DISTRIBUCIONES COMUNMENTE USADAS

A continuación se presentan algoritmos de generación de variables aleatorias de distribuciones comúnmente usadas.

1. BERNOULLI

Esta es la más simple de las distribuciones discretas. Toma solo dos valores que se denotan como fracaso ($x = 0$) o éxito ($x = 1$), con probabilidades 1-p y p respectivamente.

Se usa para modelar la probabilidad de que un resultado sea de una clase especifica o tenga una característica especifica.

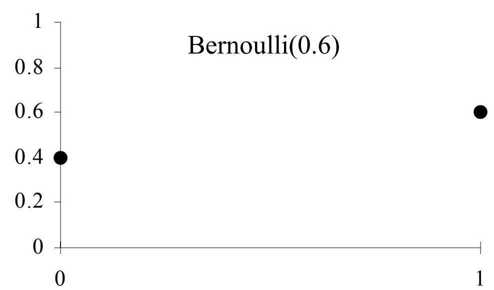
- Un sistema de computación esta funcionando o no.
- Un paquete en una red llego a su destino o no.

Esta distribución junto con sus derivadas, se puede usar solo si los ensayos son independientes e idénticamente distribuidos de forma tal que la probabilidad de éxito en cada ensayo sea p y no sea afectada por el resultado en ensayos anteriores.

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

Media: p

Varianza: $p(1-p)$



Generación:

Use transformación inversa. Genere $u \sim U(0,1)$. Si $u \le p$ retorne 1, de otra forma retorne 0.

2. BETA

Se usa para representar variables que están acotadas, por ejemplo, entre 0 y 1.

El rango de la variable puede ser cambiado por otro rango $[x_{\min}, x_{\max}]$ sustituyendo x en la ecuación siguiente por $(x - x_{\min}) / (x_{\max} - x_{\min})$.

Se usa para modelar:

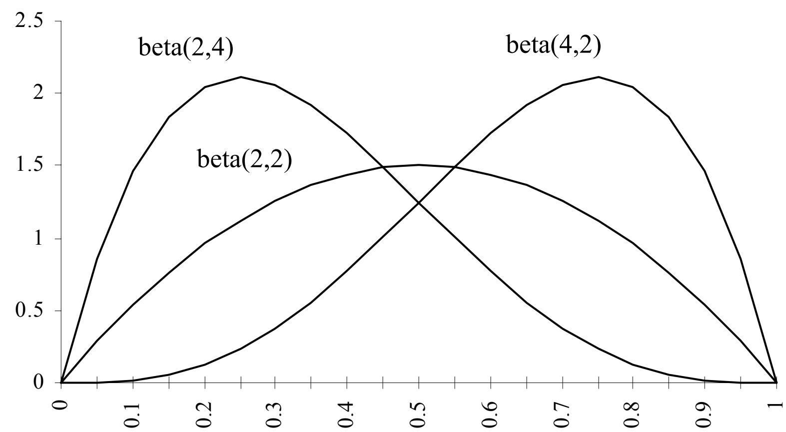
- La fracción de paquetes que requieren retransmisión
- La fracción de llamadas a procedimientos remotos que tardan más de cierto tiempo.

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)} \qquad 0 \le x \le 1,\ a > 0,\ b > 0$$

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Media: $a/(a+b)$

Varianza: $ab/[(a+b)^2(a+b+1)]$



Generación:

1. Genere dos gamas y tome la razón:

$$\text{beta}(a,b) = \frac{\gamma(1,a)}{\gamma(1,a) + \gamma(1,b)}$$

2. Si a y b son enteros:

   - Genere $a + b + 1$ números uniformes $U(0,1)$.
   - Retorne el a-esimo menor número como beta(a , b).

3. Si a y b son ambos menores que 1:
   - Genere u_1 y u_2 ambos $U(0,1)$.
   - Haga $x = u_1^{1/a}$ y $y = u_2^{1/b}$. Si $x + y > 1$ vaya al paso previo, de lo contrario retorne $x/(x + y)$ como el valor de beta(a, b).
4. Si a y b son ambos mayores que 1, un algoritmo basado en el método del rechazo puede ser fácilmente implementado.

3. BINOMIAL

Se usa para modelar el número de éxitos en una secuencia de n ensayos Bernoulli, por ejemplo:

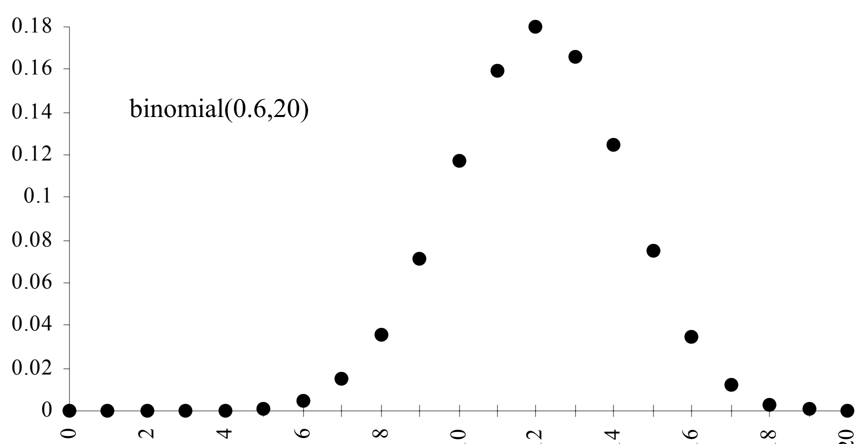
- El número de procesadores que están funcionando en un sistema multiprocesador.
- El número de paquetes que alcanzaron su destino sin errores.
- El número de elementos en una cola que tienen determinada característica.

La varianza de la binomial es siempre menor que la media. Si en las aplicaciones anteriores la varianza es mayor o igual a la media, se pueden usar distribuciones binomial negativa o Poisson.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,\ldots,n$$

Media: np

Varianza: $np(1-p)$



Generación:

1. Genere n $U(0,1)$. El cantidad de números aleatorios menor que p es binomial(n , p).Este es el método de composición basado en que la binomial es suma de variables Bernoulli.
2. Si p es pequeño, una técnica más rápida es:
   - Genere números aleatorios geométricos $G_i(p) = \left\lceil \ln(u_i) / \ln(1-p) \right\rceil$.
   - Si para el momento $\sum_{i=1}^{m} G_i(p) > n$, retorne m - 1, de lo contrario regrese al paso anterior.
3. Método de la transformación inversa: calcule la FDA $F(x)$ para x = 0, 1, ..., n y almacénela en un arreglo. Para cada variable binomial genere $u \sim U(0,1)$ y cheque el arreglo hasta encontrar x tal que $F(x) \le u < F(x + 1)$ y retorne x.

4. BINOMIAL NEGATIVA

El número de fracasos x antes del m-esimo éxito en una secuencia Bernoulli es binomial negativa: por ejemplo:

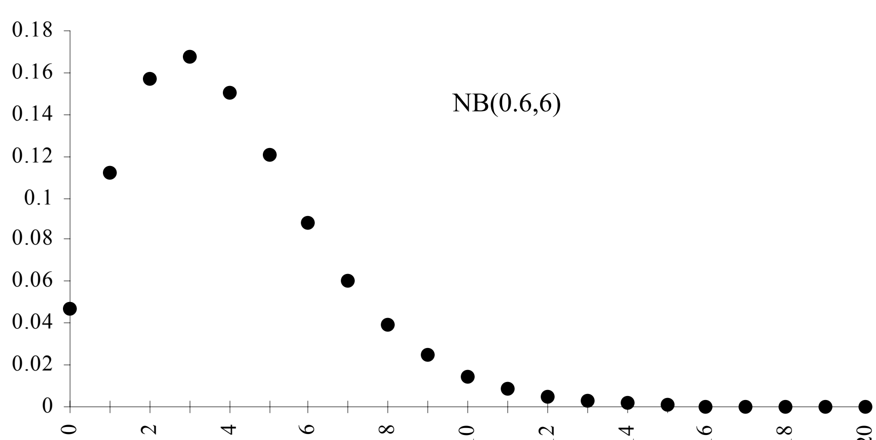
- Número de consultas locales a una base de datos antes de la m-esima consulta remota.
- Número de retransmisiones de un mensaje de m paquetes.

$$f(x) = \binom{m+x-1}{m-1} p^m (1-p)^x = \frac{\Gamma(m+x)}{\Gamma(m)\Gamma(x)} p^m (1-p)^x \qquad x = 0,1,2,\ldots,\infty; \qquad 0 < p < 1;$$

m entero positivo para la primera expersion y $m > 0$ en la segunda.

Media: $m(1-p)/p$

Varianza: $m(1-p)/p^2$



Generación:

1. Genere $u_i \sim U(0,1)$ hasta que m de los u_i sean menor o igual a p. Retorne el número de los u_i que fueron mayores que p como BN(p,m).
2. La suma de m geométricas $G(p)$ da el número total de ensayos para m éxitos, por lo tanto:

$$\text{BN}(p,m) \sim \left(\sum_{i=1}^{m} G(p)\right) - m$$

3. El siguiente método puede ser usado para valores de m enteros o no-enteros:
   a) Genere $y \sim \Gamma(p/(1\text{-}p),m)$.
   b) Genere $x \sim$ Poisson(y).
   c) Retorne x como BN(p,m).

5. CHI-CUADRADO

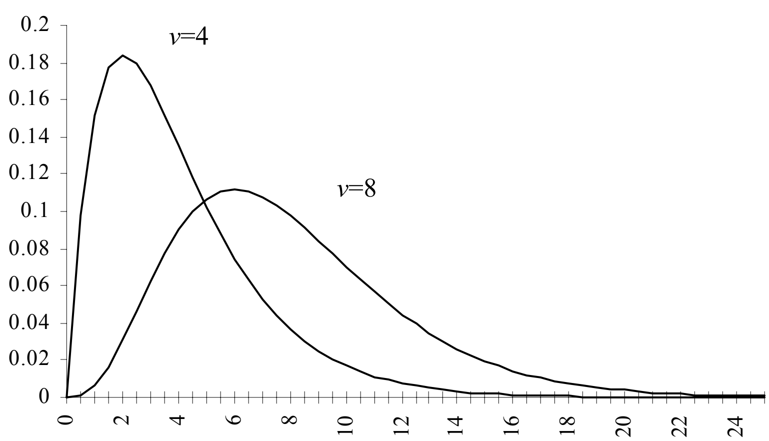
Se usa cuando tenemos una suma de cuadrados de normales estándar, por ejemplo, para modelar varianzas muéstrales.

$$f(x) = \frac{x^{(v-2)/2} e^{-x/2}}{2^{v/2}\Gamma(v/2)} \qquad 0 \le x < \infty$$

$$\Gamma(b) = \int_0^\infty e^{-x} x^{b-1} dx, \quad \Gamma(b+1) = b\Gamma(b), \quad \Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(b+1) = b! \text{ si } b = 0,1,2,...$$

Media: v

Varianza: $2v$



Generación:

1. El siguiente método se basa en el hecho de que la $\chi^2(v)$ es una $\gamma(2, v/2)$.
   - Para *v* par:

$$\chi^2(v) = -\frac{1}{2}\ln\left(\prod_{i=1}^{v/2} u_i\right)$$

   - Para *v* impar:

$$\chi^2(v) = \chi^2(v-1) + \left[N(0,1)\right]^2$$

2. Genere *v* $N(0,1)$ y retorne la suma de sus cuadrados.

6. ERLANG

Es generalmente usada en modelos de cola como una extensión de la exponencial cuando el coeficiente de variación (razón entre la desviación estándar y la media) es menor que 1, por ejemplo:

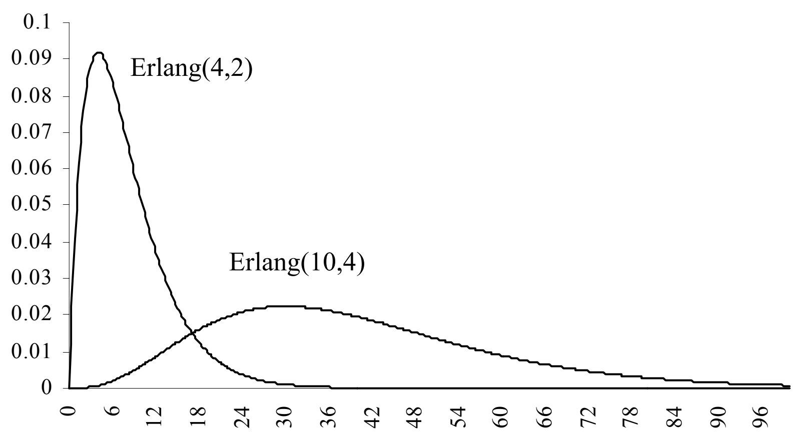
- Modelar tiempos de servicio: una taquilla con tiempo de servicios ~ Erlang(a , m) puede ser representada como *m* taquillas con tiempos de servicio exponenciales.
- Modelar el tiempo de reparación y el tiempo entre fallas.

$$f(x) = \frac{x^{m-1}e^{-x/a}}{(m-1)!a^m} \qquad 0 \le x < \infty, \quad a > 0, \quad m \text{ entero positivo.}$$

$$F(x) = 1 - e^{-x/a}\left[\sum_{i=0}^{m-1}\frac{(x/a)^i}{i!}\right]$$

Media: am

Varianza: a^2m



Generación:

Por convolución. Genere m $U(0,1)$ y calcule:

$$\text{Erlang}(a, m) \sim \text{-}a \ln\left(\prod_{i=1}^{m} u_i\right)$$

7. EXPONENCIAL

Es usada extensivamente en modelos de cola. Es la única distribución continua con la propiedad de perdida de memoria: recordar el tiempo desde el último evento no ayuda a predecir el tiempo hasta el próximo evento. Es usada para modelar el tiempo entre eventos sucesivos, por ejemplo:

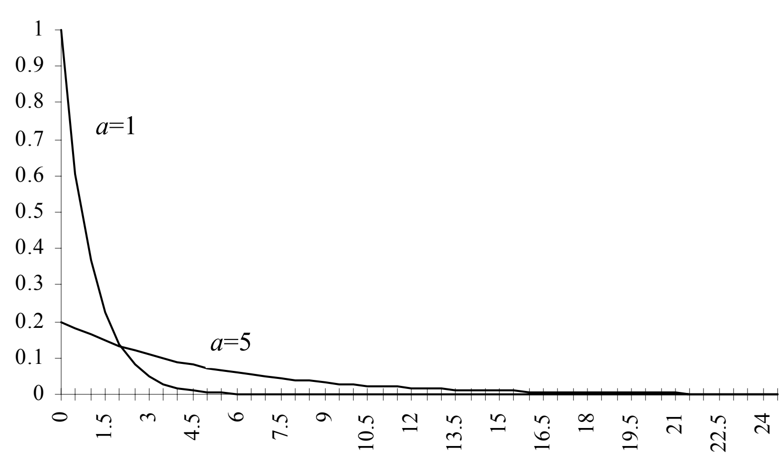
- El tiempo entre llegadas.
- El tiempo entre fallas.

$f(x) = \frac{1}{a} e^{-x/a} \qquad 0 \le x < \infty, \qquad \text{a} > 0.$

$F(x) = 1 - e^{-x/a}$

Media: a

Varianza: a^2



Generación:

Por transformación inversa. Genere $u \sim U(0,1)$ y retorne $-\ a \ln(u)$.

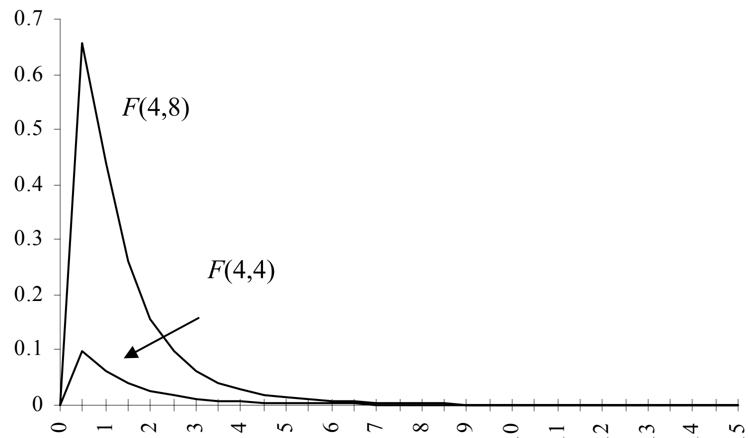
8. F

La F es la razón entre dos chi-cuadrado. Se usa para modelar la razón entre varianzas muéstrales como por ejemplo en la prueba-F en regresión y análisis de varianza.

$$f(x) = \frac{(n/m)^{n/2}}{\beta(n/2, m/2)} x^{(n-2)/2} \left(1 + \frac{n}{m}x\right)^{-(n+m)/2} \qquad 0 \le x < \infty, \qquad n \text{ y } m \text{ enteros positivos.}$$

Media: $\dfrac{m}{m-2}$ y $m > 2$

Varianza: $\dfrac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ y $m > 4$



Generación:

Por caracterizaron. Genere dos chi-cuadrados $\chi^2(n)$ y $\chi^2(m)$ y calcule:

$$F(n,m) = \frac{\chi^2(n)/n}{\chi^2(m)/m}$$

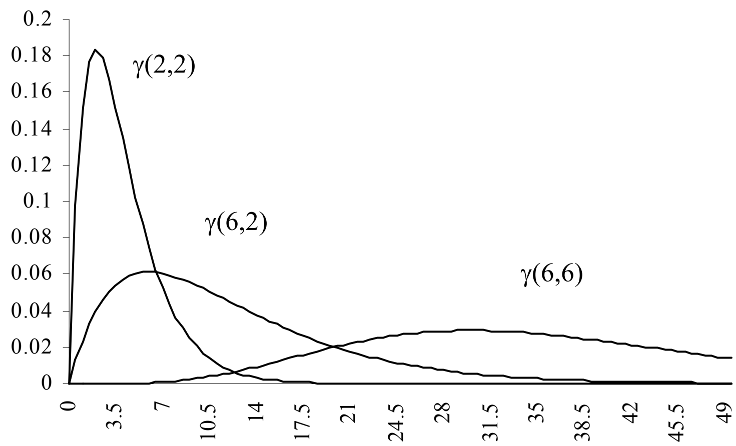
9. GAMMA

Es una generalización de la Erlang y tiene parámetros no enteros. Se usa en modelos de colas para modelar tiempos de servicio y tiempos de reparación.

$$f(x) = \frac{(x/a)^{b-1} e^{-x/a}}{a\Gamma(b)} \qquad 0 \le x < \infty, \quad a > 0, \quad b > 0.$$

Media: ab

Varianza: a^2b



Generación:

1. Si b es un entero, la suma de b exponenciales es una gamma. Si $u_i \sim U(0,1)$:

$$\gamma(a,b) \sim -a \ln\left[\prod_{i=1}^{b} u_i\right]$$

2. Para $b < 1$, genere una beta $x \sim$ beta(b, 1-b) y una exponencial $y \sim$ exp(1). El producto axy es γ(a,b).

3. Para otros valores de b, se generan dos gamma y se suman como se muestra a continuación:

$$\gamma(a,b) \sim \gamma(a,\lfloor b \rfloor) + \gamma(a, b - \lfloor b \rfloor) \qquad \text{donde } \lfloor b \rfloor \text{ es el mayor entero menor o igual a } b.$$

10. GEOMÉTRICA

El número de ensayos hasta e incluyendo el primer éxito en una secuencia Bernoulli es una geométrica. Es la equivalente discreta de la exponencial en cuanto a la propiedad de perdida de memoria: recordar el pasado no ayuda a predecir el futuro.

Se usa para modelar el número de intentos entre fallas (o éxitos) sucesivas:

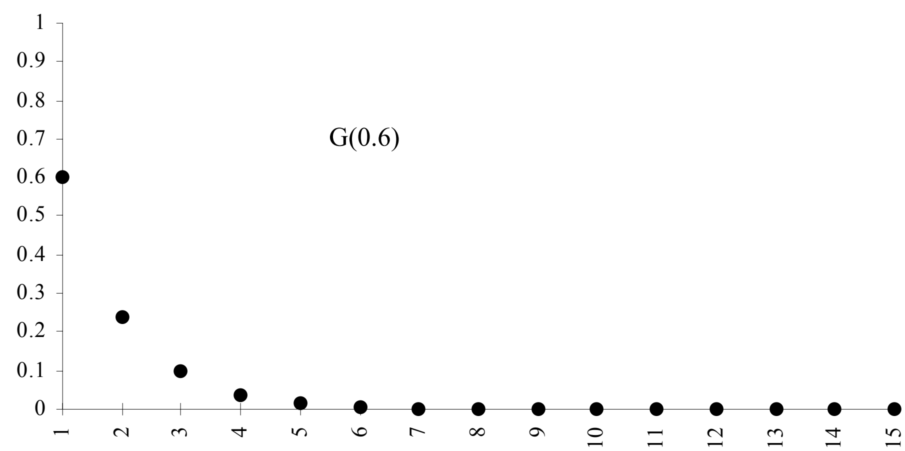
- El número de consultas locales a una base de datos entre accesos sucesivos a una base de datos remota.
- El número de paquetes trasmitidos exitosamente entre aquellos que requieren retransmisión.

$f(x) = (1-p)^{x-1} p \quad x = 1,2,...,\infty \quad 0 < p < 1$

$F(x) = 1 - (1-p)^x$

Media: $1/p$

Varianza: $\dfrac{1-p}{p^2}$



Generación:

Por transformación inversa. Genere $u \sim U(0,1)$ y calcule:

$$G(p) = \left\lceil \frac{\ln(u)}{\ln(1-p)} \right\rceil$$ donde $\lceil x \rceil$ denota el menor entero mayor o igual a x.

11. LOGNORMAL

Es el logaritmo de una normal. Se usa frecuentemente en modelos de regresión y análisis de experimentos cuando donde se aplican transformaciones logarítmicas.

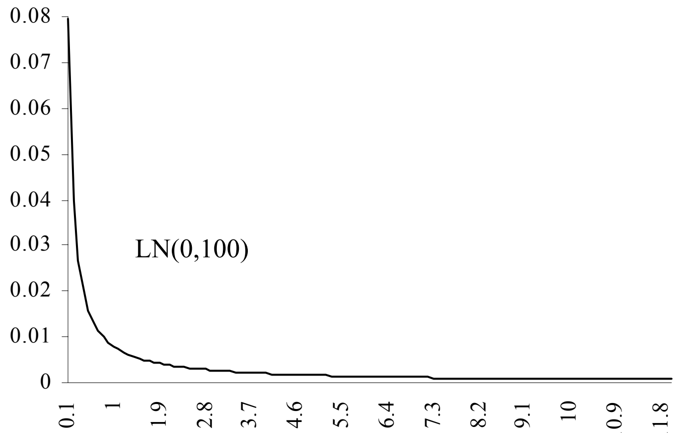
El producto de un gran número de variables aleatorias positivas tiende a la lognormal. Por lo tanto, también se usa para modelar errores que son el producto de efectos de un gran número de factores.

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-(\ln x-\mu)^2/2\sigma^2} \qquad 0 < x < \infty, \quad \mu \text{ y } \sigma > 0.$$

μ y σ son la media y la desviacion de $\log(x)$ y NO de x.

Media: $e^{\mu+\sigma^2/2}$

Varianza: $e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$



Generación:

Genere $x \sim N(0,1)$ y retorne $e^{\mu+\sigma x}$.

12. NORMAL

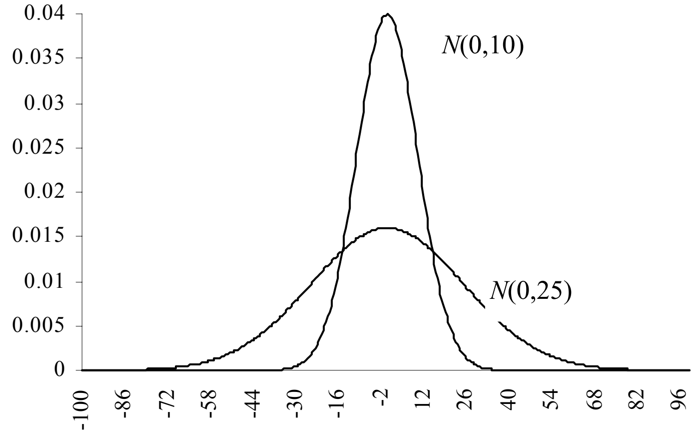
La $N(0,1)$ es llamada **normal unitaria** o **normal estándar**. Se usa cuando la aleatoriedad es causada por varias fuentes independientes actuando en forma aditiva:

- Errores de medición
- Medias muéstrales de un gran número de observaciones independientes de una distribución dada.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty, \quad \sigma > 0$$

Media: μ

Varianza: σ^2



Generación:

1. *Convolución*. Genere un número de $u_i \sim U(0,1)$ y calcule:

$$N(\mu,\sigma) \sim \mu + \sigma \frac{\left(\sum_{i=1}^{n} u_i\right) - n/2}{\left(n/12\right)^{1/2}}$$

   Generalmente se usa $n = 12$.

2. *Método de Box-Muller*. Genere dos uniformes u_1 y u_2 y calcule dos variables $N(\mu,\sigma)$ independientes usando:

$$x_1 = \mu + \sigma \cos(2\pi u_1)\sqrt{-2\ln(u_2)}$$
$$x_2 = \mu + \sigma \sin(2\pi u_1)\sqrt{-2\ln(u_2)}$$

3. *Método Polar*.

   a) Genere dos uniformes u_1 y u_2 $U(0,1)$.

   b) Haga $v_1 = 2u_1 - 1$, $v_2 = 2u_2 - 1$, y $r = v_1^2 + v_2^2$.

   c) Si $r \geq 1$, vaya a al paso 3a; de lo contrario haga $s = \left[(-2\ln r) / r \right]^{1/2}$ y retorne:

$$x_1 = \mu + \sigma\, v_1 s$$
$$x_2 = \mu + \sigma\, v_2 s$$

   como dos $N(\mu,\sigma)$ independientes.

4. *Método del rechazo*.

   a) Genere dos uniformes u_1 y u_2 $U(0,1)$.

   b) Haga x = -ln u_1.

   c) Si $u_2 > e^{-(x-1)^2/2}$, regrese a 4a.

   d) Genere u_3.

   e) Si $u_3 > 0.5$, retorne $\mu + \sigma x$; de lo contrario retorne $\mu - \sigma x$.

13. PARETO

La F(x) de Pareto es una curva que fácilmente ajusta las observaciones. Dada una muestra de tamaño n $\{x_1, ..., x_n\}$, el estimador máximo verosímil del parámetro a es:

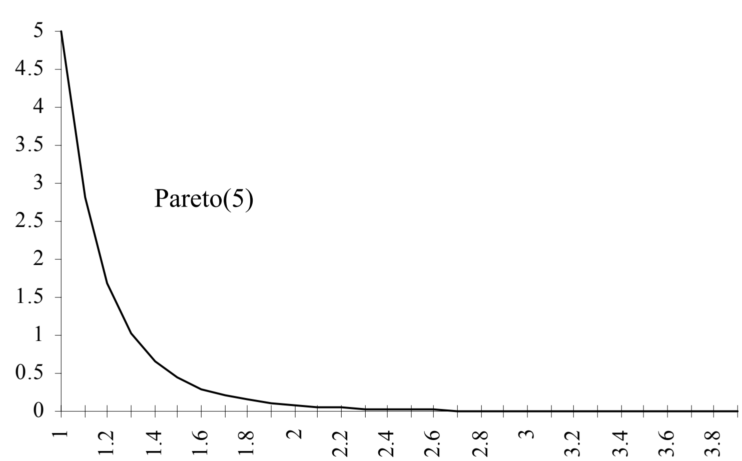
$$a = \frac{1}{\left(\frac{1}{n}\right)\sum_{i=1}^{n} \ln x_i}$$

$f(x) = ax^{-(a+1)} \quad 1 \le x < \infty, \quad a > 0$

$F(x) = 1 - x^{-a}$

Media: $\dfrac{a}{a-1}$, para $a > 1$

Varianza: $\dfrac{a}{(a-1)^2(a-2)}$, para $a > 2$



Generación:

Por transformación inversa: Genere $u \sim U(0,1)$ y retorne $1/u^{1/a}$.

14. PASCAL

Es una extensión de la geométrica. En una secuencia de ensayos Bernoulli, el número de ensayos hasta e incluyendo el *m*-esimo éxito tiene distribución de Pascal.

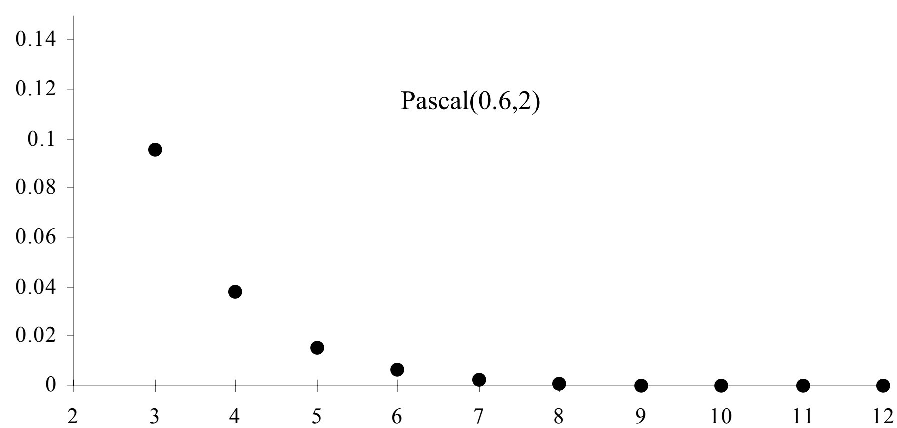
Es útil para modelar el número de intentos para obtener cierto número de éxitos:

- Número de intentos para transmitir un mensaje de *m* paquetes.
- Número de bits a enviar para recibir exitosamente una señal de *m* bits.

$$f(x) = \binom{x-1}{m-1} p^m (1-p)^{x-m}, \quad x = m, m+1, \ldots, \infty; \quad 0 < p < 1; \quad m \text{ entero positivo.}$$

Media: m/p.

Varianza: $m(1-p)/p^2$.



Generación:

Genere *m* geométricas *G*(*p*) y retorne la suma como una Pascal(*p* , *m*).

15. POISSON

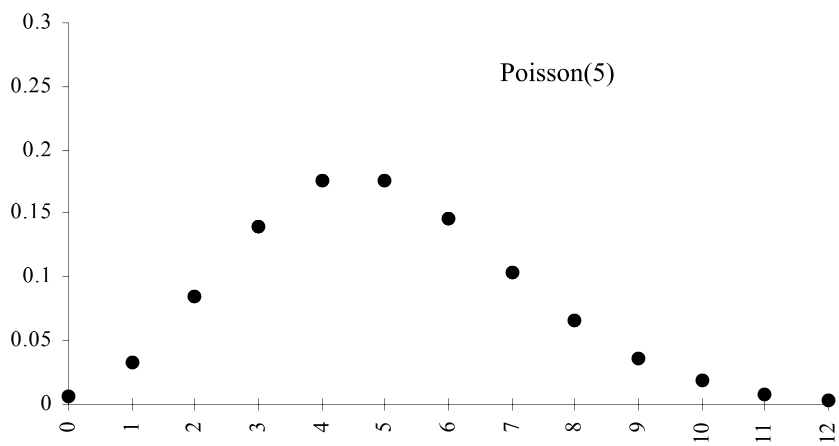
Se usa extensivamente en modelos de colas para modelar el número de llegadas en cierto intervalo:

- Número de consultas a un servidor en un intervalo *t*.
- Número de fallas en componentes por unidad de tiempo.
- Número de consultas a una base de datos en *t* segundos.
- Número de errores de tipeo por forma.

$$f(x) = \lambda^x \frac{e^{-\lambda}}{x!}, \quad x = 0,1,\ldots,\infty; \quad \lambda > 0.$$

Media: λ.

Varianza: λ.



Generación:

1. *Transformación inversa*. Calcule la FDA $F(x)$ para $x = 0,1,...$ hasta un número adecuadamente grande y guárdela en un arreglo. Para cada variable Poisson, genere $u \sim U(0,1)$ y encuentre el x tal que $F(x) \le u < F(x+1)$; retorne x.
2. Comenzando con $n = 0$, genere $u_n \sim U(0,1)$ y calcule el producto $\prod_{i=0}^{n} u_i$. A lo que no más el producto se vuelva menor que $e^{-\lambda}$, retorne n como la variable Poisson(λ). Note que n es tal que $u_0 u_1 \ldots u_{n-1} > e^{-\lambda} \ge u_0 u_1 \ldots u_n$. En promedio se requieren λ+1 variable uniformes para generar una Poisson.

3. t DE STUDENT

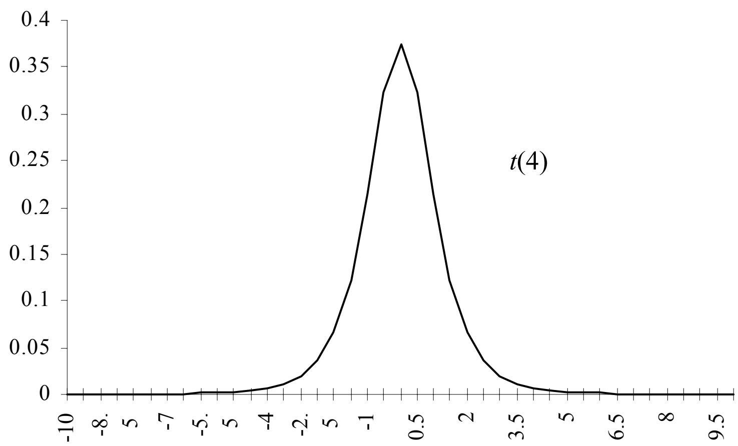
Se aplica cundo se tiene la razón entre una normal y la raíz de una chi-cuadrado y comúnmente se usa en el calculo de intervalos de confianza. Si $x \sim N(0,1)$ y $y \sim \chi^2(v)$, entonces $x/\sqrt{y/v}$ tiene distribución t con v grados de libertad.

$$\frac{N(0,1)}{\sqrt{\chi^2(v)/v}} \sim t(v)$$

La $f(x)$ de la t es muy similar a la de la normal estándar: tiene forma de campana y es simétrica respecto a cero. Para grados de libertad grandes (v>30), la t se puede aproximar por la normal estándar.

$$f(x) = \frac{\Gamma\left[(v+1)/2\right]\left[1+(x^2/v)\right]^{-(v+1)/2}}{(\pi v)^{1/2}\Gamma(v/2)}, \quad -\infty < x < \infty, \quad v \text{ entero positivo.}$$

Varianza: $v/(v-2)$, para $v > 2$.



Generación:

Por caracterizaron. Genere $x \sim N(0,1)$, $y \sim \chi^2(v)$, y retorne $x/\sqrt{y/v}$ como $t(v)$.

17. UNIFORME (CONTINUA)

Es comúnmente usada cuando una variable esta acotada y no se tiene más información de ella:

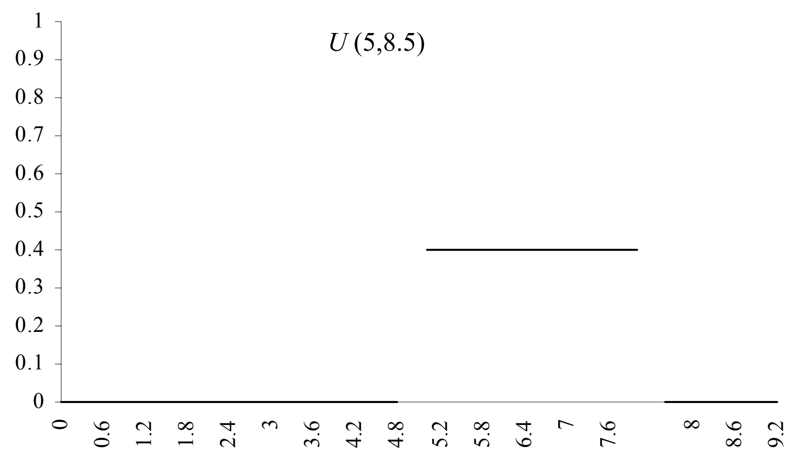
- Distancia entre origen y destinos de mensajes en una red.
- Tiempo de acceso en un disco.

$$f(x) = \frac{1}{b-a}, \quad a \le x \le b, \quad b > a.$$

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } a \le x < b \\ 1 & \text{si } b \le x \end{cases}$$

Media: $\dfrac{a+b}{2}$

Varianza: $(b-a)^2 / 12$



Generación:

Genere $u \sim U(0,1)$ y retorne $a + (b - a)u$.

18. UNIFORME (DISCRETA)

Toma un número finito de valores, todos con la misma probabilidad. Se usa cuando se cree que los valores sobre un intervalo son equiprobables:

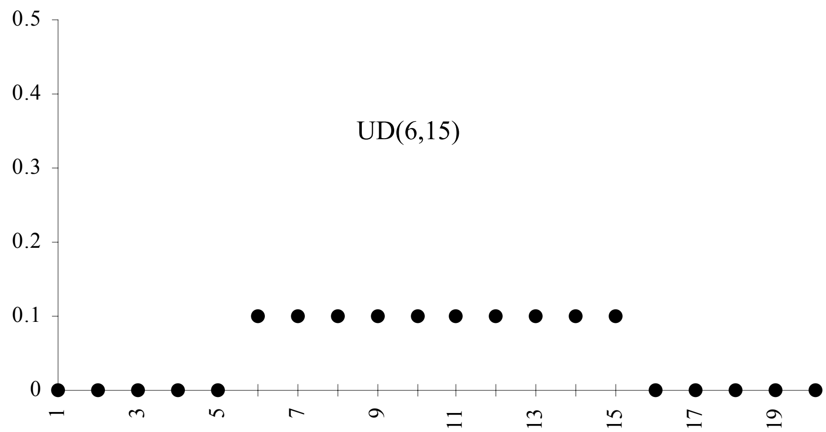
- Número de pistas a accesar en un disco.
- El número del dispositivo de entrada/salida seleccionado para la próxima operación.
- El nodo de origen y destino del próximo paquete en una red.

$$f(x) = \frac{1}{n-m+1}, \quad x = m, m+1, \ldots, n; \quad m \text{ y } n \text{ enteros y } n > m.$$

$$F(x) = \begin{cases} 0 & \text{si } x < m \\ \dfrac{x-m+1}{n-m+1} & \text{si } m \le x < n \\ 1 & \text{si } n \le x \end{cases}$$

Media: $(n + m)/2$

Varianza: $\dfrac{(n-m+1)^2 - 1}{12}$



Generación:

Genere $u \sim U(0,1)$ y retorne $\lfloor m + (n - m + 1)u \rfloor$.

19. WEIBULL

Es usada comúnmente en análisis de confiabilidad. Se usa para modelar tiempo de vida de componentes. Si $b < 1$, la tasa de fallas incrementa con el tiempo. Si $b > 1$, la tasa de fallas decrementa con el tiempo. Si $b = 1$, la tasa de falla es constante y los tiempos de vida son exponenciales.

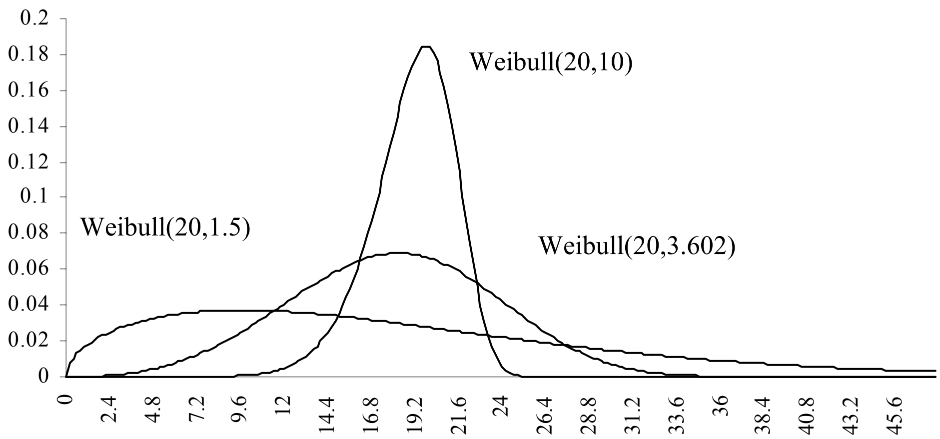
Si $a = 3.602$, la Weibull se aproxima a la normal. Para $a > 3.602$, tiene una larga cola a la izquierda. Para $a < 3.602$, la cola es para la derecha. Para $b \le 1$, tiene forma de L. Para $b > 1$, tiene forma de campana.

$$f(x) = \frac{bx^{b-1}}{a^b} e^{-(x/a)^b}, \quad 0 \le x < \infty, \quad a > 0, \quad b > 0.$$

$$F(x) = 1 - e^{-(x/a)^b}$$

Media: $\frac{a}{b}\Gamma(1/b)$

Varianza: $\frac{a^2}{b^2}\left\{2b\Gamma(2/b) - \left[\Gamma(1/b)\right]^2\right\}$



Generación:

Por transformación inversa. Genere $u \sim U(0,1)$ y retorne $a(\text{-ln } u)^{1/b}$ como la Weibull(a , b).