1 Funciones generadoras de probabilidad

Las `funciónes 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 sum de variables aleatorias.

1.1 Funciones generadoras

Una manera de registrar una secuencia de números reales $u_0, u_1, u_2, \ldots$ 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:

$$u_0 + u_1 s + u_2 s^2 + \cdots$$

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

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

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

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

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 $u_0, u_1, u_2, \ldots$ podemos encontrar una función generadora y viceversa, si tenemos una función generadora $\mathbb{U}(s)$ con una serie de Taylor convergente:

$$\mathbb{U}(s) = u_0 + u_1 s + u_2 s^2 + \cdots$$

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

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

$$u_0 + u_1 s + \frac{1}{2!} u_2 s^2 + \frac{1}{3!} u_3 s^3 + \cdots$$

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 $1, 2, 3, \ldots$, sin referenciar explícitamente al hecho de que la serie que le corresponde (1) converge absolutamente cuando $|s| < 1$.

Ejemplo 1.1 La secuencia dada por

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

tiene una función generadora

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

1.2 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 $p_0, p_1, p_2, \ldots$ de números, donde

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

satisfaciendo

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

Definición 1.1 La `función generadora de probabilidad` o `pgf` de X es la función $\mathbb{G}_X(s)$ definida por

$$\mathbb{G}_X(s) = p_0 + p_1 s + p_2 s^2 + \cdots,$$

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

En otras palabras la función generadora de probabilidad $\mathbb{G}_X(s)$ de X es la función generadora de la secuencia $p_0, p_1, p_2, \ldots$, que cumple con ciertas propiedades:

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

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

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

Ejemplo 1.2 Sea X una variable aleatoria teniendo una distribución geométrica con parámetro p. Entonces: $\mathbb{P}(X = k) = pq^{k-1}$ para $k = 1, 2, 3, \ldots$, con $p + q = 1$. Entonces X tiene un `pgf` dado por:

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

Proposición 1.1 Si X e Y tienen `pgf` $\mathbb{G}_X$ y $\mathbb{G}_Y$ respectivamente, entonces

$$\mathbb{G}_X(s) = \mathbb{G}_Y(s) \quad \text{para todo} \quad s$$

si y sólo si $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ para $k = 0, 1, 2, \ldots$. 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 $\mathbb{G}_X = \mathbb{G}_Y$ implica que $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$ para todo k. Como $\mathbb{G}_X$ y $\mathbb{G}_Y$ tienen radio de convergencia a lo más 1 y por tanto tienen una única expansión alrededor del origen:

$$\mathbb{G}_X(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k), \quad \mathbb{G}_Y(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}(Y = k)$$

Si $\mathbb{G}_X = \mathbb{G}_Y$, 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 geómetrica, de parámetro p.

Ejemplo 1.3 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 $\mathbb{G}_X(s) = q + ps$.
2. Si X tiene una distribucioń Binomial con parámetros n y p entonces

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

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

$$\mathbb{G}_X(s) = \sum_{k=0}^{n} \frac{1}{k!} \lambda^k e^{-\lambda} s^k = e^{\lambda(s-1)}.$$

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

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

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 $\mathbb{E}(X), \mathbb{E}(X^2), \mathbb{E}(X^3), \ldots$ es llamada `momentos` de X.

Definición 1.2 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 $\mathbb{E}(X)$ de X y la varianza de X :

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

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

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

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

Proposición 1.2 Sea X una variable aleatoria con una función generadora de probabilidad, $\mathbb{G}_X(s)$, la `r-ésima` derivada e $\mathbb{G}_X(s)$ en el punto $s = 1$ es igual $\mathbb{E}(X[X-1]\cdots[X-r+1])$ para $r = 1, 2, \ldots$. Es decir:

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

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

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

Ejemplo 1.4 Sea X una variable aleatoria con una distribución geómetrica con parámetro $p \in (0, 1)$. El `pgf` de X es $\mathbb{G}_X(s) = ps(1-qs)^{-1}$ para $|s| < q^{-1}$, cuando $p + q = 1$. Así:

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

1.4 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 1.3 Si X y Y son variables aleatorias independientes, cada una de ellas con valores en $\{0, 1, \ldots\}$, entonces su suma tiene una función generadora de probabilidad:

$$\mathbb{G}_{X+Y}(s) = \mathbb{G}_X(s)\mathbb{G}_Y(s)$$

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

$$\mathbb{G}_{S_n}(s) = \mathbb{G}_{X_1}(s) G_{X_2}(s) \cdots \mathbb{G}_{X_n}(s)$$

Proposición 1.4 Sea N y $X_1, X_2, \dots$ variables aleatorias identicamente distribuidas eindependientes, con valores en $\{0, 1, 2, \dots\}$ y un `pgf` común $\mathbb{G}_X$, entonces la suma

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

tienen una función generadora de probabilidad:

$$\mathbb{G}_S(s) = \mathbb{G}_N(\mathbb{G}_X(s))$$

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 preposición, calculamos $\mathbb{G}_S^{'}$ como sigue

$$\mathbb{G}_S^{'}(s) = \mathbb{G}_N^{'}(\mathbb{G}_X(s)) \mathbb{G}_X^{'}(s)$$

Para $s = 1$ obtenemos:

$$\mathbb{G}_S^{'}(1) = \mathbb{G}_N^{'}(\mathbb{G}_X(1)) \mathbb{G}_X^{'}(1) = \mathbb{G}_N^{'}(1) \mathbb{G}_X^{'}(1)$$

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

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

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

1.5 Caso general de momentos

Para una variable aleatoria X el `k-ésimo momento` de X es definido para $k = 0, 1, 2, \cdots$ como el número $\mathbb{E}(X^k)$ que es la esperanza de la `k-ésima potencia` de X siempre que la esperanza exista.

La secuencia $\mathbb{E}(X), \mathbb{E}(X^2), \dots$ contiene mucha información acerca de la distribución de X.

Ejemplo 1.5 Si X tienen una distribución exponencial con paramétro λ, entonces:

$$\begin{aligned} \mathbb{E}(X^k) &= \int_0^{\infty} x^k \lambda e^{-\lambda x} dx \\ &= \frac{k}{\lambda} \mathbb{E}(X^{k-1}) \end{aligned}$$

Si $k \geq 1$, tenemos el siguiente resultado:

$$
\begin{aligned}
\mathbb{E}(X^k) = \frac{k}{\lambda}\mathbb{E}(X^{k-1}) = \frac{k(k-1)}{\lambda^2}\mathbb{E}(X^{k-2}) = \cdots \\
= \frac{k!}{\lambda^k}\mathbb{E}(X^0) = \frac{k!}{\lambda^k}\mathbb{E}(1) = \frac{k!}{\lambda^k}
\end{aligned}
$$

En particular, la distribución exponencial tienen momentos de todos los órdenes.

Ejemplo 1.6 Si X tiene una distribución de Cauchy, entonces

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} \frac{x^k}{\pi(1+x^2)} dx$$

para valores de k para el cual, la integral converge absolutamente. Pero se cumple que:

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

si $k \geq 1$. Entonces la distribución de Cauchy no tiene momentos.

Podemos adaptar este ejemplo para encontrar una función densidad con algunos momentos, pero no con todos. Considera la función de densidad:

$$f_X(x) = \frac{c}{1+|x|^m} \text{ para } x \in \mathbb{R}.$$

donde $m(\geq 2)$ es un entero y c es escogido de manera que, en efecto es una función densidad:

$$c = \left(\int_{-\infty}^{\infty} \frac{dx}{1+|x|^m} \right)^{-1}.$$

Se puede verificar que esta función densidad tiene un k-ésimo momento, para aquellos valores k satisfaciendo $1 \leq k \leq m-2$ solamente.

Dada la función de distribución F_X de la variable aleatoria X, podemos calcular los momento siempre y cuando esta función exista, ya sea X discreta o continua. Lo contrario; es decir ¿dada una secuencia $\mathbb{E}(X), \mathbb{E}(X^2), \ldots$ de momentos(finitos) de X es posible reconstruir X?. La respuesta en general no se cumple en general, salvo algunas condiciones extras.

Proposición 1.5 Supongamos que todos los momentos $\mathbb{E}(X), \mathbb{E}(X^2) \ldots$ 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 para determinar una distribución subyacente. Mostremos que una distribución que no es determinada por momentos de manera única

Ejemplo 1.7 Si X tiene una distribución normal con media 0 y varianza 1, entonces $Y = e^X$ tiene una distribución `log-normal` con función densidad

$$f_X(x) = \begin{cases} \frac{1}{x\sqrt{2\pi}} \exp[-\frac{1}{2}(\log x)^2] & \text{si } x > 0, \\ 0 & \text{en otros casos} \end{cases}$$

Suponiendo que $-1 \leq a \leq 1$, definimos:

$$f_a x = [1 + a \sin(2\pi \log x)] f(x)$$

entonces se cumple:

- f_a es una función densidad.
- f tiene momentos finitos de todos los órdenes.
- Si f_a y f tienen momentos iguales de todos los órdenes, cuando

$$\int_{-\infty}^{\infty} x^k f(x) dx = \int_{-\infty}^{\infty} x^k f_a(x) dx \;\; \text{para } k = 1, 2, \ldots$$

Así $\{f_a : -1 \leq a \leq 1\}$ es una colección de distintas funciones densidad teniendo los mismos momentos.

Varianza y covarianza

La varianza de una variable aleatoria X es definida como:

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

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 por lo tanto $\mathbb{E}([X - \mu]^2)$ 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 es `concentrado` en algún punto. En este caso, para una variable Y:

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

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

$$\mathbb{V}(X) \;\; \text{si sólo si} \;\; \mathbb{P}(X = \mu) = 1$$

de esta forma la varianza cero significa que `no hay dispersión en lo absoluto`. Algunas propiedades de la varianza son las siguientes:

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

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 $\sqrt{\mathbb{V}(X)}$.

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

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

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 1.3 La `covarianza` de las variables aleatorias X y Y es la cantidad denotada por $cov(X,Y)$ y es dada por:

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

siempre que las esperanzas existan.

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

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

La covarianza se utiliza a menudo como una medida de la dependencia de X e Y y la razón de esto es que $cov(X,Y)$ 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 $cov(X,Y) > 0$, 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 estaán medidas en pulgadas y U y V tienen las misma medida en centímetros (tal que $U = \alpha X$ y $V = \alpha Y$, donde $\alpha \sim 2.54$), entonces $cov(U,V) \sim 6cov(X,Y)$ a pesar del hecho de que el par (X,Y) y (U,V) miden la misma cantidad. Para tratar con esta `re-escala` en la covarianza se define lo siguiente:

Definición 1.4 `El coeficiente de correlación` de las variables aleatorias X y Y es la cantidad $\rho(X,Y)$ definida por:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

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

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 1.1 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 1.2 (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 1.1`, se coloca $U = X - \mathbb{E}(X)$ y $V = Y - \mathbb{E}(Y)$, en la desigualdad de Cauchy-Schwarz, para hallar que:

$$cov(X,Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y).$$

Lo que concluye con el resultado.

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

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

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 exahustivo 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 varianza 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`.

2 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:

$$\mathbb{G}_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k) \tag{2}$$

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 2.1 La función generadora de momentos (mgf) de la variable aleatoria de X es la función $\mathbb{M}_X$ definida como:

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

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

$$\mathbb{M}_X(t) = \mathbb{E}(e^{tX}) = \mathbb{G}_X(e^t)$$

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

$$\mathbb{M}_X(t) = \mathbb{E}(e^{tX}) = \begin{cases} \displaystyle\sum_x e^{tx}\mathbb{P}(X = x) & \text{si } X \text{ es discreta} \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f_X(x)dx & \text{si } X \text{ es continua} \end{cases}$$

siempre que 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 2.1 Si X tiene una distribución normal con media 0 y varianza 1, entonces

$$\begin{aligned} \mathbb{M}_X(t) &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx \\ &= e^{\frac{1}{2}t^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-t)^2}dx \\ &= e^{\frac{1}{2}t^2} \end{aligned}$$

desde que el integrando en la última integral es la función densidad de la distribución normal con media t y varianza 1, la integral tiene que ser 1. El momento $\mathbb{M}_X(t)$ existe para todo $t \in \mathbb{R}$.

Ejemplo 2.2 Si X tiene una distribución de Poisson con parámetro λ y función de masa de probabilidad:

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

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

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

Ejemplo 2.3 Si X tiene una distribución exponencial con paramétro λ, entonces

$$M_X(t) = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}dx = \begin{cases} \dfrac{\lambda}{\lambda - t} & \text{si } t < \lambda \\ \infty & \text{si } t \geq \lambda. \end{cases}$$

así existe $M_X(t)$ existe sólo para valores t satisfaciendo $t < \lambda$.

Ejemplo 2.4 Si X tiene una distribución de Cauchy, entonces

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{\pi(1+x^2)}dx = \begin{cases} 1 & \text{si } t = 0 \\ \infty & \text{si } t \neq 0. \end{cases}$$

así $M_X(t)$ existe sólo cuando $t = 0$.

La dificultad de la existencia de $\mathbb{E}(e^{tX})$ puede ser evitada usando la función de variable compleja $\phi_X(t) = \mathbb{E}(e^{itX})$ llamada `función característica` de X que existe para todo $t \in \mathbb{R}$. Es importante decir que $\mathbb{E}(e^{tX})$ existe en alguna vecindad del origen y la razón está contenida en el teorema de unicidad de funciones generadores de momentos. Generalmente utilizaremos las funciones generadoras de momentos libremente, pero siempre sujetas a la asunción implícita de la existencia de una vecindad del origen.

La razón del nombre de función generadora de momentos es la siguiente la expansión que muestra que $\mathbb{M}_X(t)$ es la función generadora exponencial de los momentos de X:

$$\begin{aligned} \mathbb{M}_X(t) = \mathbb{E}(e^{tX}) &= \mathbb{E}\Big(1 + tX + \frac{1}{2!}(tX)^2 + \cdots\Big) \\ &= 1 + t\mathbb{E}(X) + \frac{1}{2!}t^2\mathbb{E}(X^2) + \cdots \end{aligned}$$

Teorema 2.1 Si $\mathbb{M}_X(t)$ existe en una vecindad de 0, entonces para $k = 1, 2 \ldots$

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

la `k-ésima` derivada de $\mathbb{M}_X(t)$ evaluada en $t = 0$.

Como se ha señalado antes, mucho de la teoria 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 $a, b \in \mathbb{R}$

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

lo que produce que $\mathbb{M}_{aX+b}(t) = e^{tb}\mathbb{M}_X(at)$.

Teorema 2.2 Si X e Y son variables aleatorias independientes, entonces $X + Y$ tiene una función generadora de momentos:

$$\mathbb{M}_{X+Y}(t) = \mathbb{M}_X(t)\mathbb{M}_Y(t)$$

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

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

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

$$\mathbb{M}_S(t) = \mathbb{M}_{X_1}(t) \cdots \mathbb{M}_{X_n}(t)$$

Teorema 2.3 (Unicidad) Si la función generadora de momentos $\mathbb{M}_X$ satisface $\mathbb{M}_X(t) = \mathbb{E}(e^{tX}) < \infty$ para todo t satisfaciendo $-\delta < t < \delta$ para algún $\delta > 0$, entonces existe una única distribución con función generadora de momentos $\mathbb{M}_X$. Además bajo esta condición, se tiene que $\mathbb{E}(X^k) < \infty$ para $k = 1, 2, \dots$ y

$$\mathbb{M}_X(t) = \sum_{k=0}^{\infty} \frac{1}{k!} t^k \mathbb{E}(X^k) \text{ para } |t| < \delta.$$

Ejemplo 2.5 Sean X e Y dos variables aleatorias independiendientes, teniendo X una distribución normal con parámetros μ_1 y σ_1^2 y Y una distribución normal con parámetros μ_2 y σ_2^2, veamos que propiedades tiene la suma $Z = X + Y$:

Sea U una variable aleatoria que tiene una distribución normal con parámetros μ y σ^2. La función generadora de momentos de U es:

$$\begin{aligned}
\mathbb{M}_U(t) = &\int_{-\infty}^{\infty} e^{tu} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(u-\mu)^2\right) du \\
= &e^{\mu t} \int_{-\infty}^{\infty} e^{x\sigma t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \quad \text{sustituyendo} \quad x = \frac{u-\mu}{\sigma} \\
= &\exp(\mu t + \frac{1}{2}\sigma^2 t^2)
\end{aligned}$$

Por el teorema de unicidad:

$$\begin{aligned}
\mathbb{M}_Z(t) &= \mathbb{M}_X(t)\mathbb{M}_Y(t) \\
&= \exp(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2) \exp(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2) \\
&= \exp\left[(\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\right]
\end{aligned}$$

Luego se deduce que es la función generadora de momentos de la distribución normal con parámetros $\mu_1 + \mu_2$ y $\sigma_1^2 + \sigma_2^2$. Así Z tiene esta distribución apelando al teorema de unicidad.

2.1 Transformada de Laplace

Una función que está estrechamente relacionada con la función generadora de momentos es la transformada de Laplace. Para una función real valorada $h(t), t \geq 0$, la transformada de Laplace, denotada por $h^*(s)$, es definida como:

$$h^*(s) = \int_0^{\infty} e^{-sx} h(x) dx$$

Si $h(t), t \geq 0$, es una función que corresponde a una función de densidad de una variable aleatoria no negativa X entonces por la definición de una función generadora de momentos, la transformada de Laplace de la función $h^*(s)$ es igual a la función generadora de momentos de X evaluado en $-s$, es decir:

$$h^*(s) = \mathbb{M}_X(-s). \tag{3}$$

Para preservar la semántica asociada con variables aleatorias utilizamos la notación $\mathcal{L}_X$ cuando se trata de la transformada de Laplace de una variable aleatoria X:

$$\mathcal{L}_X(s) = \int_0^\infty e^{-sx} f_X(x) dx.$$

Una interpretación de la transformada para el caso de una variable aleatoria es que es la esperanza de $\exp[-sX]$ y por tanto, puede escribirse como:

$$\mathcal{L}_X(s) = \mathbb{E}[e^{-sX}].$$

La transformada de Laplace es generalmente más útil que la función generadora de momentos, ya que puede aplicarse a funciones que no son densidades. La transformada de Laplace es la contrapartida continua de la función de generadora ordinaria que se vió anteriormente. Para ver la relación entre estas dos funciones, supongamos que la función $h(t)$ toma valores no nulos para el entero t, es decir, que $h(t) = h_i$, para $t = i$ e $i = 0, 1, \ldots$ y sea $H(x)$ la función generadora:

$$H(x) = \sum_{i=0}^{\infty} h_i x^i.$$

Entonces, se sigue que:

$$h^*(s) = \int_0^\infty e^{-sx} h(x) dx = \sum_{j=0}^{\infty} e^{-sj} h_j = H(e^{-s}). \tag{4}$$

Y así la transformación de Laplace es igual a la función generadora evaluado en e^{-s}. Si h_i representa, a función de densidad de probabilidad, entonces las ecuaciones (3) y (4) muestran que simplemente hemos establecido una forma diferente de relación entre las funciones generadoras de momentos y sus correspondientes funciones generadoras de probabilidad para variables aleatorias discretas.

Dado que las transformadas de Laplace son equivalentes a funciones generadoras de momentos cuando la función subyacente es una densidad, se sigue que ellas satisfacen propiedades análogas a las que se dan para las funciones generadoras de momentos. Por ejemplo, se cumple lo siguiente:

$$\mathcal{L}_X^{(n)}(0) = (-1)^n \mathbb{E}(X^n).$$

La transformada de Laplace se usa frecuentemente cuando se forma la <u>convolución de dos variables aleatorias</u>. Para derivar la transformada de Laplace para una convolución, se $h(t)$ la función densidad que resulta cuando convolucionamos dos funciones de densidad continuas f y g. El operador de convolución para funciones de densidad se define como:

$$h(t) = \int_0^t f(x-t) g(x) dx.$$

De manera equivalente, escribimos la notación de convolución como:

$$h(t) = f \odot g(t).$$

Para calcula $h^*(s)$ escribimos:

$$
\begin{aligned}
h^*(s) &= \int_0^\infty e^{-st}h(t)dt \\
&= \int_0^\infty e^{-st}dt \int_0^t f(t-x)g(x)dx \\
&= \int_0^\infty e^{-sx}g(x)dx \int_0^\infty e^{-s(t-x)}f(t-x)dx \\
&= g^*(s)f^*(s).
\end{aligned}
$$

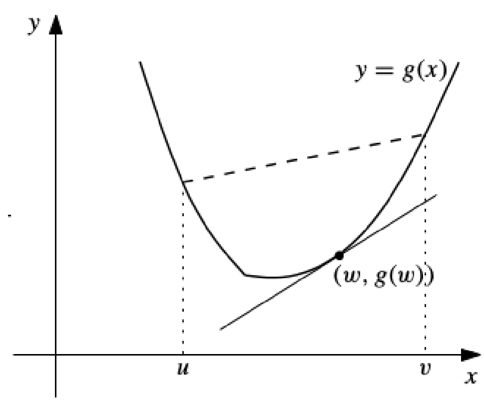
2.2 Algunos resultados

Sea $-\infty \leq a < b < \infty$ una función $g : (a,b) \to \mathbb{R}$ 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 2.1 Sea $g : (a,b) \to \mathbb{R}$ convexa y sea $w \in (u,v)$. Entonces existe un $\alpha \in \mathbb{R}$ tal que

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

Teorema 2.4 `(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 $g : (a,b) \to \mathbb{R}$ 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 2.6 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 $x_1, x_2, \ldots, x_n$. Entonces:

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

donde

$$\gamma = \Big(\prod_{i=1}^{n} x_i\Big)^{1/n}, \qquad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

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 2.5 (`Desigualdad de Markov`) Para una variable no negativa X

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

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 2.7 Sea X una variable aleatoria. El número m es llamado una `mediana` de X, si

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