1 Esperanza y varianza de una variable aleatoria

Al igual que las variables en álgebra, las variables aleatorias tienen valores, pero estos valores sólo pueden determinarse a partir de experimentos aleatorios. Para caracterizar todos los resultados posibles de tal experimento usamos funciones de distribución, que proporcionan una especificación completa de una variable aleatoria sin la necesidad de definir un espacio de probabilidad subyacente. Especificar funciones de distribución no plantea dificultades matemáticas inherentes.

Sin embargo, en las aplicaciones con frecuencia no es posible determinar la función de distribución completamente. A menudo una representación más simple es posible, en lugar de especificar toda la función especificamos un conjunto de `valores estadísticos`. Un valor estadístico es el promedio de una función de una variable aleatoria. Esto puede ser pensado como el valor que se obtiene de promediar los resultados de un gran número de experimentos aleatorios en el enfoque de frecuencia de la probabilidad y en el marco axiomático dicho promedio corresponde a la operación de la `esperanza` aplicado a una función de una variable aleatoria.

Como se sabe en el marco axiomático, una variable aleatoria se caracteriza completamente por su función de distribución acumulativa o de forma equivalente, por su función de masa o de densidad de probabilidad. Sin embargo, no siempre es posible y en muchos casos no es necesario, tener esta información completa. En algunos casos, un solo número que captura alguna característica promedio es suficiente. Tales números o valores estadísticos, incluyen la `esperanza`, la `mediana` y la `moda` de la variable aleatoria. La esperanza de una variable aleatoria se calcula de manera similar cuando se calcula el valor promedio de un conjunto de números. El valor promedio de un conjunto de n números se forma multiplicando cada número por $1/n$ y añadiendo los resultados. La esperanza de una variable aleatoria que toma los valores $\{x_1, x_2, \ldots, x_n\}$ es obtenida multiplicando, cada proporción de veces que ocurre y agregándo al resultado, es decir, se forma $\sum_{i=1}^{n} x_i p_i$.

El `valor medio` de una variable aleatoria X también se le llama `esperanza` de X. La mediana de un conjunto de números es el número (o números) que se encuentra en el centro del conjunto una vez que los números están dispuestos en orden ascendente o descendente. La `mediana` de una variable aleatoria X es cualquier número m que satisface:

$$\mathbb{P}(X < m) \leq 1/2 \ \text{ y } \ \mathbb{P}(X > m) \leq 1/2,$$

o equivalentemente,

$$\mathbb{P}(X < m) = \mathbb{P}(X > m)$$

La `moda` es el valor posible más probable. Para una variable aleatoria, es el valor para el cual la función de masa de probabilidad o la función de densidad de probabilidad alcanza su máximo. Si m es un modo de una variable aleatoria X, entonces $\mathbb{P}(X = m) \geq \mathbb{P}(X = x)$ para todos los valores de x . Al igual que la mediana y a diferencia del valor medio, una variable aleatoria puede tener múltiples modas.

Ejemplo 1.1 Si X es la variable aleatoria que denota el número de puntos obtenidos en el lanzamiento de un dado, entonces hay seis modas (ya que cada número de puntos es igualmente probable y todos los resultados alcanzan el mismo valor máximo de $1/6$), la mediana es cualquier número en el intervalo $(3, 4)$. Sólo hay un valor medio, que es igual a $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$.

La definición de la media y el hecho de que sea un número único lo convierte, desde un punto de vista matemático, en un valor mucho más atractivo para trabajar que la mediana o la moda. Este valor puede ampliarse fácilmente para proporcionar una caracterización completa de una variable aleatoria.

De hecho, una variable aleatoria puede ser completamente caracterizada por un conjunto de valores, llamados `momentos`. Estos momentos se definen en términos de la función de distribución, pero generalmente no se necesita calcular la función de distribución. Puede demostrarse que, si dos variables aleatorias tienen los mismos momentos de todos los órdenes, entonces también tienen la misma distribución.

El primer momento, es la media o esperanza de una variable aleatoria X y es dado por la integral de Riemann-Stieltjes:

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

Para una variable aleatoria continua, se tiene que:

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

donde $f_X(x)$ es la función de densidad de probabilidad de X. Si X es una variable aleatoria discreta cuyos valores $\{x_1, x_2, \dots\}$, tienen las correspondientes probabilidades $\{p_1, p_2, \dots\}$, entonces la integral, es una suma y como hemos indicado anteriormente:

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_i.$$

Cuando una variable aleatoria discreta toma valores enteros sobre los enteros no negativos, podemos escribir, la esperanza como:

$$\begin{aligned}\mathbb{E}(X) &= p_1 + 2p_2 + 3p_3 + 4p_4 + \cdots \\ &= (p_1 + p_2 + p_3 + p_4 + \cdots) + (p_2 + p_3 + p_4 + \cdots) + (p_3 + p_4 + \cdots) + \cdots,\end{aligned}$$

Lo que conduce a una interesante fórmula:

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$$

El n-ésimo momento de una variable aleatoria X es definida de manera similar. Para una variable aleatoria continua es dado por:

$$\mathbb{E}(X^n) = \int_{\infty}^{\infty} x^n f_X(x) dx,$$

mientras que para una variable aleatoria discreta X, se tiene que:

$$\mathbb{E}(X^n) = \sum_i x_i^n p_i.$$

Los primeros dos momentos son los más importantes. El primero, es conocido como la esperanza o valor medio de X, el segundo momento es llamado `media cuadrada` de X ambos están definidos como:

$$\mathbb{E}(X^2) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x^2 f_X(x) dx & \text{si X es continua} \\ \displaystyle\sum_i x_i^2 p_i & \text{si X es discreta.} \end{cases}$$

Los momentos centrales son también importantes. Ellos son los momentos de la diferencia entre la variable aleatoria X y el valor medio $\mathbb{E}(X)$. El n-ésimo momento central es dado por:

$$\mathbb{E}[(X-\mathbb{E}(X))^n] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} (X-\mathbb{E}(X))^n f_X(x)dx & \text{si X es continua} \\ \displaystyle\sum_i (x_i-\mathbb{E}(X))^n p_i & \text{si X es discreta} \end{cases}$$

El segundo momento central es llamada `varianza` de la variable aleatoria X y es escrita como:

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mathbb{E}(X))^2] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} (x-\mathbb{E}(X))^2 f_X(x)dx & \text{si X es continua} \\ \displaystyle\sum_i (x_i-\mathbb{E}(X))^2 p_i & \text{si X es discreta} \end{cases}$$

Una fórmula usada para la varianza es:

$$\mathrm{Var}(x) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

De esta ecuación, se tiene que la varianza, no puede ser negativa y caracteriza la dispersión de la variable aleatoria X con respecto a la media. La raíz cuadrada de la varianza, es la `desviación estándar` de X, denotada como σ_X. La desviación estándar produce un número cuyas unidades son iguales a la de la variable aleatoria y por esta razón proporciona una imagen más clara de la dispersión de la variable aleatoria respecto de su media. Una caracterización final de una variable aleatoria es el `coeficiente de variación`, escrito como C_X y definido como:

$$C_X = \frac{\sigma_X}{\mathbb{E}(X)},$$

y su utilidad reside en el hecho de que es una medida adimensional de la variabilidad de X . A menudo, en lugar del coeficiente de variación, se usa el `coeficiente de variación cuadrático`, que se define como:

$$C_X^2 = \frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2}.$$

Ejemplo 1.2 Considera la variable aleatoria X que denota el número total de puntos obtenidos cuando dos dados son lanzados simultáneamente. Anteriormente, vimos que la función de masa de probabilidad de X está dada por:

x_i	2	3	4	5	6	7	8	9	10	11	12
$p_X(x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

En este caso la esperanza de X es dado por:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^{\infty} x_i p_i = \sum_{i=2}^{12} x_i p_i \\ &= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} \\ &= \frac{252}{36} = 7.0 \end{aligned}$$

El cálculo de la varianza de X, se procede como sigue:

$$
\begin{aligned}
\operatorname{Var}(X) &= \sum_{i=2}^{12} (x_i - 7)^2 p_i \\
&= 5^2\frac{1}{36} + 4^2\frac{2}{36} + 3^2\frac{3}{36} + 2^2\frac{4}{36} + 1^2\frac{5}{36} + 0^2\frac{6}{36} + 1^2\frac{5}{36} + 2^2\frac{4}{36} + 3^3\frac{3}{36} + 4^2\frac{2}{36} + 5^2\frac{1}{36} = 5.8333.
\end{aligned}
$$

Se sigue que la desviación estándar X es $\sqrt{5.8333} = 2.4152$.

Ejemplo 1.3 El tiempo transcurrido en minutos entre la colocación de un pedido de pizza y su entrega es aleatorio con la función densidad:

$$
f_X(x) = \begin{cases} 1/15 & \text{si } 25 < x < 40 \\ 0 & \text{en otros casos.} \end{cases}
$$

- Determina la media y la desviación estándar del tiempo que tarda la pizzería en entregar la pizza.
- Supongamos que la tienda de pizza tarda 12 minutos en hornear la pizza. Determina la media y la desviación estándar del tiempo que tarda la persona de entrega en entregar la pizza.

Sea el tiempo transcurrido entre la colocación de una orden y su entrega sea en X minutos. Entonces:

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{25}^{40} x\frac{1}{15}dx = 32.5 \\
\mathbb{E}(X^2) &= \int_{25}^{40} x^2\frac{1}{15}dx = 1075.
\end{aligned}
$$

Por tanto $\operatorname{Var}(X) = 1075 - (32.5)^2 = 18.75$ y así $\delta_X = \sqrt{\operatorname{Var}(X)} = 4.33$.

El tiempo que tarda la persona que entrega la pizza en entregarla es $X - 12$. Por lo tanto, las cantidades deseadas son:

$$
\begin{aligned}
\mathbb{E}(X-12) &= \mathbb{E}(X) - 12 = 32.5 - 12 = 20.5 \\
\sigma_{X-12} &= |1|\sigma_X = \sigma_X = 4.33
\end{aligned}
$$

Ejemplo 1.4 En un juego, un jugador lanza una moneda sucesivamente hasta que consigue una `Cara`. Si esto ocurre en el k-ésimo lanzamiento, el jugador gana 2^k soles. Por lo tanto, si el resultado del primer lanzamiento es cara, el jugador gana 2 soles. Si el resultado del primer lanzamiento es sello y el segundo lanzamiento es cara, el gana 4. Si los resultados de los dos primeros lanzamientos son sellos, pero el tercero sale cara, ganará 8 soles y así sucesivamente. La pregunta es, para jugar este juego, ¿ cuánto debe pagar una persona, que está dispuesta a jugar este juego?.

Para responder a esta pregunta, sea X la cantidad de dinero que el jugador gana. Entonces X es una variable aleatoria con el conjunto de valores posibles $\{2, 4, 8, \dots, 2^k, \dots\}$ y

$$
\mathbb{P}(X = 2^k) = \left(\frac{1}{2}\right)^k, \quad k = 1, 2, 3, \dots
$$

Por tanto,

$$
\mathbb{E}(X) = \sum_{k=1}^{\infty} 2^k \left(\frac{1}{2}\right)^k = \sum_{k=1}^{\infty} 1 = 1 + 1 + 1 + \cdots = \infty.
$$

Este resultado muestra que el juego sigue siendo injusto, incluso si una persona paga la mayor cantidad posible para jugar. En otras palabras, este es un juego en el que uno siempre gana, no importa lo caro que es jugar. Para ver cuál es la falla, tenga en cuenta que teóricamente este juego no es factible de jugar porque requiere una enorme cantidad de dinero. En la práctica, sin embargo, la probabilidad de que un jugador gane 2^k soles para un valor $k \to 0$. Incluso para valores pequeños de k, ganar es altamente improbable. Por ejemplo, para ganar $2^{30} = 1.073.741.824$, debes conseguir 29 sellos en una fila seguida por una cara. La probabilidad de que esto suceda es 1 en 1.073.741.824, mucho menos de 1 en un billón.

Ejemplo 1.5 Sea X_0, la cantidad de lluvia que caerá en los Estados Unidos el próximo dia de Navidad. Para $n > 0$, sea X_n la cantidad de lluvia que caerá en los Estados Unidos en Navidad, n años después. Sea N el menor número de años que transcurren antes de una lluvia de Navidad mayor que X_0. Supongamos que $P(X_i = X_j) = 0$ si $i \neq j$, los sucesos relativos a la cantidad de lluvia en dias de Navidad de diferentes años son independientes y los X_n son idénticamente distribuidos. Hallar el valor esperado de N.

Como el valor N es el primer valor de n, para que se cumpla $X_n > X_0$,

$$\begin{aligned}\mathbb{P}(N > n) &= \mathbb{P}(X_0 > X_1, X_0 > X_2, \dots, X_0 > X_n)\\ &= \mathbb{P}(\max(X_0, X_1, X_2, \dots, X_n) = X_0) = \frac{1}{n+1}\end{aligned}$$

donde la última igualdad procede de la simetría, no hay más razón para que el máximo esté en x_0 que estar en $X_i, 0 \leq i \leq n$. Por tanto,

$$\mathbb{P}(N = n) = \mathbb{P}(N > n-1) - \mathbb{P}(N > n) = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)}.$$

Entonces se sigue que:

$$\mathbb{E}(N) = \sum_{n=1}^{\infty} n\mathbb{P}(N = n) = \sum_{n=1}^{\infty} \frac{n}{n(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty.$$

Se debe notar que $\mathbb{P}(N > n-1) = 1/n$ da la probabilidad que, en los Estados Unidos, tendremos que esperar más, digamos, tres años para una lluvia de navidad que sea mayor que X_0 y que la probabilidad sea sólo $1/4$ y la probabilidad de que debamos esperar más de nueve años es sólo $1/10$. Incluso con probabilidades tan bajas en promedio, debería tomar infinitamente muchos años antes de que tengamos más lluvia en un d?a de navidad de la que tendremos el próximo dia de Navidad.

Ejemplo 1.6 Considera la variable aleatoria continua M, cuya función de densidad de probabilidad es dada por:

$$f_M(x) = \begin{cases} e^{-x} & \text{si } x \geq 0, \\ 0 & \text{en otros casos.} \end{cases}$$

En este caso, la esperanza es dada por:

$$\mathbb{E}(M) = \int_{-\infty}^{\infty} x f_M(x) dx = \int_0^{\infty} x e^{-x} dx = -e^{-x}\Big|_0^{\infty} = 1,$$

y su varianza se calcula como:

$$\text{Var}(M) = \int_0^{\infty} (x - \mathbb{E}(M))^2 e^{-x} dx = \int_0^{\infty} (x-1)^2 e^{-x} dx$$

$$= \int_0^{\infty} x^2 e^{-x} dx - 2\int_0^{\infty} x e^{-x} dx + \int_0^{\infty} e^{-x} dx = 2 - 2 + 1,$$

y se muestra que ambas valores son iguales.

Ejemplo 1.7 La función

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty,$$

es una función densidad de probabilidad, pero no tiene un valor para la esperanza. Para ver esto, observa que:

$$\int_{\infty}^{\infty} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi}\arctan(x)\Big|_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} - \frac{-\pi}{2}\right) = 1,$$

y así es una función densidad. Sin embargo, no converge aboslutamente, desde que:

$$\int_{-\infty}^{\infty} \left|\frac{dx}{\pi(1+x^2)}\right| dx = \frac{1}{\pi}\int_{0}^{\infty} \frac{2x}{1+x^2}dx = \frac{1}{\pi}\ln(1+x^2)\Big|_{0}^{\infty} = \infty.$$

Afirmamos anteriormente que la desviación estándar de una variable aleatoria es un número cuyas unidades son las mismas que las de la variable aleatoria y es una medida apropiada de la dispersión de la variable aleatoria sobre su media. Si un valor particular para X, digamos x, está en un intervalo de σ_X alrededor de $\mathbb{E}(X)$, esto es $x \in [\mathbb{E}(x) - \sigma_X, \mathbb{E}(X) + \sigma_X]$, se dice que es "cercano" al valor esperado de X. En caso contrario se dice que "está lejos de la media".

La desigualdad de Chebychev, hace estos conceptos más precisos. Sea X una variable aleatoria cuya esperanza $\mathbb{E}(X) = \mu_X$ y cuya desviación estándar sea dada por σ_X. La desigualdad de Chebychev indica que:

$$\mathbb{P}(|X - \mu_X| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}.$$

En palabras, esto dice que la probabilidad de que una variable aleatoria esté dentro de $\pm k$ desviaciones estándar de su valor medio es al menos $1 - 1/k^2$. Alternativamente, la probabilidad de que una variable aleatoria esté fuera de $\pm k$ desviaciones estándar de su media es como máximo $1/k^2$.

Esto significa que el 75%$(= 1 - 1/2^2)$ de las veces, una variable aleatoria está dentro de dos desviaciones estándar de su media y 89%$(= 1 - 1/3^2)$ de las veces , cae dentro de tres desviaciones estándar . De hecho resulta que en la mayoría de los casos, estos números son subestimaciones de lo que generalmente ocurre. De hecho, para muchas distribuciones, la probabilidad de que una variable aleatoria se encuentre dentro de dos desviaciones estándar puede ser tan alta como el 95% o mayor.

Las variables aleatorias que consideraremos, todas tienen esperanza finita. Se dice que la esperanza no existe a menos que la integral o sumatoria correspondiente sean absolutamente convergente, es decir, la espernza sólo existe si:

$$\int_{-\infty}^{\infty} |x| f_X(x)dx < \infty \quad \text{y} \quad \sum_{i=1}^{\infty} |x_i| p_i < \infty.$$

Proposición 1.1 Para una variable aleatoria continua X con una función de distribución F_X y una función densidad de f_x,

$$\mathbb{E}(X) = \int_0^{\infty} [1 - F_X(t)]dt - \int_0^{\infty} F_X(-t)dt.$$

Note que:

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{0} x f_X(x)dx + \int_{0}^{\infty} x f_X(x)dx \\
&= -\int_{-\infty}^{0}\left(\int_{0}^{-x} dt\right) f_X(x)dx + \int_{0}^{\infty}\left(\int_{0}^{x} dt\right) f_X(x)dx \\
&= -\int_{0}^{\infty}\left(\int_{-\infty}^{-t} f_X(x)dx\right)dt + \int_{0}^{\infty}\left(\int_{t}^{\infty} f_X(x)dx\right)dt,
\end{aligned}
$$

Donde la última igualdad, es obtenida cambiando el orden de la integración. El teorema sigue desde que : $\int_{-\infty}^{-t} f_X(x)dx = F_X(-t)$ y $\int_{t}^{\infty} f_X(x)dx = \mathbb{P}(X > t) = 1 - F_X(t)$. En general, para alguna variable aleatoria X,

$$
\mathbb{E}(X) = \int_{0}^{\infty} \mathbb{P}(X > t)dt - \int_{0}^{\infty} \mathbb{P}(X \leq -t)dt.
$$

En particular, si X es no negativa, esto es $\mathbb{P}(X < 0) = 0$, este teorema, establece que:

$$
\mathbb{E}(X) = \int_{0}^{\infty} [1 - F_X(t)]dt = \int_{0}^{\infty} \mathbb{P}(X > t)dt.
$$

1.1 Esperanza de funciones de variables aleatorias

Dado una variable aleatoria discreta X con un conjunto posible de valores A y la función de masa de probabilidad $p_X(x)$, la esperanza de un nueva variable aleatoria Y, que es una función de X, es decir $Y = h(X)$ se puede escribir como:

$$
\mathbb{E}(Y) = \mathbb{E}(h(X)) = \sum_{x \in A} h(x) p_X(x). \tag{1}
$$

En efecto, sea Ω un espacio muestral. Sea $h : \mathbb{R} \to \mathbb{R}$, una función de variable real y $X : \Omega \to A \subseteq \mathbb{R}$ es una variable aleatoria con el conjunto de posibles de valores A. Como se sabe, $h(X)$, la composición de g y X es una función desde Ω al conjunto $h(A) = \{h(x) : x \in A\}$. Así $h(X)$ es una variable aleatoria con un posible conjunto de valores $g(A)$. Ahora por definición de esperanza:

$$
\mathbb{E}(h(X)) = \sum_{z \in h(A)} z \mathbb{P}(h(X) = z)
$$

Sea $h^{-1}(\{z\}) = \{x : h(x) = z\}$. Debemos notar que no aseveramos que h tiene una función inversa y en este contexto, consideramos el conjunto $\{x : h(x) = z\}$, que es llamado la imagén inversa de z y es denotada por $h^{-1}(\{z\})$. Ahora:

$$
\mathbb{P}(h(X) = z) = \mathbb{P}\Big(X \in h^{-1}(\{z\})\Big) = \sum_{\{x : x \in h^{-1}(\{z\})\}} \mathbb{P}(X = x) = \sum_{\{x : h(x) = z\}} p_X(x).
$$

$$
\begin{aligned}
\mathbb{E}(h(X)) &= \sum_{z\in h(A)} z\mathbb{P}(h(X)=z) = \sum_{z\in h(A)} z \sum_{\{x:h(x)=z\}} p_X(x) \\
&= \sum_{z\in h(A)} \sum_{\{x:h(x)=z\}} z p_X(x) = \sum_{z\in h(A)} \sum_{\{x:h(x)=z\}} h(x)p_X(x) \\
&= \sum_{x\in A} h(x)p_X(x),
\end{aligned}
$$

Donde la última igualdad se sigue del hecho que la suma es sobre A puede ser llevados en dos escenarios: Podemos sumar primero todos los x con $h(x) = z$ y entonces para todo z.

Dos útiles propiedades de la esperanza se sigue a continuación:

Proposición 1.2 Sea X una variable aleatoria discreta y sea $a, b \in \mathbb{R}$.

1. Si $\mathbb{P}(X \geq 0) = 1$ y $\mathbb{E}(X) = 0$, entonces $\mathbb{P}(X = 0) = 1$.
2. Tenemos que $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

Ejemplo 1.8 La función de masa de probabilidad de una variable aleatoria X es dado por:

$$
p_X(x) = \begin{cases} x/15 & x = 1,2,3,4,5 \\ 0 & \text{en otros casos} \end{cases}
$$

Calculemos la esperanza de $X(6 - X)$.

$$
\mathbb{E}[X(6-X)] = 5\cdot\frac{1}{15} + 8\cdot\frac{2}{15} + 9\cdot\frac{3}{15} + 8\cdot\frac{4}{15} + 5\cdot\frac{1}{15} + 5\cdot\frac{5}{15} = 7.
$$

Ejemplo 1.9 Para la variable aleatoria discreta R que denota el número de caras obtenidas en tres lanzamientos de una moneda, la función de masa de probabilidad está dada por:

x_k	0	1	2	3
$p_R(x_k)$	0.125	0.375	0.375	0.125

Consideremos ahora una situación de juego en la que un jugador pierde 3 soles si no aparecen caras, pierde 2 soles si aparece una cara y no gana nada si aparecen dos caras, pero gana 7 soles si aparecen tres caras. Calculemos el número medio de soles ganados o pérdidos en el juego. Para ello, definimos la variable aleatoria derivada Y como sigue:

$$
Y = h(X) = \begin{cases} -3, & X = 0, \\ -2, & X = 1, \\ 0 & X = 2, \\ 7, & X = 3, \\ 0 & \text{en otros caso.} \end{cases}
$$

Calculemos $\mathbb{E}(Y)$ y descubramos que el jugador pierde 25 céntimos cada vez que juega:

$$
\mathbb{E}(Y) = \sum_{x=0,1,2,3} h(X)p_X(x) = -3\times 1/8 - 2\times 3/8 + 7\times 1/8 = -1/4.
$$

Similarmente, para variables aleatorias continuas X, la esperanza de alguna función $h(X)$ de X es dado por:

$$\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(X) f_X(x) dx. \tag{2}$$

En efecto, para verificar el caso continuo, sea $h^{-1}(t,\infty) = \{x : h(x) \in (t,\infty)\} = \{x : h(x) > t\}$ y con una representación similar para $h^{-1}(-\infty,-t)$. No estamos asumiendo que h tiene una función invesa. Estamos considerando el conjunto $\{x : h(x) \in (t,\infty)\}$, que es llamada la inversa de (t,∞) y es denotada por $h^{-1}(t,\infty)$. Por la proposición (1.1) tenemos:

$$\begin{aligned}
\mathbb{E}(h(X)) &= \int_0^{\infty} \mathbb{P}(h(X) > t) dt - \int_0^{\infty} \mathbb{P}(h(X) \leq -t) dt \\
&= \int_0^{\infty} \mathbb{P}(X \in h^{-1}(t,\infty)) dt - \int_0^{\infty} \mathbb{P}(X \in h^{-1}(-\infty,-t]) dt \\
&= \int_0^{\infty} \left(\int_{\{x : x \in h^{-1}(t,\infty)\}} f_X(x) dx \right) dt - \int_0^{\infty} \left(\int_{\{x : x \in h^{-1}(-\infty,-t]\}} f_X(x) dx \right) dt \\
&= \int_0^{\infty} \left(\int_{\{x : h(x) > t\}} f_X(x) dx \right) dt - \int_0^{\infty} \left(\int_{\{x : h(x) \leq -t\}} f_X(x) dx \right) dt.
\end{aligned}$$

Ahora cambiamos el orden de integración para ambas integrables dobles. Desde que

$$\{(t,x) : 0 < t < \infty,\ h(x) > t\} = \{(t,x) : h(x) > 0,\ 0 < t < h(x)\},$$

y

$$\{(t,x) : 0 < t < \infty,\ h(x) \leq -t\} = \{(t,x) : h(x) < 0,\ 0 < t \leq -h(x)\},$$

Conseguimos :

$$\begin{aligned}
\mathbb{E}(h(X)) &= \int_{\{x : h(x) > 0\}} \left(\int_0^{h(x)} dt \right) f(x) dx - \int_{\{x : h(x) < 0\}} \left(\int_0^{-h(x)} dt \right) f_X(x) dx \\
&= \int_{\{x : h(x) > 0\}} h(x) f_X(x) dx + \int_{\{x : h(x) < 0\}} h(x) f_X(x) dx \\
&= \int_{-\infty}^{\infty} h(x) f_X(x) dx.
\end{aligned}$$

Note que la última ecuación se sigue, debido a que $\displaystyle\int_{\{x : h(x) = 0\}} h(x) f_X(x) dx = 0.$

Proposición 1.3 Sea X una variable aleatoria con función densidad de probabilidad $f_X(x)$. Sean $h_1, h_2, \ldots, h_n$, funciones de variable real y $\alpha_1, \alpha_2, \ldots, \alpha_n$ números reales. Entonces:

$$\mathbb{E}(\alpha_1 h_1(X) + \alpha_2 h_2(X) + \cdots + \alpha_n h_n(X)) = \alpha_1 \mathbb{E}(h_1(X)) + \alpha_2 \mathbb{E}(h_2(X)) + \cdots + \alpha_n \mathbb{E}(h_n(X)).$$

Ejemplo 1.10 Un punto X es escogido desde el intervalo $(0, \pi/4)$ aleatoriamente. Calcula $\mathbb{E}(\cos 2X)$ y $\mathbb{E}(\cos^2 X)$.

Primero calculamos la función distribución de X:

$$F_X(t) = \mathbb{P}(X \leq t) = \begin{cases} 0 & t < 0 \\ \dfrac{t-0}{\frac{\pi}{4}-0} = \dfrac{4}{\pi} & 0 \leq t < \dfrac{\pi}{4} \\ 1 & t \geq \dfrac{\pi}{4}. \end{cases}$$

Así, la función densidad f de X es:

$$f_X(t) = \begin{cases} \dfrac{4}{\pi} & 0 < t < \dfrac{\pi}{4} \\ 0 & \text{en otros casos.} \end{cases}$$

Ahora por (2),

$$\mathbb{E}(\cos 2X) = \int_0^{\pi/4} \frac{4}{\pi} \cos 2x dx = \frac{2}{\pi}.$$

Para calcular $\mathbb{E}(\cos^2 X)$, notamos que $\cos^2 X = (1 + \cos 2X)/2$. Luego por la proposición 1.3, tenemos:

$$\mathbb{E}(\cos^2 X) = \mathbb{E}\left(\frac{1}{2} + \frac{1}{2}\cos 2X\right) = \frac{1}{2} + \frac{1}{2}\mathbb{E}(\cos 2X) = \frac{1}{2} + \frac{1}{2} \cdot \frac{2}{\pi} = \frac{1}{2} + \frac{1}{\pi}.$$

Ejemplo 1.11 Sea X una variable aleatoria continua con función densidad de probabilidad dada por:

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1, \\ 0 & \text{en otros casos.} \end{cases}$$

y sea la variable Y, definida como $Y = 4X + 2$. La esperanza de X es dada por:

$$\mathbb{E}(X) = \int_0^1 2x^2 dx = \left.\frac{2x^3}{3}\right|_0^1 = 2/3.$$

Ahora calculemos la esperanza de Y de dos maneras: primero integrando la función de densidad sobre un intervalo apropiado y segundo usando el enfoque ofrecido por la ecuación (2).

Para el primer caso, encontremos la función de densidad de probabilidad de Y al encontrar primero su función de distribuci?n acumulativa. Sin embargo, primero necesitamos encontrar la función de distribución acumulativa de X. Esto es:

$$F_X(x) = \int_0^x 2u du = x^2 \quad \text{para} \quad 0 \leq x \leq 1;$$

es igual a 0 para $x < 0$ y es igual a 1 para $x > 1$. Ahora calculemos $F_Y(y)$ como:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(4X + 2 \leq y) = \mathbb{P}\left(X \leq \frac{y-2}{4}\right) = F_X\left(\frac{y-2}{4}\right).$$

Para encontrar los límites, observamos que $F_X(x) = 0$ para $x < 0$, lo que nos permite afirmar que $F_Y(y) = 0$ para $(y-2)/4$, esto es, $y < 2$. También $F_X(x) = 1$ para $x > 1$, así $F_Y(y) = 1$ para $(y-2)/4 > 1$, esto es, para $y > 6$. Luego de la ecuación anterior para $F_Y(y)$, tenemos que:

$$F_Y(y) = \left(\frac{y-2}{4}\right)^2, \text{ para } 2 \leq y \leq 6.$$

Finalmente, diferenciando $F_Y(y)$ con respecto a y, podemos tener la función densidad de Y como:

$$f_Y(y) = \begin{cases} 2(y-2)/16, & 2 \leq y \leq 6, \\ 0 & \text{en otros casos.} \end{cases}$$

$$\mathbb{E}(Y) = \int_2^6 2y(y-2)/16 dy = (2y^3/3 - 2y^2)/16\Big|_2^6 = 4.6667.$$

Usando la ecuación (2), obtenemos:

$$\mathbb{E}(Y) = \mathbb{E}((\mathbb{X})) = \int_{-\infty}^{\infty} h(X) f_X(x) dx = \int_0^1 (4x+2)2x dx = \int_0^1 (8x^2 + 4x) dx$$
$$= (8x^3/3 + 2x^2)\Big|_0^1 = 14/3.$$

Este resultado coincide con el resultado anterior, pero implica menos cálculos.

La extensión a funciones de n variables aleatorias es inmediata. Si Y es una variable aleatoria definida como una función h de n variables aleatorias $X_1, X_2, \ldots, X_n$, la esperanza de Y puede ser encontrada utilizando la fórmula:

$$\mathbb{E}(Y) = \mathbb{E}[h(X_1, X_2, \ldots, X_n)] = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, x_2, \ldots, x_n) f_X(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n \\ \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} h(x_1, x_2, \ldots, x_n) p_X(x_1, x_2, \ldots, x_n). \end{cases}$$

para el caso continuo y discreto, respectivamente, Si la función h es una función lineal, esto es:

$$h(x_1, x_2, \ldots, x_n) = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n,$$

donde $\alpha_i,\ i = 1, 2, \ldots, n$ son constantes, entonces:

$$\mathbb{E}(Y) = \mathbb{E}[h(X_1, X_2, \ldots, X_n)] = h(\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n)).$$

Se debe observar que las variables aleatorias $X_1, X_2, \ldots X_n$ no se necesitan ser independientes. Sin embargo, este resultado no se aplica generalmente a funciones h que no son lineales.