Capítulo 3

Transformaciones y esperanza

3.1. Introducción

Por lo general estamos en condiciones de modelar un fenómeno en términos de una variable aleatoria X cuya función de distribución acumulada es $F_X(x)$. La pregunta es si estaremos también en condiciones de estudiar el comportamiento de aquellas variables aleatorias que son trasformaciones de X. En este capítulo se estudian algunas técnicas que permiten conocer el comportamiento probabilístico de $Y = g(X)$.

3.2. Transformaciones para variables aleatorias

Si X es una variable aleatoria con función de distribución acumulada $F_X(x)$ entonces cualquier función de X es también una variable aleatoria. Si se define $Y = g(X)$ es posible describir el comportamiento probabilistico de Y en términos de X.

Formalmente $y = g(x)$ define un mapa desde el espacio muestral de X al espacio muestral de Y. Es decir:

$$g(x) : \mathcal{X} \longrightarrow \mathcal{Y}$$

Se asocia a g un mapa inverso, denotado por g^{-1}, definido por:

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\} \tag{3.2.1}$$

Para cualquier conjunto $A \subset \mathcal{Y}$:

$$\begin{aligned} \Pr(Y \in A) &= \Pr(g(X) \in A) \\ &= \Pr(\{x \in \mathcal{X} : g(x) \in A\}) \\ &= \Pr\left(X \in g^{-1}(A)\right) \end{aligned} \tag{3.2.2}$$

3.2.1. Caso discreto

Si X es una variable aleatoria discreta entonces $\mathcal{X}$ es numerable. El espacio muestral para $Y = g(X)$ es:

$$\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$$

el cual también es un conjunto numerable. Usando la ecuación 3.2.2 la función de probabilidad de Y es:

$$f_Y(y) = \Pr(Y = y) = \sum_{x \in g^{-1}(y)} f_X(x)$$

Ejemplo 3.2.1 Sea $X \sim \mathcal{BI}(n, p)$. Hallar la función de probabilidad de $Y = g(X) = n - X$.

Ejemplo 3.2.2 Sea X una variable aleatoria con función de probabilidad:

$$f_X(x) = \frac{1}{2n} \qquad x = \pm 1, \pm 2, \cdots, \pm n$$

Hallar la función de probabilidad de $Y = g(X) = X^2$.

3.2.2. Caso continuo

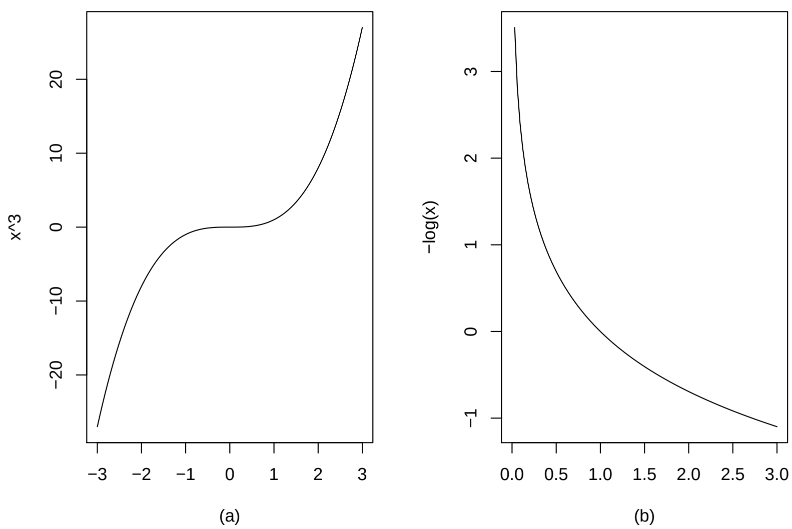
La función de distribución acumulada de $Y = g(X)$ es:

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) \\ &= \Pr(g(X) \leq y) \\ &= \Pr(\{x \in \mathcal{X} : g(x) \leq y\}) \\ &= \int_{\{x \in \mathcal{X} : g(x) \leq y\}} f_X(x)\, dx \end{aligned} \tag{3.2.3}$$

Es sencillo trabajar con funciones $g(x)$ que son *monótonas*, es decir aquellas que satisfacen alguna de las siguientes relaciones:

$$u > v \Rightarrow g(u) > g(v)\,\text{(creciente)} \quad \text{o} \quad u < v \Rightarrow g(u) > g(v)\,\text{(decreciente)}$$

Figura 3.1: Función monótona creciente (a) y decreciente (b)



Si la transformación $x \to g(x)$ es monótona entonces es *uno a uno* y *sobreyectiva*. Si g es creciente:

$$\begin{aligned} \{x \in \mathcal{X} : g(x) \leq y\} &= \left\{x \in \mathcal{X} : g^{-1}(g(x)) \leq g^{-1}(y)\right\} \\ &= \left\{x \in \mathcal{X} : x \leq g^{-1}(y)\right\} \end{aligned} \tag{3.2.4}$$

y usando 3.2.3, se tiene que:

$$F_Y(y) = \int_{\{x \in \mathcal{X} : x \leq g^{-1}(y)\}} f_X(x)\,dx = \int_{-\infty}^{g^{-1}(y)} f_X(x)\,dx = F_X\left(g^{-1}(y)\right)$$

Si g es decreciente, entonces:

$$\begin{aligned}\{x \in \mathcal{X} : g(x) \leq y\} &= \{x \in \mathcal{X} : g^{-1}(g(x)) \geq g^{-1}(y)\} \\ &= \{x \in \mathcal{X} : x \geq g^{-1}(y)\} \qquad (3.2.5)\end{aligned}$$

se tiene que:

$$F_Y(y) = \int_{g^{-1}(y)}^{\infty} f_X(x)\, dx = 1 - F_X\left(g^{-1}(y)\right)$$

Se resumen los resultados anteriores en el siguiente teorema.

Teorema 3.2.1 Sea X una variable aleatoria con función de distribución acumulada $F_X(x)$. Se define $Y = g(X)$ y los espacios muestrales $\mathcal{X}$ y $\mathcal{Y}$.

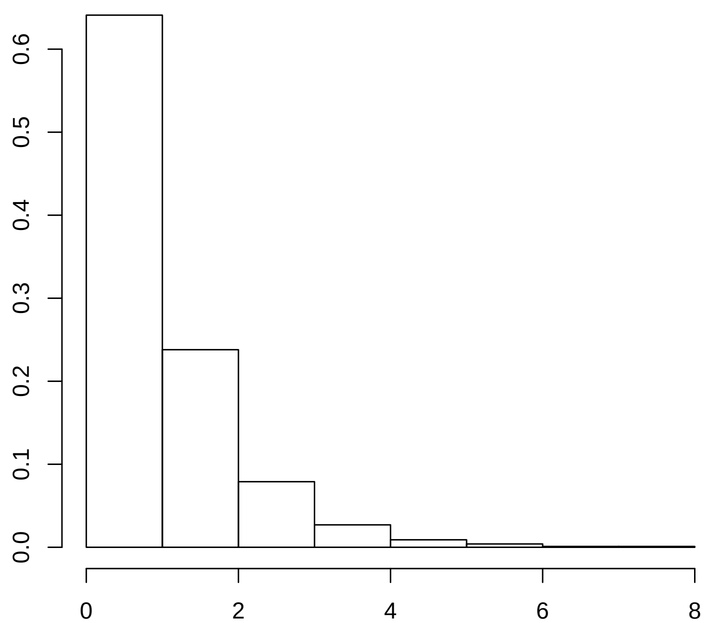
- **a.** Si g es una función creciente sobre $\mathcal{X}$, entonces $F_Y(y) = F_X(g^{-1}(y))$ para $y \in \mathcal{Y}$.
- **b.** Si g es una función decreciente sobre $\mathcal{X}$, entonces $F_Y(y) = 1 - F_X(g^{-1}(y))$ para $y \in \mathcal{Y}$.

Ejemplo 3.2.3 Suponga que X tiene distribución uniforme en el intervalo (0,1). Demostrar que $Y = g(X) = -\log X$ tiene distribución exponencial con parámetro $\beta = 1$.

En R se pueden generar números aleatorios sobre la distribución uniforme y luego obtener un histograma para los valores transformados según $g(X)$.

```
> set.seed(200)
> x <- runif(n=1000, min=0, max=1)
> y <- -log(x)
> hist(y, freq=F, xlab="", ylab="", main="")
```

Figure 3.2: Histograma para $Y = -\log X$



En el ejemplo anterior la función de densidad de Y se obtiene derivando su función de distribución acumulada. La expresión resultante se presenta en el siguiente teorema.

Teorema 3.2.2 Sean X con función de densidad $f_X(x)$ y $Y = g(X)$ donde g es una función monótona. Suponga que $f_X(x)$ es continua sobre $\mathcal{X}$ y que $g^{-1}(y)$ tiene una derivada continua sobre $\mathcal{Y}$. Entonces la función de densidad de Y es:

$$f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right| & y \in \mathcal{Y} \\ 0 & \text{de otro modo} \end{cases} \tag{3.2.6}$$

Demostración: Usando el teorema 3.2.1 y la regla de la cadena se tiene:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\frac{d}{dy}g^{-1}(y) & \text{si } g \text{ es creciente} \\ -f_X\left(g^{-1}(y)\right)\frac{d}{dy}g^{-1}(y) & \text{si } g \text{ es decreciente} \end{cases}$$

Ejemplo 3.2.4 Sea $X \sim \mathcal{W}(\gamma, \beta)$ donde $\gamma > 0$. Hallar la distribución de $Y = g(X) = X^{\gamma}$.

En muchas aplicaciones la función g podría no ser creciente ni decreciente, por consiguiente no prodrian aplicarse los resultados anteriores. Sin embargo es común el caso en el que la función g es monótona sobre ciertos subintervalos, los que permiten obtener una expresión para $Y = g(X)$.

Ejemplo 3.2.5 Suponga que X es una variable aleatoria continua. La función de distribución acumulada de $Y = X^2$, para $y > 0$, es:

$$F_Y(y) = \Pr(Y \le y) = \Pr\left(X^2 \le y\right) = \Pr(-\sqrt{y} \le X \le \sqrt{y})$$

Como X es variable aleatoria continua se tiene:

$$\begin{aligned} F_Y(y) &= \Pr(X \le \sqrt{y}) - \Pr(X \le -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

La función de densidad de Y puede obtenerse derivando su función de distribución acumulada:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) \\ &= \frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(-\sqrt{y})\right] \end{aligned}$$

y usando la regla de la cadena para derivar $F_X(\sqrt{y})$ y $F_X(-\sqrt{y})$ se tiene:

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{1}{2\sqrt{y}} \tag{3.2.7}$$

Notar que la función de densidad anterior esta expresada como la suma de dos componentes sobre los intervalos donde $g(x) = x^2$ es monótona.

Teorema 3.2.3 Sea X con función de densidad $f_X(x)$, $Y = g(X)$ y espacio muestral $\mathcal{X}$. Suponga que existe una partición $A_0, A_1, \cdots, A_k$ de $\mathcal{X}$ tal que $\Pr(X \in A_0) = 0$ y $f_X(x)$ es continua sobre cada A_i. Suponga además que existen funciones $g_1(x), \cdots, g_k(x)$ definidas sobre $A_1, \cdots, A_k$ respectivamente, que satisfacen:

- **a.** $g(x) = g_i(x)$ para $x \in A_i$,
- **b.** $g_i(x)$ es monótona sobre A_i,
- **c.** El conjunto $\mathcal{Y} = \{y : y = g_i(x) \text{ para algún } x \in A_i\}$ es el mismo para cada $i = 1, \cdots, k$.
- **d.** $g_i^{-1}(y)$ tiene una derivada continua en $\mathcal{Y}$, para cada $i = 1, \cdots, k$.

Entonces:

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X\left(g_i^{-1}(y)\right)\left|\frac{d}{dy}g_i^{-1}(y)\right| & y \in \mathcal{Y} \\ 0 & \text{de otro modo} \end{cases}$$

Es importante notar que cada $g_i(x)$ es una transformación uno a uno desde A_i hacia $\mathcal{Y}$. Además, $g_i^{-1}(y)$ es una función uno a uno desde $\mathcal{Y}$ hacia A_i, tal que, para $y \in \mathcal{Y}$, $g_i^{-1}(y)$ permite obtener un único $x = g_i^{-1}(y) \in A_i$ para el cual $g_i(x) = y$.

Ejemplo 3.2.6 Sea $X \sim \mathcal{N}(0, 1)$. Demostrar que $Y = g(X) = X^2$ tiene distribución *chi-cuadrado* con 1 grado de libertad.

En algunos ocasiones la partición $A_0, A_1, \cdots, A_k$ de $\mathcal{X}$ forma también una partición del conjunto $\mathcal{Y}$. En este caso todavía es posible usar 3.2.3.

Ejemplo 3.2.7 Sea $X \sim \mathcal{E}(\beta)$. Hallar la función de densidad de $Y = g(X) = (X - 2)^2$.

Teorema 3.2.4 Sea X cuya función de distribución acumulada, $F_X(x)$, es continua. Si se define la variable aleatoria $Y = F_X(x)$, entonces Y tiene distribución uniforme en el intervalo $(0, 1)$.

Demostración: Si $Y = F_X(x)$ entonces $0 < y < 1$,

$$\begin{aligned}
\Pr(Y \leq y) &= \Pr(F_X(X) \leq y) \\
&= \Pr\left(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)\right) \\
&= \Pr\left(X \leq F_X^{-1}(y)\right) \\
&= F_X\left(F_X^{-1}(y)\right) \\
&= y
\end{aligned}$$

3.3. Valores esperados

Definición 3.3.1 El *valor esperado* o *media* de una variable aleatoria $g(X)$, denotado por $\mathrm{E}[g(X)]$, es:

$$\mathrm{E}[g(X)] = \begin{cases} \sum_{x\in\mathcal{X}} g(x) f_X(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} g(x) f_X(x)\, dx & \text{si } X \text{ es continua} \end{cases} \tag{3.3.1}$$

siempre que la integral o suma exista. Si $\mathrm{E}[|g(X)|] = \infty$ se dice que $\mathrm{E}[g(X)]$ no existe.

Ejemplo 3.3.1 Suponga que $X \sim \mathcal{E}(\beta)$, entonces:

$$\mathrm{E}[X] = \int_0^{\infty} x \frac{1}{\beta} e^{-x/\beta} dx = \beta$$

Ejemplo 3.3.2 Si $X \sim \mathcal{BI}(n,p)$, entonces:

$$\mathrm{E}[X] = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

Usando la identidad $x\binom{n}{x} = n\binom{n-1}{x-1}$ se tiene:

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{x=1}^{n} n \binom{n-1}{x-1} p^x (1-p)^{n-x} \\
&= \sum_{y=0}^{n-1} n \binom{n-1}{y} p^{y+1} (1-p)^{n-(y+1)} \\
&= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \\
&= np
\end{aligned}$$

Ejemplo 3.3.3 Un ejemplo clásico de una variable aleatoria cuyo valor esperado no existe corresponde a la *distribución de Cauchy*.

Teorema 3.3.1 Sea X una variable aleatoria y sean a, b y c constantes. Entonces para funciones cualesquiera $g_1(x)$ y $g_2(x)$ cuyos valores esperados existan,

a. $\mathrm{E}[ag_1(X) + bg_2(X) + c] = a\mathrm{E}[g_1(X)] + b\mathrm{E}[g_2(X)] + c$.

b. Si $g_1(x) \geq 0$ para todo x, entonces $\mathrm{E}[g_1(X)] \geq 0$.

c. Si $g_1(x) \geq g_2(x)$ para todo x, entonces $\mathrm{E}[g_1(X)] \geq \mathrm{E}[g_2(X)]$.

d. Si $a \leq g_1(x) \leq b$ para todo x, entonces $a \leq \mathrm{E}[g_1(X)] \leq b$.

Ejemplo 3.3.4 Suponga que se mide la distancia entre una variable aleatoria X y una constante b mediante $(X-b)^2$. Mientras más cerca esté b de X más pequeña sera dicha cantidad. El objetivo es determinar el valor de b que minimize $\mathrm{E}[(X-b)^2]$.

$$\begin{aligned}
\mathrm{E}\left[(X-b)^2\right] &= \mathrm{E}\left[(X-\mathrm{E}[X]+\mathrm{E}[X]-b)^2\right] \\
&= \mathrm{E}\left[((X-\mathrm{E}[X])+(\mathrm{E}[X]-b))^2\right] \\
&= \mathrm{E}\left[(X-\mathrm{E}[X])^2\right]+\mathrm{E}\left[(\mathrm{E}[X]-b)^2\right]
\end{aligned}$$

ya que $\mathrm{E}[(X-\mathrm{E}[X])(\mathrm{E}[X]-b)] = 0$. Además $(\mathrm{E}[X]-b)$ es una constante. Luego:

$$\mathrm{E}\left[(X-b)^2\right] = \mathrm{E}\left[(X-\mathrm{E}[X])^2\right] + (\mathrm{E}[X]-b)^2$$

Como no se tiene control sobre el primer término del lado derecho y el segundo término puede ser mayor o igual a 0, el menor valor se obtiene cuando $b = \mathrm{E}[X]$. Entonces:

$$\min_b \mathrm{E}\left[(X-b)^2\right] = \mathrm{E}\left[(X-\mathrm{E}[X])^2\right]$$

3.4. Momentos y función generatriz de momentos

Definición 3.4.1 Para cada entero n, el *n-ésimo momento de X*, $\mu_n^{'}$, es:

$$\mu_n^{'} = \mathrm{E}\left[X^n\right]$$

El *n-ésimo momento central de X*, μ_n, es:

$$\mu_n = \mathrm{E}\left[(X-\mu)^n\right]$$

donde $\mu = \mu_1^{'} = \mathrm{E}\left[X\right]$.

Definición 3.4.2 La *varianza* de una variable aleatoria X es su segundo momento central, $\mathrm{Var}\left(X\right) = \mathrm{E}\left[(X-\mu)^2\right]$. La raíz cuadrada positiva de la varianza es conocida como *desviación estándar*.

Ejemplo 3.4.1 Si $X \sim \mathcal{E}\left(\beta\right)$, entonces:

$$\mathrm{Var}\left(X\right) = \mathrm{E}\left[(X-\mu)^2\right] = \int_0^{\infty} (x-\beta)^2 \frac{1}{\beta} e^{-x/\beta} dx = \beta^2$$

Teorema 3.4.1 *Si X es una variable aleatoria con varianza finita, entonces para constantes cualesquiera a y b:*

$$\mathrm{Var}\left(aX+b\right) = a^2 \mathrm{Var}\left(X\right)$$

Demostración: Usando la definición de varianza:

$$\begin{aligned} \mathrm{Var}\left(aX+b\right) &= \mathrm{E}\left[\left((aX+b) - \mathrm{E}\left[(aX+b)\right]\right)^2\right] \\ &= \mathrm{E}\left[\left(aX - a\mathrm{E}\left[X\right]\right)^2\right] \\ &= a^2 \mathrm{E}\left[\left(X - \mathrm{E}\left[X\right]\right)^2\right] \\ &= a^2 \mathrm{Var}\left(X\right) \end{aligned}$$

La siguiente forma de calcular la varianza es bastante útil:

$$\mathrm{Var}\left(X\right) = \mathrm{E}[X^2] - \mathrm{E}^2[X] \tag{3.4.1}$$

Ejemplo 3.4.2 Si $X \sim \mathcal{BI}(n,p)$, entonces:

$$\mathrm{E}[X^2] = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x}$$

pero:

$$x^2 \binom{n}{x} = x \frac{n!}{(x-1)!(n-x)!} = xn \binom{n-1}{x-1}$$

luego,

$$\begin{aligned} \mathrm{E}[X^2] &= \sum_{x=1}^{n} xn \binom{n-1}{x-1} p^x (1-p)^{n-x} \\ &= n \sum_{y=0}^{n-1} (y+1) \binom{n-1}{y} p^{y+1} (1-p)^{n-1-y} \\ &= np(n-1)p + np \end{aligned}$$

Finalmente:

$$\mathrm{Var}[X] = n^2p^2 - np^2 + np - (np)^2 = np(1-p)$$

.

Definición 3.4.3 Sea X una variable aleatoria. La *función generatriz de momentos de* X, denotada por $M_X(t)$, es:

$$M_X(t) = \mathrm{E}\left[e^{tX}\right]$$

sujeto a que el valor esperado exista. Entonces:

$$M_X(t) = \begin{cases} \sum_x e^{tx} f_X(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx & \text{si } X \text{ es continua} \end{cases}$$

Teorema 3.4.2 Si X tiene función generatríz de momentos $M_X(t)$ entonces:

$$\mathrm{E}[X] = M_X^{(1)}(0)$$

donde $M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$.

Prueba: Asumiendo que es posible intercambiar la derivada con la integral, se tiene:

$$\begin{aligned}\frac{d}{dt}M_X(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f_X(x)dx \\ &= \int_{-\infty}^{\infty}\left(\frac{d}{dt}e^{tx}\right) f_X(x)dx \\ &= \int_{-\infty}^{\infty} xe^{tx} f_X(x)dx \\ &= \mathrm{E}\left[Xe^{tX}\right]\end{aligned}$$

luego $\frac{d}{dt}M_X(t)\Big|_{t=0} = \mathrm{E}\left[Xe^{tX}\right]\Big|_{t=0} = \mathrm{E}\left[X\right]$. Trabajando de manera análoga, se puede establecer que:

$$\frac{d^n}{dt^n}M_X(t)\Big|_{t=0} = \mathrm{E}\left[X^n e^{tX}\right]\Big|_{t=0} = \mathrm{E}\left[X^n\right]$$

Ejemplo 3.4.3 Sea $X \sim \mathcal{G}(\alpha, \beta)$. La función generatriz de momentos esta dada por:

$$\begin{aligned}M_X(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^{\infty} e^{tx}x^{\alpha-1}e^{-x/\beta}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^{\infty} x^{\alpha-1}e^{-x((1/\beta)-t)}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^{\infty} x^{\alpha-1}e^{-x/\left(\frac{\beta}{1-\beta t}\right)}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^\alpha \\ &= \left(\frac{1}{1-\beta t}\right)^\alpha\end{aligned}$$

y existe solo si $t < 1/\beta$. La media de la distribución gamma es:

$$\mathrm{E}\left[X\right] = \frac{d}{dt}M_X(t)\Big|_{t=0} = \frac{\alpha\beta}{(1-\beta t)^{\alpha+1}}\Big|_{t=0} = \alpha\beta$$

Los otros momentos pueden calcularse de forma similar.

Ejemplo 3.4.4 Si $X \sim \mathcal{BI}\,(n, p)$, entonces:

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= [pe^t + (1-p)]^n \end{aligned}$$

recordando que $\sum_{x=0}^{n} \binom{n}{x} u^x v^{n-x} = (u+v)^n$.

Teorema 3.4.3 Suponga $\{X_i, i = 1, 2, \cdots\}$ es una secuencia de variables aleatorias cuya función generatriz de momentos es $M_{X_i}(t)$. Además:

$$\lim_{i \to \infty} M_{X_i}(t) = M_X(t)$$

donde $M_X(t)$ es una función generatriz de momentos. Entonces existe una única función de distribución acumulada F_X cuyos momentos estan definidos por $M_X(t)$ tal que:

$$\lim_{i \to \infty} F_{X_i}(x) = F_X(x)$$

Ejemplo 3.4.5 Una aproximación usada en cursos elementales de estadística permite aproximar las probabilidades binomiales usando la distribución de Poisson. Esta aproximación es válida *cuando n es grande y np es pequeño*.

La función de probabilidad de Poisson es:

$$\Pr(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \qquad y = 0, 1, 2, \cdots$$

donde λ es una constante positiva. La aproximacion es tal que si $X \sim \mathcal{BI}(n, p)$ y $Y \sim \mathcal{P}(\lambda)$ con $\lambda = np$, entonces:

$$\Pr(X = x) \approx \Pr(Y = x)$$

Recordar que:

$$M_X(t) = [pe^t + (1-p)]^n$$

es la función generatriz de momentos de la distribución binomial. Para la distribución de Poisson se puede demostrar que su función generatriz de momentos es:

$$M_Y(t) = e^{\lambda(e^t - 1)}$$

Como $\lambda = np$, entonces:

$$\begin{aligned} M_X(t) &= \left[1 + p(e^t - 1)\right]^n \\ &= \left[1 + \frac{(e^t - 1)\lambda}{n}\right]^n \end{aligned}$$

Lema 3.4.1 Sean $a_1, a_2, \cdots$ una secuencia de números que convergen hacia a, es decir $\lim_{n\to\infty} a_n = a$, entonces:

$$\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$$

Demostración: La demostración de este lema puede encontrarse en los textos de cálculo.

Luego, si se toma $a_n = \lambda(e^t - 1) = a$ entonces:

$$\lim_{n\to\infty} M_X(t) = e^{\lambda(e^t-1)} = M_Y(t)$$

es la función generatriz de momentos de la distribución de Poisson.

Teorema 3.4.4 Sean a y b constantes, la función generatriz de momentos de la variable aleatoria $aX + b$ está dada por:

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Prueba: Por definición:

$$\begin{aligned} M_{aX+b}(t) &= \mathrm{E}\left[e^{(aX+b)t}\right] \\ &= \mathrm{E}\left[e^{aXt} e^{bt}\right] \\ &= e^{bt}\mathrm{E}\left[e^{X(at)}\right] \\ &= e^{bt} M_X(at) \end{aligned}$$