sucesión creciente (en sentido amplio), de números reales.

   En efecto, observamos que si aumentamos n en una unidad, es decir, en lugar de n ponemos $n+1$, lo que hacemos es subdividir cada uno de los intervalos
$$\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$$
en dos partes de igual longitud:
$$I_1 = \left[\frac{2k+1}{2^{n+1}}, \frac{k+1}{2^n}\right) \qquad I_2 = \left[\frac{k}{2^n}, \frac{2k+1}{2^{n+1}}\right).$$
Si tenemos que $X_n(\omega) = k/2^n$, lo que quiere decir que $\omega \in A_{k,n}$, o aún, de manera equivalente, que
$$\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n},$$
tenemos dos alternativas posibles (ver figura 5.2):

   En primer lugar, que $X(\omega) \in I_1$, lo que implica que
$$X_{n+1}(\omega) = \frac{2k+1}{2^{n+1}} > \frac{k}{2^n} = X_n(\omega).$$
En segundo lugar, que $X(\omega) \in I_2$, lo que implica que
$$X_{n+1}(\omega) = \frac{k}{2^n} = X_n(\omega).$$

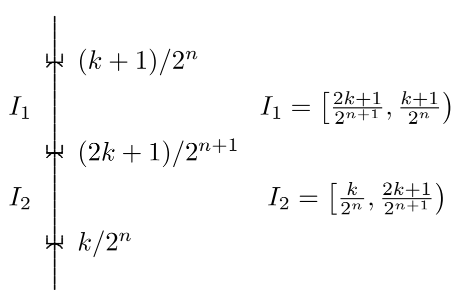


Figura 5.2

En cualquiera de los dos casos se verifica que $X_{n+1}(\omega) \geq X_n(\omega)$.

3. Se tiene
$$0 \leq X(\omega) - X_n(\omega) < \frac{1}{2^n} \qquad \text{para todo } \omega \in \Omega. \tag{5.39}$$
De aquí se deduce que
$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \quad \text{para todo } \omega \in \Omega.$$

4. Puesto que X_n es una variable aleatoria discreta y no-negativa, está definida su esperanza matemática
$$\begin{aligned} \mathrm{E}(X_n) &= \sum_{k=0}^{\infty} \frac{k}{2^n} P(X_n = \frac{k}{2^n}) = \sum_{k=0}^{\infty} \frac{k}{2^n} P(A_{k,n}) \\ &= \sum_{k=0}^{\infty} \frac{k}{2^n} P(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}). \end{aligned} \tag{5.40}$$
Observamos que, cuando definimos la esperanza matemática de una variable aleatoria discreta, dijimos que ésta existe cuando la serie que aparece en la fórmula (5.18) es convergente. En caso contrario, y para simplificar la exposición que sigue, vamos a convenir en que $\mathrm{E}(X_n)$ es $+\infty$. Esta convención es útil a los fines de la presentación del tema.

Estamos ahora en condiciones de definir, de manera general, la esperanza matemática de una variable aleatoria. En una primera instancia consideremos una variable no-negativa X, y la sucesión aproximante X_n definida en (5.38). Tenemos dos alternativas posibles:

- O bien, para algún valor de n, $\mathrm{E}(X_n) = +\infty$, en cuyo caso definimos $\mathrm{E}(X) = +\infty$.
- O bien, para todo $n = 1, 2, \ldots$, $\mathrm{E}(X_n) < \infty$. En este caso, la sucesión de números reales no-negativos $\mathrm{E}(X_n)$ es monótona creciente, ya que, de acuerdo a la propiedad 8 de la sección 5.1, del hecho de que $X_n \leq X_{n+1}$ se deduce que $\mathrm{E}(X_n) \leq \mathrm{E}(X_{n+1})$. Pero, una sucesión creciente de números reales tiene límite, finito o infinito. Definimos entonces
$$\mathrm{E}(X) = \lim_{n\to\infty} \mathrm{E}(X_n) \tag{5.41}$$
y esto asigna un valor a la esperanza matemática $\mathrm{E}(X)$, para variables aleatorias no-negativas.

Consideremos ahora una variable aleatoria cualquiera X. Definimos otras dos variables X^+ y X^-, denominadas respectivamente, la parte positiva y la parte negativa de X, de la siguiente forma:
$$X^+(\omega) = \begin{cases} X(\omega) & \text{si } X(\omega) \geq 0, \\ 0 & \text{si } X(\omega) < 0. \end{cases}$$
$$X^-(\omega) = \begin{cases} 0 & \text{si } X(\omega) \geq 0, \\ -X(\omega) & \text{si } X(\omega) < 0. \end{cases}$$
Se verifica fácilmente que
$$\begin{aligned} X(\omega) &= X^+(\omega) - X^-(\omega) \\ |X(\omega)| &= X^+(\omega) + X^-(\omega). \end{aligned}$$

En efecto, si $X(\omega) \geq 0$, de acuerdo a las definiciones

$$X(\omega) = X^+(\omega) \qquad \text{y} \qquad |X(\omega)| = X^+(\omega),$$

mientras que si $X(\omega) < 0$,

$$X(\omega) = -X^-(\omega) \qquad \text{y} \qquad |X(\omega)| = X^-(\omega),$$

y en ambos casos se satisfacen ambas igualdades.

Ahora bien, X^+ y X^- son ambas no-negativas y podemos definir $\mathrm{E}(X^+)$ y $\mathrm{E}(X^-)$ utilizando (5.19). Basándonos en esto, definimos $\mathrm{E}(X)$ de la siguiente manera:

1. Si $\mathrm{E}(X^+) < \infty$ y $\mathrm{E}(X^-) < \infty$ definimos $\mathrm{E}(X) = E(X^+) - E(X^-)$.

2. Si $\mathrm{E}(X^+) = +\infty$ y $\mathrm{E}(X^-) < \infty$ definimos $\mathrm{E}(X) = +\infty$.

3. Si $\mathrm{E}(X^+) < \infty$ y $\mathrm{E}(X^-) = +\infty$ definimos $\mathrm{E}(X) = -\infty$.

4. Si $\mathrm{E}(X^+) = +\infty$ y $\mathrm{E}(X^-) = +\infty$ diremos que el valor esperado de la variable aleatoria X no está definido.

Lo primero que verificaremos es que nuestra nueva definición coincide con la anterior (sección 5.1) en el caso de variables aleatorias discretas. Para ello, basta con hacerlo para variables aleatorias que son no-negativas ya que, en caso contrario, podemos considerar por separado X^+ y X^-, que si lo son, y luego sumar.

Sea entonces X una variable aleatoria discreta no-negativa, es decir que

$$P(X = x_k) = p_k, \qquad x_k \geq 0, \quad k = 1, 2, \dots \qquad \sum_{k=1}^{\infty} p_k = 1.$$

Vamos a aplicarle a esta variable aleatoria X nuestra nueva definición. Supongamos primero que

$$\sum_{k=1}^{\infty} x_k p_k < \infty.$$

Para cada n, $n = 1, 2, \dots$ definimos la variable aleatoria aproximante X_n correspondiente a X de acuerdo a (5.16). Sabemos, por la desigualdad (5.17), que

$$0 \leq X - X_n < \frac{1}{2^n}$$

y puesto que X y X_n son variables aleatorias discretas, aplicando la propiedad 8 de la sección 5.1 para valores esperados de variables aleatorias discretas, se tiene

$$0 \leq \sum_{k=1}^{\infty} x_k p_k - \mathrm{E}(X_n) \leq \frac{1}{2^n}$$

Haciendo tender n a ∞, el tercer miembro tiende a cero y $\mathrm{E}(X_n)$ converge a $\mathrm{E}(X)$ (de acuerdo a nuestra nueva definición). Por lo tanto

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} x_k p_k,$$

o sea que las definiciones vieja y nueva de la esperanza matemática de X coinciden.

Si

$$\sum_{k=1}^{\infty} x_k p_k$$

es una serie divergente, dejamos como ejercicio para el lector probar que $E(X) = +\infty$, de acuerdo a la nueva definición. (Sugerencia: muestre que, en este caso, $E(X_n) = +\infty$ para cualquier valor de n).

5.10.2. Esperanza Matemática de Variables Aleatorias con Densidad.

Teorema 5.4 *Sea X una variable aleatoria, F_X su función de distribución y supongamos que ésta posee una densidad f_X. El valor esperado de X existe y es finito si y sólo si la integral*

$$\int_{-\infty}^{\infty} |x| f_X(x) dx \tag{5.42}$$

es finita, y en este caso $\mathrm{E}(X)$ *se calcula mediante la fórmula*

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{5.43}$$

La validez de la fórmula anterior es completamente general cuando la integral en (5.16) es convergente. A continuación procedemos a probar la fórmula (5.17) cuando la función f_X es bastante regular como para que las integrales que aquí aparecen tengan sentido para nosotros, es decir, que puedan ser manejadas con las herramientas usuales del Cálculo. Por ejemplo, ese es el caso si f_X es una función continua, salvo a lo sumo en un número finito de puntos. El lector no especialmente interesado puede saltarse la demostración de (5.21) en una primera lectura. De todos modos, la aplicación de (5.21) es directa, como veremos en los ejemplos, cuando f_X es una función conocida.

Demostración del teorema 5.4

Consideremos X^+ y X^-, la parte positiva y la parte negativa de la variable aleatoria X y sus variables aleatorias aproximantes X_n^+ y X_n^-. Tenemos

$$\begin{aligned}\mathrm{E}(X^+) &= \sum_{k=0}^{\infty} \frac{k}{2^n} P(\frac{k}{2^n} \leq X^+ < \frac{k+1}{2^n}) = \sum_{k=0}^{\infty} \frac{k}{2^n} P(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) \\ &= \sum_{k=0}^{\infty} \frac{k}{2^n} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f_X(x)\, dx = \sum_{k=0}^{\infty} \left[\int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} x f_X(x)\, dx + \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} (\frac{k}{2^n} - x) f_X(x)\, dx \right]\end{aligned}$$

Observe el lector que hemos tenido en cuenta que, dado que f_X es la densidad de X,

$$P(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) = \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f_X(x)\, dx.$$

Por lo tanto,

$$\mathrm{E}(X^+) - \int_0^{+\infty} x\, f_X(x)\, dx = \sum_{k=0}^{\infty} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} (\frac{k}{2^n} - x) f_X(x)\, dx$$

y se deduce que

$$\begin{aligned}\left|\mathrm{E}(X^+) - \int_0^{+\infty} x\, f_X(x)\, dx\right| &\leq \sum_{k=0}^{\infty} \frac{1}{2^n} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f_X(x)\, dx \\ &= \frac{1}{2^n} \int_0^{+\infty} f_X(x)\, dx \leq \frac{1}{2^n}\end{aligned}$$

que converge a cero cuando n tiende a $+\infty$.

En el razonamiento anterior hemos utilizado la desigualdad triangular, la acotación

$$\left| \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} (\frac{k}{2^n} - x) f_X(x)\, dx \right| \leq \frac{1}{2^n} \int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} f_X(x)\, dx,$$

y el hecho de que

$$\int_0^{+\infty} f_X(x)\,dx \le \int_{-\infty}^{+\infty} f_X(x)\,dx = 1.$$

Se concluye entonces que

$$E(X^+) = \lim_{n\to\infty} \mathrm{E}(X_n^+) = \int_0^{+\infty} x\, f_X(x)\,dx \tag{5.44}$$

Dejamos como ejercicio la verificación de este resultado si $\mathrm{E}(X^+) = +\infty$.

De manera enteramente similar a lo que hicimos para probar (5.44) se demuestra que

$$E(X^-) = \lim_{n\to\infty} \mathrm{E}(X_n^-) = \int_0^{+\infty} (-x)\, f_X(x)\,dx \tag{5.45}$$

De (5.44) y (5.45) resulta inmediatamente (5.17). Observamos además, que $\mathrm{E}(X^+)$ y $\mathrm{E}(X^-)$ son ambas finitas, y por lo tanto también lo es $\mathrm{E}(|X|)$, si y sólo si las dos integrales en (5.44) y (5.45) son finitas, es decir, si y sólo si

$$\int_{-\infty}^{+\infty} |x| f_X(x)\,dx < \infty$$

■

5.11. Cambio de Variables. Momentos de Orden Superior.

Sea $f(x_1,\dots,x_n)$ la densidad conjunta de las variables aleatorias $X_1,\dots,X_n$ y

$$g:\mathbb{R}^n \to \mathbb{R}$$

una función de n variables que tiene alguna regularidad. Consideremos la variable aleatoria

$$Y = g(X_1, X_2, \dots, X_n).$$

Entonces, el valor esperado de Y se puede calcular mediante la fórmula

$$\mathrm{E}(Y) = \int\cdots\int_{\mathbb{R}^n} g(x_1,\dots,x_n) f(x_1,\dots,x_n) dx_1\cdots dx_n \tag{5.46}$$

siempre que la integral de la derecha sea absolutamente convergente, es decir que

$$\int\cdots\int_{\mathbb{R}^n} |g(x_1,\dots,x_n)| f(x_1,\dots,x_n) dx_1\cdots dx_n < \infty$$

La integral en la fórmula (5.24) la indicaremos, a los efectos de abreviar la notación, por

$$\int_{\mathbb{R}^n} g(x)f(x)dx,$$

donde debe entenderse que x representa el vector $(x_1, x_2, \dots, x_n)$.

A continuación demostraremos la fórmula (5.46). Al igual que en algunas situaciones anteriores, el lector puede saltar esta demostración en una primera lectura y pasar directamente a los ejemplos de aplicación de la fórmula (5.46).

Demostración de la fórmula (5.24)

Basta considerar el caso en que $g \ge 0$, ya que si esto no ocurre, separamos el estudio para g^+ y para g^-, y luego sumamos, donde definimos

$$g^+(x) = \begin{cases} g(x) & \text{si } g(x)\ge 0 \\ 0 & \text{si } g(x) < 0. \end{cases} \qquad g^-(x) = \begin{cases} 0 & \text{si } g(x)\ge 0 \\ -g(x) & \text{si } g(x) < 0. \end{cases}$$

Sea entonces $g(x) \geq 0$ para todo $x \in \mathbb{R}^n$. Definimos (analogamente a lo hecho para variables aleatorias), para $m = 1, 2, \ldots$

$$g_m(x) = \frac{k}{2^m} \quad \text{si } \frac{k}{2^m} \leq g(x) < \frac{k+1}{2^m} \quad k = 0, 1, \ldots$$

Es claro que

$$g(x) - \frac{1}{2^m} \leq g_m(x) \leq g(x) \quad \text{para todo } x \in \mathbb{R}^n \tag{5.47}$$

y por lo tanto

$$g(X(\omega)) - \frac{1}{2^m} \leq g_m(X(\omega)) \leq g(X(\omega)),$$

donde

$$X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$$

es nuestro vector aleatorio, cuyas coordenadas son las variables aleatorias dadas.

Tomando valores esperados (ver las propiedades en la sección siguiente), resulta

$$\mathrm{E}(g(X)) - \frac{1}{2^m} \leq \mathrm{E}(g_m(X)) \leq E(g(X)). \tag{5.48}$$

Ahora bien,

$$g_m(X) = g_m(X_1, X_2, \ldots, X_n)$$

es una variable aleatoria discreta, que puede tomar sólo valores de la forma $k/2^m$ para $k = 0, 1, 2, \ldots$, y por lo tanto su valor esperado se calcula directamente

$$\begin{aligned}
\mathrm{E}(g_m(X)) &= \sum_{k=0}^{\infty} \frac{k}{2^m} P(g_m(X) = \frac{k}{2^m}) \\
&= \sum_{k=0}^{\infty} \frac{k}{2^m} P(\frac{k}{2^m} \leq g_m(X) < \frac{k+1}{2^m}) \\
&= \sum_{k=0}^{\infty} \frac{k}{2^m} P\left(X \in g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right)\right)
\end{aligned} \tag{5.49}$$

donde hemos indicado, como es usual

$$g^{-1}(I) = \{x : x \in \mathbb{R}^n, g(x) \in I\}.$$

En consecuencia, (5.27) no es sino

$$\begin{aligned}
\mathrm{E}(g_m(X)) &= \sum_{k=0}^{\infty} \frac{k}{2^m} \int_{g^{-1}([\frac{k}{2^m}, \frac{k+1}{2^m}))} f(x)dx \\
&= \sum_{k=0}^{\infty} \int_{g^{-1}([\frac{k}{2^m}, \frac{k+1}{2^m}))} g_m(x)\, f(x)dx \\
&= \int_{\mathbb{R}^n} g_m(x) f(x)dx
\end{aligned} \tag{5.50}$$

En virtud de (5.48), cuando $m \to \infty$, el primer miembro de (5.50) tiende a $E(g(X))$. En cuanto al último miembro, tenemos, haciendo intervenir (5.47),

$$\int_{\mathbb{R}^n} g(x)f(x)dx - \frac{1}{2^m} \leq \int_{\mathbb{R}^n} g_m(x)f(x)dx \leq \int_{\mathbb{R}^n} g(x)f(x)dx$$

y por lo tanto, cuando $m \to \infty$, resulta que el último miembro de (5.50) tiende a

$$\int_{\mathbb{R}^n} g(x)f(x)dx.$$

Resumiendo:

$$E(g(X)) = \int_{\mathbb{R}^n} g(x)f(x)dx,$$

que es justamente (5.46). ■

Observación 5.10 Si el lector interesado observa con cuidado la demostración que hemos hecho de la fórmula (5.46), verá que algunas afirmaciones debieran merecer una mayor atención. Por lo pronto, en (5.48)

$$\left\{\omega : X(\omega) \in g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right)\right\},$$

tiene que ser un evento en el espacio de probabilidad subyacente. Para ello, se requiere que la función $g : \mathbb{R}^n \to \mathbb{R}$ sea lo que se denomina una función "boreliana", lo cual significa que la preimagen por g de un intervalo en $\mathbb{R}$, es un conjunto de Borel en $\mathbb{R}^n$. Por cierto, todas las funciones corrientes que aparecen en el cálculo, son funciones borelianas, de modo que, en este contexto, esto no implica ninguna limitación efectiva.

En segundo término, tiene que estar bien definida la integral

$$\int_{g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right)} f(x)dx$$

que figura en la fórmula (5.50).

En tercer término, la última igualdad que aparece en (5.50), dice que, si denotamos por

$$B_{k,m} = g^{-1}\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right) = \left\{x : x \in \mathbb{R}^n, g(x) \in \left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right\}$$

entonces

$$\sum_{k=0}^{\infty} \int_{B_{k,m}} h(x)\, dx = \int_{\mathbb{R}^n} h(x)\, dx,$$

donde $h(x) = g_m(x)f(x)$, es una cierta función no-negativa.

Todas estas afirmaciones son correctas, pero requieren un mayor desarrollo, y la tercera de ellas debe ser probada. El lector puede consultar, si desea comprender con la profundidad adecuada estos temas que están más allá del nivel de esta exposición, los libros de M. Loéve, D. Billingsley y R. M. Dudley incluidos en la bibliografía.

5.12. Propiedades del Valor Esperado en el Caso General.

Tal como hemos mencionado anteriormente, con la definición general el valor esperado sigue teniendo todas las propiedades enunciadas y probadas en la sección 5.1 para el caso discreto. La mayor parte de ellas son de demostración muy sencilla, y el lector puede probarlas por si mismo, sin mayor dificultad. Las únicas dos que requieren una mayor atención son las propiedades 7 y 9, a saber:

> **Propiedad 7.** Si X e Y son variables aleatorias que tienen valor esperado, entonces también existe el valor esperado de $X + Y$ y se tiene
>
> $$\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y). \tag{5.51}$$

Propiedad 9. Si X e Y son variables aleatorias independientes con valor esperado, entonces existe $\mathrm{E}(XY)$ y
$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y). \tag{5.52}$$

El lector que no esté especialmente interesado puede saltar las demostraciones de estas dos propiedades, en una primera lectura.

Demostración de la propiedad 7

Son obvias las desigualdades

$$(X+Y)^+ \leq |X+Y| \leq |X| + |Y|$$
$$(X+Y)^- \leq |X+Y| \leq |X| + |Y|$$

y como por hipótesis se cumplen

$$\mathrm{E}(|X|) < \infty, \qquad \mathrm{E}(|Y|) < \infty,$$

se deducen
$$\mathrm{E}((X+Y)^+) < \infty, \qquad \mathrm{E}((X+Y)^-) < \infty \Rightarrow E(X+Y) < \infty.$$

Probaremos la igualdad (5.51) cuando X e Y son no-negativas; cuando esto no ocurre, descomponemos Ω en los conjuntos donde $X,\ Y$ y $X+Y$ tienen signo constante, y aplicamos el resultado que vamos a demostrar. Los detalles quedan a cargo del lector.

Supongamos entonces, $X \geq 0,\ Y \geq 0$. Por lo tanto, también $X+Y \geq 0$ y usamos la notación de la sección 5.4, llamando
$$X_n,\ Y_n,\ (X+Y)_n$$

las variables aleatorias aproximantes de $X,\ Y,\ X+Y$ respectivamente. Sabemos que

$$X_n \leq X < X_n + \frac{1}{2^n}$$
$$Y_n \leq Y < Y_n + \frac{1}{2^n}$$
$$(X+Y)_n \leq X+Y < (X+Y)_n + \frac{1}{2^n}$$

Se deduce que

$$\begin{aligned}\frac{-2}{2^n} &= X - \frac{1}{2^n} + Y - \frac{1}{2^n} - (X+Y) < X_n + Y_n - (X+Y)_n \\ &< X + Y - (X+Y) - \frac{1}{2^n} = \frac{1}{2^n}\end{aligned}$$

es decir,
$$\frac{-2}{2^n} < X_n + Y_n - (X+Y)_n < \frac{1}{2^n} \tag{5.53}$$

Como $X_n,\ Y_n,\ (X+Y)_n$ son variables aleatorias discretas, sabemos que (5.33) implica que

$$\frac{-2}{2^n} < \mathrm{E}(X_n) + \mathrm{E}(Y_n) - \mathrm{E}((X+Y)_n) < \frac{1}{2^n}$$

Si ahora hacemos tender n a infinito, en virtud de la definición de la esperanza matemática en el caso general, se tiene que

$$\mathrm{E}(X_n) \to \mathrm{E}(X), \qquad \mathrm{E}(Y_n) \to \mathrm{E}(Y), \qquad \mathrm{E}((X+Y)_n) \to \mathrm{E}(X+Y)$$

y por lo tanto,

$$0 \leq \mathrm{E}(X) + \mathrm{E}(Y) - \mathrm{E}(X+Y) \leq 0,$$

lo cual prueba (5.51). ■

Para demostrar la propiedad 9 podemos seguir un método similar de aproximación por variables discretas, haciendo valer nuevamente el hecho de que ya sabemos que la propiedad es válida para el caso de las variables aleatorias discretas.

5.12.1. Demostración de (5.23)

Para demostrar (5.23), apliquemos la definición del valor esperado. Sea X_n la variable aleatoria aproximante de X definida en la sección 5.4. Tenemos

$$\begin{aligned}
\mathrm{E}(X) &= \lim_{n\to\infty} \mathrm{E}(X_n) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} P(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}) \\
&= \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} \left[P(X \geq \frac{k}{2^n}) - P(X \geq \frac{k+1}{2^n}) \right].
\end{aligned}$$

Introducimos la notación

$$G(x) = 1 - F_X(x^-) = P(X \geq x)$$

y reemplazando, resulta

$$\mathrm{E}(X) = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^n} \left[G(\frac{k}{2^n}) - G(\frac{k+1}{2^n}) \right].$$

Mas aún, si ponemos

$$p_k = G(\frac{k}{2^n}) - G(\frac{k+1}{2^n}), \qquad q_k = \sum_{j=k+1}^{\infty} p_j G(\frac{k+1}{2^n}).$$

entonces podemos aplicar la fórmula (5.21), de la forma siguiente:

$$\begin{aligned}
\mathrm{E}(X_n) &= \sum_{k=0}^{\infty} \frac{k}{2^n} p_k = \frac{1}{2^n} \sum_{k=0}^{\infty} k p_k = \frac{1}{2^n} \sum_{k=0}^{\infty} q_k \\
&= \sum_{k=0}^{\infty} \frac{1}{2^n} G(\frac{k+1}{2^n}) \leq \int_0^{+\infty} G(x)\,dx. \qquad (5.54)
\end{aligned}$$

Para comprender la última desigualdad, hay que tener en cuenta que la función G es decreciente (en sentido amplio), ya que F_X es creciente (en sentido amplio), y que estamos comparando una suma de Riemann calculada evaluando en el extremo derecho de cada intervalo, con la integral (ver figura 5.4).

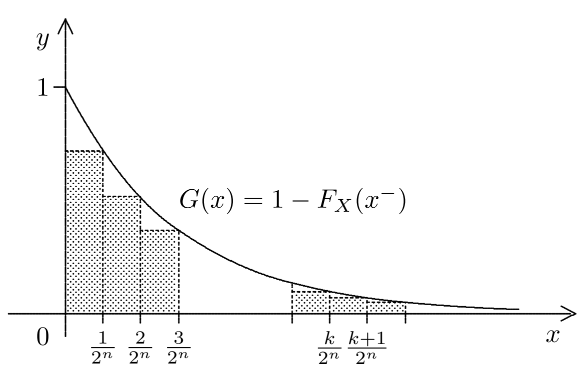


Figura 5.4

De (5.54) ya resulta que si $\mathrm{E}(X) = +\infty$, como el primer miembro tiende a $+\infty$ cuando $n \to \infty$, entonces

$$\int_0^{+\infty} G(x)\,dx \quad \text{diverge.}$$

Consideremos entonces el caso en el cual $\mathrm{E}(X) < \infty$. Pasando al límite cuando $n \to \infty$ en (5.54), se obtiene

$$\mathrm{E}(X) \leq \int_0^{+\infty} G(x)\,dx. \tag{5.55}$$

Por otro lado, si N es un número natural positivo cualquiera, la integral en el intervalo $[0, N]$ es el límite de las sumas de Riemann, cuando la partición se afina, y entonces

$$\begin{aligned}\int_0^{N} G(x)\,dx &= \lim_{n\to\infty} \sum_{k=0}^{N2^n-1} \frac{1}{2^n} G(\frac{k+1}{2^n}) \leq \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{1}{2^n} G(\frac{k+1}{2^n}) \\ &= \lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X).\end{aligned}$$

Haciendo $N \to +\infty$, resulta

$$\int_0^{+\infty} G(x)\,dx \leq \mathrm{E}(X),$$

que conjuntamente con (5.55), implica (5.23). ■

Observación 5.11 En la fórmula (5.23), si en lugar de $1 - F_X(x^-)$ ponemos en el integrando la función $1 - F_X(x)$, el resultado es el mismo, es decir

$$\mathrm{E}(X) = \int_0^{+\infty} (1 - F_X(x))\,dx. \tag{5.56}$$

El lector puede verificar esto por su cuenta. También es posible ver que si F_X no tiene saltos entonces (5.23) y (5.56) son la misma cosa, pero si hay saltos, ambos integrandos no coinciden en todo punto.