Aproximación de Binomial y Poisson por Normal

Para calcular probabilidades de distribuciones discretas con números grandes, es preciso sumar muchos términos, lo cual puede resultar poco práctico.

Sin embargo, las características de algunas distribuciones, como la Binomial y la de Poisson, permiten muy buenas aproximaciones mediante la distribución Normal.

Y como la distribución Normal se puede obtener de una tabla, el problema de sumar una gran cantidad de términos, queda reducido a buscar uno o dos valores en una tabla.

Aproximación de la distribución binomial por la distribución normal

Si X es una variable distribuida Binomialmente, $X{\sim}Bi(n\,,p)$, con $n \geq 10$ y p cercano a 0,5, entonces se puede aproximar utilizando la distribución Normal. Para ello podemos obtener la media y la desviación estándar de la distribución Normal de la siguiente manera:

$$\mu = n.p$$

$$\sigma = \sqrt{n.p.q}$$

Esto es válido por que si p es cercano a 0,5 y n es lo suficientemente grande (generalmente se pide $n \geq 10$) entonces la forma de la distribución binomial, a pesar de ser discreta, se parece mucho a la de una distribución normal.

Esta propiedad nos permite utilizar una variable normal estándar que se encuentra tabulada, para ahorrarnos la engorrosa tarea de sumar una cantidad elevada de términos de probabilidades binomiales, especialmente cuando n es muy grande y la cantidad de éxitos está lejos de 0 y lejos de n.

Teorema de Moivre: Supongamos una distribución binomial $X{\sim}Bi(\,n\,,p\,)$ en la que se cumplan simultáneamente las condiciones:

$$np > 5 \quad \text{y} \quad nq > 5$$

Entonces:

$$X{\sim}Bi(\,n\,,p\,) \quad \longrightarrow \quad Y{\sim}N(np\,,\sqrt{npq}\,)$$

Es decir, si X es una variable binomial de parámetros n y p, $X{\sim}Bi(n\,,p)$, se puede considerar que X sigue aproximadamente una distribución normal Y de media $\mu = np$ y desviación típica
$\sigma = \sqrt{n.p.q}$.

Debido a que la distribución normal es continua, y en consecuencia entre dos valores existirá una serie infinita de valores posibles, para estimar una variable aleatoria discreta se requiere de un leve ajuste, denominado factor de corrección de continuidad o corrección de Yates, sumando o restando **1/2** al valor de **x**. De esta forma el valor de **z** se obtiene mediante la fórmula:

$$z = \frac{\left(x + \frac{1}{2}\right) - \mu}{\sigma}$$

Entonces, para utilizar correctamente esta transformación de una variable discreta, *X*, en una continua, *Y*, es necesario realizar una "**corrección de continuidad**". Así la probabilidad de que $X = k$ como discreta, es la probabilidad de que $k - 0{,}5 \leq Y \leq k + 0{,}5$ como continua.

Ejemplo:

Se arrojan dos dados 600 veces ¿cuál es la probabilidad de obtener 95 o más veces la suma 7?

Solución

X= cantidad de veces que suma 7

$X \sim Bi(600\,, \frac{1}{6})$

$E(x) = np \quad \Rightarrow \quad E(x) = 600.\frac{1}{6} = 100$

Buscamos $P(X \geq 95)$

$$P(X \geq 95) = \binom{600}{95}\left(\frac{1}{6}\right)^{95}\left(\frac{5}{6}\right)^{600-95} + \ldots\ldots + \ldots\ldots + \ldots\ldots$$

Observamos que $n \;\geq\; 10$

Además se cumple el teorema de Moivre: $n.p = 100 \;\Rightarrow\; np > 5 \quad y \quad nq = 500 \;\Rightarrow np > 5$

Calculemos el desvío.

$\sigma = \sqrt{n.p.q} \;=\; \sqrt{\frac{500}{6}} \;= 9{,}13$

Entonces, podemos decir que $X \sim Bi\left(600\,, \frac{1}{6}\right) \longrightarrow Y \sim N(100\,, \sqrt{\frac{500}{6}})$

Recordemos que debemos aplicar la corrección de yates, entonces:

Hago el cálculo $95 - 0{,}5 = 94{,}5$

$$P(X \geq 95) \rightarrow P(Y > 94{,}5)$$

Y ahora podemos tipificar para utilizar la tabla de la normal estándar:

$$Z = \frac{94{,}5 - 100}{\sqrt{\frac{500}{6}}} = -\,0{,}60$$

$$P(Y > 94{,}5) = P(Z > -\,0{,}6) = 1 - P(Z \leq -\,0{,}6) = 0{,}7258$$

La probabilidad de obtener la suma 7, 95 o más veces es de 0,7258.

Aproximación de la distribución de Poisson por la distribución normal

Si X es una variable de Poisson, con μ > 10, $X \sim Po(\mu)$, entonces la variable aleatoria $Y = \frac{X-\mu}{\sqrt{\mu}}$ tiene una distribución aproximadamente normal estándar. Esto es válido porque si μ es mucho mayor que 10, entonces la forma de la distribución de Poisson, a pesar de ser discreta, se parece mucho a la de una distribución normal.

El cambio de variable Y no es otra cosa que la estandarización de esa variable aproximadamente normal (ya que μ es a la vez la media y la varianza de X).

Esta propiedad nos permite utilizar una variable normal estándar, que se encuentra tabulada, para ahorrarnos la tarea de sumar una cantidad elevada de términos de probabilidades de Poisson al calcular probabilidades acumuladas, especialmente cuando necesitamos calcular la probabilidad acumulada para un valor que esté lejos del cero.

Debido a que utilizaremos una variable continua por aproximar una discreta, también aquí debemos utilizar la corrección de Yates.

Ejemplo

Un banco recibe en promedio 6 cheques falsos al día. Suponiendo que el número de cheques falsos sigue una distribución de Poisson, hallar:

a) *Probabilidad de que se reciban cuatro cheques falsos en un día*
b) *Probabilidad de que se reciban más de 30 cheques en una semana*

Solución

a) X=cantidad de cheques falsos por día $X \sim Po(6)$
$P(X = 4) = \frac{e^{-6}+6^4}{4!} = 0{,}1339$
b) En 1 día son 6 cheques falsos entonces en 7 días (una semana) son 42 cheques.
X=cantidad de cheques falsos por semana $X \sim Po(42)$
Como μ > 10, entonces: $X \sim Po(42) \quad \rightarrow \quad N(42, \sqrt{42})$
Queremos $P(X > 30)$
Aplicamos la corrección de Yates:
$$30 + 0{,}5 = 30{,}5$$
Y tipificamos:
$$P(X > 30) = P\left(Z > \frac{30{,}5 - 42}{\sqrt{42}}\right)$$
Ahora solo queda hacer las cuentas y buscar en la tabla. (se los dejo)