

La distribución normal

La distribución normal o gaussiana es la distribución continua más importante.

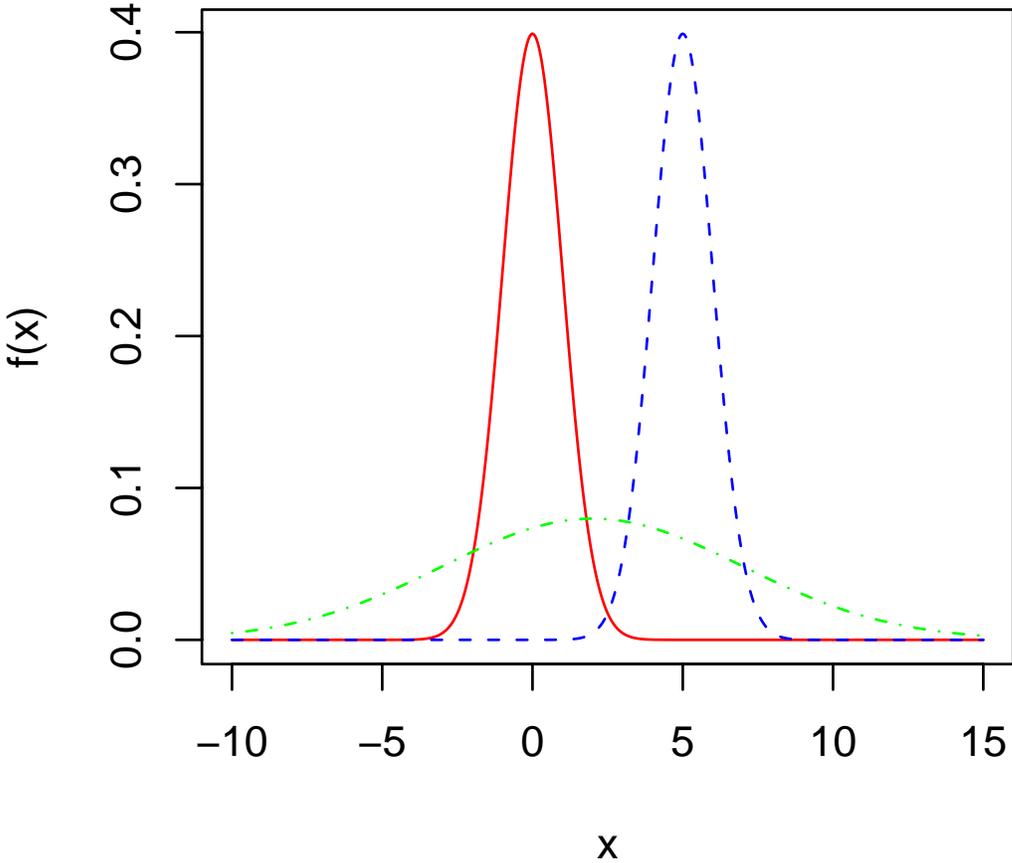
Definición 42 *Se dice que una variable X se distribuye como normal con parámetros μ y σ si*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

En este caso, se escribe $X \sim N(\mu, \sigma)$.

La media de la distribución normal es μ y la desviación típica es σ . El siguiente gráfico muestra la función de densidad de tres distribuciones normales con distintas medias y desviaciones típicas.

La función de densidad normal



Se ve que la densidad es simétrica en torno de la media.

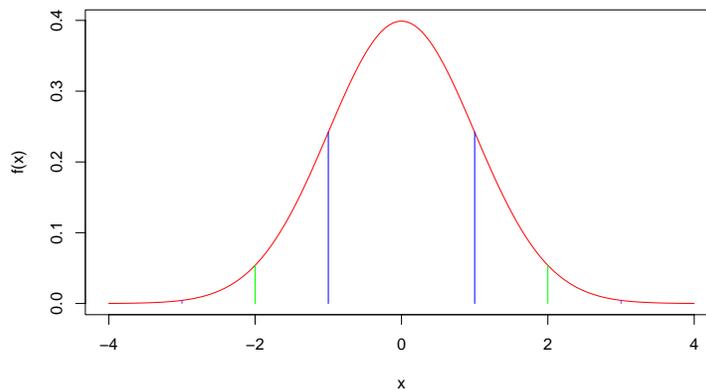
Una propiedad de la distribución normal

Si $X \sim N(\mu, \sigma)$, entonces

$$P(\mu - \sigma < X < \mu + \sigma) \approx 0,683$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0,955$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0,997$$



La **regla de Chebyshev** dice que para cualquiera variable X

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

El resultado para la normal justifica la regla empírica del Tema 1.

Transformación de una distribución normal

Si $X \sim N(\mu, \sigma)$ e $Y = a + bX$ es una transformación lineal, luego

$$Y \sim N(a + b\mu, b\sigma).$$

En particular, definiendo la **transformación tipificante** $Z = \frac{X - \mu}{\sigma}$, se tiene

$$Z \sim N(0, 1)$$

que es la **distribución normal estándar**.

Existen tablas de esta distribución que se emplean para hallar probabilidades.

Cálculo de probabilidades

Ejemplo 173 Sea $Z \sim N(0, 1)$. Hallar

- $P(Z < 1,5)$

- $P(Z > -1,5)$

- $P(-1,5 < Z < 1,5)$

- $P(1 < Z < 2)$

Usamos las tablas de la distribución normal para sacar los resultados.

Tenemos

$$P(Z < 1,5) = 0,9332$$

$$\begin{aligned} P(Z > -1,5) &= P(Z > 1,5) \quad \text{por simetría} \\ &= 0,9332 \end{aligned}$$

$$\begin{aligned} P(-1,5 < Z < 1,5) &= P(Z < 1,5) - P(Z < -1,5) \\ &= P(Z < 1,5) - P(Z > 1,5) \\ &= P(Z < 1,5) - \\ &\quad (1 - P(Z < 1,5)) \\ &= 2P(Z < 1,5) - 1 \\ &= 1 \times 0,9332 - 1 = 0,8664 \\ P(1 < Z < 2) &= P(Z < 2) - P(Z < 1) \\ &= 0,9772 - 0,8413 = 0,1359 \end{aligned}$$

Ejemplo 174 Sea $X \sim N(\mu = 2, \sigma = 3)$. Calculamos $P(X < 4)$.

En este caso, transformamos la variable original usando la transformación tipificante.

$$\begin{aligned} P(X < 4) &= P(X - 2 < 4 - 2) \\ &= P\left(\frac{X - 2}{3} < \frac{4 - 2}{3}\right) \\ &= P(Z < 0,66\dot{6}) \quad \text{donde } Z \sim N(0, 1) \\ &= 0,7475 \end{aligned}$$

¿Cuál es $P(-1 < X < 3,5)$?

$$\begin{aligned} P(-1 < X < 3,5) &= P(-1 - 2 < X - 2 < 3,5 - 2) \\ &= P\left(\frac{-1 - 2}{3} < \frac{X - 2}{3} < \frac{3,5 - 2}{3}\right) \\ &= P(-1 < Z < 0,5) \\ &\quad \text{donde } Z \sim N(0, 1) \\ &= P(Z < 0,5) - P(Z < -1) \\ &= 0,6915 - 0,1587 = 0,5328 \end{aligned}$$

Ejemplo 175 *Es difícil etiquetar la carne empaquetada con su peso correcto debido a los efectos de pérdida de líquido (definido como porcentaje del peso original de la carne). Supongamos que la pérdida de líquido en un paquete de pechuga de pollo se distribuye como normal con media 4% y desviación típica 1%*

Sea X la pérdida de líquido de un paquete de pechuga de pollo elegido al azar.

¿Cuál es la probabilidad de que $3\% < X < 5\%$?

¿Cuál es el valor de x para que un 90% de paquetes tienen pérdidas de líquido menores de x ?

En una muestra de 4 paquetes, hallar la probabilidad de que todos tengan pérdidas de peso de entre 3 y 5%.

Sexauer, B. (1980). *Journal of Consumer Affairs*, **14**, 307-325.

$$\begin{aligned}
P(3 < X < 5) &= P\left(\frac{3-4}{1} < \frac{X-4}{1} < \frac{5-4}{1}\right) \\
&= P(-1 < Z < 1) \\
&= P(Z < 1) - P(Z < -1) \\
&= 0,8413 - 0,1587 = 0,6827
\end{aligned}$$

Queremos $P(X < x) = 0,9$. Entonces

$$P\left(\frac{X-4}{1} < \frac{x-4}{1}\right) = P(Z < x-4) = 0,9$$

Mirando las tablas, tenemos $x-4 \approx 1,282$ que implica que un 90 % de las paquetes tienen pérdidas de menos de $x = 5,282\%$.

Para un paquete $p = P(3 < X < 5) = 0,6827$. Sea Y el número de paquetes en la muestra que tienen pérdidas de entre 3 % y 5 %. Luego $Y \sim B(4, 0,6827)$.

$$P(Y = 4) = \binom{4}{4} 0,6827^4 (1-0,6827)^0 = 0,2172$$

Sumas y diferencias de dos variables normales

Si $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu, \sigma_Y)$ son independientes, entonces la distribución de la suma o diferencia de ambas variables es también normal con las siguientes medias y desviaciones típicas.

$$X + Y \sim N\left(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$
$$X - Y \sim N\left(\mu_X - \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$

Ejemplo 176 (Jun 04, Ej. 2) Un determinado establecimiento vende tres marcas diferentes de coches. Sean X_1 , X_2 y X_3 variables aleatorias independientes normales, que representan el volumen semanal de ventas para cada una de las marcas. Las ventas medias semanales de estas marcas son 42, 60 y 78 mil euros, respectivamente, y sus desviaciones típicas respectivas son 12, 18 y 10 mil euros.

a) ¿Cuál es la probabilidad de que la primera marca no supere los 30 mil euros en una semana? ¿Y la probabilidad de que la segunda marca supere en un semana la mediana de la tercera marca?

b) Calcular la probabilidad de que, en una semana determinada, las ventas del establecimiento sean superiores a los 120 mil euros.

c) ¿Cuál es la probabilidad de que la suma de las ventas de la primera marca y de la tercera superen a las ventas de la segunda marca en más de 18 mil euros, en una semana?

a) $X_1 \sim N(42, 12)$. Luego

$$\begin{aligned} P(X_1 < 30) &= P\left(\frac{X_1 - 42}{12} < \frac{30 - 42}{12}\right) \\ &= P(Z < -1) \quad \text{donde } Z \sim N(0, 1) \\ &= 0,1587 \end{aligned}$$

La distribución normal es simétrica en torno de su media, es decir que la moda, media y mediana son iguales. Entonces, queremos calcular $P(X_2 > 78)$.

$$\begin{aligned} P(X_2 > 78) &= P\left(\frac{X_2 - 60}{18} > \frac{78 - 60}{18}\right) \\ &= P(Z > 1) \\ &= 0,1587 \end{aligned}$$

b) Las ventas totales son $Y = X_1 + X_2 + X_3$.
Luego

$$E[Y] = 42 + 60 + 78 = 180$$

$$V[Y] = 12^2 + 18^2 + 10^2 = 568$$

$$DT[Y] \approx 23,833$$

$$Y \sim N(180, 23,833)$$

$$\begin{aligned} P(Y > 120) &= P\left(\frac{Y - 180}{23,833} > \frac{120 - 180}{23,833}\right) \\ &= P(Z > -2,52) \\ &= P(Z < 2,52) = 0,994 \end{aligned}$$

c) Se quiere $P(X_1 + X_3 - X_2 > 18)$. Sea $D = X_1 + X_3 - X_2$. Luego $E[D] = 42 + 78 - 60 = 60$ y $V[D] = 568$ como calculamos anteriormente.

Luego

$$\begin{aligned} P(D > 18) &= P\left(\frac{D - 60}{23,833} > \frac{18 - 60}{23,833}\right) \\ &= P(Z > -1,762) \\ &= 0,961 \end{aligned}$$

Ejemplo 177 (Jun 02, Ej. 4) Las calificaciones de 0 a 100 obtenidas en dos pruebas distintas A y B por los alumnos presentados a la Selectividad, son independientes y siguen las distribuciones normales: $N_A(\mu = 62; \sigma = 20)$ y $N_B(\mu = 52; \sigma = 10)$; respectivamente. La prueba se considera superada con 50 puntos. Calcular:

(a) La probabilidad de que un alumno en la prueba A haya obtenido una puntuación menor que 40.

(b) La probabilidad que haya superado la prueba B.

(c) Si un alumno ha obtenido una puntuación de 68 en la primera prueba y de 62,5 en la segunda ¿en qué prueba ha obtenido mejor resultado respecto de los demás alumnos?

(d) Sea M la variable definida por $M = \frac{1}{2}(X + Y)$; donde X e Y representan las variables aleatorias anteriores. Calcular la media y la desviación típica.

(e) Si para el acceso a una Universidad se necesita que la media aritmética de las dos notas anteriores sea mayor que 70, ¿cuál es la probabilidad de que un alumno escogido al azar pueda acceder a dicha Universidad?

a) Sea X la nota del alumno de la prueba A. Luego $X \sim N(62, 20)$.

$$\begin{aligned} P(X < 40) &= P(X - 62 < 40 - 62) \\ &= P\left(\frac{X - 62}{20} < \frac{40 - 62}{20}\right) \\ &= P(Z < -1,1) \quad \text{donde } Z \sim N(0, 1) \\ &\approx 0,1357 \end{aligned}$$

b) Sea Y la nota del alumno de la prueba B. Entonces $Y \sim N(52, 10)$.

$$\begin{aligned} P(Y > 50) &= P\left(\frac{Y - 52}{10} > \frac{50 - 52}{10}\right) \\ &= P(Z > -0,2) \quad \text{donde } Z \sim N(0, 1) \\ &= P(Z < 0,2) \quad \text{por simetría} \\ &\approx 0,5793 \end{aligned}$$

c) Si su resultado en la primera prueba es 68, calculamos el valor de Z

$$Z_1 = \frac{68 - 62}{20} = 0,3$$

En el segundo caso su resultado es 62,5 y

$$Z_2 = \frac{62,5 - 52}{10} = 1,05$$

Luego hace mucho mejor en la segunda prueba con respecto a sus compañeros.

d)

$$\begin{aligned} E[M] &= E \left[\frac{1}{2}(X + Y) \right] \\ &= \frac{1}{2} (E[X] + E[Y]) \\ &= \frac{1}{2}(62 + 52) = 57 \\ V[M] &= V \left[\frac{1}{2}(X + Y) \right] \\ &= \frac{1}{4} (V[X] + V[Y]) \\ &= \frac{1}{4} (20^2 + 10^2) = 125 \end{aligned}$$

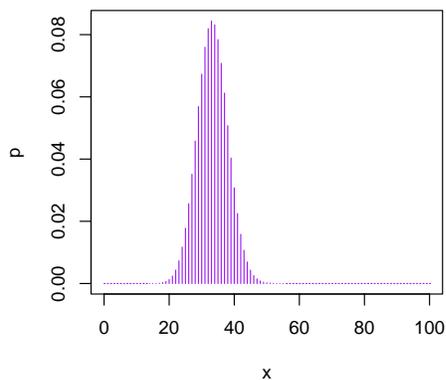
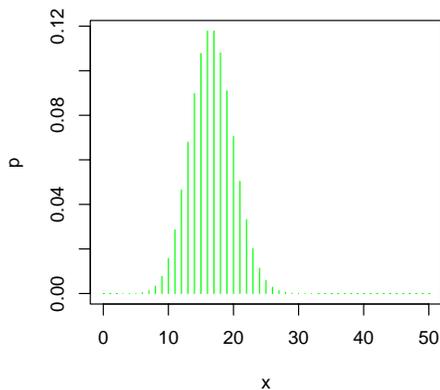
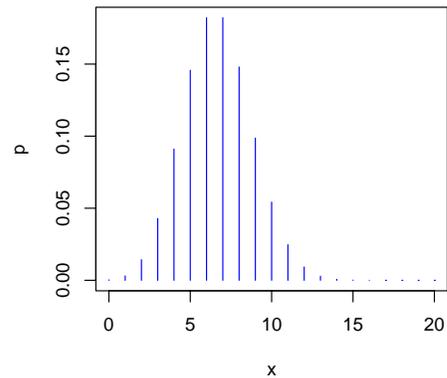
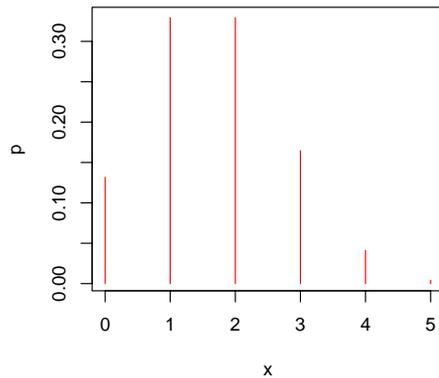
La desviación típica es $DT[M] = \sqrt{125} \approx 11,18$.

e) Luego $M \sim N(57, \sqrt{125})$ y

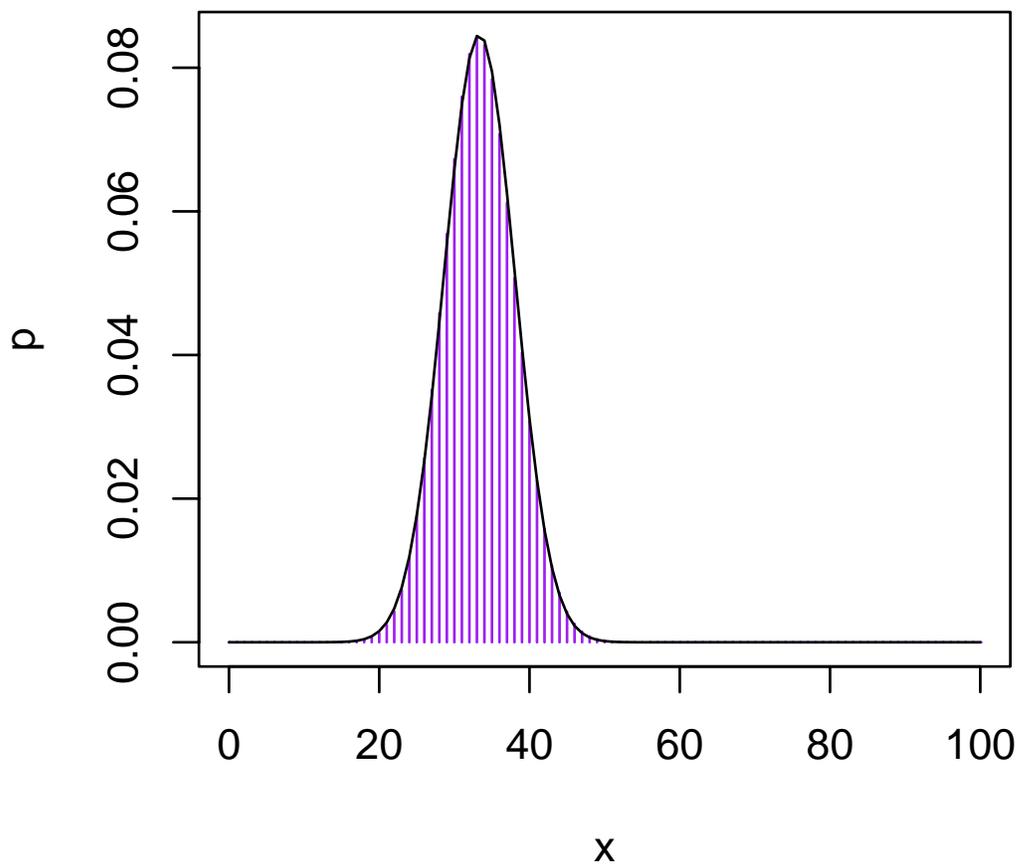
$$\begin{aligned} P(M > 70) &= P\left(\frac{M - 57}{\sqrt{125}} > \frac{70 - 57}{\sqrt{125}}\right) \\ &= P(Z > 1,163) \\ &= P(Z < -1,163) = 0,1225 \end{aligned}$$

Aproximación mediante la distribución normal

Hacemos el experimento de tirar una moneda con $p = 1/3$ un número n de veces. Dibujamos la función de probabilidad de $X = \#$ cruces en los casos $n = 5, 20, 50$ y 100 .



Se ve que para n grande, la función de probabilidad binomial tiene una forma parecida a la densidad normal.



Aproximación de la distribución binomial

Teorema 19 Si $X \sim B(n, p)$, entonces si n (y np y $np(1 - p)$) es grande,

$$\frac{X - np}{\sqrt{np(1 - p)}} \approx N(0, 1)$$

Esta aproximación funciona bastante bien si tanto n ($n > 30$) como np y $n(1 - p)$ son bastante grandes. Si np o $n(1 - p)$ es pequeño, (< 5) la aproximación Poisson funciona mejor.

Ejemplo 178 Sea $X \sim B(100, 1/3)$. Estimar $P(X < 40)$.

Calculamos primero a media y varianza de X .

$$\begin{aligned} E[X] &= 100 \times \frac{1}{3} = 33.3 \\ V[X] &= 100 \times \frac{1}{3} \times \frac{2}{3} \\ &= 22.2 \\ DT[X] &\approx 4,714 \end{aligned}$$

Ahora usamos la aproximación normal

$$\begin{aligned} P(X < 40) &= P\left(\frac{X - 33.3}{4,714} < \frac{40 - 33.3}{4,714}\right) \\ &\approx P(Z < 1,414) \quad \text{donde } Z \sim N(0, 1) \\ &\approx 0,921 \end{aligned}$$

La probabilidad correcta es

$$\sum_{x=0}^{39} \binom{100}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{100-x} = 0,903$$

La aproximación no parece gran cosa pero lo podemos mejorar.

La corrección de continuidad

Si $X \sim B(n, p)$, entonces X es una variable discreta y luego $P(X \leq x) = P(X < x + 1)$ y igualmente $P(X \geq x) = P(X > x - 1)$.

Luego cuando implementamos la aproximación normal, usamos **la corrección de continuidad**

$$P(X \leq x) = P(X < x + 0,5)$$

$$P(X \geq x) = P(X > x - 0,5)$$

$$P(x_1 \leq X \leq x_2) = P(x_1 - 0,5 < X < x_2 + 0,5)$$

Ejemplo 179 *Volvemos al Ejemplo 178. Ahora usamos la corrección de continuidad.*

$$\begin{aligned} P(X < 40) &= P(X \leq 39) \\ &= P(X < 39,5) \\ &\approx P\left(Z < \frac{39,5 - 33,3}{4,714}\right) \\ &= P(Z < 1,308) = 0,905 \end{aligned}$$

La aproximación es algo mejor usando la corrección de continuidad.

Ejemplo 180 *El 35 % de los habitantes de una ciudad votan a cierto partido político. Se encuesta a 200 personas. Llamemos X al número de personas que votan a dicho partido.*

¿Cuál es la distribución de X ?

Calcular la probabilidad de que entre la gente de la encuesta haya entre 70 y 80 votantes de ese partido.

La verdadera distribución de X es binomial

$$X \sim B(200, 0,35).$$

La media de la distribución es 70 y la desviación típica es 6,745.

Para calcular $P(70 \leq X \leq 80)$ usamos una aproximación normal

$$\begin{aligned}
P(70 \leq X \leq 80) &= P(69,5 < X < 80,5) \\
&= P\left(\frac{69,5 - 70}{6,745} < \frac{X - 70}{6,745} < \frac{80,5 - 70}{6,745}\right) \\
&\approx P(-0,074 < Z < 1,557) \\
&= P(Z < 1,557) - P(Z < -0,074) \\
&= 0,940 - 0,470 = 0,47
\end{aligned}$$

La distribución binomial no es la única distribución que se puede aproximar mediante una distribución normal. Cualquiera distribución la que se puede representar como la distribución de una media (o suma) de variables independientes y idénticamente distribuidas

$$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n)$$

puede estar aproximada por una normal.

El teorema central del límite

Teorema 20 *Sea $X_1, \dots, X_n \sim f(\cdot)$ con media μ y desviación típica σ . Luego si n es grande,*

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0, 1)$$

El teorema también implica que si n es grande, la suma $\sum_{i=1}^n X_i$ tiene aproximadamente una distribución normal

$$\sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2)$$

Aproximación de la distribución Poisson

Sea $X \sim P(\lambda)$ el número de sucesos raros en una unidad de tiempo. Definimos Y como el número de sucesos en n unidades de tiempo. Luego podemos escribir

$$Y = X_1 + X_2 + \dots + X_n$$

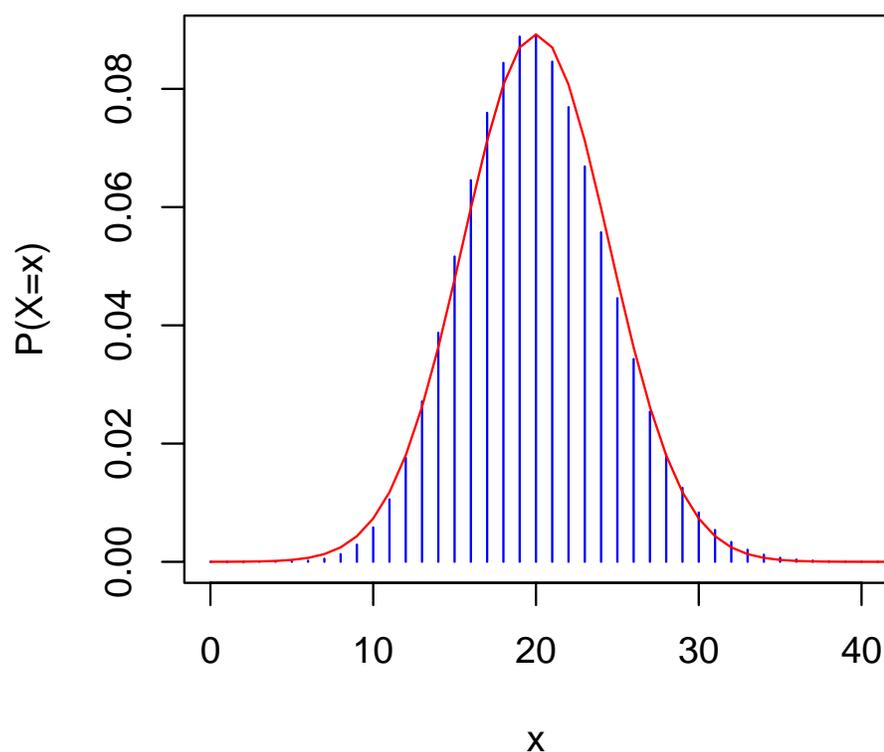
donde $X_i \sim P(\lambda)$ es el número de sucesos en la i -ésima unidad de tiempo.

Así, podemos aplicar el teorema central del límite a aproximar la distribución Poisson con una distribución normal.

Teorema 21 *Sea $X \sim P(\lambda)$. Para λ grande ($\lambda > 20$), entonces*

$$X \approx N(\lambda, \sqrt{\lambda})$$

El gráfico muestra la función de probabilidad de la distribución Poisson con $\lambda = 20$ y la densidad normal con media y varianza λ .



La aproximación se mejora si el valor de λ es más grande.

Cuando se utiliza la aproximación a la distribución Poisson, es importante aplicar la corrección de continuidad.

Ejemplo 181 Sea $X \sim P(49)$.

Estimar $P(45 \leq X \leq 52)$.

$$\begin{aligned}
 P(45 \leq X \leq 52) &= P(44,5 < X < 52,5) \\
 &\quad \text{por la corrección de continuidad} \\
 &= P\left(\frac{44,5 - 49}{\sqrt{49}} < \frac{X - 49}{\sqrt{49}} < \frac{52,5 - 49}{\sqrt{49}}\right) \\
 &\approx P(-0,643 < Z < 0,5) \text{ donde } Z \sim N(0, 1) \\
 &= P(Z < 0,5) - P(Z < -0,643) \\
 &= 0,6915 - 0,2602 \\
 &\approx 0,431
 \end{aligned}$$

La solución exacta calculada a través de la distribución Poisson es

$$\begin{aligned}
 P(45 \leq X \leq 52) &= \sum_{x=45}^{52} \frac{49^x e^{-49}}{x!} \\
 &= 0,433
 \end{aligned}$$

La distribución logarítmico normal

Si $X \sim N(\mu, \sigma)$ y se define $Y = e^X$, luego se dice que Y se distribuye como logarítmico normal con parámetros μ, σ . La distribución logarítmico normal es un modelo empleado típicamente para tiempos de funcionamiento de máquinas y para variables asimétricas como ingresos o gastos.

