Probabilidad y estadística

MAT 041, Primer semestre

Francisco Cuevas Pacheco

3 de octubre de 2022

Contenidos

Aproximación de De Moivre-Laplace:

Para n suficientemente grande y k en una vecindad de np, donde $0 < p < 1$ y $q = 1 - p$, tenemos la aproximación de De Moivre-Laplace,

$$\frac{n!}{k!(n-k)!} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right).$$

Observación

En otras palabras, la densidad de una distribución normal (con esperanza np y varianza npq) aproxima la función de cuantía de la distribución binomial con parámetros n y p.

Para justificar la aproximación de De Moivre-Laplace,

- utilizaremos la fórmula de Stirling:

$$\boxed{n! \approx n^n e^{-n}\sqrt{2\pi n}, \qquad n \to \infty.}$$

- consideramos la siguiente expansión en serie de potencias:

$$\boxed{\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots}$$

Tenemos la siguiente aproximación

$$
\begin{aligned}
\frac{n!}{k!(n-k)!}p^k q^{n-k} &\approx \frac{n^n \exp(-n)\sqrt{2\pi n}}{k^k \exp(-k)\sqrt{2\pi k}(n-k)^{n-k}\exp(-n+k)\sqrt{2\pi(n-k)}}p^k q^{n-k} \\
&= \sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k} \\
&= \sqrt{\frac{n}{2\pi k(n-k)}}\exp\left\{\log\left[\left(\frac{np}{k}\right)^k\right]+\log\left[\left(\frac{nq}{n-k}\right)^{n-k}\right]\right\} \\
&= \sqrt{\frac{n}{2\pi k(n-k)}}\exp\left\{-k\log\left(\frac{k}{np}\right)+(k-n)\log\left(\frac{n-k}{nq}\right)\right\}.
\end{aligned}
$$

Además, como estamos considerando $k \approx np$, tenemos que

$$
\sqrt{\frac{n}{2\pi k(n-k)}} \approx \frac{1}{\sqrt{2\pi npq}}.
$$

Por otro lado, el término $-k\log\left(\frac{k}{np}\right)$ puede ser reescrito a través del cambio de variable $x = (k-np)/\sqrt{npq}$. Como resultado, tenemos que

$$
\begin{aligned}
-k\log\left(\frac{k}{np}\right) &= -(np + x\sqrt{npq})\log\left(\frac{np + x\sqrt{npq}}{np}\right) \\
&= -(np + x\sqrt{npq})\log\left(1 + x\sqrt{\frac{q}{np}}\right) \\
&= -(np + x\sqrt{npq})\left(x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + \cdots\right) \\
&= -x\sqrt{npq} + \frac{x^2 q}{2} - x^2 q + \cdots
\end{aligned}
$$

Los demás términos en la expansión tienden a 0 a medida que $n \to \infty$, asi que son omitidos.

De manera completamente análoga, se obtiene la igualdad

$$(k-n)\log\left(\frac{n-k}{nq}\right) = x\sqrt{npq} + \frac{x^2 p}{2} - x^2 p + \cdots$$

Juntando todos estos cálculos se obtiene la igualdad

$$-k\log\left(\frac{k}{np}\right) + (k-n)\log\left(\frac{n-k}{nq}\right) = -\frac{x^2}{2}.$$

Con estos argumentos hemos justificado la aproximación de De Moivre-Laplace.

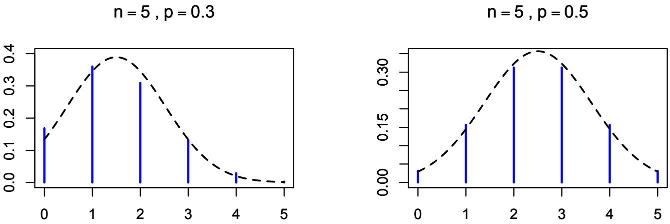

Figura 1: Aproximación de De Moivre-Laplace para diferentes valores de n y p. Las líneas verticales representan la función de cuantía de la distribución $\text{Bin}(n, p)$, mientras que la línea segmentada es la función de densidad de la distribución Normal con media np y varianza $np(1-p)$.

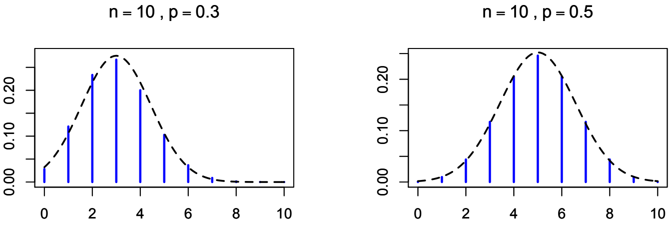

Figura 2: Aproximación de De Moivre-Laplace para diferentes valores de n y p. Las líneas verticales representan la función de cuantía de la distribución $\text{Bin}(n, p)$, mientras que la línea segmentada es la función de densidad de la distribución Normal con media np y varianza $np(1-p)$.

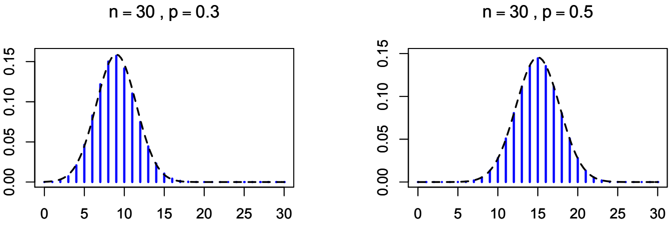

Figura 3: Aproximación de De Moivre-Laplace para diferentes valores de n y p. Las líneas verticales representan la función de cuantía de la distribución $\text{Bin}(n, p)$, mientras que la línea segmentada es la función de densidad de la distribución Normal con media np y varianza $np(1-p)$.

Teorema (Teorema Central del Límite)

Sea $X_1, X_2, \dots$ una sucesión de variables aleatorias independientes e identicamente distribuídas, con $E(X_i) = \mu$ y $var(X_i) = \sigma^2 > 0$. Si $Z_n = X_1 + \cdots + X_n$, entonces

$$\frac{Z_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{Z}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

tiene aproximadamente una distribución normal estándar cuando n es grande.

Idea de la Demostración:

La FGM de $\frac{Z_n - n\mu}{\sigma\sqrt{n}}$ está dada por

$$E\left(e^{t\left[\frac{Z_n - n\mu}{\sigma\sqrt{n}}\right]}\right) = e^{-t\frac{\sqrt{n}\mu}{\sigma}} \prod_{i=1}^{n} \psi_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right) = e^{-t\frac{\sqrt{n}\mu}{\sigma}} \left[\psi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Sin pérdida de generalidad, asumamos que $\mu = 0$. Luego,

$$\begin{aligned} \psi_{X_1}(t) &= \psi_{X_1}(0) + \psi'_{X_1}(0)t + \frac{\psi''_{X_1}(0)}{2!}t^2 + R(t) \\ &= 1 + E(X_1)t + \frac{E(X_1^2)t^2}{2} + R(t) \\ &= 1 + \frac{\sigma^2 t^2}{2} + R(t) \end{aligned}$$

De este modo,

$$E\left(e^{t\left[\frac{Z_n - n\mu}{\sigma\sqrt{n}}\right]}\right) = \left[1 + \frac{t^2}{2n} + R\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Cuando $n \to \infty$, esta cantidad converge a $e^{t^2/2}$ (la FGM de la Normal estándar).

Ejemplo

Suponga que el número de errores por programa de computadora sigue una distribución Poisson con media 5*. Conseguimos* 125 *programas. Sea* $X_1, \ldots, X_{125}$ *el número de errores de los programas. Definimos el promedio* $\overline{X} = (X_1 + \cdots + X_{125})/125$*. Queremos aproximar* $P(\overline{X} < 5.5)$*. Sea* $\mu = E(X_1) = \lambda = 5$ *y* $\sigma^2 = var(X_1) = \lambda = 5$*. Entonces,*

$$\begin{aligned} P(\overline{X} < 5.5) &= P\left(\frac{\sqrt{125}(\overline{X} - \mu)}{\sigma} < \frac{\sqrt{125}(5.5 - \mu)}{\sigma}\right) \\ &\approx P(Z < 2.5) \\ &= 0.9938 \end{aligned}$$

donde $Z \sim N(0,1)$*.*

Ejemplo

Una persona que vive en placeres usa la micro 604 para ir a Viña del mar. Basado en su experiencia, propone que la distribución de los tiempos de espera en el paradero se puede modelar mediante la siguiente distribución (en horas)

$$f(x) = 6x(1-x)\mathbb{I}[0,1](x)$$

- *Calcule el valor esperado y su varianza.*
- *Si esta persona tomó la micro durante 100 días. ¿Cuál es la probabilidad de que la espera promedio sea entre 27 y 33 minutos?*

Solución:

- $$E(X) = \int_0^1 x \cdot 6x(1-x)\mathrm{d}x = \frac{1}{2}$$
  $$E(X^2) = \int_0^1 x^2 \cdot 6x(1-x)^2\mathrm{d}x = \frac{3}{10}$$
  Luego, $V(X) = \frac{1}{20}$

- $$\begin{aligned} P(0.45 < \overline{X} < 0.55) &= P\left(\frac{\sqrt{100}(0.45-\mu)}{\sigma} < \frac{\sqrt{100}(\overline{X}-\mu)}{\sigma} < \frac{\sqrt{100}(0.55-\mu)}{\sigma}\right) \\ &\approx P(-2.236 < Z < 2.236) \\ &= 0.9747 \end{aligned}$$
  donde $Z \sim \mathrm{N}(0,1)$.