

Capítulo 3 Variables Aleatorias Discretas y sus Distribuciones de Probabilidad

Definición Básica

Como se vio en el capítulo anterior una variable aleatoria es una función de valor real definida sobre un espacio muestral. Por lo tanto, una variable aleatoria se puede usar para identificar eventos numéricos que son de interés en un experimento. Existen dos tipos de variables aleatorias, variables aleatorias discretas y continuas. En este capítulo se explorarán las discretas.

Definición 3.1

Se dice que una variable aleatoria Y es *discreta* si puede tomar sólo un número finito o contablemente infinito de valores distintos.

Un conjunto de elementos es contablemente infinito si los elementos del conjunto se pueden poner en correspondencia biunívoca con los enteros positivos. Se necesita la probabilidad de un evento observado para hacer inferencias acerca de una población. Los eventos de interés son con frecuencia numéricos y corresponden a valores de variables aleatorias discretas. Como ciertos tipos de variables aleatorias se presentan frecuentemente en la práctica, es útil tener a la mano la probabilidad para cada valor de una variable aleatoria. Este conjunto de probabilidades recibe el nombre de *distribución de probabilidad* de la variable aleatoria discreta. Numerosos experimentos muestran características similares y general variables aleatorias con el mismo tipo de distribución de probabilidad.

La distribución de probabilidad para una Variable Aleatoria Discreta

De modo notacional, usaremos una *letra mayúscula* para denotar una *variable aleatoria* y una *letra minúscula* para denotar un *valor particular* que puede tomar una variable aleatoria. Entonces la expresión ($Y = y$) se puede leer como *el conjunto de todos los puntos en S a los que la variable aleatoria Y asigna el valor y* . Con esto se llega a la siguiente definición.

Definición 3.2

La probabilidad de que Y tome el valor y , $P(Y = y)$, se define como la *suma de las probabilidades de todos los puntos muestrales en S a los que se asigna el valor y* . A veces denotaremos $P(Y = y)$ por $p(y)$.

Como $p(y)$ es una función que asigna probabilidad a cada valor y de la variable aleatoria Y , a veces recibe el nombre de *función de probabilidad* para Y .

Definición 3.3

La *distribución de probabilidad* para una variable discreta Y puede ser representada por una fórmula, una tabla o una gráfica que produzca $p(y) = P(Y = y)$ para toda y .

Nótese que $p(y) \geq 0$ para toda y , pero la distribución de probabilidad para una variable aleatoria discreta asigna probabilidades diferentes de cero a sólo un número contable de valores y distintos. Se entiende que cualquier valor y no explícitamente asignado a una posibilidad positiva es tal que $p(y) = 0$.

Teorema 3.1

Para cualquier distribución de probabilidad discreta, lo siguiente debe ser verdadero:

1. $0 \leq p(y) \leq 1$ para toda y .
2. $\sum_y p(y) = 1$, donde la sumatoria es para todos los valores de y con probabilidad diferente de cero.

El valor esperado de una variable aleatoria o una función de una variable aleatoria

Definición 3.4

Sea Y una variable aleatoria discreta con la función de probabilidad $p(y)$. Entonces el valor esperado de Y , $E(Y)$ se define como

$$E(Y) = \sum_y yp(y)$$

Para ser precisos, se dice que el valor esperado de una variable aleatoria discreta existe si la suma, como se dio antes, es absolutamente convergente, es decir, si

$$\sum_y |y|p(y) < \infty$$

Si $p(y)$ es una caracterización precisa de la distribución de frecuencia poblacional, entonces $E(Y) = \mu$ es la media poblacional. A la media también se le puede llamar promedio. La definición anterior se puede extender cuando se está tratando de una función de valor real de alguna variable aleatoria.

Teorema 3.2

Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$ y sea $g(Y)$ una función de valor real de Y . Entonces, el valor esperado de $g(Y)$ está dado por

$$E[g(Y)] = \sum_{\text{toda } y} g(y)p(y)$$

Demostración

Demostramos el resultado en el caso que la variable aleatoria Y toma el número finito de valores y_1, y_2, \dots, y_n . Como la función $g(y)$ puede no ser biunívoca, suponga que $g(Y)$ toma valores g_1, g_2, \dots, g_m (donde $m \leq n$). Se deduce que $g(Y)$ es una variable aleatoria tal que para $i = 1, 2, \dots, m$,

$$P[g(Y) = g_i] = \sum_{\substack{\text{toda } y_i \\ \text{tal que} \\ g(y_i) = g_i}} p(y_i) = p^*(g_i)$$

Entonces, por la Definición 3.4,

$$\begin{aligned} E[g(Y)] &= \sum_{i=1}^m g_i p^*(g_i) \\ &= \sum_{i=1}^m g_i \left\{ \sum_{\substack{\text{toda } y_i \\ \text{tal que} \\ g(y_i) = g_i}} p(y_i) \right\} \\ &= \sum_{i=1}^m \left\{ \sum_{\substack{\text{toda } y_i \\ \text{tal que} \\ g(y_i) = g_i}} g_i p(y_i) \right\} \\ &= \sum_{j=1}^n g(y_j) p(y_j) \end{aligned}$$

El objetivo inmediato es hallar medidas numéricas descriptivas (o parámetros) para caracterizar $p(y)$. Como ya se vio $E(Y)$ proporciona la media de una población con distribución dada por $p(y)$. A continuación se obtendrá la definición de varianza y desviación estándar de una población.

Definición 3.5

Si Y es una variable aleatoria con media $E(Y) = \mu$, la *varianza* de una variable aleatoria Y se define como el valor esperado de $(Y - \mu)^2$. Esto es,

$$V(Y) = E[(Y - \mu)^2] = \sigma^2$$

La *desviación estándar* de Y es la raíz cuadrada positiva de $V(Y)$.

A continuación se presentan tres teoremas útiles de expectativa que se deduce directamente de la teoría de sumatoria.

Teorema 3.3

Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$ y sea c una constante. Entonces $E(c) = c$.

Demostración

Considere la función $g(Y) \equiv c$. Por el Teorema 3.2,

$$E(c) = \sum_y cp(y) = c \sum_y p(y)$$

Pero $\sum_y p(y) = 1$ (Teorema 3.1) y, por tanto, $E(c) = c(1) = c$.

Teorema 3.4

Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$, $g(y)$ una función de Y y c una constante. Entonces

$$E[cg(Y)] = cE[g(Y)]$$

Demostración

Por el teorema 3.2,

$$E[cg(Y)] = \sum_y cg(y)p(y) = c \sum_y g(y)p(y) = cE[g(Y)]$$

Teorema 3.5

Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$ y sean $g_1(Y), g_2(Y), \dots, g_k(Y)$ k funciones de Y . Entonces

$$E[g_1(Y) + g_2(Y) + \dots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \dots + E[g_k(Y)]$$

Demostración

Demostraremos la prueba sólo para el caso $k = 2$, pero pasos análogos se cumplirán para cualquier k finita. Por el Teorema 3.2,

$$\begin{aligned} E[g_1(Y) + g_2(Y)] &= \sum_y [g_1(y) + g_2(y)]p(y) \\ &= \sum_y g_1(y)p(y) + \sum_y g_2(y)p(y) \\ &= E[g_1(Y)] + E[g_2(Y)] \end{aligned}$$

Estos últimos teoremas se pueden usar de inmediato para desarrollar un teorema útil para hallar la varianza de una variable aleatoria discreta.

Teorema 3.6

Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$ y media $E(y) = \mu$; entonces

$$V(Y) = \sigma^2 = E[(Y - \mu)^2] = E(Y^2) - \mu^2$$

Demostración

$$\begin{aligned} \sigma^2 &= E[(Y - \mu)^2] = E(Y^2 - 2\mu Y + \mu^2) \\ &= E(Y^2) - E(2\mu Y) + E(\mu^2) \quad \text{por el Teorema 3.5} \end{aligned}$$

Si se observa que μ es una constante y se aplican los Teoremas 3.4 y 3.3 a los términos segundo y tercero, respectivamente, tenemos

$$\sigma^2 = E(Y^2) - 2\mu E(Y) + \mu^2$$

Pero $\mu = E(Y)$ y por lo tanto,

$$\sigma^2 = E(Y^2) - 2\mu^2 + \mu^2 = E(Y^2) - \mu^2$$

La distribución de probabilidad binomial

La distribución de probabilidad más simple que existe es conocida como la distribución bernoulli. Consiste en la probabilidad de un éxito de algún evento y la probabilidad de fracaso del evento. Existen experimentos que solo consisten en la observación de una secuencia de intentos independientes, cada uno de estos intentos tiene distribución bernoulli. Cuando se observan todos los experimentos en conjunto entonces se está tratando con un experimento binomial y se definen de la siguiente manera.

Definición 3.6

Un *experimento binomial* presenta las siguientes propiedades:

1. Consiste en un número fijo, n , de pruebas idénticas.
2. Cada prueba resulta en uno de dos resultados: éxito, S , o fracaso, F .
3. La probabilidad de éxito en una sola prueba es igual a algún valor p y es el mismo de una prueba a la otra. La probabilidad de fracaso es igual a $q = (1 - p)$.
4. Las pruebas son independientes.
5. La variable aleatoria de interés es Y , el número de éxitos observados durante las n pruebas.

Como la variable aleatoria de interés es el número de éxitos observados entonces se puede deducir la distribución $p(y)$ de probabilidad binomial al aplicar el método de punto muestral para hallar la probabilidad de que el experimento produzca y éxitos. Cada punto muestral del espacio muestral tiene n pruebas, algunas son éxitos y otras son fracasos. Entonces un punto muestral típico se pudiera

ver como

$$SSSFSSFFFSFSFFFFSSSFF \dots FFSSF$$

Cada uno de los éxitos tiene la misma probabilidad de haber sucedido, esto es lo mismo para los fracasos. Entonces, considerando el caso particular de que hayan $Y = y$ éxitos esto significa que hay $n - y$ fracasos. Con esta información se puede encontrar la probabilidad de este punto muestra que consiste de y éxitos y $n - y$ fracasos, como cada prueba es independiente entonces la probabilidad conjunta es el producto de las probabilidades. Lo que da lo siguiente:

$$ppppp \dots pqqqq \dots qq = p^y q^{n-y}$$

Ahora hay que considerar todos los posibles arreglos que contienen y éxitos ya que cada uno de estos puntos muestrales tiene la misma probabilidad. Esta cantidad está dada por el Teorema 2.3 y resulta en lo siguiente.

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

Con esta información se deduce que la distribución de probabilidad binomial es la siguiente.

Definición 3.7

Se dice que una variable aleatoria Y tiene una *distribución binomial* basada en n pruebas con probabilidad p de éxito si y sólo si

$$p(y) = \binom{n}{y} p^y q^{n-y} \quad y = 0, 1, 2, \dots, n \quad 0 \leq p \leq 1$$

Teorema 3.7

Sea Y una variable aleatoria binomial basada en n pruebas y probabilidad p de éxito. Entonces

$$\mu = E(Y) = np \quad \sigma^2 = V(Y) = npq$$

Demostración

Por las Definiciones 3.4 y 3.7

$$E(Y) = \sum_y yp(y) = \sum_{y=0}^n y \binom{n}{y} p^y q^{n-y}$$

Observe que el primer término de la suma es 0 y por lo tanto

$$\begin{aligned} E(Y) &= \sum_{y=1}^n y \frac{n!}{(n-y)!y!} p^y q^{n-y} \\ &= \sum_{y=1}^n \frac{n!}{(n-y)!(y-1)!} p^y q^{n-y} \end{aligned}$$

Demostración

Los sumandos de esta última expresión presentan un sorprendente parecido con las probabilidades binomiales. De hecho, si factorizamos np en cada término de la suma y hacemos $z = y - 1$,

$$\begin{aligned} E(Y) &= np \sum_{y=1}^n \frac{(n-1)!}{(n-y)!(y-1)!} p^{y-1} q^{n-y} \\ &= np \sum_{z=0}^{n-1} \frac{(n-1)!}{(n-1-z)!z!} p^z q^{n-1-z} \\ &= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^z q^{n-1-z} \end{aligned}$$

Observe que $p(z) = \binom{n-1}{z} p^z q^{n-1-z}$ es la función de probabilidad binomial basada en $(n-1)$ pruebas. Así, $\sum_z p(z) = 1$ y se deduce que

$$\mu = E(Y) = np$$

Del Teorema 3.6 sabemos que $\sigma^2 = V(Y) = E(Y^2) - \mu^2$. Entonces, σ^2 se puede calcular si hallamos $E(Y^2)$. Hallar $E(Y^2)$ directamente es difícil porque

$$\begin{aligned} E(Y^2) &= \sum_{y=0}^n y^2 p(y) = \sum_{y=0}^n y^2 \binom{n}{y} p^y q^{n-y} \\ &= \sum_{y=0}^n y^2 \frac{n!}{y!(n-y)!} p^y q^{n-y} \end{aligned}$$

y la cantidad y^2 no aparece como factor de $y!$. Sin embargo, observe que

$$E[Y(Y-1)] = E(Y^2 - Y) = E(Y^2) - E(Y)$$

Por lo tanto

$$E(Y^2) = E[Y(Y-1)] + E(Y) = E[Y(Y-1)] + \mu$$

En este caso, $E[Y(Y-1)] = \sum_{y=0}^n y(y-1) \frac{n!}{y!(n-y)!} p^y q^{n-y}$. Los términos primero y segundo de esta suma son iguales a cero (cuando $y = 0$ y $y = 1$). Entonces

$$E[Y(Y-1)] = \sum_{y=2}^n \frac{n!}{(y-2)!(n-y)!} p^y q^{n-y}$$

(Observe la cancelación que llevó a este último resultado. La anticipación de esta cancelación es lo que en realidad motivó la consideración de $E[Y(Y-1)]$.) De nuevo, los sumandos de la última expresión se asemejan mucho a probabilidades binomiales. Factorice $n(n-1)p^2$ en cada término de la suma y sea $z = y - 2$ para obtener

Demostración

$$\begin{aligned} E[Y(Y-1)] &= n(n-1)p^2 \sum_{y=2}^n \frac{(n-2)!}{(y-2)!(n-y)!} p^{y-2} q^{n-y} \\ &= n(n-1)p^2 \sum_{z=0}^{n-2} \frac{(n-2)!}{z!(n-2-z)!} p^z q^{n-2-z} \\ &= n(n-1)p^2 \sum_{z=0}^{n-2} \binom{n-2}{z} p^z q^{n-2-z} \end{aligned}$$

De nuevo observe que $p(z) = \binom{n-2}{z} p^z q^{n-2-z}$ es la función de probabilidad binomial basada en $(n-2)$ pruebas.

Así $\sum_{z=0}^{n-2} p(z) = 1$ (de nuevo usando el dispositivo ilustrado en la obtención de la media) y

$$E[Y(Y-1)] = n(n-1)p^2$$

Entonces,

$$E(Y^2) = E[Y(Y-1)] + \mu = n(n-1)p^2 + np$$

y

$$\begin{aligned} \sigma^2 &= E(Y^2) - \mu^2 = n(n-1)p^2 - n^2p^2 \\ &= np[(n-1)p + 1 - np] \\ &= np(1-p) = npq \end{aligned}$$

Este teorema ilustra el uso de el hecho de que $\sum p(y) = 1$ si $p(y)$ es una función de probabilidad válida. Esta técnica se ocupará para otras distribuciones de probabilidad vistos en este capítulo.

La distribución de probabilidad geométrica

La variable aleatoria con distribución de probabilidad geométrica está asociada con un experimento que comparte algunas características de un experimento binomial. Al igual que un experimento binomial el experimento también tiene pruebas idénticas e independientes. Cada una de estas pruebas solo tiene dos resultados, éxito o fracaso. La diferencia entre estos dos experimentos es que en el geométrico la variable aleatoria Y es el número de pruebas en que ocurre el primer éxito. Esto significa que si el primer éxito ocurre en y intentos entonces hubo $y-1$ fracasos. Como solo hay un punto muestral en que esto ocurre no hay que considerar los demás casos en los que solo ocurre un éxito en y intentos. Con esto se obtiene la siguiente definición.

Definición 3.8

Se dice que una variable aleatoria Y tiene una *distribución de probabilidad geométrica* si y solo si

$$p(y) = q^{y-1}p \quad y = 1, 2, 3, \dots, \quad 0 \leq p \leq 1$$

Al igual que la distribución anterior es importante conocer la media y varianza de esta distribución, la cual se observa en el siguiente teorema.

Teorema 3.8

Si Y es una variable aleatoria con una distribución geométrica,

$$\mu = E(Y) = \frac{1}{p} \quad \sigma^2 = V(Y) = \frac{1-p}{p^2}$$

Demostración

$$E(Y) = \sum_{y=1}^{\infty} yq^{y-1}p = p \sum_{y=1}^{\infty} yq^{y-1}$$

Esta serie podría parecer difícil de sumarse directamente. En realidad, se puede sumar con facilidad si tomamos en cuenta que, para $y \geq 1$,

$$\frac{d}{dq}(q^y) = yq^{y-1}$$

y por lo tanto,

$$\frac{d}{dq} \left(\sum_{y=1}^{\infty} q^y \right) = \sum_{y=1}^{\infty} yq^{y-1}$$

(El intercambio de derivada y suma aquí se puede justificar.) Sustituyendo, obtenemos

$$E(Y) = p \sum_{y=1}^{\infty} yq^{y-1} = p \frac{d}{dq} \left(\sum_{y=1}^{\infty} q^y \right)$$

La última suma es la serie geométrica, $q + q^2 + q^3 + \dots$, que es igual a $q/(1-q)$ (vea el Apéndice A1.11). Por lo tanto,

$$E(Y) = p \frac{d}{dq} \left(\frac{q}{1-q} \right) = p \left[\frac{1}{(1-q)^2} \right] = \frac{p}{p^2} = \frac{1}{p}$$

Para resumir, nuestro método es expresar una serie que no se puede sumar directamente como la derivada de una serie para la cual la suma se puede obtener con facilidad. Una vez que evaluemos la serie más fácilmente manejable, derivamos para completar el proceso. La derivación de la varianza se deja como Ejercicio 3.85.

La distribución de probabilidad binomial negativa

Esta distribución al igual que las últimas dos distribuciones utiliza intentos independientes e idénticos que solo tienen dos resultados, fracasos o éxitos. La variable aleatoria que nos interesa en esta distribución es parecida a la variable aleatoria vista en la distribución geométrica, solo que vez del número de intentos para el primer éxito ahora nos interesa el número de intentos para obtener r éxitos. Esto significa que habrá r éxitos en un punto muestral y $y - r$ fracasos, ahora hay que considerar todos los puntos muestrales que tienen estas mismas cantidades de fracasos y éxitos. Esto se hace de la misma manera que en la distribución binomial. Por lo tanto se tiene la siguiente definición.

Definición 3.9

Se dice que una variable aleatoria Y tiene una *distribución de probabilidad binomial negativa* si y sólo si

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r} \quad y = r, r+1, r+2, \dots, \quad 0 \leq p \leq 1$$

Teorema 3.9

Si Y es una variable aleatoria con una distribución binomial negativa,

$$\mu = E(Y) = \frac{r}{p} \quad \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}$$

La distribución de probabilidad hipergeométrica

Suponga que una población contiene un número finito N de elementos, ahora suponga que se toma una muestra considerable de n elementos. Ahora considere que r elementos podrían ser rojos y $b = N - r$ negros. Si la muestra de n elementos es seleccionada al azar entonces la variable aleatoria de interés es el número de elementos rojos en la muestra. Esta variable tiene una *distribución de probabilidad hipergeométrica*. La distribución de probabilidad se puede obtener de manera similar al de la distribución binomial.

Definición 3.10

Se dice que una variable aleatoria Y tiene una *distribución de probabilidad hipergeométrica* si y sólo si

$$p(y) = \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}}$$

donde y es un entero $0, 1, 2, \dots, n$, sujeto a las restricciones $y \leq r$ y $n - y \leq N - r$.

Teorema 3.10

Si Y es una variable aleatoria con distribución hipergeométrica,

$$\mu = E(Y) = \frac{nr}{N} \quad \sigma^2 = V(Y) = n \left(\frac{r}{N} \right) \left(\frac{N-r}{N} \right) \left(\frac{N-n}{N-1} \right)$$

La distribución de probabilidad de Poisson

La distribución Poisson se le conoce como una distribución de conteos. La variable aleatoria de esta distribución mide la cantidad de éxitos de un evento en un periodo particular. Esta distribución se asemeja a la distribución binomial. Esto se puede demostrar de la siguiente manera, sea $\lambda = np$ y tomando el límite de la probabilidad binomial $p(y) = \binom{n}{y} p^y (1-p)^{n-y}$ cuando $n \rightarrow \infty$, tenemos

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{y} p^y (1-p)^{n-y} &= \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-y+1)}{y!} \left(\frac{\lambda}{n} \right)^y \\ &\quad \left(1 - \frac{\lambda}{n} \right)^{n-y} \\ &= \lim_{n \rightarrow \infty} \frac{\lambda^y}{y!} \left(1 - \frac{\lambda}{n} \right)^n \frac{n(n-1) \cdots (n-y+1)}{n^y} \\ &\quad \left(1 - \frac{\lambda}{n} \right)^{-y} \\ &= \frac{\lambda^y}{y!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n} \right)^n \left(1 - \frac{\lambda}{n} \right)^{-y} \left(1 - \frac{1}{n} \right) \\ &\quad \times \left(1 - \frac{2}{n} \right) \times \cdots \times \left(1 - \frac{y-1}{n} \right) \end{aligned}$$

Si observamos que $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n} \right)^n = e^{-\lambda}$ y todos los otros términos a la derecha del límite tienen un límite de 1, obtenemos

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}$$

Esto nos dice que la probabilidad binomial converge a la de Poisson. Con esto se llega a la siguiente definición.

Definición 3.11

Se dice que una variable aleatoria Y tiene *distribución de probabilidad de Poisson* si y sólo si

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda} \quad y = 0, 1, 2, \dots \quad \lambda > 0$$

Teorema 3.11

Si Y es una variable aleatoria que posee una distribución de Poisson con parámetro λ , entonces

$$\mu = E(Y) = \lambda \quad \sigma^2 = V(Y) = \lambda$$

Distribución

Por definición,

$$E(Y) = \sum_y y p(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!}$$

Observe que el primer término de esta suma es igual a 0 (cuando $y = 0$) y por lo tanto,

$$E(Y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} = \sum_{y=1}^{\infty} \frac{\lambda^y e^{-\lambda}}{(y-1)!}$$

Así como está, esta cantidad no es igual a la suma de los valores de una función de probabilidad $p(y)$ para todos los valores de y , pero podemos cambiarla a la forma apropiada al factorizar λ de la expresión y haciendo $z = y - 1$. Entonces los límites de sumatoria se convierten en $z = 0$ (cuando $y = 1$) y $z = \infty$ (cuando $y = \infty$), y

$$E(Y) = \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1} e^{-\lambda}}{(y-1)!} = \lambda \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!}$$

Observe que $p(z) = \lambda^z e^{-\lambda} / z!$ es la función de probabilidad par una variable aleatoria de Poisson, y $\sum_{z=0}^{\infty} p(z) = 1$.

Por lo tanto $E(Y) = \lambda$. Entonces, la media de una variable aleatoria de Poisson es el parámetro individual λ que aparece en la expresión para la función de probabilidad de Poisson.

Momentos y funciones generadoras de momento

En ocasiones es difícil encontrar la media y varianza de una variable aleatoria. Estas medidas son importantes para cualquier distribución de probabilidad ya que en algunas ocasiones tienen el mismo valor. A continuación consideramos un conjunto de medidas descriptivas numéricas que en ciertas condiciones determinan $p(y)$ de manera única.

Definición 3.12

El k -ésimo momento de una variable aleatoria Y tomada alrededor del origen se define como $E(Y^k)$ y se denota con μ'_k .

Definición 3.13

El k -ésimo momento de una variable aleatoria Y tomado alrededor de su media o el k -ésimo momento central de Y , se define como $E[(Y - \mu)^k]$ y está denotado por μ_k .

Un uso importante de los momentos es para calcular la distribución

de probabilidad de una variable aleatoria. Por lo tanto los momentos μ'_k , donde $k = 1, 2, 3, \dots$, son principalmente de valor teórica para $k > 3$.

Definición 3.14

La *función generadora de momento* $m(t)$ para una variable aleatoria Y se define como $m(t) = E(e^{tY})$. Decimos que una función generadora de momento para Y existe si existe una constante positiva b tal que $m(t)$ es finita para $|t| \leq b$.

¿Por qué $E(e^{tY})$ recibe el nombre de *función generadora de momento* para Y ? De una expansión de serie para e^{tY} se tiene lo siguiente.

$$e^{tY} = 1 + ty + \frac{(ty)^2}{2!} + \frac{(ty)^3}{3!} + \frac{(ty)^4}{4!} + \dots$$

Entonces suponiendo que μ'_k es finita para $k = 1, 2, 3, \dots$, tenemos

$$\begin{aligned} E(e^{tY}) &= \sum_y e^{tY} p(y) \\ &= \sum_y \left[1 + ty + \frac{(ty)^2}{2!} + \frac{(ty)^3}{3!} + \frac{(ty)^4}{4!} + \dots \right] p(y) \\ &= \sum_y p(y) + t \sum_y yp(y) + \frac{t^2}{2!} \sum_y y^2 p(y) + \dots \\ &= 1 + t\mu'_1 + \frac{t^2}{2!} \mu'_2 + \frac{t^3}{3!} \mu'_3 + \dots \end{aligned}$$

Entonces, $E(e^{tY})$ es una función de todos los momentos μ'_k alrededor del origen, para $k = 1, 2, 3, \dots$. En particular, μ'_k es el coeficiente de $t^k/k!$ en la expansión de serie de $m(t)$. La función generadora de momentos posee dos aplicaciones importantes, la primera esta dada por el siguiente teorema.

Teorema 3.12

Si $m(t)$ existe, entonces para cualquier entero positivo k ,

$$\left. \frac{d^k m(t)}{dt^k} \right|_{t=0} = m^{(k)}(0) = \mu'_k(0) = \mu'_k$$

En otras palabras, si el estudiante encuentra la k -ésima derivada de $m(t)$ con respecto a t y luego hace $t = 0$, el resultado será μ'_k .

Demostración

$\frac{d^k m(t)}{dt^k}$, o $m^{(k)}(t)$, es la k -ésima derivada de $m(t)$ con respecto a t . Como

$$m(t) = E(e^{tY}) = 1 + t\mu'_1 + \frac{t^2}{2!} \mu'_2 + \frac{t^3}{3!} \mu'_3 + \dots$$

se deduce que

$$m^{(1)}(t) = \mu'_1 + \frac{2t}{2!} \mu'_2 + \frac{3t^2}{3!} \mu'_3 + \dots$$

$$m^{(2)}(t) = \mu'_2 + \frac{2t}{2!} \mu'_3 + \frac{3t^2}{3!} \mu'_4 + \dots$$

y en general,

$$m^{(k)}(t) = \mu'_k + \frac{2t}{2!} \mu'_{k+1} + \frac{3t^2}{3!} \mu'_{k+2} + \dots$$

Si hacemos $t = 0$ en cada una de las derivadas anteriores, obtenemos

$$m^{(1)}(0) = \mu'_1 \quad m^{(2)}(0) = \mu'_2$$

y en general,

$$m^{(k)}(0) = \mu'_k$$

Estas operación implican intercambiar derivadas y sumas infinitas, que se pueden justificar si existe $m(t)$.

La segunda aplicación de una función generadora de momento es demostrar que una variable aleatoria posee una distribución de probabilidad particular $p(y)$. Si $m(t)$ existe para una distribución de probabilidad $p(y)$, es única. Esto significa que si dos variables aleatorias tienen la misma función generadora de momentos entonces deben tener la misma distribución de probabilidad.

Funciones generadoras de probabilidad

Una clase importante de variables aleatorias discretas es aquella en la que Y representa una cantidad y , en consecuencia toma valores enteros: $Y = 0, 1, 2, \dots$. Las variables aleatorias binomiales, geométricas, hipergeométricas y de Poisson caen todas en esta clase. Un procedimiento matemático útil para hallar las distribuciones de probabilidad y otras propiedades de variables aleatorias de valor entero es la función generadora de probabilidad.

Definición 3.15

Sea Y una variable aleatoria de valor entero para la cual $P(Y = i) = p_i$, donde $i = 0, 1, 2, \dots$. La *función generadora de probabilidad* $P(t)$ para Y se define como

$$P(t) = E(t^Y) = p_0 + p_1 t + p_2 t^2 + \dots = \sum_{i=0}^{\infty} p_i t^i$$

para todos los valores de t tales que $P(t)$ sea finita.

Definición 3.16

El k -ésimo *momento factorial* para una variable aleatoria Y se define como

$$\mu_{[k]} = E[Y(Y-1)(Y-2)\dots(Y-k+1)]$$

donde k es un entero positivo.

Teorema 3.13

Si $P(t)$ es la función generadora de probabilidad para una variable aleatoria de valor entero, Y , entonces el k -ésimo momento factorial de Y está dado por

$$\left. \frac{d^k P(t)}{dt^k} \right|_{t=1} = P^{(k)}(1) = \mu_{[k]}$$

Demostración

Como

$$P(t) = p_0 + p_1 t + p_2 t^2 + p_3 t^3 + p_4 t^4 + \dots$$

se deduce que

$$P^{(1)}(t) = \frac{dP(t)}{dt} = p_1 + 2p_2 t + 3p_3 t^2 + 4p_4 t^3 + \dots$$

$$P^{(2)}(t) = \frac{d^2 P(t)}{dt^2} = (2)(1)p_2 + (3)(2)p_3 t + (4)(3)p_4 t^2 + \dots$$

y en general,

$$\begin{aligned} P^{(k)}(t) &= \frac{d^k P(t)}{dt^k} \\ &= \sum_{y=k}^{\infty} y(y-1)(y-2)\dots(y-k+1)p(y)t^{y-k} \end{aligned}$$

Haciendo $t = 1$ en cada una de estas derivadas, obtenemos

Demostración

$$P^{(1)}(1) = p_1 + 2p_2 + 3p_3 + 4p_4 + \cdots = mu_{[1]} = E(Y)$$

$$P^{(2)}(1) = (2)(1)p_2 + (3)(2)p_3 + (4)(3)p_4 + \cdots = mu_{[2]} \\ = E[Y(Y-1)]$$

y en general,

$$P^{(k)}(1) = \sum_{y=k}^{\infty} y(y-1)(y-2) \cdots (y-k+1)p(y) \\ = E[Y(Y-1)(Y-2) \cdots (Y-k+1)] = \mu_{[k]}$$

Teorema de Tchebysheff

Teorema 3.14

Teorema de Tchebysheff Sea Y una variable aleatoria con media μ y varianza finita σ^2 . Entonces, para cualquier constante $k > 0$,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{o} \quad P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Deben señalarse dos aspectos importantes de este resultado. Primero, el resultado se aplica a *cualquier* distribución de probabilidad. En segundo término, los resultados del teorema son muy conservadores en el sentido de que la probabilidad real de que Y esté en el intervalo $\mu \pm k\sigma$ por lo general excede del límite inferior para la

probabilidad, $1 - \frac{1}{k^2}$, por una cantidad considerable. No obstante, para cualquier $k > 1$, es posible construir una distribución de probabilidad tal que, para esa k , el límite provisto por el teorema se alcance en realidad.