

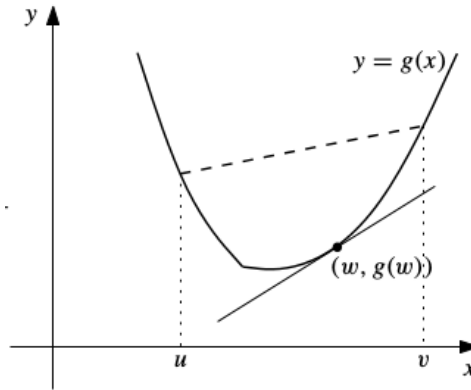
1 Resultados

Sea $-\infty \leq a < b < \infty$ una función $g : (a, b) \rightarrow \mathbb{R}$ es convexa si

$$g([1-t]u + tv) \geq (1-t)g(u) + tg(v)$$

para $t \in [0, 1]$ y $u, v \in (a, b)$.

Presentamos el teorema de soporte de hiperplanos que servirá para la desigualdad de Jensen, basado en el gráfico siguiente



Proposición 1.1 Sea $g : (a, b) \rightarrow \mathbb{R}$ convexa y sea $w \in (u, v)$. Entonces existe un $\alpha \in \mathbb{R}$ tal que

$$g(x) \geq g(w) + \alpha(x - w) \quad \text{para } x \in (a, b)$$

Teorema 1.1 (Desigualdad de Jensen) Sea X una variable aleatoria que toma valores en el intervalo (posiblemente infinito) (a, b) tal que $E(X)$ existe y sea $g : (a, b) \rightarrow \mathbb{R}$ una función convexa tal que $\mathbb{E}|g(X)| < \infty$. Entonces:

$$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X)).$$

Sea X que toma valores en (a, b) con media $\mu = \mathbb{E}(X)$. Sea g una función convexa sobre este intervalo satisfaciendo $\mathbb{E}|g(X)| < \infty$. Por la proposición (2.1) con $w = \mu$ existe un $\alpha \in \mathbb{R}$ tal que $g(x) \geq g(\mu) + \alpha(x - \mu)$. Por tanto $g(X) \geq g(\mu) + \alpha(X - \mu)$. Tomando esperanza en ambos lados, obtenemos $\mathbb{E}(g(X)) \geq g(\mu)$.

Ejemplo 1.1 La función $g(x) = -\log x$ es convexa en el intervalo $(0, \infty)$. Por la desigualdad de Jensen aplicada a una variable aleatoria X positiva con media finita,

$$\mathbb{E}(\log X) \leq \log(\mathbb{E}(X))$$

Supongamos que X es una variable aleatoria discreta que es igual de probable de tener los valores x_1, x_2, \dots, x_n . Entonces:

$$\mathbb{E}(\log X) = \frac{1}{n} \sum_{i=1}^n \log x_i = \log \gamma, \quad \mathbb{E}(X) = \bar{X}$$

donde:

$$\gamma = \left(\prod_{i=1}^n x_i \right)^{1/n}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

son la media geométrica y aritmética de los valores x_i respectivamente, con la propiedad que $\log \gamma \leq \log \bar{X}$ que conduce a que $\gamma \leq \bar{x}$.

La desigualdad de Markov se refiere a la siguiente pregunta: si sabes que la esperanza $\mathbb{E}(X)$ de una variable aleatoria no negativa X es finita, ¿qué dice esto acerca de la distribución de X ? Las llamadas colas derecha e izquierda de X son las probabilidades de que X sea grande y positiva (respectivamente, grande y negativa). Más específicamente, la cola (derecha) es la función $\mathbb{P}(X \geq t)$ para t grande, con una definición correspondiente para la cola izquierda. Si X es positiva y $\mathbb{E}(X) < \infty$, entonces la cola derecha de X no puede ser demasiada 'grande'.

Teorema 1.2 (Desigualdad de Markov) Para una variable no negativa X

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t} \quad \text{para } t > 0.$$

Sea X una variable aleatoria no negativa y sea $t > 0$. Como X es una función desde Ω a $[0, \infty)$, tenemos la desigualdad trivial:

$$X(\omega) = \begin{cases} t & \text{Si } X(\omega) \geq t, \\ 0 & \text{Si } X(\omega) < t \end{cases}$$

para $\omega \in \Omega$. Podemos escribir, esto como una desigualdad entre la función X y la función indicador del evento $A = \{X \geq t\}$:

$$X \geq tI_A$$

Ahora tomando, esperanza y recordando que $\mathbb{E}(I_A) = \mathbb{P}(A)$, tenemos la desigualdad requerida $\mathbb{E}(X) \geq t\mathbb{P}(X \geq t)$.

Ejemplo 1.2 Sea X una variable aleatoria. El número m es llamado una mediana de X , si

$$\mathbb{P}(X < m) \leq \frac{1}{2} \leq \mathbb{P}(X \leq m)$$

2 Funciones generadoras de probabilidad

Las funciones generadoras son una herramienta útil para variables aleatorias que toman valores enteros. Los momentos de una variable aleatoria son derivadas de las funciones generadoras. Las funciones generadoras son útiles en el entendimiento de la suma de variables aleatorias.

Una manera de registrar una secuencia de números reales u_0, u_1, u_2, \dots es escribir una fórmula del n -ésimo término. Otra manera es escribir, la función generadora de la secuencia, definida como la suma de la serie de potencias:

$$u_0 + u_1s + u_2s^2 + \dots$$

Por ejemplo la secuencia $1, 2, 4, 8, \dots$ tiene una función generadora :

$$1 + 2s + 4s^2 + \dots = \sum_{n=0}^{\infty} (2s)^n = \frac{1}{1-2s}$$

válida siempre que $|s| < \frac{1}{2}$. De manera similar, la secuencia $1, 2, 3, \dots$ tiene una función generadora:

$$1 + 2s + 3s^2 + \dots = \sum_{n=0}^{\infty} (n+1)s^n = \frac{1}{1-s}^2 \quad (1)$$

válida siempre que $|s| < 1$.

Tales funciones generadoras son útiles para tratar con secuencias reales desde que ellas especifican la secuencia de manera única, es decir dada la secuencia u_0, u_1, u_2, \dots podemos encontrar una función generadora y viceversa, si tenemos una función generadora $U(s)$ con una serie de Taylor convergente:

$$U(s) = u_0 + u_1s + u_2s^2 + \dots$$

para todo s pequeño, entonces esta secuencia es única y $U(s)$ genera la secuencia u_0, u_1, u_2, \dots de manera única.

Podemos definir otro tipos de funciones generadoras, como es el caso de la función generadora exponencial de la secuencia real u_0, u_1, u_2, \dots , definida como la serie de potencia:

$$u_0 + u_1s + \frac{1}{2!}u_2s^2 + \frac{1}{3!}u_3s^3 + \dots$$

siempre que la series converga.

Cuando operamos con funciones generadoras de secuencias reales, es importante que la serie de potencia subyacente converja para cierto $s \neq 0$ y mientras este sea el caso, normalmente no vamos a decir para qué valores de s la serie de potencia es absolutamente convergente.

Por ejemplo, decimos que $(1-s)^{-2}$ es la función generadora de la secuencia $1, 2, 3, \dots$, sin referenciar explícitamente al hecho de que la serie que le corresponde (1) converge absolutamente cuando $|s| < 1$.

Ejemplo 2.1 La secuencia dada por:

$$u_n = \begin{cases} \binom{N}{n} & \text{si } n = 0, 1, 2, \dots, N \\ 0 & \text{en otros casos.} \end{cases}$$

tiene una función generadora:

$$U(s) = \sum_{n=0}^N \binom{N}{n} s^n = (1+s)^N.$$

2.1 Variables aleatorias de valor entero

Muchas variables aleatorias toman valores en el conjunto de los enteros no negativos. Podemos escribir la función masa de esta variable X como un secuencia p_0, p_1, p_2, \dots de números, donde

$$p_k = \mathbb{P}(X = k) \quad \text{para todo } k = 0, 1, 2, \dots$$

satisfaciendo

$$p_k \geq 0 \text{ para } k, \text{ y } \sum_{k=0}^{\infty} p_k = 1$$

Definición 2.1 La función generadora de probabilidad o pgf de X es la función $G_X(s)$ definida por

$$G_X(s) = p_0 + p_1s + p_2s^2 + \cdots,$$

para todo valor s para el cual la sumatoria converge absolutamente.

En otras palabras la función generadora de probabilidad $G_X(s)$ de X es la función generadora de la secuencia p_0, p_1, p_2, \dots , que cumple con ciertas propiedades:

- $G_X(0) = p_0$ y $G_X(1) = 1$
- $G_X(s) = \mathbb{E}(s^X)$, siempre que la esperanza exista.

Desde este último resultado $G_X(s)$ existe para todos los valores de s satisfaciendo $|s| \leq 1$, ya que en este caso, la serie G_X tiene radio de convergencia al menos 1.

$$\sum_{k=0}^{\infty} |p_k s^k| \leq \sum_{k=0}^{\infty} p_k = 1$$

Ejemplo 2.2 Sea X una variable aleatoria teniendo una distribución geométrica con parámetro p . Entonces: $\mathbb{P}(X = k) = pq^{k-1}$ para $k = 1, 2, 3, \dots$, con $p + q = 1$. Entonces X tiene un pgf dado por:

$$\begin{aligned} G_X(s) &= \sum_{k=1}^{\infty} pq^{k-1}s^k \\ &= ps \sum_{k=0}^{\infty} (qs)^k = \frac{ps}{1 - qs} \text{ si } |s| < q^{-1}. \end{aligned}$$

Proposición 2.1 Si X e Y tienen pgf G_X y G_Y respectivamente, entonces:

$$G_X(s) = G_Y(s) \text{ para todo } s$$

si y sólo si $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ para $k = 0, 1, 2, \dots$. En otras palabras, las variables aleatorias de valores enteros, tienen la misma función generadora de probabilidad si y sólo si tienen la misma función de masa de probabilidad.

Para probar esta proposición, sólo se necesita probar que $G_X = G_Y$ implica que $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ para todo k . Como G_X y G_Y tienen radio de convergencia a lo más 1 y por tanto tienen una única expansión alrededor del origen:

$$G_X(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k), \quad G_Y(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}(Y = k)$$

Si $G_X = G_Y$, esas dos series tienen idénticos coeficientes.

Por la proposición anterior, se puede deducir para el ejemplo anterior que: si X tiene un pgf igual a $ps(1 - qs)^{-1}$, entonces X tiene una distribución geométrica, de parámetro p .

Ejemplo 2.3 Sea $p = 1 - q \in [0, 1]$ entonces los pgf de algunas distribuciones discretas son:

1. Si X tiene una distribución Bernoulli con parámetro p entonces $G_X(s) = q + ps$.
2. Si X tiene una distribución Binomial con parámetros n y p entonces

$$G_X(s) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} s^k = (q + ps)^n.$$

3. Si X tiene una distribución de Poisson con parámetro λ entonces:

$$G_X(s) = \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k e^{-\lambda} s^k = e^{\lambda(s-1)}.$$

4. Si X tiene una distribución binomial negativa, con parámetros n y p , entonces

$$G_X(s) = \sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n q^{k-n} s^k = \left(\frac{ps}{1-qs} \right)^n \quad \text{si } |s| < q^{-1}.$$

3 Momentos

Para una variable aleatoria discreta X , la esperanza $\mathbb{E}(X)$ indica el centro de la distribución de X . Este es el primer elemento de una secuencia de números conteniendo información acerca de la distribución de X . Esta secuencia de números $\mathbb{E}(X), \mathbb{E}(X^2), \mathbb{E}(X^3), \dots$ es llamada momentos de X .

Definición 3.1 Sea $k \geq 1$. El k -ésimo momento de la variable aleatoria discreta X es el valor de $\mathbb{E}(X^k)$.

Posiblemente los valores más importantes que surgen a partir de la definición de momentos es la esperanza $\mathbb{E}(X)$ y la varianza de X :

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

Una relación importante entre la varianza y el momento de X , se puede distinguir a partir de los siguientes resultados:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2. \end{aligned}$$

Si X es una variable aleatoria con valores enteros no negativos, los momentos de X se pueden encontrar a partir de la función generadora de momentos de X , calculando la derivada de la función en el punto $s = 1$.

Proposición 3.1 Sea X una variable aleatoria con una función generadora de probabilidad, $G_X(s)$, la r -ésima derivada e $G_X(s)$ en el punto $s = 1$ es igual $\mathbb{E}(X[X-1] \cdots [X-r+1])$ para $r = 1, 2, \dots$. Es decir:

$$G_X^{(r)}(1) = \mathbb{E}(X[X-1] \cdots [X-r+1]).$$

De la proposición anterior, calcular los momentos de X se pueden realizar de la siguientes maneras:

1. $\mathbb{E}(X) = G'_X(1)$.
2. $\mathbb{E}(X^2) = \mathbb{E}(X[X-1] + X) = \mathbb{E}(X[X-1]) + \mathbb{E}(X) = G''_X(1) + G'_X(1)$.
3. $\text{Var}(X) = G''_X(1) + G'_X(1) - G_X(1)^2$.

Ejemplo 3.1 Sea X una variable aleatoria con una distribución geométrica con parámetro $p \in (0, 1)$. El pgf de X es $G_X(s) = ps(1 - qs)^{-1}$ para $|s| < q^{-1}$, cuando $p + q = 1$. Así:

$$\begin{aligned}\mathbb{E}(X) &= G'_X(1) = \frac{1}{p} \\ \mathbb{E}(X^2) &= G''_X(1) + G'_X(1) = \frac{q+1}{p^2} \\ \text{Var}(X) &= \frac{q}{p^2}\end{aligned}$$

4 Suma de variables aleatorias independientes

Muchos de los problemas en Probabilidad, están relacionados con la suma de variables aleatorias. La fórmula de convolución es inconveniente, desde que $n - 1$ convoluciones son requeridas para encontrar la función de masa de probabilidad de la suma de n variables aleatorias independientes y cada operación puede ser muy complicada. En este aspecto las funciones generadoras de momentos son un herramienta importante.

Proposición 4.1 Si X y Y son variables aleatorias independientes, cada una de ellas con valores en $\{0, 1, \dots\}$, entonces su suma tiene una función generadora de probabilidad:

$$G_{X+Y}(s) = G_X(s)G_Y(s)$$

En general, la suma $S_n = X_1 + X_2 + \dots + X_n$ de n variables aleatorias independientes, cada una de ellas con valores en $\{0, 1, \dots\}$, tiene una función generadora de probabilidad:

$$G_{S_n}(s) = G_{X_1}(s)G_{X_2}(s) \dots G_{X_n}(s)$$

Proposición 4.2 Sea N y X_1, X_2, \dots variables aleatorias idénticamente distribuidas e independientes, con valores en $\{0, 1, 2, \dots\}$ y un pgf común G_X , entonces la suma:

$$S = X_1 + X_2 + \dots + X_N$$

tienen una función generadora de probabilidad:

$$G_S(s) = G_N(G_X(s))$$

La fórmula anterior proporciona mucha información acerca de la suma de variables aleatorias. Por ejemplo, para encontrar la esperanza de S en la notación del anterior proposición, calculamos G'_S como sigue:

$$G'_S(s) = G'_N(G_X(s))G'_X(s)$$

Para $s = 1$ obtenemos:

$$G'_S(1) = G'_N(G_X(1))G'_X(1) = G'_N(1)G'_X(1)$$

desde que $G_X(1) = 1$. Luego como $\mathbb{E}(X) = G'_X(1)$, se tiene:

$$\mathbb{E}(S) = G'_S(1) = \mathbb{E}(N)\mathbb{E}(X)$$

donde $\mathbb{E}(X)$ es la esperanza de una variable aleatoria X_i .

Proposición 4.3 Supongamos que todos los momentos $\mathbb{E}(X), \mathbb{E}(X^2) \dots$ de la variable aleatoria X existen y que la serie:

$$\sum_{k=0}^{\infty} \frac{1}{k!} t^k \mathbb{E}(X^k)$$

es absolutamente convergente para algún $t > 0$. Entonces la secuencia de momentos determina únicamente la distribución de X .

La absoluta convergencia para algún t es una condición suficiente más no necesaria, para los momentos de determinar una distribución subyacente.

5 Varianza y covarianza

La varianza de una variable aleatoria X es definida como:

$$\text{Var}(X) = \mathbb{E}([X - \mu]^2),$$

donde $\mu = \mathbb{E}(X)$.

La varianza es una medida de dispersión con respecto a μ , en el sentido de que si X toma valores que difieren considerablemente de μ , entonces el valor de $|X - \mu|$ es grande y $\mathbb{E}([X - \mu]^2)$ tiene un valor grande, mientras que si X es cercano al valor de μ , entonces $|X - \mu|$ tiene un valor que usualmente pequeño de la misma forma que $\mathbb{E}([X - \mu]^2)$.

En un caso extremo cuando X está concentrado en algún punto, para una variable Y se tiene el siguiente resultado:

$$\mathbb{E}(Y^2) = 0 \text{ si y sólo si } \mathbb{P}(Y = 0) = 1$$

si aplicamos la ecuación anterior a $Y = X - \mu$, tenemos el siguiente resultado:

$$\text{Var}(X) = 0 \text{ si sólo si } \mathbb{P}(X = \mu) = 1$$

de esta forma la varianza cero significa que no hay dispersión en lo absoluto.

5.1 Propiedades de la varianza

Algunas propiedades de la varianza son:

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

La varianza como medida de dispersión tiene una propiedad no deseable: no es lineal en el sentido de que la varianza de aX es a^2 veces la varianza de X , por esta razón es preferible trabajar con la desviación estándar de X , definido como $\sqrt{\text{Var}(X)}$.

En términos de $\text{Var}(X)$ y $\text{Var}(Y)$, la expresión de $\text{Var}(X + Y)$ es de la forma:

$$\text{Var}(X + Y) = \text{Var}(X) + 2\mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) + \text{Var}(Y)$$

Es conveniente tener una palabra especial para el término medio en la última expresión y para ello definimos la covarianza de X e Y .

Definición 5.1 La covarianza de las variables aleatorias X y Y es la cantidad denotada por $\text{cov}(X, Y)$ y es dada por:

$$\text{cov}(X, Y) = \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

siempre que las esperanzas existan.

Podemos agregar algunas propiedades más, de la varianza y covarianza:

- $\text{Var}(X + Y) = \text{Var}(X) + 2\text{cov}(X, Y) + \text{Var}(Y)$
- Si X y Y son independientes : $\text{cov}(X, Y) = 0$ y $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

La covarianza se utiliza a menudo como una medida de la dependencia de X e Y y la razón de esto es que $\text{cov}(X, Y)$ es un sólo un número (en lugar de un objeto complicado como la función densidad conjunta) que contiene información útil sobre el comportamiento conjunto de X e Y .

Por ejemplo, si $\text{cov}(X, Y) > 0$, entonces $X - \mathbb{E}(X)$ y $Y - \mathbb{E}(Y)$ pueden tener una buena chance (en algún sentido) de tener el mismo signo.

La principal desventaja de la covarianza como medida de dependencia de X e Y es que no es invariante en la escala: Si X y Y están medidas en pulgadas y U y V tienen la misma medida en centímetros (tal que $U = \alpha X$ y $V = \alpha Y$, donde $\alpha \sim 2.54$), entonces $\text{cov}(U, V) \sim 6\text{cov}(X, Y)$ a pesar del hecho de que el par (X, Y) y (U, V) miden la misma cantidad.

Para tratar con esta reescala en la covarianza se define lo siguiente:

Definición 5.2 El coeficiente de correlación de las variables aleatorias X y Y es la cantidad $\rho(X, Y)$ definida por:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

siempre que la última cantidad exista y $\text{Var}(X)\text{Var}(Y) \neq 0$.

Una propiedad del coeficiente de correlación que se pueden deducir de la definición es: $\rho(aX + b, cY + d) = \rho(X, Y)$, para $a, b, c, d \in \mathbb{R}$ tal que $ac \neq 0$, es decir el coeficiente de correlación es invariante a escala.

La correlación tiene una propiedad importante como medida de dependencia:

Teorema 5.1 Si X y Y son variables aleatorias, entonces $-1 \leq \rho(X, Y) \leq 1$, siempre que la correlación exista.

La prueba de este teorema, es una aplicación directa de la siguiente desigualdad:

Teorema 5.2 (Teorema de Cauchy-Schwarz) Si U y V son variables aleatorias, entonces

$$[\mathbb{E}(UV)]^2 \leq \mathbb{E}(U^2)\mathbb{E}(V^2)$$

siempre que las esperanzas existan.

Para probar el Teorema 5.1, se coloca $U = X - \mathbb{E}(X)$ y $V = Y - \mathbb{E}(Y)$, en la desigualdad de Cauchy-Schwarz, para hallar que:

$$\text{cov}(X, Y)^2 \leq \text{Var}(X)\text{Var}(Y).$$

Lo que concluye con el resultado.

Sea $a = \text{Var}(X)$, $b = 2\text{cov}(X, Y)$, $c = \text{Var}(Y)$ y supongamos que $\rho(X, Y) \neq \pm 1$. Entonces tenemos que $\text{Var}(X)\text{Var}(Y) \neq 0$ y

$$b^2 - 4ac = 4\text{Var}(X)\text{Var}(Y)[\rho(X, Y)^2 - 1] = 0$$

y así la ecuación cuadrática: $as^2 + bs + c = 0$ tiene dos raíces iguales en $s = \alpha$. Por tanto, $W = \alpha[X - \mathbb{E}(X)] + [Y - \mathbb{E}(Y)]$ satisface

$$\mathbb{E}(W^2) = a\alpha^2 + b\alpha + c = 0$$

dando que $\mathbb{P}(W = 0) = 1$ y probando que $Y = -\alpha X + \beta$, donde $\beta = \alpha\mathbb{E}(X) + \mathbb{E}(Y)$.

Un tratamiento más exhaustivo distingue los valores de ± 1 :

- $\rho(X, Y) = 1$ si y sólo si $P(Y = \alpha X + \beta) = 1$ para algún real α y β con $\alpha > 0$.
- $\rho(X, Y) = -1$ si y sólo si $P(Y = \alpha X + \beta) = 1$ para algún real α y β con $\alpha < 0$.

Se usa $\rho(X, Y)$ como una medida de las dependencias de X y Y . Si X y Y tienen varianzas distintas de cero, entonces $\rho(X, Y)$ toma algún valor en el intervalo $[-1, 1]$ y este valor debe ser interpretado de la forma en que los valores $-1, 0, 1$ surgen:

- Si X y Y son independientes entonces $\rho(X, Y) = 0$.
- Y es una función creciente de X si y sólo si $\rho(X, Y) = 1$.
- Y es una función decreciente de X si y sólo si $\rho(X, Y) = -1$.

Si $\rho(X, Y) = 0$ se dice que X y Y son no correlacionados.

6 Funciones generadoras de momentos

Si X es una variable aleatoria tomando los valores en $\{0, 1, 2, \dots\}$, una función generadora de probabilidad es definida como:

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k) \quad (2)$$

Las funciones generadoras de probabilidad son muy útiles, pero sólo cuando las variables aleatorias toman valores enteros no negativos. Para variables aleatorias mucho más generales, es necesario hacer una modificación de la ecuación anterior.

Definición 6.1 La función generadora de momentos (MGF) de la variable aleatoria de X es la función M_X definida como:

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

para todo $t \in \mathbb{R}$ para el cual la esperanza existe.

Esta es una modificación de (2), en el sentido que, si X toma valores en $\{0, 1, 2, \dots\}$, entonces

$$M_X(t) = \mathbb{E}(e^{tX}) = G_X(e^t)$$

con la sustitución de $s = e^t$. En general:

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_x e^{tx} \mathbb{P}(X = x) \text{ si } X \text{ es discreta}$$

Con esta suma o integral converga absolutamente en algunos casos. La existencia de $M_X(t)$ puede plantear un problema para valores t distintos de cero.

Ejemplo 6.1 Si X tiene una distribución de Poisson con parámetro λ y función de masa de probabilidad:

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, \dots$$

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!}$$

Si realizamos el cambio de variable $a = e^t \lambda$, entonces obtenemos $M_X(t) = e^{\lambda(e^t - 1)}$.

La razón del nombre de función generadora de momentos es la siguiente la expansión que muestra que $M_X(t)$ es la función generadora exponencial de los momentos de X :

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}) = \mathbb{E}\left(1 + tX + \frac{1}{2!}(tX)^2 + \dots\right) \\ &= 1 + t\mathbb{E}(X) + \frac{1}{2!}t^2\mathbb{E}(X^2) + \dots \end{aligned}$$

Teorema 6.1 Si $M_X(t)$ existe en una vecindad de 0, entonces para $k = 1, 2, \dots$

$$\mathbb{E}(X^k) = M_X^{(k)}(0)$$

la k -ésima derivada de $M_X(t)$ evaluada en $t = 0$.

Como se ha señalado antes, mucho de la teoría se refiere a la suma de variables aleatorias, que puede ser difícil en la práctica calcular de la distribución de una suma a partir del conocimiento de las distribuciones de los sumandos y es aquí donde las funciones generadoras de momentos son muy útiles.

Consideremos primero la función lineal $aX + b$ de la variable aleatoria X . Si $a, b \in \mathbb{R}$

$$\begin{aligned} M_{aX+b}(t) &= \mathbb{E}(e^{t(aX+b)}) = \mathbb{E}(e^{atX}e^{tb}) \\ &= e^{tb}\mathbb{E}(e^{(at)X}) \end{aligned}$$

lo que produce que $M_{aX+b}(t) = e^{tb}M_X(at)$.

Teorema 6.2 Si X e Y son variables aleatorias independientes, entonces $X + Y$ tiene una función generadora de momentos:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

Tenemos por independencia y propiedad de la esperanza para la variable aleatoria X y Y :

$$\begin{aligned} M_{X+Y}(t) &= \mathbb{E}(e^{t(x+y)}) = \mathbb{E}(e^{tX}e^{tY}) \\ &= \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) \end{aligned}$$

En general la suma $S = X_1 + \cdots + X_n$ de n variables aleatorias independientes tiene una función generadora de momentos de la forma:

$$M_S(t) = M_{X_1}(t) \cdots M_{X_n}(t)$$