RECORDATORIO DE LAS DISTRIBUCIONES DE PROBABILIDAD

La Distribución Bernoulli.

La distribución Bernoulli está asociada con un experimento aleatorio denominado experimento Bernoulli. Un experimento Bernoulli es aquél, en donde son sólo dos los resultados posibles del experimento, a estos resultados posibles los llamaremos "éxito" y "fracaso". Por ejemplo, lanzar una moneda y registrar el lado que cae hacia arriba es un experimento Bernoulli, porque sólo hay dos resultados posibles. Otro ejemplo, al analizar un producto para saber si está defectuoso o no está defectuoso es un experimento Bernoulli, ya que hay dos resultados posibles: "defectuoso" y "no defectuoso".

Si en un experimento Bernoulli al resultado "éxito" se le asigna el número 1 y al resultado "fracaso" se le asigna el número 0, entonces se podría definir la v.a. X con rango $X = \{0, 1\}$, donde X("éxito")=1 y X("fracaso")=0. Si conocemos la probabilidad de que ocurra el "éxito" denotada por p, entonces la probabilidad de que ocurra el "fracaso" es 1-p, de tal manera que el valor de p define la función de probabilidad de la v.a. X,y a esta función de probabilidad se le llama distribución Bernoulli como se indica en la siguiente definición.

Definición de la Distribución Bernoulli.

La v.a. discreta X tiene distribución Bernoulli si su función de probabilidad es:

$$f(x) = \begin{cases} p & \text{si } x = 1 \\ 1-p & \text{si } x = 0 \\ 0 & \text{de otro modo} \end{cases}$$

además: E(X) = p y Var(X) = p(1-p).

Ejemplo

Se analizará un producto para determinar si es "defectuoso" o "no defectuoso". Se sabe que el 11% de la producción sale defectuosa. Si se define la v.a. X = número de artículos defectuosos que resultan, obtener la función de probabilidad de X. Para obtener $f(x)$, observe que el rango de X es {1, 0}, solo dos resultados posibles, entonces X tiene distribución Bernoulli, donde P(X=1) = P("defectuoso") = 0.11 y P(X=0) = P("no defectuoso") = 1-0.11 = 0.89; luego:

$$f(x) = \begin{cases} 0.11 & \text{si } x = 1 \\ 0.89 & \text{si } x = 0 \\ 0 & \text{de otro modo.} \end{cases}$$

además, E(X) = 0.11 y Var(X) = 0.11(0.89) = 0.0979.

La Distribución Binomial.

La distribución binomial se aplica en la siguiente situación. Suponer que un experimento Bernoulli se repite n veces <u>de manera independiente</u>, donde la probabilidad de que ocurra "éxito" es p. Sea la v.a. X definida como X = al número de "éxitos" ocurridos en las n repeticiones del experimento Bernoulli. Entonces, la v.a. X tiene distribución binomial con parámetros n, p. Observe que el rango de X es: $\{0, 1, 2, \cdots, n\}$.

La situación anterior se presenta cuando, por ejemplo, se inspecciona un lote de 50 artículos y se determina la cantidad de artículos defectuosos que hay en el lote. El experimento Bernoulli es inspeccionar un artículo para determinar si está defectuoso o no y este experimento se va a repetir 50 veces (son 50 artículos). Si se sabe que el 10% de los artículos son defectuosos, entonces la probabilidad de que un artículo sea defectuoso es 0.1 y si los experimentos Bernoulli (inspección de un artículo) son independientes, (el hecho de que un artículo salió o no salió defectuoso no afecta a la probabilidad de que el siguiente salga defectuoso), entonces la v.a. definida como el número de artículos defectuosos en el lote tiene distribución binomial con parámetros n = 50, p = 0.1.

Definición de Distribución Binomial.

La v.a. X tiene distribución binomial, si su función de probabilidad es:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{para } x = 0, 1, 2, \cdots, n.$$

donde $f(x) = 0$ de otro modo. Además, $E(X) = np$ y $\mathrm{Var}(X) = np(1-p)$.

El símbolo $\binom{n}{x}$ (se lee: "combinaciones de n en x") y se define como $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ con $0 \le x \le n$, donde $r! = 1\times2\times3\times4\times\cdots\times r$, y $0!=1$. El símbolo $r!$, se lee "r factorial" o "el factorial de r". Además, $E(X) = np$ y $\mathrm{Var}(X) = np(1-p)$.

El cálculo de $\binom{n}{x}$ ya está integrado en las calculadoras científicas de bolsillo y no debe ser problema obtener el resultado de estos cálculos.

Ejemplo

Se tiene un lote de 50 artículos, de los cuales se sabe que el 10% de la producción es defectuoso. Al hacer la inspección del lote, obtener la probabilidad de que haya a) 3 artículos defectuosos, b) a lo más 4 artículos defectuosos, c) más de 3 artículos defectuosos, d) el promedio del número de artículos defectuosos, e) la varianza y la desviación estándar del número de artículos defectuosos.

Solución: Sea X = número de artículos defectuosos en el lote. Suponiendo que el resultado de las inspecciones son independientes, entonces X tiene distribución binomial con parámetros n=50 p=0.1 (en símbolos: X~binomial(n=50, p=0.1)). Luego, de la definición de distribución binomial tenemos que la función de probabilidad de X es:

$$f(x) = \binom{50}{x} (0.1)^x (0.9)^{50-x} \quad \text{para } x = 0, 1, 2, \cdots, 50.$$

Para resolver el a), tenemos que:

$$P(X=3) = f(3) = \binom{50}{3} (0.1)^3 (0.9)^{47} = (19600)(0.1)^3(0.9)^{47} \approx 0.1386.$$

Para resolver el b) tenemos que:

$$P(X \le 4) = \sum_{i=0}^{4} f(i) = \sum_{i=0}^{4} \binom{50}{i} (0.1)^i (0.9)^{50-i} = \binom{50}{0}(0.1)^0(0.9)^{50} + \binom{50}{1}(0.1)^1(0.9)^{49} +$$

$$\binom{50}{2}(0.1)^2(0.9)^{48} + \binom{50}{3}(0.1)^3(0.9)^{47} + \binom{50}{4}(0.1)^4(0.9)^{46} \approx 0.4312.$$

Estas probabilidades se pueden obtener en una hoja de cálculo como excel.

Para resolver el c), tenemos que:

$$P(X > 3) = 1 - P(X \le 3) = 1 - \left[\sum_{i=0}^{3} f(i)\right] = 1 - \left[\binom{50}{0}(0.1)^0(0.9)^{50} + \binom{50}{1}(0.1)^1(0.9)^{49} +\right.$$

$$\left.\binom{50}{2}(0.1)^2(0.9)^{48} + \binom{50}{3}(0.1)^3(0.9)^{47}\right] \approx 1 - 0.2503 = 0.7497.$$

Para resolver el d), de la definición de distribución binomial tenemos que: E(X) = np = (50)(0.1) = 5.0; es decir, en promedio salen 5 artículos defectuosos por lote. Igualmente, para resolver el e), tenemos de la definición de distribución binomial, que:

$$\text{Var}(X) = np(1\text{-}p) = (50)(0.1)(0.9) = 4.5 \ \text{ y } \ \sigma_X = \sqrt{4.5} \approx 2.121$$

Las figuras 1 y 2 ilustran el aspecto de la gráfica de una función de probabilidad para una distribución binomial.

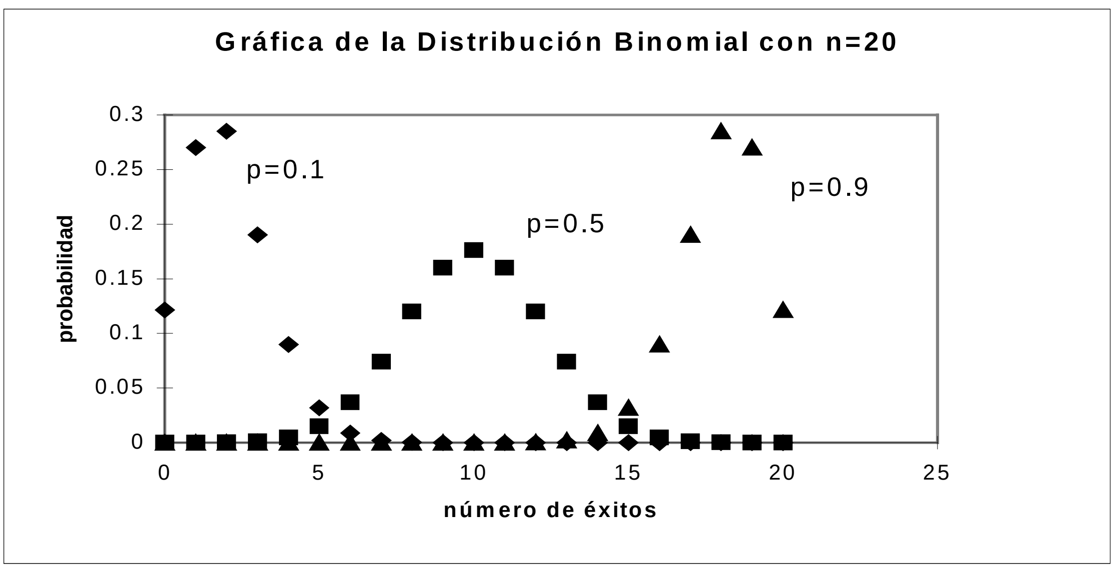


Figura 1 Gráfica de $f(x)$ de una distribución binomial variando el valor de p.

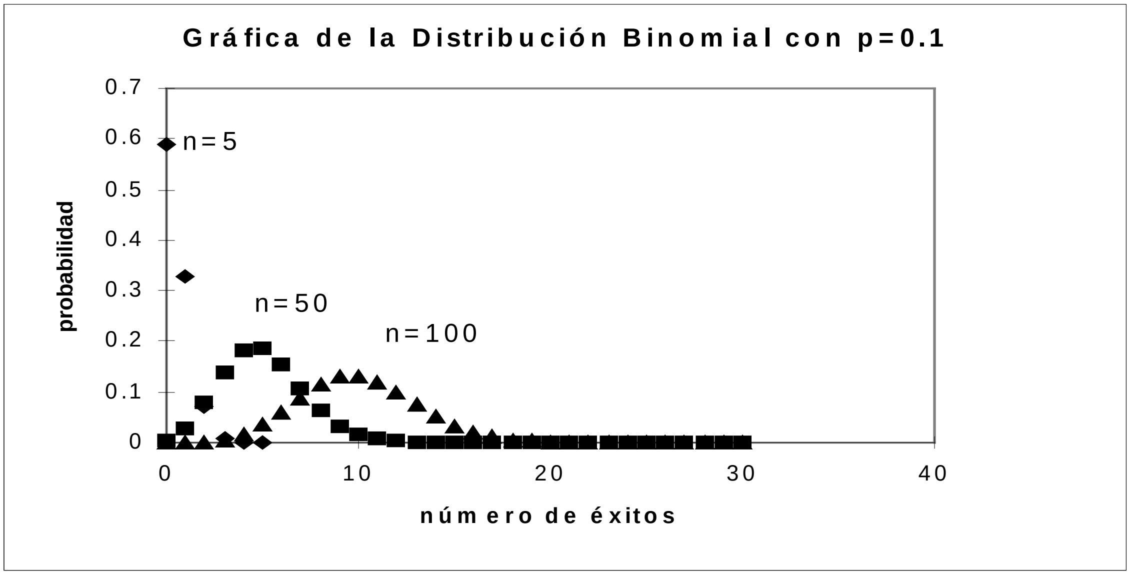


Figura 2 Gráfica de $f(x)$ de una distribución binomial variando el valor de n.

NOTA: Las probabilidades de la distribución binomial se pueden calcular en una hoja de cálculo como excel.

La Distribución Geométrica.

La distribución geométrica se aplica en la siguiente situación. Suponer que se tiene un experimento Bernoulli con probabilidad de "éxito" p. Suponer que este experimento de Bernoulli se repite independientemente hasta que por primera vez ocurre el "éxito". Sea la v.a. X = número de repeticiones realizadas hasta que ocurre el "éxito" por primera vez. Entonces, X tiene distribución geométrica con parámetro p. Observe que el rango de X es $\{1, 2, 3, \cdots\}$.

Por ejemplo, suponga que se tienen 5 llaves, de las cuales sólo una llave pone a funcionar una máquina. Una persona saca con reposición una llave e intenta encender la máquina con dicha llave, entonces, el número de intentos realizados hasta que se enciende la máquina tiene distribución geométrica con parámetro p = 1/5. Observe que, como se saca una llave con reposición, entonces los intentos son independientes y son experimentos de Bernoulli, ya que hay dos resultados posibles: enciende o no enciende la máquina.

Definición de Distribución Geométrica.

La v.a X tiene distribución geométrica si su función de probabilidad es:

$$f(x) = (1 - p)^{x-1}\, p \qquad \text{para } x = 1, 2, 3, \cdots$$

además, $\quad E(X) = \dfrac{1}{p}$ y $\mathrm{Var}(X) = \dfrac{1}{p}\left(\dfrac{1}{p} - 1\right)$.

Ejemplo

Se tiene un equipo de transmisión. Se sabe que la señal enviada se recibe incorrectamente el 5.7% de las veces. Se envía una señal consecutivamente hasta que se recibe en forma incorrecta. Calcular la probabilidad de que se hagan; a) 5 envíos, b) a lo más 5 envíos, c) más de 3 envíos, d) más de 2 y menos de 7 envíos, e) calcular el promedio del número de envíos, f) calcular la varianza y desviación estándar del número de envíos.

Sea X = número de envíos hechos hasta que se recibe incorrectamente la señal. Entonces, suponiendo resultados independientes en cada señal para el a), tenemos que: $P(X = 5) = f(5) = (0.943)^4(0.057) \approx 0.045$, para el b) tenemos que

$P(X \leq 5) = \sum_{i=1}^{5} f(i) = (0.943)^0(0.057) + (0.943)(0.057) + (0.943)^2(0.057) + (0.943)^3(0.057) +$
$(0.943)^4(0.057) \approx 0.2543$, para el c) tenemos que:

$P(X > 3) = 1 - P(X \leq 3) = 1 - \mathbf{[}(0.943)^0(0.057) + (0.943)(0.057) + (0.943)^2(0.057)\mathbf{]} \approx 0.83856$. Las probabilidades acumuladas $P(X \leq x)$ se puede calcular con la fórmula $P(X \leq x) = 1 - (1 - p)^x$. Entonces $P(X \leq 3) = 1 - (1 - 0.057)^3 = 0.1614$.

Para el d) tenemos que: $P(2 < X < 7) = \sum_{i=3}^{6} f(i) = (0.943)^2(0.057) + (0.943)^3(0.057) + (0.943)^4(0.057) + (0.943)^5(0.057) \approx 0.18606$, para el e), de la definición de distribución geométrica, tenemos que $E(X) = 1/0.057 \approx 17.54$; entonces, en promedio se hacen 17.5 envíos hasta que se recibe la señal incorrectamente. También de la definición de distribución geométrica se tiene que para el f), $Var(X) = (1/0.057)[(1/0.057) - 1] \approx 290.24$ y $\sigma_X \approx 17.037$.

La figura 3 muestra el aspecto de la función de probabilidad de una distribución geométrica variando el valor de p.

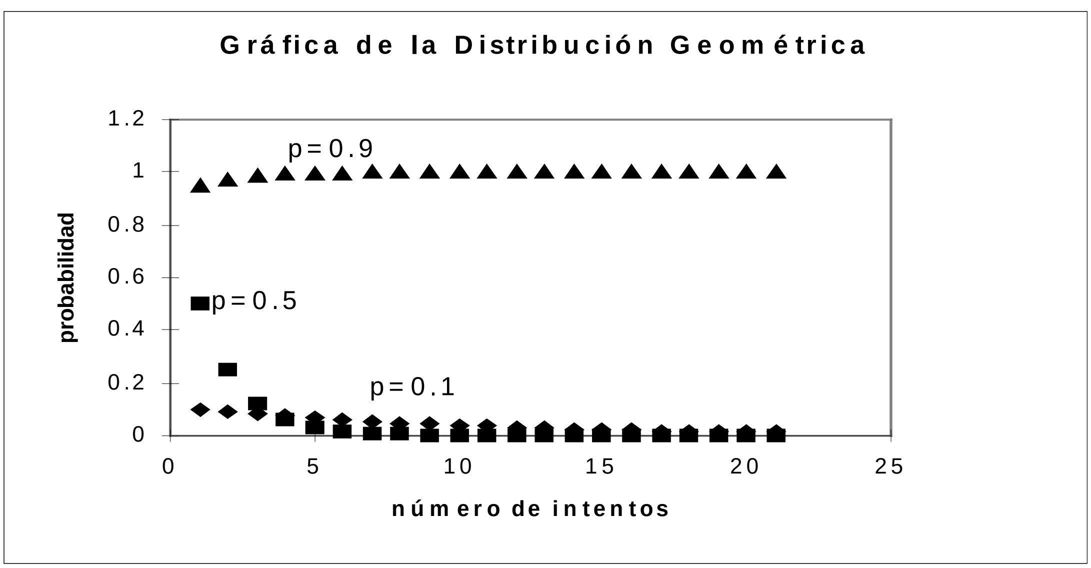


Figura 3 Gráfica de $f(x)$ de una distribución Geométrica variando p.

La Distribución Poisson

La distribución Poisson se usa para modelar el número de ocurrencias de un evento que denominaremos "éxito" en un intervalo específico de tiempo o de espacio. Por ejemplo, la distribución Poisson podría utilizarse para modelar el número de clientes por hora que llegan a un negocio, el número de imperfecciones por 100 m en un cable de acero, el número de defectos por 5 m^2 de lámina, el número de partículas de plomo por m^3 de aire, etc.

Definición de la Distribución Poisson

La v.a. X tiene distribución Poisson si su función de probabilidad es:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{para } x = 0, 1, 2, 3, \cdots$$

donde $\lambda > 0$. Además $E(X) = \lambda$ y $Var(X) = \lambda$.

La figura 4 muestra el aspecto de la función de probabilidad de la distribución Poisson para diferentes valores de λ.

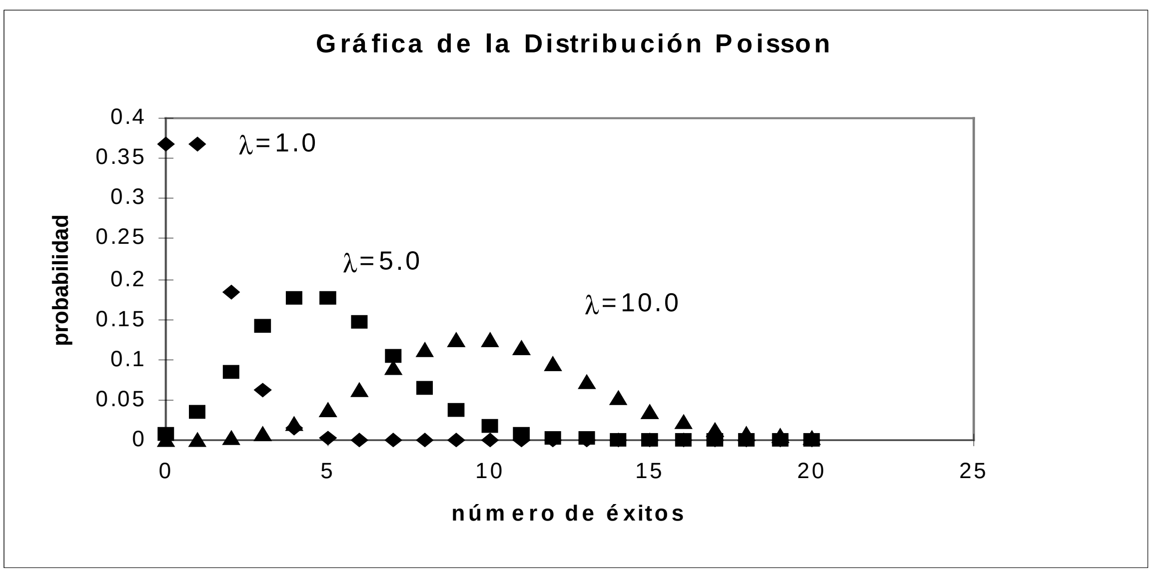


Figura 4 Gráfica de $f(x)$ para la distribución Poisson variando el valor de λ.

Ejemplo

Suponer que el número de imperfecciones en una lámina de acero tiene una distribución Poisson con un promedio de 5 imperfecciones por 7 m^2. Calcular la probabilidad de que en 7 m^2 haya; a) 3 imperfecciones, b) más de 5 imperfecciones, c) menos de 6 imperfecciones, d) por lo menos 4 y a lo mucho 8 imperfecciones.

De la definición de distribución Poisson tenemos que para el a) λ = 5, entonces: P(X = 3) = $f(3) = \frac{(5)^{-3}e^{-5}}{3!} \approx$ 0.00000898; para el b) tenemos que P(X > 5)=1 - F(5)=1 - $\sum_{i=0}^{5}\left\{\frac{5^{-i}e^{-5}}{i!}\right\}$ ≈1 - 0.616, entonces P(X > 5) ≈ 0.384, para el c) tenemos que:

P(X < 6) = F(5) = $\sum_{i=0}^{5}\left\{\frac{5^{-i}e^{-5}}{i!}\right\}$ ≈ 0.616, para el d) tenemos que:

P(4 ≤ X ≤ 8) = F(8) - F(3) = $\sum_{i=0}^{8}\left\{\frac{5^{-i}e^{-5}}{i!}\right\} - \sum_{i=0}^{3}\left\{\frac{5^{-i}e^{-5}}{i!}\right\} = \sum_{i=4}^{8}\left\{\frac{5^{-i}e^{-5}}{i!}\right\} \approx 0.6669$ ■

NOTA: Las probabilidades de una distribución poisson se pueden obtener en una hoja de cálculo como excel.

La Distribución Normal

La distribución normal se aplica para modelar errores en las mediciones, entre otras muchas aplicaciones. La distribución normal también es fundamental en el desarrollo de la estadística matemática, de tal manera que es la base para diseñar muchas pruebas de hipótesis y estimación de parámetros poblacionales. Por supuesto que la distribución normal no escapa a las aplicaciones de la estadística en la industria.

Definición de la Distribución Normal

La v.a. X tiene distribución normal si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2} \quad \text{para} \quad -\infty < x < +\infty$$

donde, $-\infty<\mu<+\infty$ y $\sigma > 0$. Además; E(X) = μ y Var(X) = σ^2.

Cuando $\mu = 0$ y $\sigma = 1$, entonces se le da el nombre de distribución normal estándar, denotada por $\phi(z)$, quedando como: $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ para $-\infty < z < +\infty$. La función de distribución acumulada de z denotada por $\Phi(z)$, ya está calculada para algunos valores de z en las tablas 2.6 y 2.7 que aparecen al final de esta unidad, donde $\Phi(z) = P(Z \le z)$.

Las probabilidades de una v.a. con distribución normal se pueden calcular mediante las probabilidades de la distribución normal estándar mediante la transformación: $Z = \frac{X-\mu}{\sigma}$ como se muestra a continuación.

Ejemplo

Se sabe que el peso (en kg.) de carga de un camión, es una v.a. con distribución normal con media 3000 kg y desviación estándar 500 kg. Calcular la probabilidad de que la carga del camión sea: a) menor a 2500 kg., b) mayor a 3600 kg., c) entre 2600 y 4000 kg.

Solución: tenemos que los parámetros de la distribución de X son; $\mu = 3000$ y $\sigma = 500$. Luego, para el a), tenemos que P(X < 2500) = P(X - 3000 < 2500 - 3000) = $P\left(\frac{X-3000}{500} < \frac{2500-3000}{500}\right)$ = P(Z < -1.0) = Φ(-1.0) ≈ 0.1587. Observe las transformaciones algebraicas para obtener $Z = \frac{X-\mu}{\sigma} = \frac{X-3000}{500}$ y obtener el valor de la tabla de probabilidades de la distribución normal estándar que está al final del capítulo. Para el b), tenemos que;

P(X > 3600) = $P\left(\frac{X-3000}{500} < \frac{3600-3000}{500}\right)$ = P(Z < 1.2) = Φ(1.2) ≈ 0.8849,

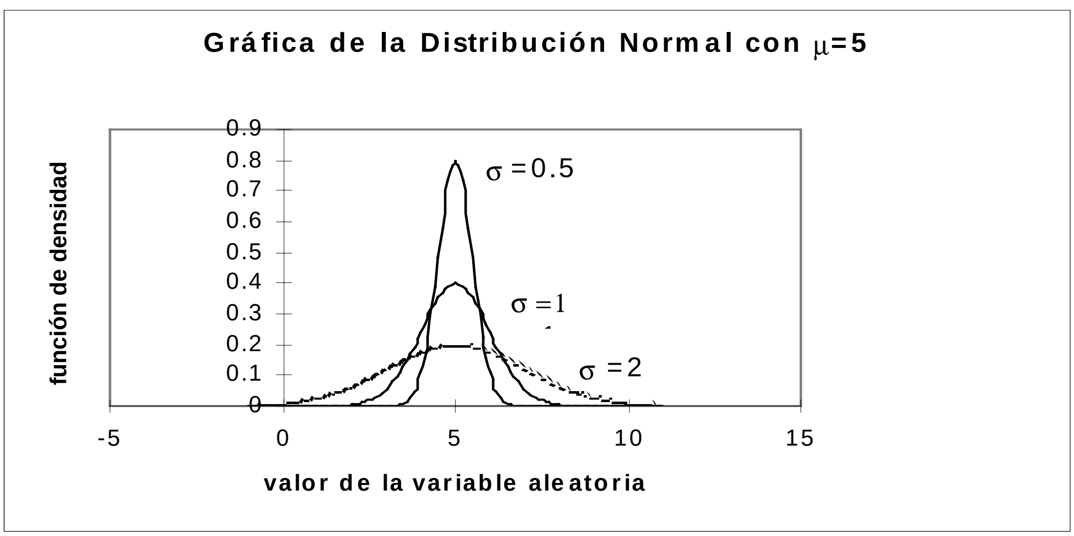
para el c), tenemos que;

P(2600 < X < 4000) = $P\left(\frac{2600-3000}{500} < \frac{X-3000}{500} < \frac{4000-3000}{500}\right)$, luego,

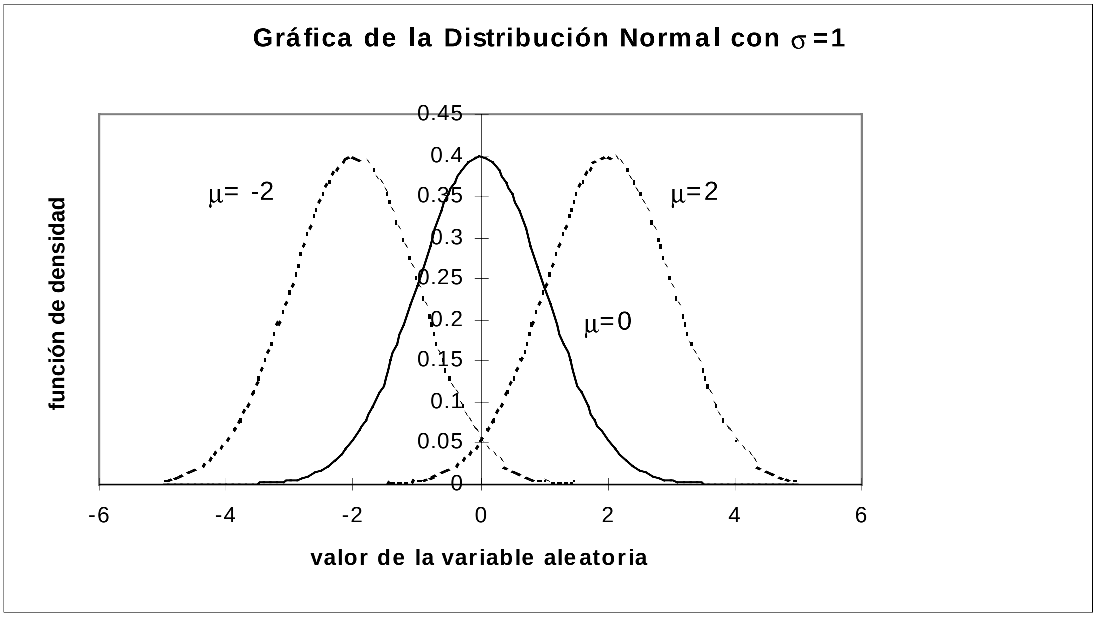
P(2600 < X < 4000) = P(-0.8 < Z < 2.0) = Φ(2.0) - Φ(-0.8) ≈ 0.9772 - 0.2119 = 0.7653

NOTA: Las probabilidad acumulada de una distribución normal se pueden obtener directamente de una hoja de cálculo como excel.

Las siguientes figuras muestran el aspecto de la función de densidad de la distribución normal para diferentes valores de μ y σ.



Gráfica de $f(x)$ para la distribución normal para varios valores de σ.



Gráfica de $f(x)$ para la distribución normal para varios valores de μ.

La Distribución Exponencial.

Esta distribución se aplica para modelar el tiempo transcurrido entre la ocurrencia de dos "éxitos" consecutivos; estos éxitos pueden ser: llegadas de clientes a un negocio, llegadas de piezas de producción a una estación de trabajo, fallas consecutivas de una máquina. En general, esta distribución se utiliza en la modelación de sistemas de líneas de espera y en la teoría de confiabilidad.

Definición de la Distribución Exponencial.

La v.a. X tiene distribución exponencial, si su función de densidad es:

$$f(x) = \begin{cases} \theta e^{-\theta x} & \text{para } x > 0 \\ 0 & \text{de otro modo} \end{cases}$$

donde, $\theta > 0$. Además, $E(X) = \dfrac{1}{\theta}$ y $\mathrm{Var}(X) = \dfrac{1}{\theta^2}$.

Ejemplo

El tiempo en horas que transcurre entre dos fallas consecutivas de una máquina es una v.a. con distribución exponencial: $f(t) = \begin{cases} (1/1000)\, e^{-t/1000} & \text{para } t > 0 \\ 0 & \text{de otro modo.} \end{cases}$

Calcular la probabilidad de que el tiempo que transcurre para que la máquina falle sea: a) menos de 100 hrs, b) más de 1000 hrs, c) entre 800 y 1000 hrs, d) el promedio del tiempo entre fallas consecutivas, d) la varianza y desviación estándar del tiempo entre fallas consecutivas.

Solución: de la definición de $f(x)$ se tiene que $\theta = 1/1000$, luego, para el a), tenemos que:

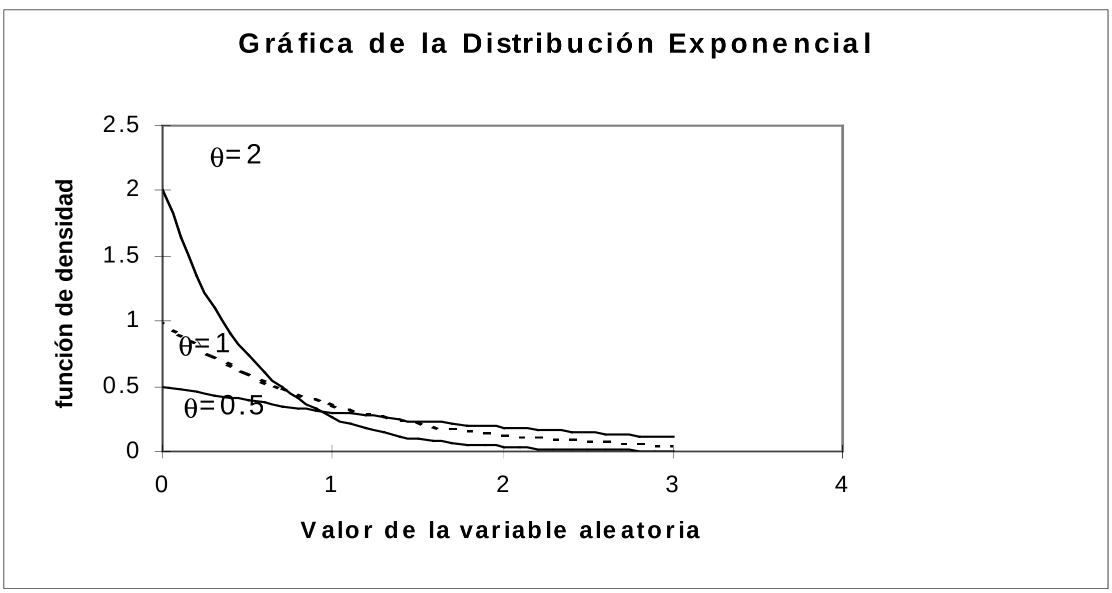
$P(T < 100) = \int_0^{100} (1/1000)\, e^{-t/1000} dt = 1 - e^{-1/10} \approx 0.0952$, para el b), tenemos que:

$P(T > 1000) = \int_{1000}^{+\infty} (1/1000)\, e^{-t/1000} dt = e^{-1} \approx 0.3679$, para el c), tenemos que:

$P(800 < T < 1000) = \int_{800}^{1000} (1/1000)\, e^{-t/1000} dt = e^{-8/10} - e^{-1} \approx 0.0814$, para el d), tenemos que:

$E(T) = 1/\theta = \dfrac{1}{1/100} = 1000$; luego, en promedio transcurren 1000 horas entre fallas consecutivas. Para el e), tenemos que: $\mathrm{Var}(T) = \dfrac{1}{\theta^2} = \dfrac{1}{(1/1000)^2} = 1000^2$ y $\sigma_T = 1000$ hrs. La probabilidad acumulada se puede obtener directamente con la fórmula $F(x) = 1 - e^{-x/\theta}$.

La siguiente figura muestra el aspecto de la función de densidad de una distribución exponencialpara diferentes valores de θ.



Gráfica de $f(x)$ de la distribución exponencial para varios valores de θ.

La Distribución χ^2 (Ji cuadrada)

Una variable aleatoria X tiene distribución χ^2 con ν grados de libertad si su función de densidad es $f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu-2)/2} e^{-x/2}$ para $x > 0$, donde $\Gamma(w)$ es la función gamma evaluada en w y ν es una constante positiva. Además $E(X) = \nu$, y $V(X) = 2\nu$.

Por ejemplo, si X sigue una distribución χ^2 con 7 grados de libertad, entonces de una hoja de cálculo como excel se puede obtener que $P(X < 5) = 0.5481$, y $P(X < 2) = 0.1509$. Luego $P(2 < X < 5) = 0.5481 - 0.1509 = 0.3972$. Consultar en libros de estadística el aspecto que tiene la curva de f(x) correspondiente a la distribución χ^2 .

La Distribución F

Una variable aleatoria tiene distribución F con ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador si su función de densidad es;

$$f(x) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_2}{2}-1}\left(1+\frac{\nu_1 x}{\nu_2}\right)^{-(\nu_1+\nu_2)/2} \quad \text{para } x > 0.$$

Por ejemplo, si X es una variable aleatoria que sigue una distribución F con 5 grados de libertad en el numerador y 10 grados de libertad en el denominador entonces de excel se puede obtener que $P(X < 15) = 0.9998$ y $P(X < 6) = 0.9919$. Luego $P(6 < X < 15) = 0.9998 - 0.9919 = 0.0079$. Consultar en libros de estadística el aspecto que tiene la curva de f(x) correspondiente a la distribución F .