5. Modelos probabilísticos de fenómenos aleatorios continuos
Objetivo: El alumno aplicará algunas de las distribuciones más utilizadas en la práctica de la ingeniería y elegir la más adecuada para analizar algún fenómeno aleatorio continuo en particular.

Contenido:
5.1 Distribuciones continuas, distribución uniforme continua, cálculo de su media y varianza, generación de números aleatorios y el uso de paquetería de cómputo para la generación de números aleatorios con distribución discreta o continua, utilizando el método de la transformación inversa.
5.2 Distribución Gamma, sus parámetros, momentos y funciones generatrices, distribución exponencial, sus parámetros, momentos y funciones generatrices.
5.3 Distribuciones normal y normal estándar, uso de tablas de distribución normal estándar, la aproximación de la distribución binomial a la distribución normal.
5.4 Distribuciones Chi-Cuadrada, T de Student, F de Fisher, Weibull y distribución Log normal, como modelos teóricos para la estadística aplicada, sus parámetros, momentos y funciones generatrices.

TEMA 5

Modelos probabilísticos de fenómenos aleatorios continuos

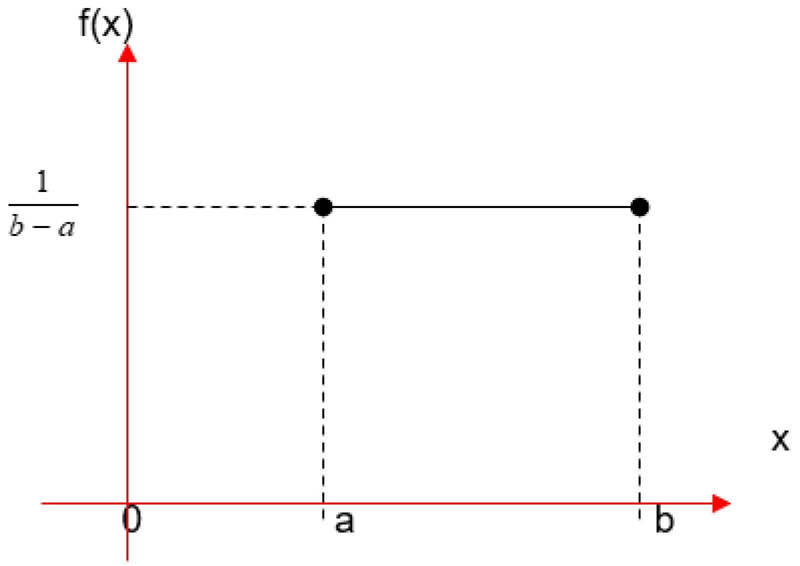
Objetivo: El alumno aplicará algunas de las distribuciones más utilizadas en la práctica de la ingeniería y elegirá la más adecuada para analizar algún fenómeno aleatorio continuo en particular.

5.1 Distribuciones continuas, distribución uniforme continua, cálculo de su media y varianza, generación de números aleatorios y el uso de paquetería de cómputo para la generación de números aleatorios con distribución discreta o continua, utilizando el método de la transformación inversa.

DISTRIBUCIÓN UNIFORME CONTINUA O RECTANGULAR.

Su función densidad de probabilidad es:

$$f(x)\begin{cases} \dfrac{1}{b-a} & a \leq x \leq \mathrm{b} \\ 0 & C.O.C. \end{cases}$$



Media o Valor Esperado:

$$\mu_x = F(x) = \int_a^b xf(x) = \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)}$$

$$\boxed{\mu_x = \frac{a+b}{2}} \Rightarrow \text{ Media o valor esperado}$$

Varianza

$$\sigma_x^2 = E[x^2] - E^2[x]$$

$$E[x^2] = \frac{1}{b-a}\int_a^b x^2dx = \frac{1}{b-a}\left.\frac{x^3}{3}\right|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

$$\sigma_x^2 = \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{b^2+a^2-2ab}{12} = \boxed{\frac{(b-a)^2}{12}}$$

$$\boxed{\sigma_x = \frac{b-a}{\sqrt{12}}}$$

SIMULACIÓN

Simulación es el proceso de diseñar un modelo de un sistema real y desarrollar experimentos con este modelo con el propósito de entender el comportamiento del sistema o de evaluar varias estrategias (dentro de los límites impuestos por un criterio o conjunto de criterios) para la operación del sistema.

Aunque la construcción de modelos arranca desde el Renacimiento, el uso de la palabra simulación data de 1949, cuando los científicos Von Neuman y Ulam que trabajaban en el proyecto Monte Carlo, durante la Segunda Guerra Mundial, resolvieron problemas de reacciones nucleares cuya solución experimental sería muy cara y el análisis matemático demasiado complicado.

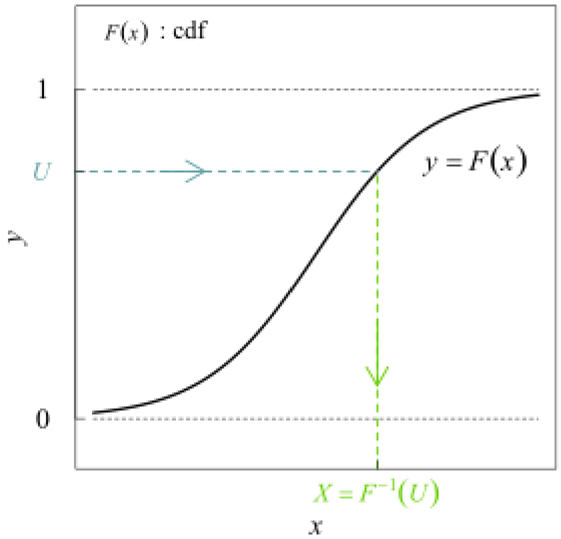
Definición de simulación:
Thomas H. Naylor la define así: Simulación es una técnica numérica para conducir experimentos en una computadora digital. Estos experimentos comprenden ciertos tipos de relaciones matemáticas y lógicas, las cuales son necesarias para describir el comportamiento y la estructura de sistemas complejos del mundo real a través de largos períodos de tiempo.

Modelado de simulación:
Es una metodología que busca:
Describir el comportamiento del sistema
Construir teorías o probar hipótesis con base en el comportamiento observado.
Usar estas teorías para predecir comportamientos futuros.

Ventajas de la simulación:
Se recomienda usar la simulación cuando existan una o más de las siguientes condiciones:
No haya una formulación matemática completa del problema o los métodos de solución del modelo matemático no hayan sido desarrollados. Muchos modelos de líneas de espera están en esta categoría.
Los métodos analíticos están disponibles, pero los procedimientos matemáticos son tan complejos y tediosos que la simulación proporciona un método de solución más simple.

Las soluciones analíticas existen y son posibles pero la habilidad matemática del personal disponibles no es suficiente.
El costo de diseñar, probar y correr una simulación deberá evaluarse contra el costo de obtener ayuda externa.
Si se desea observar una historia simulada del proceso en un período de tiempo para estimar ciertos parámetros.
La simulación puede ser la única posibilidad debido a la dificultad en conducir experimentos y observar fenómenos en su medio ambiente real. Por ejemplo: daños ecológicos, plagas, vehículos espaciales, etc.
El manejo de experimentos en tiempo acelerado o retardado según la naturaleza del fenómeno.
Una ventaja adicional de la simulación es su potencial educativo y su aplicación en la capacitación.

Método de la transformada inversa



El **método de la transformada** (o **transformación**) **inversa**, también conocido como **método de la inversa de la transformada**,[1] es un método para la generación de números aleatorios de cualquier distribución de probabilidad continua cuando se conoce la inversa de su función de distribución (cdf). Este método es en general aplicable, pero puede resultar muy complicado obtener una expresión analítica de la inversa para algunas distribuciones de probabilidad. El método de Box-Muller es un ejemplo de algoritmo que aunque menos general, es más eficiente desde el punto de vista computacional.

El método se utiliza para simular valores de las distribuciones normal, binomial, Poisson, exponencial, Cauchy, triangular, de Pareto y Weibull.

Obtención del método

El método de la transformada inversa se basa en el siguiente teorema:

Teorema de inversión. Sea X una variable aleatoria con función de distribución de probabilidad acumulada F, continua e invertible, y sea F^{-1} su función inversa. Entonces, la variable aleatoria $U = F(X)$

tiene distribución uniforme en $(0,1)$. Como consecuencia, si U es una variable aleatoria uniforme en$(0,1)$ entonces la variable aleatoria $X = F^{-1}(U)$ satisface la distribución F.

El método

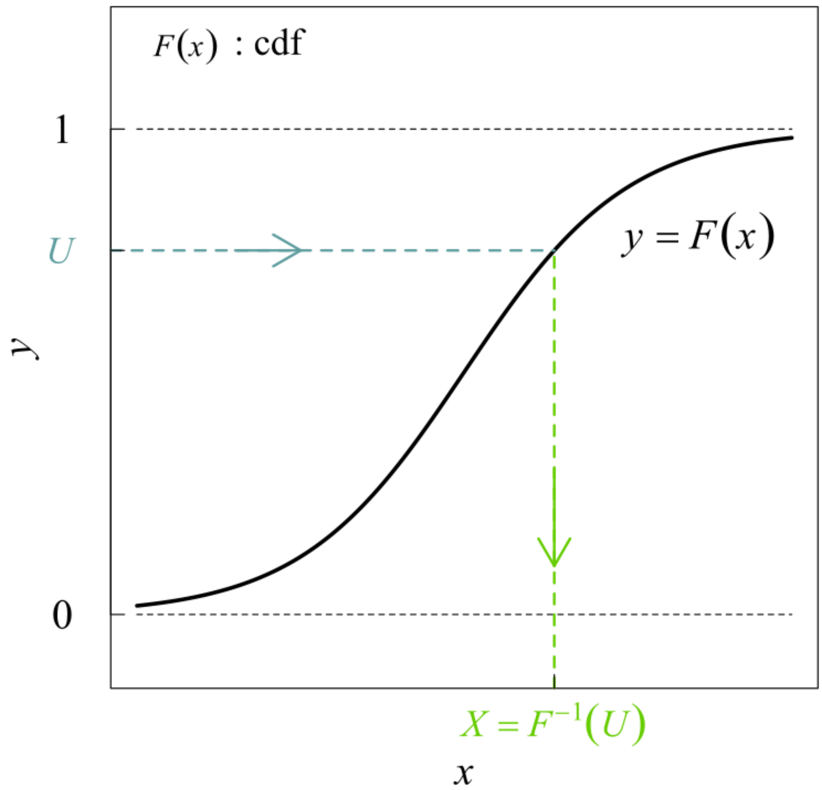
El problema que resuelve el método de la transformada inversa es el siguiente:

- Sea X una variable aleatoria cuya distribución puede ser descrita por la cdf F. (función de probabilidad acumulada)
- Se desea generar valores de X que están distribuidos según dicha distribución.

Numerosos lenguajes de programación poseen la capacidad de generar números pseudo-aleatorios que se encuentran distribuidos de acuerdo con una distribución uniforme estándar. Si una variable aleatoria posee ese tipo de distribución, entonces la probabilidad de que el número caiga dentro de cualquier subintervalo (a, b) del intervalo entre 0 a 1 es la longitud del subintervalo, o sea $b - a$.

El método de la transformada inversa funciona de la siguiente manera:

1. Se genera un número aleatorio a partir de la distribución uniforme estándar; se lo llama u.
2. Se calcula el valor x tal que $F(x) = u$; y se le llama x_{elegido}.
3. Se toma x_{elegido} como el número aleatorio extraído de la distribución caracterizada por F.



Demostración del teorema

Sea

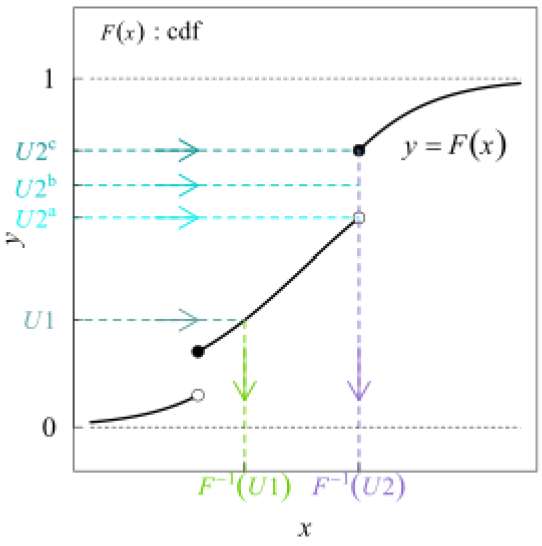
$F^{-1}(u) = \inf\{x \mid F(x) = u, 0 < u < 1\}$

$\Pr(F^{-1}(U) \leq x)$

$= \Pr(\inf\{x \mid F(x) = U\} \leq x)$ (por definición de F^{-1})

$= \Pr(U \leq F(x))$ (aplicando F , que es monótona, a ambos lados)

$= F(x)$ (porque $\Pr(U \leq y) = y$, dado que U es uniforme en el intervalo unitario)



Ejemplos de obtención de números aleatorios con diferentes distribuciones de probabilidad utilizando Excel.

5.2 Distribución Gamma, sus parámetros, momentos y funciones generatrices, distribución exponencial, sus parámetros, momentos y funciones generatrices.

Función Gamma

Ahora estudiaremos una función conocida como la función gamma $\Gamma(x)$, la cual es de gran importancia en análisis y en aplicaciones. Esta función se define en términos de una integral impropia, la cual no puede calcularse en términos de funciones elementales.

Definición [Función Gamma]

La función $\Gamma:[0,\infty]\to\mathbb{R}$ dada por

$$\Gamma(x)=\int_0^{\infty} e^{-t}t^{x-1}dt$$

se conoce como la función gamma. Su gráfica se muestra en la figura 1.

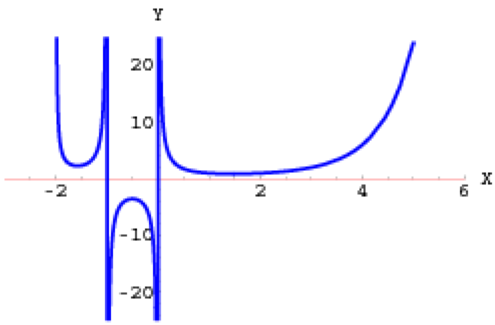


Figura 1.

El siguiente teorema establece una de las propiedades más importantes de la función gamma.

Teorema [Recursividad de gamma]

Para toda $x>0$ se tiene que

$$\Gamma(x+1)=x\Gamma(x)$$

o bien $\Gamma(x)=(x-1)\Gamma(x-1)$

Demostración

Integrando por partes

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$$

$$= \underbrace{\left. \frac{e^{-t} t^x}{x} \right|_0^\infty}_{0} + \frac{1}{x} \int_0^\infty e^{-t} t^x dt$$

$$= \frac{1}{x} \Gamma(x+1)$$

Ejemplo

Calcule $\Gamma\left(\frac{1}{2}\right)$

Solución

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \underbrace{t^{-\frac{1}{2}} e^{-t}}_{u^2=t} dt$$

$$= 2\int_0^\infty e^{-u^2} du$$

$$= \sqrt{\pi}$$

El resultado anterior puede generalizarse, como muestra en el siguiente corolario.

Corolario [Recursividad de Gamma]

Para $x > 0$ $p = 1,2, \ldots$ y $n = 1,2, \ldots$ se tiene que

$$\Gamma(x+p) = (x+p-1)(x+p-2)\ldots x\Gamma(x)$$

Si $x = n$ donde n es entero y positivo y $p = 1$
entonces $\Gamma(n+1) = n(n-1)(n-2)\cdots(2)(1)n\Gamma(1) = n!\,\Gamma(1)$

$$\Gamma(1) = \int_0^\infty e^{-t} dt = \lim_{k\to\infty} -e^{-t}\Big|_0^k = -0-(-1) = 1$$

Combinando estas dos relaciones se obtiene que el factorial de un número entero y positivo es un caso especial de la función Gamma:

$$\Gamma(n+1) = n!$$

Observación: La función $\Gamma(n+1) = n!$, se le conoce como el factorial generalizado. Se puede emplear también: $\Gamma(n) = (n-1)!$

Ejemplo

Calcular los valores de $\Gamma\left(\frac{3}{2}\right), \Gamma\left(\frac{5}{2}\right), \Gamma\left(-\frac{1}{2}\right)$

Solución

Usando la propiedad recursiva, tenemos que:

Para $\Gamma\left(\frac{3}{2}\right)$:

$$\Gamma\left(\frac{3}{2}\right) = \Gamma\left(\frac{1}{2}+1\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

- Para $\Gamma\left(\frac{5}{2}\right)$:

$$\begin{aligned}\Gamma\left(\frac{5}{2}\right) &= \Gamma\left(\frac{3}{2}+1\right)\\ &= \frac{3}{2}\Gamma\left(\frac{3}{2}\right)\\ &= \frac{3}{2}\frac{\sqrt{\pi}}{2}\\ &= \frac{3\sqrt{\pi}}{4}\end{aligned}$$

- Para $\Gamma\left(-\frac{1}{2}\right)$:

$$\Gamma\left(\frac{1}{2}\right) = \Gamma\left(-\frac{1}{2}+1\right)$$

$$\sqrt{\pi} = -\frac{1}{2}\Gamma\left(-\frac{1}{2}\right)$$

De donde

$$\Gamma\left(-\frac{1}{2}\right) = -2\sqrt{\pi}$$

Valores de la función Gamma

$$\begin{array}{lll}
\Gamma(-3/2) & = \frac{4\sqrt{\pi}}{3} & \approx 2.363 \\
\Gamma(-1/2) & = -2\sqrt{\pi} & \approx -3.545 \\
\Gamma(1/2) & = \sqrt{\pi} & \approx 1.772 \\
\Gamma(1) & = 0! & = 1 \\
\Gamma(3/2) & = \frac{\sqrt{\pi}}{2} & \approx 0.886 \\
\Gamma(2) & = 1! & = 1 \\
\Gamma(5/2) & = \frac{3\sqrt{\pi}}{4} & \approx 1.329 \\
\Gamma(3) & = 2! & = 2 \\
\Gamma(7/2) & = \frac{15\sqrt{\pi}}{8} & \approx 3.323 \\
\Gamma(4) & = 3! & = 6
\end{array}$$

La distribución Gamma

Es una distribución adecuada para modelar el comportamiento de variables aleatorias continuas con asimetría positiva. Es decir, variables que presentan una mayor densidad de sucesos a la izquierda de la media que a la derecha. En su expresión se encuentran dos parámetros, siempre positivos, (α) y (β) de los que depende su forma y alcance por la derecha, y también la función Gamma Γ(α), responsable de la convergencia de la distribución.

Los parámetros de la distribución

El primer parámetro (α) sitúa la máxima intensidad de probabilidad y por este motivo en algunas fuentes se denomina "la forma" de la distribución: cuando se toman valores próximos a cero aparece entonces un dibujo muy similar al de la distribución exponencial. Cuando se toman valores más grandes de (α) el centro de la distribución se desplaza a la derecha y va apareciendo la forma de una campana de Gauss con asimetría positiva. Es el segundo parámetro (β) el que determina la forma o alcance de esta asimetría positiva desplazando la densidad de probabilidad en la cola de la derecha. Para valores elevados de (β) la distribución acumula más densidad de probabilidad en el extremo derecho de la cola, alargando mucho su dibujo y dispersando la probabilidad a lo largo del plano. Al dispersar la probabilidad la altura máxima de densidad de probabilidad se va reduciendo; de aquí que se le denomine "escala". Valores más pequeños de (β) conducen a una figura más simétrica y concentrada, con un pico de densidad de probabilidad más elevado.

Una forma de interpretar (β) es "tiempo promedio entre ocurrencia de un suceso". Relacionándose con el parámetro de la Poisson como β=1/λ. Alternativamente λ será la razón de ocurrencia: λ=1/β.
La expresión β^{-1} también será necesaria más adelante para poder llevar a cabo el desarrollo matemático.

Relación con otras distribuciones

Si se tiene un parámetro α de valores elevados y β pequeña, entonces la función Gamma converge con la distribución normal. De media $\mu = \alpha * \beta$, y varianza $\sigma^2 = \alpha * \beta^2$.
Cuando $\alpha = 1$ y β= 0 la distribución Gamma es exactamente la distribución exponencial con parámetro (α=1).
Cuando la proporción entre parámetros es $[\alpha = \frac{v}{2} : \beta = v]$ entonces la variable aleatoria se distribuye como una Chi-cuadrado con v grados de libertad.
Si α=1, entonces se tiene la distribución exponencial negativa de parámetro λ=1/β.

Ventajas

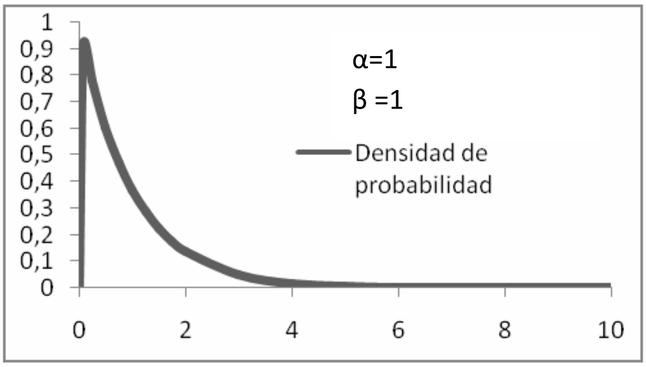
De esta forma, la distribución Gamma es una distribución flexible para modelar las formas de la asimetría positiva, de las más concentradas y puntiagudas, a las más dispersas y achatadas. Como ejemplos de variables que se comportan así:

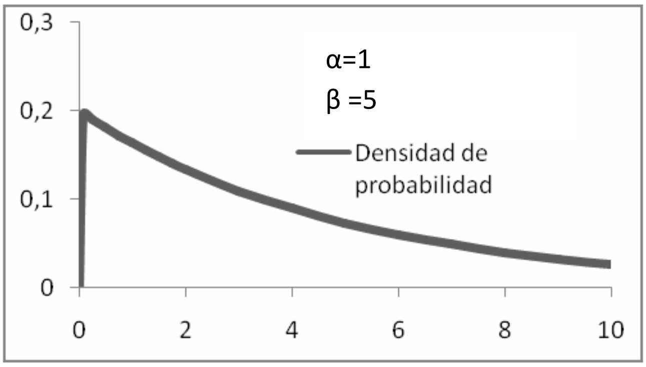
- Número de individuos involucrados en accidentes de tráfico en el área urbana: es más habitual que la mayoría de los reportes abiertos den la proporción de 1 herido por vehículo, que otras proporciones superiores.
- Altura a la que se inician las precipitaciones; sucede de forma más habitual precipitaciones iniciadas a una altura baja, que iniciadas a gran altitud.
- Tiempo o espacio necesarios para observar X sucesos que siguen una distribución de Poisson.
- Distribución de la finura de fibras de lana: la mayoría presentan una menor finura que unas pocas fibras más gruesas.

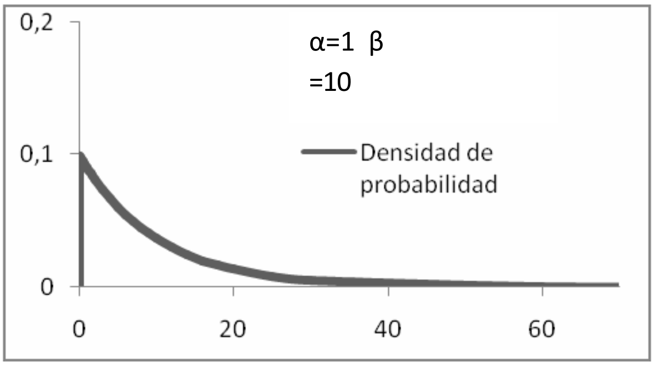
Inconvenientes

Problemas en la complejidad de algunos cálculos, especialmente respecto a la función Gamma cuando el parámetro α es un valor no entero. También problemas de cálculo en la estimación de los parámetros muestrales. Ambos inconvenientes se pueden abordar satisfactoriamente con ordenador.

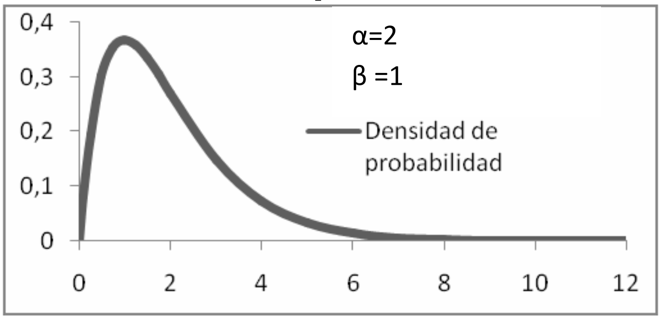
Para valorar la evolución de la distribución al variar los parámetros se tienen los siguientes gráficos. Primero se comprueba que para α=1 la distribución tiene similitudes con la exponencial.

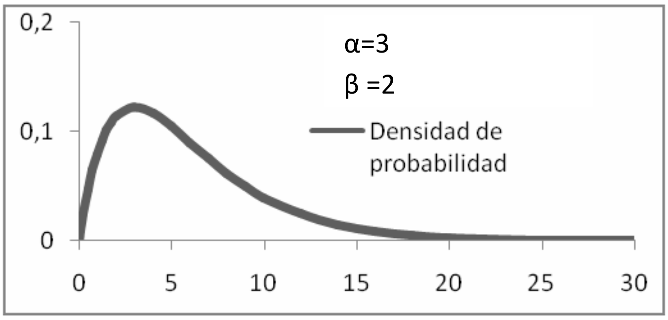


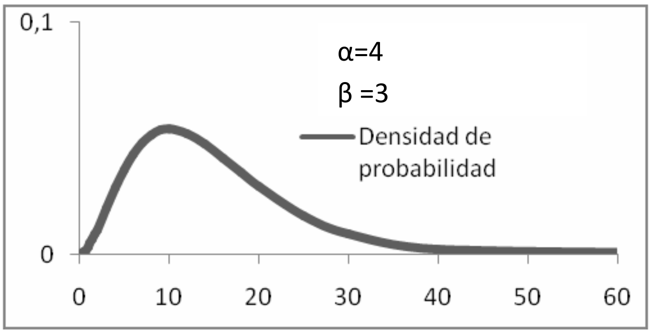




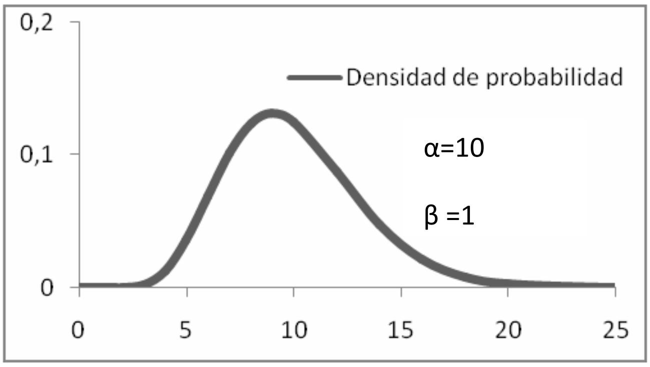
Si ahora se hace variar el parámetro alfa,







Y para valores altos de α y pequeños de β, se observa la convergencia con la normal,



Como ya se ha indicado, la expresión de la distribución Gamma incluye la propia función Gamma $\Gamma(\alpha)$, que para valores enteros de alpha se ha demostrado que $\Gamma(\alpha) = (\alpha - 1)!$ En este caso la distribución Gamma se conoce como "distribución de Erlang".

Valor α	Función Gamma
1	1
2	1
3	2
4	6
5	24
6	120
7	720
8	5040
9	40320
10	362880
11	3628800
12	39916800

Para valores no enteros, **el valor de la función Gamma** se obtiene a partir de,

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} * e^{-x} dx$$

La **función de densidad** de la distribución Gamma es,

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} * x^{\alpha-1} * e^{-\frac{x}{\beta}}$$

donde x>0 y α, β son parámetros positivos.

Se puede ver que para α=1 la función de densidad será,

$$f(x) = \frac{1}{1*1} * x^0 * e^{-\frac{x}{1}} = e^{-x}$$

lo que significa que una distribución Gamma de parámetro α=1 y β =0 es una distribución exponencial de parámetro α=1.

Se demuestra que f(x) es una función de densidad porque para f(x)≥0 y haciendo el cambio de variable $h = \frac{1}{\beta}$

$$\int_0^{\infty} f(x)dx = \frac{h^{\alpha}}{\Gamma(\alpha)} * \int_0^{\infty} x^{\alpha-1} * e^{-xh} * dx = \frac{h^{\alpha}}{\Gamma(\alpha)} * \frac{\Gamma(\alpha)}{h^{\alpha}} = 1$$

La **función de distribución** es,

$$F(x) = P(X \leq x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} * \int_0^{x} u^{\alpha-1} * e^{-\frac{u}{\beta}} * du$$

La **función característica** es, teniendo en cuenta de nuevo que$h = \frac{1}{\beta}$

$$\ell_x(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx = \int_0^{+\infty} e^{itx} \frac{h^{\alpha}}{\Gamma(\alpha)} * x^{\alpha-1} * e^{-hx} * dx = \frac{h^{\alpha}}{\Gamma(\alpha)} * \int_0^{\infty} \mathbf{h}^{\boldsymbol{\alpha}} * e^{-(h-it)x} dx =$$

$$= \left\{ \frac{\Gamma(\alpha)}{(h - i * t)^{\alpha}} = \int_0^{+\infty} x^{\alpha-1} * e^{-(h-it)x} dx \right\} = \ell_x(t) = \frac{h^{\alpha}}{\Gamma(\alpha)} * \frac{\Gamma(\alpha)}{(h - i * t)^{\alpha}} = \left(\frac{h^{\alpha}}{h - it}\right)^{\alpha} = \left(\frac{1}{1 - \frac{it}{h}}\right)^{\alpha} =$$

$$\ell_x(t) = \left(1 - \frac{it}{h}\right)^{-\alpha}$$

y volviendo a β=1/h,

$$\ell_x(t) = (1 - \beta * i * t)^{-\alpha}$$

La **esperanza matemática** será, siendo $h^{\alpha} =$ $h^{\alpha} = \dfrac{1}{\beta}$

$$E(X) = \int_0^{+\infty} x * f(x) * dx = \frac{h^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x * \boldsymbol{h^{\alpha}} * e^{-hx} * dx = \frac{h^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha} * e^{-hx} * dx =$$

$$= \frac{h^{\alpha}}{\Gamma(\alpha)} * \frac{\Gamma(\alpha+1)}{h^{\alpha+1}} = \frac{1}{\mathrm{h}} * \frac{\alpha * \Gamma(\alpha)}{\Gamma(\alpha)} = \frac{\alpha}{\mathrm{h}}$$

Volviendo a β;

$$E(X) = \alpha\beta$$

La **varianza** será,

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

donde otra vez $\mathrm{h} = \frac{1}{\beta}$

$$E(X^2) = \int_0^{\infty} x^2 f(x) dx = \frac{h^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^2 * x^{\alpha-1} * e^{-hx} * dx = \frac{h^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+1} * e^{-hx} * dx =$$

$$\frac{h^{\alpha}}{\Gamma(\alpha)} * \frac{\Gamma(\alpha+2)}{h^{\alpha+2}} = \frac{1}{h^2} * \frac{(\alpha+1) * \Gamma(\alpha)}{\Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{h^2}$$

$$V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{\alpha(\alpha+1)}{h^2} * \left(\frac{\alpha}{h}\right)^2 = \frac{\alpha^2 + \alpha - \alpha^2}{h^2} = \frac{\alpha}{h^2}$$

Y volviendo a β;

$$V(X) = \alpha\beta^2 \quad \rightarrow desv.típica = D(X) = \beta\sqrt{\alpha}$$

La moda se puede calcular como $M_o = (\alpha - 1)\beta$

Si se buscan estimadores de los parámetros de la distribución, el Método de Máxima Verosimilitud es el más adecuado ya que goza de más propiedades que el estimador obtenido por el método de los momentos (Chico, 2010). Sin embargo, la máxima verosimilitud conduce a un sistema de ecuaciones que no se puede resolver de forma analítica, y se necesita recurrir al método numérico, por ejemplo, de Newton-Raphson.

$$-ln\beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^{n} lnx_i = 0$$

$$\alpha = \frac{\sum_{i=1}^{n} x_i}{n * \beta}$$

Resolución de probabilidades de la distribución Gamma

En R se pueden usar las órdenes >dgamma y >pgamma que devuelven resultados de la función de densidad y de densidad acumulada respectivamente.

>dgamma(x, “parámetro (α)”, scale = “parámetro (β); escala”)
Y devuelve el valor de la densidad de probabilidad en cuando la variable aleatoria tiene valor = x

>pgamma(x, “parámetro (α)”, scale = “parámetro (β); escala”; lower.tail = T)
Devuelve el valor de probabilidad acumulada hasta el valor x. Con lower.tail = F devuelve la probabilidad por la derecha.

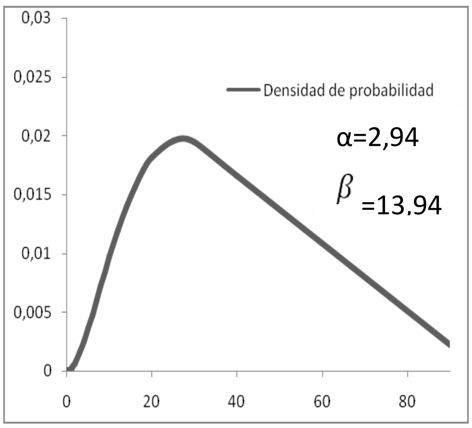
Si en lugar de parámetro escala se está usando el parámetro ratio se escribirá;
>dgamma(x, “parámetro (α)”, ratio = “parámetro (β); escala”)
>pgamma(x, “parámetro (α)”, ratio = “parámetro (β); escala”; lower.tail = T)

En excel la función es, para la intensidad de probabilidad
=DISTR.GAMMA(x;”parámetro (α);”parámetro escala (β)”;FALSO)
Y para la probabilidad acumulada,
=DISTR.GAMMA(x;”parámetro (α);”parámetro escala (β)”;VERDADERO)
Pero no permite diferenciar entre “escala” o “ratio”.

Ejemplos de distribución Gamma

Ejemplo 1.
En un estudio de la guardia urbana de Barcelona se toma una distribución gamma para modelizar el número de víctimas en accidentes de tráfico. Como es más habitual la proporción de 1 ocupante por vehículo siniestrado, y es más rara la probabilidad de 4 ó 5 ocupantes por vehículo siniestrado, se crea una distribución gamma para modelizar el número de víctimas por accidente de tráfico. El 38% de la distribución lo acumula la proporción 1 accidentado por accidente, el 36% 2:1, 16% la 3:1, 6% el 4:1 y finalmente un 3% para 5:1. La media del modelo es 1,5 víctimas por accidente, pero no indican el valor de los parámetros α y β tomados en cuenta.
(*Análisis del envío de equipos de emergencia para accidentes en las Rondas de Barcelona, pág. 40*)

Ejemplo 2.
También en el ámbito de la siniestralidad viaria, en un estudio de la ciudad de Medellín, Colombia, se usa la distribución Gamma para obtener la distribución de probabilidad de la variable aleatoria “edad de fallecimiento en accidentes de tráfico”. En este caso explican que se asignaron los parámetros α y β “a ojo”. El mejor resultado es el que parece minimizar los errores cuadráticos medios después de varias asignaciones. Finalmente obtienen α=2,94 y β =13,94.
(*Panorama de la accidentalidad vial en Medellín en 1999, pág 14*)



Ejemplo 3. Relación entre la distribución gamma y procesos de Poisson

A una centralita de teléfonos llegan 12 llamadas por minuto, siguiendo una distribución de Poisson. ¿Cuál es la probabilidad de que en menos de 1 minuto lleguen 8 llamadas?

Si $\lambda = 12\frac{llamadas}{minuto}$ entonces, $\beta = \frac{1\ minuto}{12\ llamadas}$

La **función de distribución** es,

$$F(x) = P(X \le x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} * \int_0^x u^{\alpha-1} * e^{-\frac{u}{\beta}} * du$$

$$P(x<1) = \frac{1}{\frac{1}{12}\Gamma(8)} * \int_0^1 u^{8-1} * e^{-\frac{u}{1/12}} du$$

Resolviendo en R,

```
> pgamma(1, 8, scale = 1/12, lower.tail = T)
[1] 0.9104955
>
```

Existe un 91,05% de probabilidades de recibir 8 llamadas en un plazo de tiempo de menos de 1 minuto.

Ejemplo 4.

Si un componente eléctrico falla una vez cada 5 horas, ¿cuál es el tiempo medio que transcurre hasta que fallan dos componentes? ¿Cuál es la probabilidad de que transcurran 12 horas antes de que fallen los dos componentes?

Si $\lambda = \frac{1\ fallo}{5\ horas}$ entonces, $\beta = \frac{1}{\lambda} = \frac{1}{\frac{1}{5}}$= 5 horas/fallo

$$E(X) = \alpha * \beta = 2 * 5 = 10\ horas$$

$$P(x > 12) = 1 - \frac{1}{5^2\Gamma(2)} \int_0^{12} u^{2-1} * e^{-\frac{u}{1/12}} du$$

Resolviendo en R

```
> pgamma(12, 2, scale = 5, lower.tail = F)
[1] 0.3084410
>
```

En un 30,84% de las situaciones pasarán 12 horas hasta que fallen dos componentes.

Ejemplo 5.

En una ciudad se observa que el consumo diario de energía (en millones de kilowatt-hora) es una variable aleatoria que sigue una distribución gamma con parámetros α= 3 y β =2. Si la planta de energía que suministra a la ciudad tiene una capacidad diaria de generar un máximo de 1, ¿cuál es la probabilidad de que haya un día donde no se pueda satisfacer la demanda?

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} * x^{\alpha-1} * e^{-\frac{x}{\beta}}$$

donde,

$$\alpha = 3 \quad \beta = 2$$
$$\Gamma(\alpha) = (n-1)! = \Gamma(3) = (3-1)! = 2$$

de forma que,

$$f(x) = \frac{1}{2^3 * 2} * x^2 * e^{-\frac{x}{2}}$$

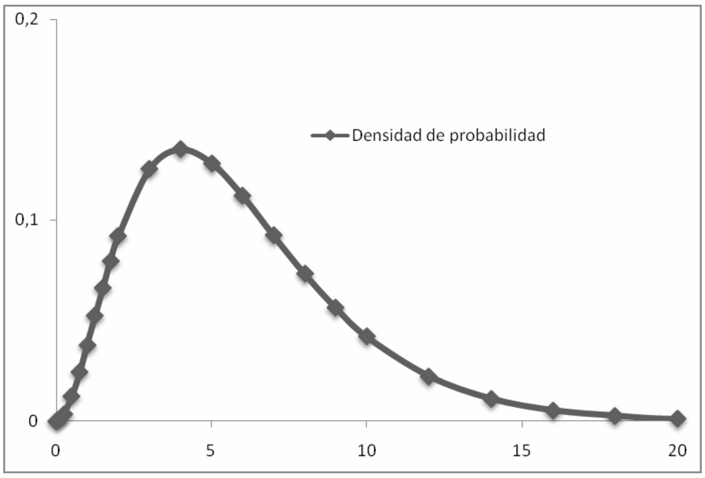
La probabilidad de que exista 1 exceso,

$$F(1) = P(X \leq 1) = \frac{1}{2^3 * 2} * \int_0^1 u^{3-1} * e^{-\frac{u}{2}} du =$$

Resolviendo la integral con ayuda del Derive, la probabilidad obtenida es,

$$= \frac{1}{16} * (16 - 26 * e^{-0,5}) = 0,01438$$

La representación gráfica y verificación con Excel del cálculo:



Valores de x	Densidad de probabilidad	Probabilidad Acumulada	Probabilidad por la derecha
0	0	0	1
0,09	0,000483974	1,4684E-05	0,999985316
0,25	0,003447254	0,000296478	0,999703522
0,5	0,012168762	0,002161497	0,997838503
0,75	0,024162514	0,006652214	0,993347786
1	0,037908166	0,014387678	0,985612322
1,25	0,052271624	0,025656931	0,974343069
1,5	0,066426546	0,040505439	0,959494561
1,75	0,079789996	0,058803721	0,941196279
2	0,09196986	0,080301396	0,919698604
3	0,125510715	0,191153169	0,808846831
4	0,135335283	0,32332358	0,67667642
5	0,12825781	0,456186873	0,543813127
6	0,112020904	0,576809919	0,423190081
7	0,092479487	0,679152801	0,320847199
8	0,073262556	0,761896694	0,238103306
9	0,056239295	0,826421929	0,173578071
10	0,042112169	0,875347981	0,124652019
12	0,02230877	0,938031196	0,061968804
14	0,011170554	0,970363836	0,029636164
16	0,005367402	0,986246032	0,013753968
18	0,002499049	0,993767805	0,006232195
20	0,001134998	0,997230604	0,002769396
25	0,000145572	0,999658545	0,000341455
30	1,7207E-05	0,999960692	3,93084E-05
40	2,06115E-07	0,999999544	4,55515E-07
90	1,44915E-17	1	0

Ejemplo 6.

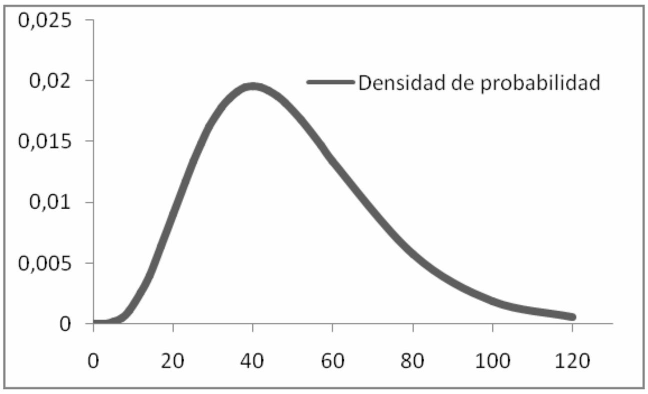
Si se sabe que el tiempo de sobrevivencia de ratas expuestas a un determinado tóxico es una variable aleatoria que sigue una distribución Gamma (5, 10), ¿cuál es la probabilidad de que una rata no supere las 60 semanas de vida?

$$P(X < 60) = \frac{1}{10^5(5-1)!} * \int_0^{60} x^{5-1} * e^{-\frac{x}{10}} * dx$$

Resolviendo en R,

> pgamma(60, 5, scale = 10, lower.tail = T)
[1] 0.7149435

Su representación gráfica en Excel


0,025
0,02
0,015
0,01
0,005
0
0
20
40
60
80
100
120
Densidad de probabilidad

Fuentes:

http://gredos.usal.es/jspui/handle/10366/83459
http://www.coini.com.ar/COINI%202009/COINI%202008/Trabajos/TC3.pdf
http://departamentos.unican.es/macc/asignaturas/informatica/estadistica/pdf/Cap4.pdf http://unbarquero.blogspot.com/2009/06/r-distribucion-gamma.html
http://www.suratep.com/articulos/27/accidentalidad_vial.pdf
http://upcommons.upc.edu/pfc/bitstream/2099.1/8321/1/Memoria%20PFC.pdf
http://www.aiaccess.net/English/Glossaries/GlosMod/e_gm_gamma_distri.htm
http://upcommons.upc.edu/revistes/bitstream/2099/5616/1/Article03a.pdf http://office.microsoft.com/es-mx/excel-help/distr-gamma-HP005209101.aspx

DISTRIBUCIÓN EXPONENCIAL

λ = número de ocurrencias en la unidad de tiempo o espacio

λ

0 0 0 0 0 | Estación de Servicio

$$\beta = \frac{1}{\lambda} = \text{ tiempo entre dos ocurrencias consecutivas}$$

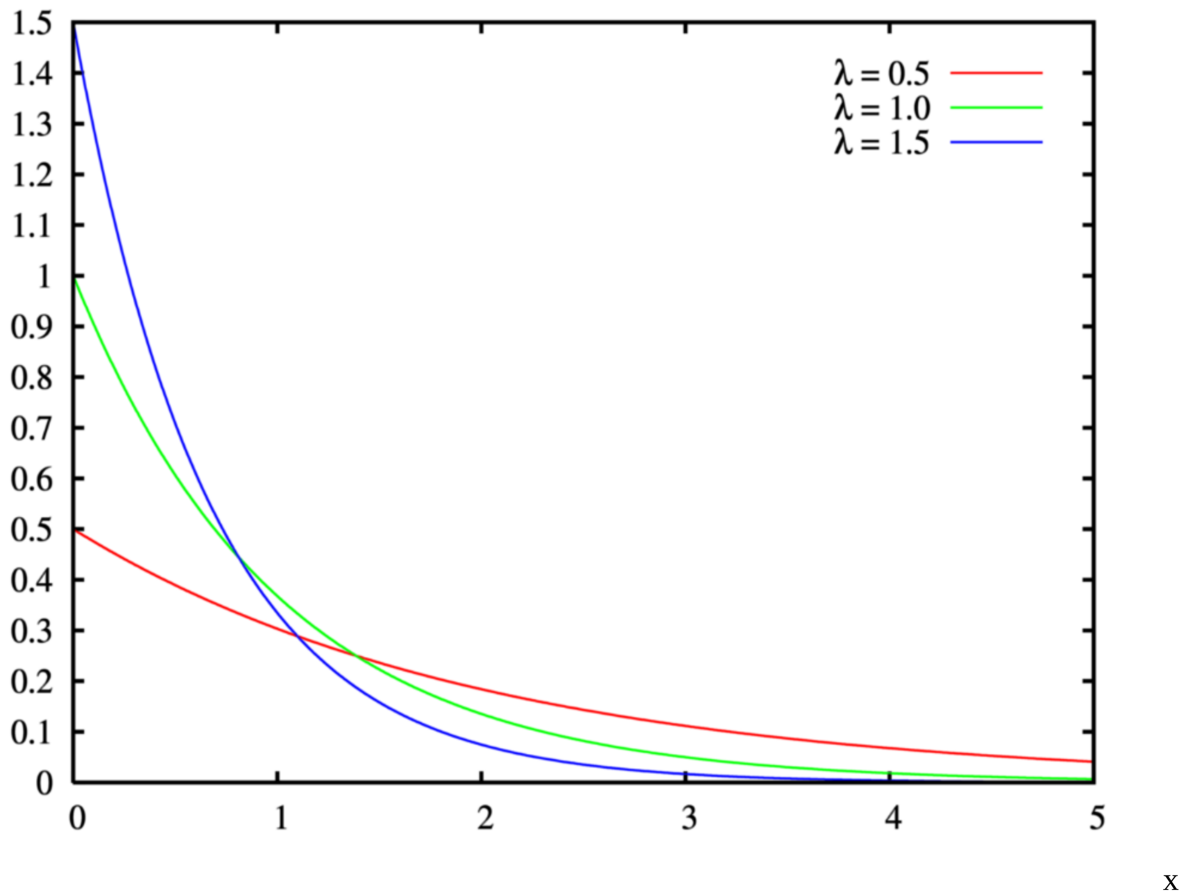
La variable aleatoria continúa x, tiene una distribución exponencial, con parámetro β si su función densidad está dada por:

$$f(x)= \begin{cases} \frac{1}{\beta} e^{\frac{-x}{\beta}}; & x \geq 0, \quad \beta > 0 \\ 0 & x < 0 \end{cases}$$

Cuando el número de sucesos por unidad de tiempo sigue una distribución de Poisson de parámetro λ (proceso de Poisson), el tiempo entre dos sucesos consecutivos sigue una distribución Exponencial de parámetro β = 1/λ.

De otra forma:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0 \end{cases}$$



La distribución exponencial tiene muchas aplicaciones en el campo de la Estadística, particularmente en las áreas de la teoría de confiabilidad y tiempos de espera o en problemas de teorías de colas.

Teorema 1. En la distribución exponencial la media o valor esperado está dada por $\mu = 1/\lambda$, la varianza se da por $\sigma^2 = 1/\lambda^2$ y por lo tanto la desviación estándar es igual a la media $\sigma = 1/\lambda$.

Teorema 2. Para un valor dado $x = t$, la probabilidad acumulada desde $x = 0$ hasta $x = t$ está dada por:

$$F(t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}$$

Por consiguiente, la probabilidad de que una variable aleatoria con distribución exponencial asuma un valor *mayor* (o mayor o igual) se da porque

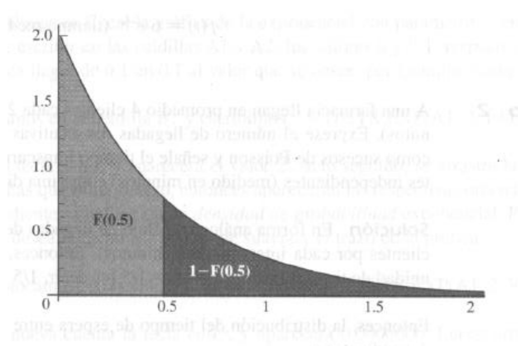
$$F(x > t) = e^{-\lambda t}$$

En la figura se ilustra la gráfica de una exponencial con parámetro $\lambda = 2$. El valor esperado o media es aquí de 0.5. Con sombreado claro aparece dibujada la probabilidad de que x tenga un valor menor a 0.5 y con sombreado oscuro la probabilidad de que asuma un valor mayor o igual a 0.5.

Área de la izquierda $= F(0.5)$ Area de la derecha $1 - F(0.5)$.

Se puede demostrar fácilmente que el área de la izquierda es aproximadamente 63.2% del área total y que el área de la derecha es aproximadamente el 36.8%.

También se puede demostrar que en la distribución exponencial la probabilidad de que X sobrepase su valor esperado es igual a $1/e$ y la probabilidad de que tome un valor menor (o menor o igual) a la media es precisamente de $1 - \frac{1}{e} = 0.63212$



En resumen:

$P(X \geq x) = \int_x^\infty \lambda e^{-\lambda t} dt = 1 - F(x) = e^{-\lambda x}$ = probabilidad a mano derecha de x

$P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$ = probabilidad a mano izquierda de x

Una propiedad importante es la denominada *carencia de memoria*, que podemos definir así: si la variable X mide el tiempo de vida y sigue una distribución Exponencial, significará que la probabilidad de que siga con vida dentro de 20 años es la misma para un individuo que a fecha de hoy tiene 25 años que para otro que tenga 60 años.

Ejemplo:

El periodo de vida en años de un interruptor eléctrico tiene una distribución exponencial con un promedio de falla de μ =2 años = β
¿cuál es la probabilidad de que al menos ocho de 10 de tales interruptores, que funcionan independientemente, fallen después del 3er año?

Distribución Exponencial

$$\mu = 2 = \beta$$

$$\lambda = \frac{1}{\beta} = \frac{1}{2} = 0.5$$

$$\int_x^\infty \lambda e^{-\lambda t} dt = -\lim_{b\to\infty} e^{-\lambda t}\Big|_x^\infty = -(0 - e^{-\lambda x}) = e^{-\lambda x}$$

$$(x > 3) = e^{-0.5(3)} = 0.2231$$

Empleando la distribución binomial:

$$n = 10 \quad p = 0.2231$$

$$P(x \geq 8) = \sum_{x=8}^{10} \binom{10}{x} (0.2231)^x \, (0.7769)^{10-x}$$

$$P(x = 8) = \binom{10}{8} (0.2231)^8 (0.7769)^2 = 0.000167$$

$$P(x = 9) = \binom{10}{9} (0.2231)^9 (0.7769)^1 = 0.000011$$

$$P(x = 10) = \binom{10}{10} (0.2231)^{10} (0.7769)^0 = 0.000000$$

Por lo que

$$P(x \geq 8) = 0.000178$$

5.3 Distribuciones normal y normal estándar, uso de tablas de distribución normal estándar, la aproximación de la distribución binomial a la distribución normal.

DISTRIBUCIÓN NORMAL

1733 De Moivre

Laplace, Gauss

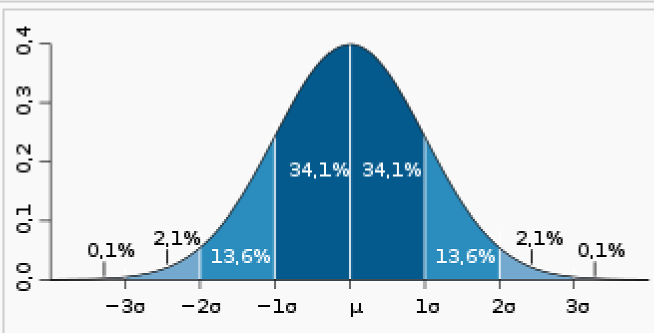
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} ; \quad -\infty < x < \infty$$

2, π constantes

μ_x - parámetro de localización

σ_x - parámetro de escala

x - variable aleatoria



Distribución de probabilidad alrededor de la media en una distribución N(μ, σ).

$$\frac{1}{\sqrt{2\pi}\sigma_x} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} dx = 1$$

*DISTRIBUCIÓN NORMAL ESTÁNDAR

$z = \dfrac{x-\mu_x}{\sigma_x}$ **..........**variable normalizada o tipificada

$\mu_z = \mu = 0$

$\sigma_z = \sigma = 1$

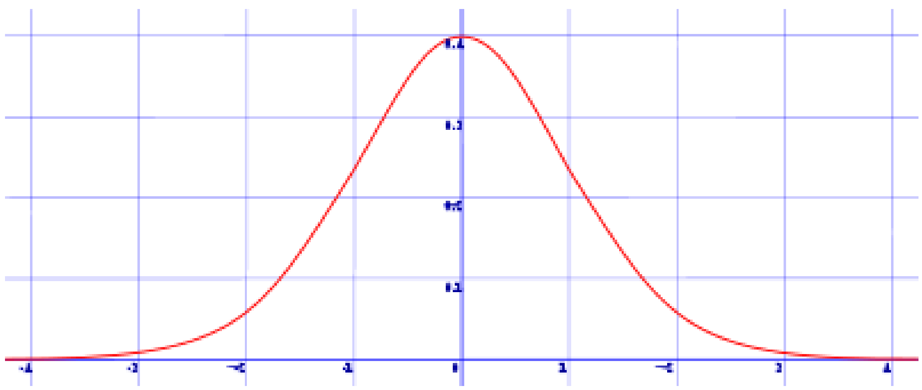
$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2};\qquad -\infty<z<\infty$$ .....Distribución normal estándar

Se puede demostrar que:

$$\int_{-\infty}^{\infty} f(z)\,dz=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}dz=1;$$

Para el cálculo de probabilidades (área bajo la curva normal)

$$P(z_1<z<z_2)=\frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-\frac{1}{2}z^2}dz=F(z_2)-F(z_1)$$

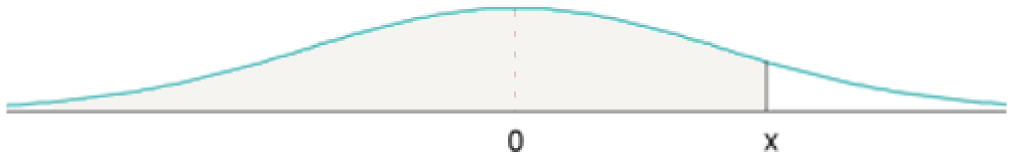


$$P(-1<z<1)=F(1)-F(-1)=0.8413-(1-0.8413)=0.8413-0.1587=0.6826$$
$$P(-2<z<2)=F(2)-F(-2)=0.9772-0.0278=0.9544$$
$$P(-3<z<3)=F(3)-F(-3)=0.9987-0.0013=0.9974$$

Distribución normal estándar

N(0, 1)

La **distribución normal estándar, o tipificada o reducida,** es aquella que tiene por **media** el valor **cero**, **μ =0**, y por **desviación típica la unidad, σ =1**.



La probabilidad de la variable X dependerá del área del recinto sombreado en la figura. Y para calcularla utilizaremos una **tabla**.

Tipificación de la variable

Para poder utilizar la tabla tenemos que transformar la variable **X** que sigue una distribución **N(μ, σ)** en otra variable **Z** que siga una distribución **N(0, 1)**.

$$Z = \frac{X - \mu}{\sigma}$$

Cálculo de probabilidades en distribuciones normales

La **tabla** nos da las **probabilidades de P(z ≤ k)**, siendo **z** la variable tipificada.

Estas probabilidades nos dan la **función de distribución Φ(k)**.

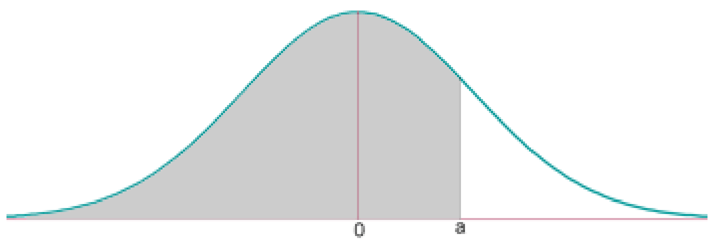
F(k) = P(z ≤ k)

Búsqueda en la tabla de valor de k

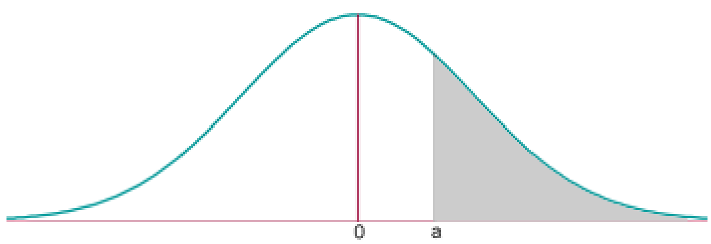
Unidades y décimas en la columna de la izquierda.

Centésimas en la fila de arriba.

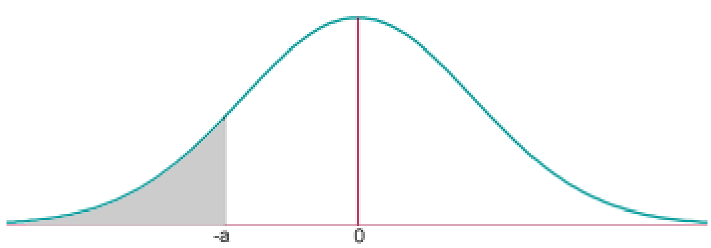
$P(Z \leq a)$



$P(Z > a) = 1 - P(Z \leq a)$

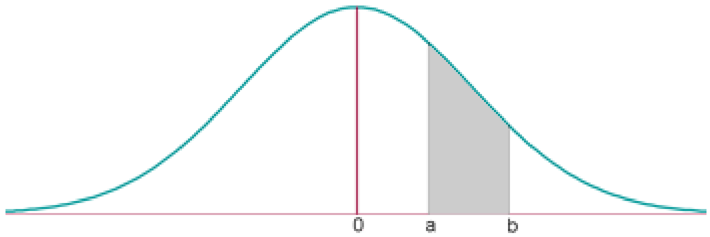


$P(Z \leq -a) = 1 - P(Z \leq a)$



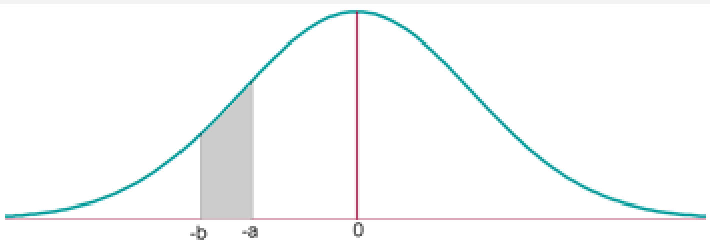
$P(Z > -a) = P(Z \leq a)$

$$P(a < Z \leq b) = P(Z \leq b) - P(Z \leq a)$$

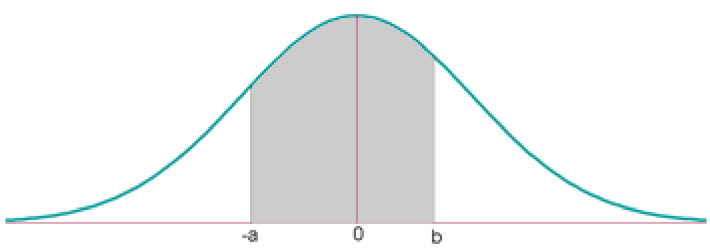


$$P(-b < Z \leq -a) = P(a < Z \leq b)$$

Nos encontramos con el caso inverso a los anteriores, conocemos el valor de la probabilidad y se trata de hallar el valor de la abscisa. Ahora tenemos que buscar en la tabla el **valor que más se aproxime a K**.



$$P(-a < Z \leq b) = P(Z \leq b) - [1 - P(Z \leq a)]$$



$$p = K$$

Para calcular la variable **X** nos vamos a la **fórmula de la tipificación.**

Ejercicio 1.

Dada una distribución normal con $\mu_x = 50$ y $\sigma_x = 10$, encuentre la probabilidad de que x tome un valor entre 45 y 62.

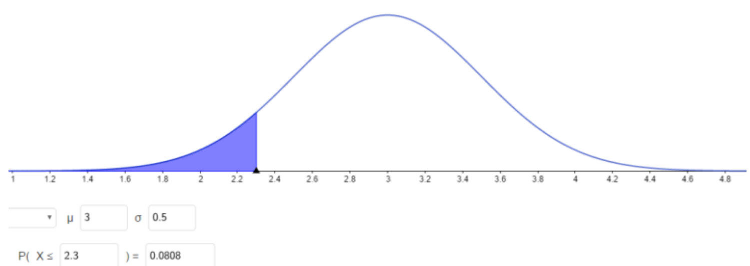
$$N = (x, \mu_x, \sigma_x) = (x, 50, 10) \qquad z = \frac{x - \mu_x}{\sigma_x}$$

$$P(45 < z < 62) = P\left(\frac{45-50}{10} < z < \frac{62-50}{10}\right) = P(-0.5 < z < 1.2)$$

$$=F(1.2)-F(-0.5)=0.8849-0.3085=0.5764$$

Ejercicio 2.

Cierto tipo de pila almacenada dura en promedio 3.0 años, con una desviación estándar de 0.5 años. Suponiendo que la vida de las pilas está distribuida normalmente. a) Encuentre la probabilidad de que una pila dada dure menos de 2.3 años. b) Si se toman 10 de estas pilas, ¿Cuál es la probabilidad de que más de una dure menos de 2.3 años?



$$P(x < 2.3) = P\left(z < \frac{2.3-3}{0.5}\right) = P(z < -1.4) = F(-1.4)$$

$$\boxed{P(x \le 2.3) = 0.0808}$$

b)

$$b(x, 10, 0.0808)$$

$$P(x > 1) = 1 - [P(0) + P(1)]$$

$$P(0) = \binom{10}{0}(0.0808)^0(0.9192)^{10} = 0.4306$$

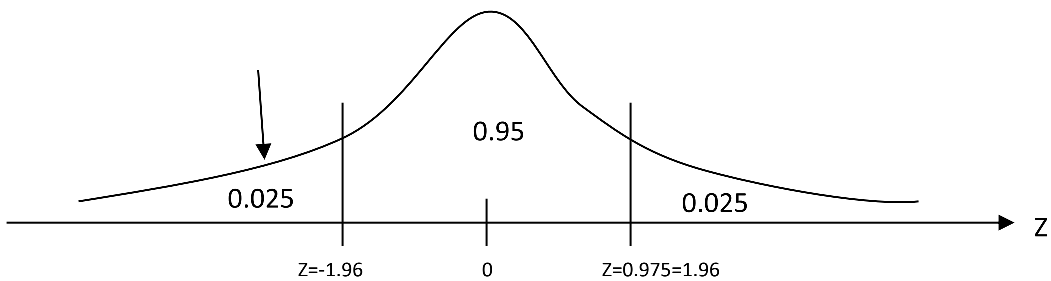
$$P(1) = \binom{10}{1}(0.0808)^1(0.9192)^9 = 0.3785$$

$$P(x > 1) = 1 - 0.8091 = \boxed{0.1909}$$

Ejercicio 3.

Un instrumento de precisión se emplea para rechazar hasta todos los componentes en los cuales cierta dimensión no cumple con la especificación 1.50 $\pm$ d. Si se sabe que esta medición está distribuida normalmente con $\mu = 1.5$ y desviación estándar de 0.2. Determine el valor d para que la especificación cubra el 95% de las mediciones.

$$\mu = 1.5 \text{ y } \sigma = 0.2; \quad 1.50 \pm \mathrm{d}$$



$$z = \frac{x - \mu_x}{\sigma_x}; \qquad x = z\sigma_x + \mu_x$$

$$x = \mu_x + d$$

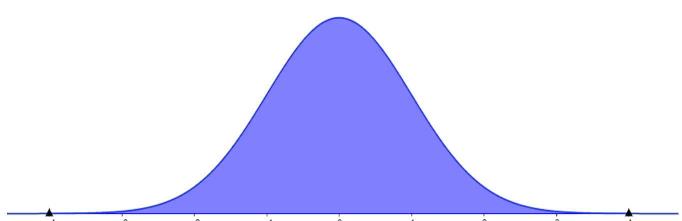
$$1.5 + d = 1.96(0.2) + 1.5$$
$$d = 0.3920$$

APROXIMACIÓN NORMAL A LA DISTRIBUCIÓN BINOMIAL.

$$\lim_{n \to \infty} b(x, n, p) = N(x, np, \sqrt{npq})$$

P=0.5

n>30



Se emplea la variable tipificada:

$$z = \frac{x - np \pm 0.5}{\sqrt{npq}}; \quad 0.5 \rightarrow \text{corrección por continuidad}$$

1. $P(x = k) \approx P\left(\frac{k - np - 0.5}{\sqrt{npq}} \leq z \leq \frac{k - np + 0.5}{\sqrt{npq}} \right)$

2. $P(x \geq k) \approx P\left(z \geq \frac{k - np - 0.5}{\sqrt{npq}} \right)$

3. $P(x \leq k) \approx P\left(z \leq \frac{k - np + 0.5}{\sqrt{npq}} \right)$

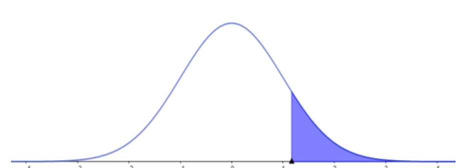
4. $P(x > k) \approx P\left(z > \frac{k - np + 0.5}{\sqrt{npq}} \right)$

5. $P(x < k) \approx P\left(z < \frac{k - np - 0.5}{\sqrt{npq}} \right)$

Ejercicio 1.

En un proceso se obtiene el 10% de piezas defectuosas. Si de dicho proceso se seleccionan al azar 100 piezas, ¿Cuál es la probabilidad de que el número de piezas defectuosas exceda de 13?

$$n = 100 \qquad p = 0.1$$
$$\mu = np = 100 * 0.1 = 10$$
$$\sigma = \sqrt{np(1-p)} = \sqrt{100 * 0.1 * 0.9} = 3$$



$$z = \frac{13 - 10 + 0.5}{3} = 1.17$$
$$P(x > 13) = P(z \geq 1.17) = 1 - F(1.17) = 1 - 0.8790 = 0.121$$

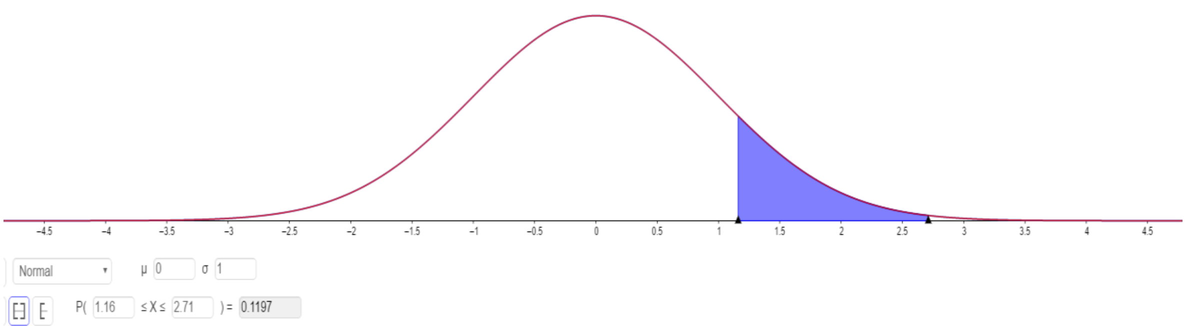
Ejercicio 2.

Un cuestionario de opción múltiple o selección múltiple contiene 200 preguntas cada una con 4 respuestas posibles y de ellas solo 1 es la correcta, ¿Cuál es la probabilidad de que por simple conjetura el alumno obtenga entre 25 y 30 respuestas correctas para 80 de las 200 preguntas cuya respuesta ignora por completo?

$$n = 80,\ p = 0.25,$$
$$\mu = np = (80)(0.25) = 20,$$
$$\sigma = \sqrt{np(1-p)} = \sqrt{(80)(0.25)(0.75)} = 3.873$$

$$z_1 = \frac{25-20-0.5}{3.873} = 1.1619$$
$$z_2 = \frac{30-20+0.5}{3.873} = 2.7111$$

$$P(25 \le x \le 30) = P(1.16 \le z \le 2.71) = F(2.71) - F(1.16) = 0.9966 - 0.8770 = \boxed{0.1196}$$



Función generatriz de momentos para una variable aleatoria con distribución normal.

$$\boxed{M_x(t) = e^{\mu t + \left(\frac{\sigma^2 t^2}{2}\right)}}$$

A partir de estas expresiones pueden calcularse la media y la varianza para cada distribución aplicando las propiedades de la función generatriz de momentos.

5.4 Distribuciones Chi-Cuadrada, T de Student, F de Fisher, Weibull y distribución Log normal, como modelos teóricos para la estadística aplicada, sus parámetros, momentos y funciones generatrices.

DISTRIBUCIONES DERIVADAS DE LA NORMAL

Son tres distribuciones derivadas de la normal, las cuales se aplican es estadística inferencial para modelas las distribuciones en el muestreo. Estas distribuciones derivadas de la distribución normal son: la ji-cuadrada (χ^2), la t de Student y la F de Fisher-Snedecor.

1. La distribución ji-cuadrada.

La distribución ji-cuadrada surge de la suma de normales estándar al cuadrado, su definición formal es la siguiente:

Definición 1.1 Sean $Z_1, Z_2, \ldots, Z_k$ variables aleatorias independientes con distribución normal estándar, entonces la variable aleatoria dada por:

$$X = Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_k^2 \tag{1.1}$$

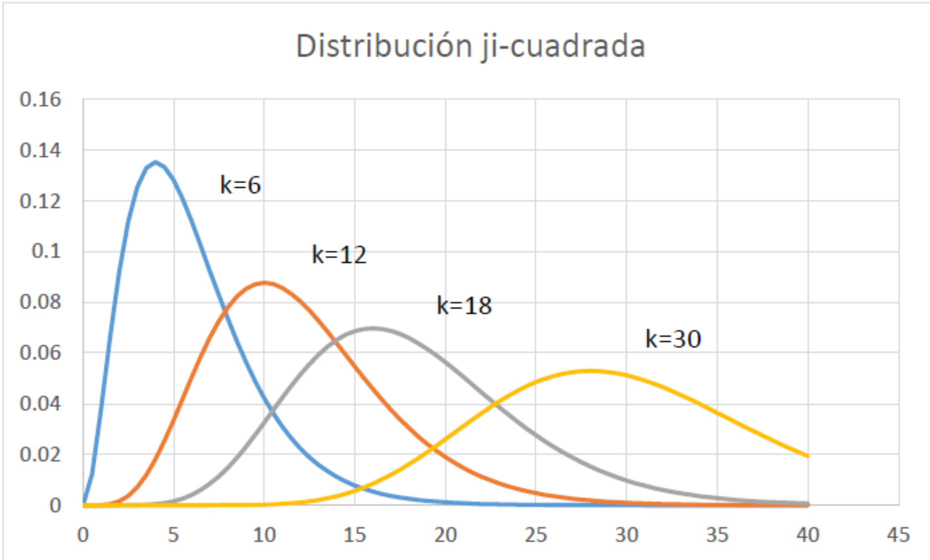
Tiene una distribución conocida como ji-cuadrada con k grados de libertad.

Teorema 1.1 Sea X una variable aleatoria ji-cuadrada con k grados de libertad, entonces su función de densidad es:

$$f_k(x) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} x^{k/2-1} e^{-x/2} \quad \text{para } x \geq 0; \quad \text{y } 0 \text{ en otro caso} \tag{1.2}$$

Obsérvese que la función de densidad ji-cuadrada corresponde a una densidad gamma con parámetros $\beta = 2$ y $\alpha = \frac{k}{2}$

La siguiente gráfica muestra la función de densidad de esta variable para distintos valores de k, observe que cuando k es pequeño la probabilidad se concentra en valores pequeños y conforme k crece la probabilidad se concentra en valores mayores.



Dado que la ji-cuadrada es l suma de variables aleatorias independientes, por el teorema del límite central, esta distribución se aproxima a un normal conforme *k* crece, como se puede ver en la gráfica.

El único parámetro de la distribución ji-cuadrada son los grados de libertad (el parámetro *k*) que corresponde al número de variables aleatorias independientes en la suma que define la variable. La variable ji-cuadrada siempre es positiva, pues corresponde a la suma de variables al cuadrado.

ORIGEN

El matemático inglés Karl Pearson (1857 / 1936), con formación en literatura medieval alemana, derecho romano, física, biología y teoría política del socialismo, hasta 1890 sobresalió por aplicar ampliamente la Estadística y la Teoría de la Probabilidad a la solución de diferentes problemas de la ingeniería industrial.
Se deriva de la distribución normal y la distribución gamma.

APLICACIONES EN EL CAMPO DE LA INGENIERIA

Sus usos son determinar la resistencia, la fuerza o la durabilidad de aleaciones, resortes, engranajes, materias primas, etc.

RELACION CON OTRAS DISTRIBUCIONES

La distribución chi cuadrado es un caso especial de la distribución gamma.

Teorema 1.2 Si X_1 y X_2 son dos variables aleatorias independientes, tales que $X_1 \sim$ ji-cuadrada con n grados de libertad y $X_2 \sim$ ji-cuadrada con m grados de libertad, entonces $X_1 + X_2 \sim$ ji-cuadrada con $n + m$ grados de libertad.

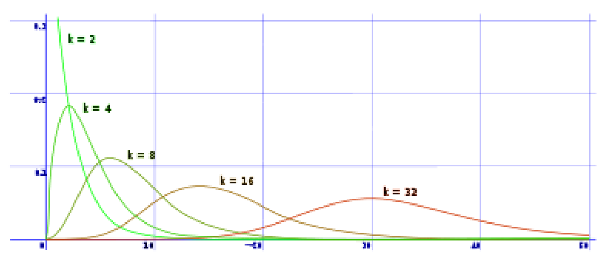
Resumen	Variable aleatoria ji-cuadrada
Función de densidad	$f_k(x) = \dfrac{1}{2^{k/2}\Gamma\left(\dfrac{k}{2}\right)} x^{k/2-1} e^{-x/2}$ para $x \geq 0$
Media	$\mu = k$
Varianza	$\sigma^2 = 2k$
Función generatriz de momentos	$M_x(t) = (1-2t)^{-k/2}, \qquad t < \dfrac{1}{2}$

Tablas estadísticas/Distribución chi-cuadrado

< Tablas estadísticas

La **Distribución chi-cuadrado**, tiene por **función de densidad**

$$\chi_k^2(x) = \frac{x^{k/2-1}\ e^{-x/2}}{2^{k/2}\ \Gamma(k/2)}$$



Donde el parámetro **k** de χ_k^2, se denomina grados de libertad de la distribución.

La Distribución chi-cuadrado no tiene sentido para valores negativos de **x**, como se puede ver en la figura.

Téngase en cuenta que para **k** = 1 y **k** = 2 la función de densidad para **x** = 0, se hace infinito:

$$\chi_1^2(0) = \infty$$

$$\chi_2^2(0) = \infty$$

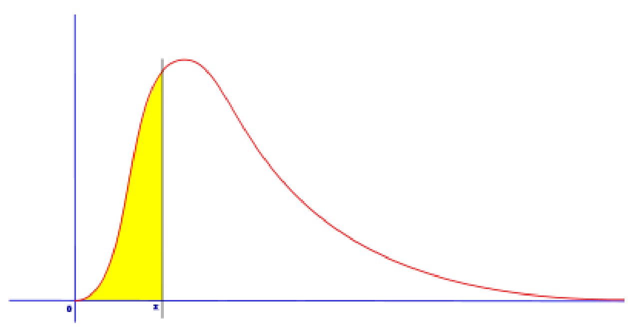
Para el resto de los valores de **k**, para **x** = 0, la función vale 0.

La **Distribución de probabilidad** de esta función para valores menores de un x dado, que representamos por $(\chi_k^2 < x)$

$$P(\chi_k^2 < x) = \int_0^x \chi_k^2\, du$$

donde:

$$\int_0^x \chi_k^2\, du = \int_0^x \frac{u^{k/2-1}\ e^{-u/2}}{2^{k/2}\ \Gamma(k/2)}\, du$$



Esta integral no tiene una solución conocida, y solo se conocen métodos numéricos para calcular sus valores, hay distintos tipos de tablas y algoritmos para ordenador con los que se pueden calcular sus soluciones, veamos una **tabla distribución chi-cuadrado** y su modo de utilización.

La mayoría de los textos de Probabilidad y Estadística contienen tablas de la distribución ji-cuadrada con los valores de x, tales que

$$P(X > x) = \alpha$$

Para

$$\alpha = 0.995, 0.99, 0.975, 0.95, 0.90, 0.10, 0.05, 0.025, 0.01, 0.005$$

Si se requiere calcular la probabilidad del complemento se usa la fórmula:

$$P(X < x) = 1 - P(X > x)$$

Ejemplo 1. Sea X una variable aleatoria ji-cuadrada con 9 grados de libertad, encontrar el valor de x, tal que:

- $P(X > x) = 0.025$
- $P(X < x) = 0.025$

Solución:

Para hallar el valor de x, tal que:

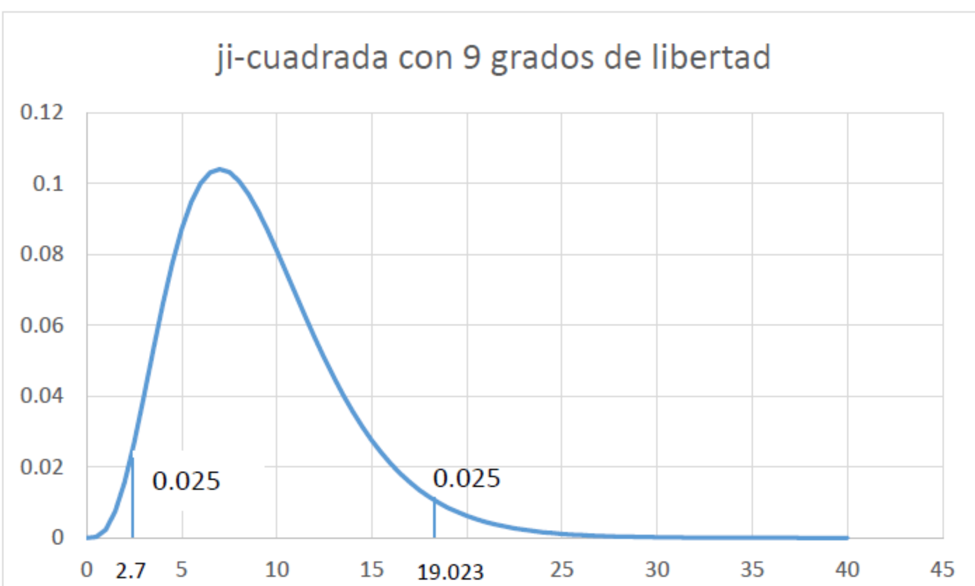
$$P(X > x) = 0.025$$

En la tabla de la ji-cuadrada busque en la columna correspondiente la probabilidad 0.025 y el renglón de 9 grados de libertad para encontrar $x = 19.023$.

Para determinar el valor de x, tal que:

$$P(X < x) = 0.025,$$

Se encuentra la probabilidad del complemento $P(X > x) = 0.975$, ahora en la columna de 0.975 busque el renglón de 9 grados de libertad y ahí se encuentra el valor de $x = 2.700$.



ejemplo:

Cual es la Distribución de probabilidad de chi-cuadrado de 4 grados de libertad de que x< 1,2

Buscando en la tabla la columna del 4 y la fila de 1,2, tenemos:

$$P(\chi_4^2 < 1,2) = 0,121901$$

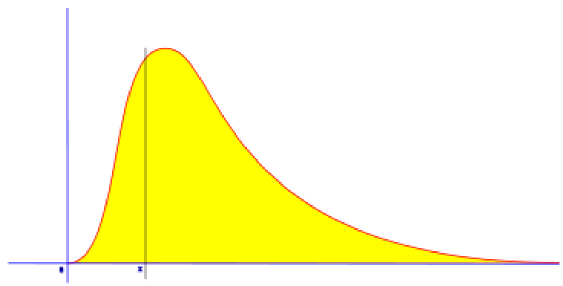
Para otros valores de x [editar]

En la tabla podemos encontrar directamente la probabilidad: $P(\chi_k^2 < x)$, pero se pueden presentar otros casos, veamos algunos.

Para la variable mayor que x [editar]

Para calcular $P(\chi_k^2 > x)$, partimos de la expresión:

$$P(\chi_k^2 < x) + P(\chi_k^2 > x) = 1$$



La probabilidad de que la variable estadística sea menor que x más la probabilidad de que sea mayor que x es la certeza, de probabilidad 1.

Operando:

$$P(\chi_k^2 > x) = 1 - P(\chi_k^2 < x)$$

Ejemplo [editar]

Calcular la distribución de probabilidad de una variable estadística chi-cuadrado, de 6 grados de libertad sea mayor de 3,4.

$$P(\chi_6^2 > 3,4)$$

según lo anterior:

$$P(\chi_6^2 > 3,4) = 1 - P(\chi_6^2 < 3,4)$$

buscando en la tabla tenemos:

$$P(\chi_6^2 < 3,4) = 0,242777$$

con lo que tenemos:

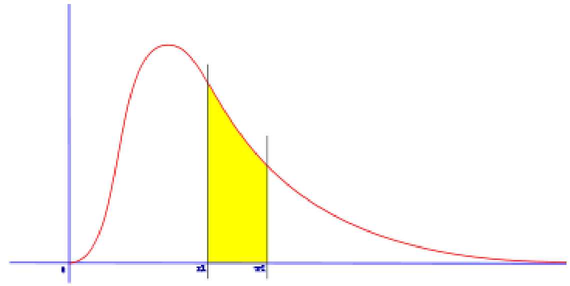
$$P(\chi_6^2 > 3,4) = 1 - 0,242777$$

operando tenemos:

$$P(\chi_6^2 > 3,4) = 0,757223$$

que es la respuesta a la pregunta.

Para la variable mayor que x1 y menor que x2 [editar]



Para calcular la probabilidad de que:

$$P(x_1 < \chi_k^2 < x_2)$$

siendo:

$$x_1 < x_2$$

tenemos que:

$$P(x_1 < \chi_k^2 < x_2) = P(\chi_k^2 < x_2) - P(\chi_k^2 < x_1)$$

Ejemplo [editar]

Cual es la probabilidad de que una variable chi-cuadrado de 8 grados de libertad este comprendida entre 3,4 y 5,6.

Esto es:

$$P(3,4 < \chi_8^2 < 5,6)$$

según la tabla tenemos:

$$\begin{cases} P(\chi_8^2 < 3,4) = 0,093189 \\ P(\chi_8^2 < 5,6) = 0,308063 \end{cases}$$

según lo anterior, tenemos que:

$$P(3,4 < \chi_8^2 < 5,6) = P(\chi_8^2 < 5,6) - P(\chi_8^2 < 3,4)$$

sustituyendo los valores:

$$P(3,4 < \chi_8^2 < 5,6) = 0,308063 - 0,093189$$

operando:

$$P(3,4 < \chi_8^2 < 5,6) = 0,214874$$

Con lo que tenemos la respuesta.

Distribución Ji cuadrada

	α																						
v	**0.999**	**0.995**	**0.99**	**0.98**	**0.975**	**0.95**	**0.9**	**0.8**	**0.75**	**0.7**	**0.6**	**0.5**	**0.4**	**0.3**	**0.25**	**0.2**	**0.1**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**
1	2E-06	.00004	.00016	0.001	0.001	0.004	0.016	0.064	0.102	0.148	0.275	0.455	0.708	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.828
2	0.002	0.010	0.020	0.040	0.051	0.103	0.211	0.446	0.575	0.713	1.022	1.386	1.833	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	0.024	0.072	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	1.869	2.366	2.946	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.267
4	0.091	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	2.753	3.357	4.045	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.467
5	0.210	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	3.656	4.351	5.132	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	0.381	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	4.570	5.348	6.211	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.458
7	0.599	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	5.493	6.346	7.283	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	0.857	1.344	1.646	2.032	2.180	2.733	3.490	4.594	5.071	5.527	6.423	7.344	8.351	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.125
9	1.152	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	7.357	8.343	9.414	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	1.479	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	8.295	9.342	10.473	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	1.834	2.603	3.054	3.609	3.816	4.575	5.578	6.989	7.584	8.148	9.237	10.341	11.530	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	2.214	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	10.182	11.340	12.584	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.299	32.909
13	2.617	3.565	4.107	4.765	5.009	5.892	7.042	8.634	9.299	9.926	11.129	12.340	13.636	15.119	15.984	16.985	19.812	22.362	24.736	25.472	27.688	29.819	34.529
14	3.041	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	12.078	13.339	14.685	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	3.483	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	13.030	14.339	15.733	17.322	18.245	19.311	22.307	24.996	27.488	28.260	30.578	32.801	37.698
16	3.942	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	13.983	15.339	16.780	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.253
17	4.416	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	14.937	16.338	17.824	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.719	40.790
18	4.905	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	15.893	17.338	18.868	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	5.407	6.844	7.633	8.567	8.906	10.117	11.651	13.716	14.562	15.352	16.850	18.338	19.910	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.821
20	5.921	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	17.809	19.337	20.951	22.775	23.828	25.037	28.412	31.410	34.170	35.020	37.566	39.997	45.315
21	6.446	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	18.768	20.337	21.992	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.797
22	6.983	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	19.729	21.337	23.031	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	7.529	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	20.690	22.337	24.069	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	8.085	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	21.652	23.337	25.106	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.559	51.179
25	8.649	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	22.616	24.337	26.143	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.620
26	9.222	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792	23.579	25.336	27.179	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.052
27	9.803	11.808	12.879	14.125	14.573	16.151	18.114	20.703	21.749	22.719	24.544	26.336	28.214	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.477
28	10.391	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647	25.509	27.336	29.249	31.391	32.621	34.027	37.916	41.337	44.461	45.419	48.278	50.993	56.892
29	10.986	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577	26.475	28.336	30.283	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.336	58.301
30	11.588	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508	27.442	29.336	31.316	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702
31	12.196	14.458	15.655	17.042	17.539	19.281	21.434	24.255	25.390	26.440	28.409	30.336	32.349	34.598	35.887	37.359	41.422	44.985	48.232	49.226	52.191	55.003	61.098
32	12.810	15.134	16.362	17.783	18.291	20.072	22.271	25.148	26.304	27.373	29.376	31.336	33.381	35.665	36.973	38.466	42.585	46.194	49.480	50.487	53.486	56.328	62.487
33	13.431	15.815	17.073	18.527	19.047	20.867	23.110	26.042	27.219	28.307	30.344	32.336	34.413	36.731	38.058	39.572	43.745	47.400	50.725	51.743	54.775	57.648	63.870
34	14.057	16.501	17.789	19.275	19.806	21.664	23.952	26.938	28.136	29.242	31.313	33.336	35.444	37.795	39.141	40.676	44.903	48.602	51.966	52.995	56.061	58.964	65.246
35	14.688	17.192	18.509	20.027	20.569	22.465	24.797	27.836	29.054	30.178	32.282	34.336	36.475	38.859	40.223	41.778	46.059	49.802	53.203	54.244	57.342	60.275	66.620
40	17.916	20.707	22.164	23.838	24.433	26.509	29.051	32.345	33.660	34.872	37.134	39.335	41.622	44.165	45.616	47.269	51.805	55.759	59.342	60.436	63.691	66.766	73.402
45	21.250	24.311	25.901	27.720	28.366	30.612	33.350	36.884	38.291	39.585	41.995	44.335	46.761	49.452	50.985	52.729	57.505	61.656	65.410	66.555	69.957	73.166	80.077
50	24.674	27.991	29.707	31.664	32.357	34.764	37.689	41.449	42.942	44.313	46.864	49.335	51.892	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	86.660
55	28.174	31.735	33.570	35.659	36.398	38.958	42.060	46.036	47.610	49.055	51.739	54.335	57.016	59.980	61.665	63.577	68.796	73.311	77.380	78.619	82.292	85.749	93.167
60	31.738	35.535	37.485	39.699	40.482	43.188	46.459	50.641	52.294	53.809	56.620	59.335	62.135	65.227	66.981	68.972	74.397	79.082	83.298	84.580	88.379	91.952	99.607
65	35.361	39.383	41.444	43.779	44.603	47.450	50.883	55.262	56.990	58.573	61.506	64.335	67.249	70.462	72.285	74.351	79.973	84.821	89.177	90.501	94.422	98.105	105.99
70	39.036	43.275	45.442	47.893	48.758	51.739	55.329	59.898	61.698	63.346	66.396	69.334	72.358	75.689	77.577	79.715	85.527	90.531	95.023	96.388	100.43	104.21	112.32
75	42.757	47.206	49.475	52.039	52.942	56.054	59.795	64.547	66.417	68.127	71.290	74.334	77.464	80.908	82.858	85.066	91.061	96.217	100.84	102.24	106.39	110.29	118.60
80	46.520	51.172	53.540	56.213	57.153	60.391	64.278	69.207	71.145	72.915	76.188	79.334	82.566	86.120	88.130	90.405	96.578	101.88	106.63	108.07	112.33	116.32	124.84
90	54.155	59.196	61.754	64.635	65.647	69.126	73.291	78.558	80.625	82.511	85.993	89.334	92.761	96.524	98.650	101.05	107.56	113.15	118.14	119.65	124.12	128.30	137.21
100	61.918	67.328	70.065	73.142	74.222	77.929	82.358	87.945	90.133	92.129	95.808	99.334	102.95	106.91	109.14	111.67	118.50	124.34	129.56	131.14	135.81	140.17	149.45
150	102.11	109.14	112.67	116.61	117.98	122.69	128.28	135.26	137.98	140.46	145.00	149.33	153.75	158.58	161.29	164.35	172.58	179.58	185.80	187.68	193.21	198.36	209.26

Teorema 8.4: Si S^2 es la varianza de una muestra aleatoria de tamaño n que se toma de una población normal que tiene la varianza σ^2, entonces el estadístico

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

tiene una distribución chi cuadrada con $v = n - 1$ grados de libertad.

Ejemplo 8.7: Un fabricante de baterías para automóvil garantiza que su producto durará, en promedio, 3 años con una desviación estándar de 1 año. Si cinco de estas baterías tienen duraciones de 1.9, 2.4, 3.0, 3.5 y 4.2 años, ¿el fabricante continuará convencido de que sus baterías tienen una desviación estándar de 1 año? Suponga que las duraciones de las baterías siguen una distribución normal.

Solución: Primero se calcula la varianza de la muestra usando el teorema 8.1,

$$s^2 = \frac{(5)(48.26) - (15)^2}{(5)(4)} = 0.815.$$

Entonces,

$$\chi^2 = \frac{(4)(0.815)}{1} = 3.26$$

es un valor de una distribución chi cuadrada con 4 grados de libertad. Como 95% de los valores χ^2 con 4 grados de libertad cae entre 0.484 y 11.143, el valor calculado con $\sigma^2 = 1$ es razonable y, por lo tanto, el fabricante no tiene razones para sospechar que la desviación estándar no sea igual a 1 año.

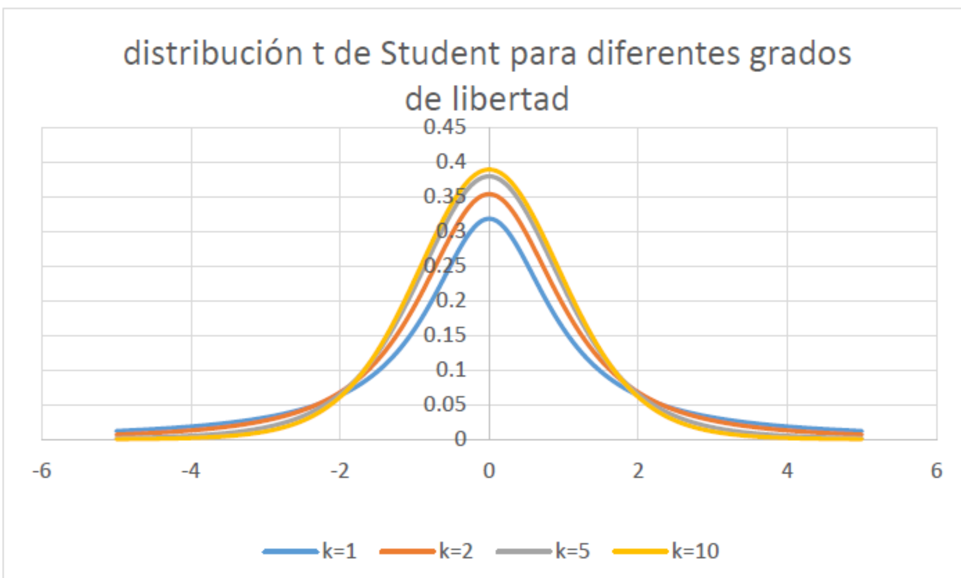
LA DISTRIBUCIÓN t DE STUDENT.

La distribución *t* de Student la propuso en 1908 William S. Gosset, quien usaba el seudónimo de Student en sus publicaciones y se define así:

Definición 1.2 Sean *X* y *Y* variables aleatorias independientes, tales que $X \sim N(0,1)$ y $Y \sim$ ji-cuadrada con *n* grados de libertad, entonces la variable:

$$W = \frac{X}{\sqrt{Y/n}} \tag{1.1}$$

Es una variable ***t*** de Student con ***n*** grados de libertad. Los grados de libertad corresponden a los grados de libertad de la ji-cuadrada del denominador.



La mayoría de las tablas reportan los valores de la distribución acumulada *t* de Student, esto es si $X \sim t$ con *n* grados de libertad, en la tabla se dan los valores de

$$P(X < x) = \alpha$$

En el renglón superior de la tabla se hallan los valores para

$$\alpha = 0.75, 0.90, 0.95, 0.975, 0.99, 0.995, 0.9975, 0.9999, 0.99995, 0.999975 \text{ y } 0.99999$$

En la primera columna están los grados de libertad y en las restantes columnas se encuentran los valores de x.

Ejemplo 2.

Sea X distribuida como *t* de Student con 10 grados de libertad, encontrar la probabilidad.

- $P(X < x) = 0.95$
- $P(X > x) = 0.05$
- $P(X < x) = 0.05$

Solución:

- La probabilidad $P(X < x) = 0.95$ se lee directamente en la tabla, de manera que se debe ir a la columna señalada como 0.95; en esta columna se observará el renglón de los grados de libertad igual que 10, ahí se encuentra el valor de $x = 1.812$.
- La probabilidad de $P(X > x) = 0.05$, no la trae la tabla, pues la variable aleatoria X está por arriba del valor x. Para poder utilizar las tablas, se determina el complemento de esta probabilidad $P(X < x) = 1 - 0.05 = 0.95$, ahora, en la columna correspondiente a 0.95 se busca el renglón de 10 grados de libertad y en esa celda se halla el valor de $x = 1.812$.
- La probabilidad de $P(X < x) = 0.05$ es una probabilidad acumulada, pero en la tabla no se encuentra el 0.05, de manera que para encontrar el valor de x se utiliza la simetría de la distribución t, entonces se satisface que $P(X < -x) = P(X > x) = 0.05$ y es el mismo valor del inciso anterior, $x = 1.812$.

Si el tamaño de la muestra es suficientemente grande, digamos $n \geq 30$, la distribución de T no difiere mucho de la normal estándar. Sin embargo, para $n < 30$ es útil tratar con la distribución exacta de T. Para desarrollar la distribución muestral de T, supondremos que nuestra muestra aleatoria se seleccionó de una población normal. Podemos escribir, entonces,

$$T = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}} = \frac{Z}{\sqrt{V/(n-1)}},$$

donde

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

tiene una distribución normal estándar y

$$V = \frac{(n-1)S^2}{\sigma^2}$$

tiene una distribución chi cuadrada con $v = n - 1$ grados de libertad. Al obtener muestras de poblaciones normales se puede demostrar que $\bar{X}$ y S^2 son independientes y, en consecuencia, también lo son Z y V. El siguiente teorema proporciona la definición de una variable aleatoria T como una función de Z (normal estándar) y χ^2. Para completar se proporciona la función de densidad de la distribución t.

Teorema 8.5: Sea Z una variable aleatoria normal estándar y V una variable aleatoria chi cuadrada con v grados de libertad. Si Z y V son independientes, entonces la distribución de la variable aleatoria T, donde

$$T = \frac{Z}{\sqrt{V/v}},$$

es dada por la función de densidad

$$h(t) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}, \quad -\infty < t < \infty.$$

Ésta se conoce como la **distribución *t*** con v grados de libertad.

A partir de lo antes expuesto, y del teorema anterior, se deriva el siguiente corolario.

Corolario 8.1: Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes normales con media μ y desviación estándar σ. Sea

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{y} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Entonces la variable aleatoria $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ tiene una distribución t con $v = n - 1$ grados de libertad.

Ejemplo 8.8: El valor t con $v = 14$ grados de libertad que deja una área de 0.025 a la izquierda y, por lo tanto, una área de 0.975 a la derecha, es

$$t_{0.975} = -t_{0.025} = -2.145.$$

Ejemplo 8.9: Calcule $P(-t_{0.025} < T < t_{0.05})$.

***Solución*:** Como $t_{0.05}$ deja una área de 0.05 a la derecha y $-t_{0.025}$ deja una área de 0.025 a la izquierda, obtenemos una área total de

$$1 - 0.05 - 0.025 = 0.925$$

entre $-t_{0.025}$ y $t_{0.05}$. En consecuencia,

$$P(-t_{0.025} < T < t_{0.05}) = 0.925.$$

Ejemplo 8.10: Calcule k tal que $P(k < T < -1.761) = 0.045$ para una muestra aleatoria de tamaño 15 que se selecciona de una distribución normal y $\frac{\overline{X} - \mu}{s/\sqrt{n}}$.

Ejemplo 8.10: Calcule k tal que $P(k < T < -1.761) = 0.045$ para una muestra aleatoria de tamaño 15 que se selecciona de una distribución normal y $\frac{\overline{X} - \mu}{s/\sqrt{n}}$.

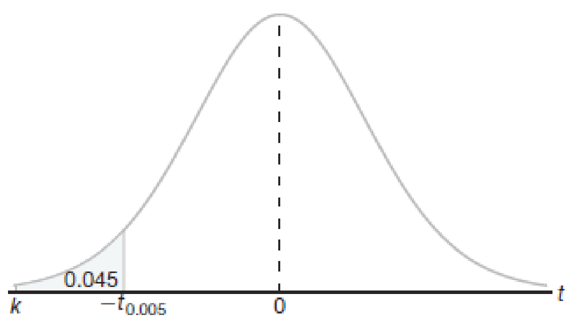


Figura 8.10: Valores t para el ejemplo 8.10.

***Solución*:** A partir de la tabla A.4 advertimos que 1.761 corresponde a $t_{0.05}$ cuando $v = 14$. Por lo tanto, $-t_{0.05} = -1.761$. Puesto que en el enunciado de probabilidad original k está a la izquierda de $-t_{0.05} = -1.761$, tenemos que $k = -t_\alpha$. Entonces, a partir de la figura 8.10, tenemos

$$0.045 = 0.05 - \alpha, \text{ o } \alpha = 0.005.$$

Así, de la tabla A.4 con $v = 14$,

$$k = -t_{0.005} = -2.977 \text{ y } P(-2.977 < T < -1.761) = 0.045.$$

Ejemplo 8.11: Un ingeniero químico afirma que el rendimiento medio de la población de un cierto proceso de lotes es 500 gramos por mililitro de materia prima. Para verificar dicha afirmación muestrea 25 lotes cada mes. Si el valor t calculado cae entre $-t_{0.05}$ y $t_{0.05}$, queda satisfecho con su afirmación. ¿Qué conclusión debería sacar de una muestra que tiene una media $\bar{x} = 518$ gramos por mililitro y una desviación estándar muestral $s = 40$ gramos? Suponga que la distribución de rendimientos es aproximadamente normal.

***Solución*:** En la tabla A.4 encontramos que $t_{0.05} = 1.711$ para 24 grados de libertad. Por lo tanto, el ingeniero quedará satisfecho con esta afirmación si una muestra de 25 lotes rinde un valor t entre -1.711 y 1.711. Si $\mu = 500$, entonces,

$$t = \frac{518 - 500}{40/\sqrt{25}} = 2.25,$$

un valor muy superior a 1.711. La probabilidad de obtener un valor t, con $v = 24$, igual o mayor que 2.25, es aproximadamente 0.02. Si $\mu > 500$, el valor de t calculado de la muestra sería más razonable. Por lo tanto, es probable que el ingeniero concluya que el proceso produce un mejor producto del que pensaba.

Distribución t de Student

v	α 0.001	0.002	0.003	0.004	0.005	0.01	0.015	0.02	0.025	0.05	0.075	0.1	0.2	0.3	0.4
1	318.29	159.14	106.10	79.572	63.656	31.821	21.205	15.894	12.706	6.314	4.165	3.078	1.376	0.727	0.325
2	22.328	15.764	12.852	11.113	9.925	6.965	5.643	4.849	4.303	2.920	2.282	1.886	1.061	0.617	0.289
3	10.214	8.052	6.994	6.322	5.841	4.541	3.896	3.482	3.182	2.353	1.924	1.638	0.978	0.584	0.277
4	7.173	5.951	5.321	4.908	4.604	3.747	3.298	2.999	2.776	2.132	1.778	1.533	0.941	0.569	0.271
5	5.894	5.030	4.570	4.262	4.032	3.365	3.003	2.757	2.571	2.015	1.699	1.476	0.920	0.559	0.267
6	5.208	4.524	4.152	3.898	3.707	3.143	2.829	2.612	2.447	1.943	1.650	1.440	0.906	0.553	0.265
7	4.785	4.207	3.887	3.667	3.499	2.998	2.715	2.517	2.365	1.895	1.617	1.415	0.896	0.549	0.263
8	4.501	3.991	3.705	3.507	3.355	2.896	2.634	2.449	2.306	1.860	1.592	1.397	0.889	0.546	0.262
9	4.297	3.835	3.573	3.390	3.250	2.821	2.574	2.398	2.262	1.833	1.574	1.383	0.883	0.543	0.261
10	4.144	3.716	3.472	3.301	3.169	2.764	2.527	2.359	2.228	1.812	1.559	1.372	0.879	0.542	0.260
11	4.025	3.624	3.393	3.231	3.106	2.718	2.491	2.328	2.201	1.796	1.548	1.363	0.876	0.540	0.260
12	3.930	3.550	3.330	3.175	3.055	2.681	2.461	2.303	2.179	1.782	1.538	1.356	0.873	0.539	0.259
13	3.852	3.489	3.278	3.128	3.012	2.650	2.436	2.282	2.160	1.771	1.530	1.350	0.870	0.538	0.259
14	3.787	3.438	3.234	3.089	2.977	2.624	2.415	2.264	2.145	1.761	1.523	1.345	0.868	0.537	0.258
15	3.733	3.395	3.197	3.056	2.947	2.602	2.397	2.249	2.131	1.753	1.517	1.341	0.866	0.536	0.258
16	3.686	3.358	3.165	3.028	2.921	2.583	2.382	2.235	2.120	1.746	1.512	1.337	0.865	0.535	0.258
17	3.646	3.326	3.138	3.003	2.898	2.567	2.368	2.224	2.110	1.740	1.508	1.333	0.863	0.534	0.257
18	3.610	3.298	3.113	2.982	2.878	2.552	2.356	2.214	2.101	1.734	1.504	1.330	0.862	0.534	0.257
19	3.579	3.273	3.092	2.962	2.861	2.539	2.346	2.205	2.093	1.729	1.500	1.328	0.861	0.533	0.257
20	3.552	3.251	3.073	2.945	2.845	2.528	2.336	2.197	2.086	1.725	1.497	1.325	0.860	0.533	0.257
21	3.527	3.231	3.056	2.930	2.831	2.518	2.328	2.189	2.080	1.721	1.494	1.323	0.859	0.532	0.257
22	3.505	3.214	3.041	2.916	2.819	2.508	2.320	2.183	2.074	1.717	1.492	1.321	0.858	0.532	0.256
23	3.485	3.198	3.027	2.904	2.807	2.500	2.313	2.177	2.069	1.714	1.489	1.319	0.858	0.532	0.256
24	3.467	3.183	3.014	2.892	2.797	2.492	2.307	2.172	2.064	1.711	1.487	1.318	0.857	0.531	0.256
25	3.450	3.170	3.003	2.882	2.787	2.485	2.301	2.167	2.060	1.708	1.485	1.316	0.856	0.531	0.256
26	3.435	3.158	2.992	2.873	2.779	2.479	2.296	2.162	2.056	1.706	1.483	1.315	0.856	0.531	0.256
27	3.421	3.146	2.982	2.864	2.771	2.473	2.291	2.158	2.052	1.703	1.482	1.314	0.855	0.531	0.256
28	3.408	3.136	2.973	2.856	2.763	2.467	2.286	2.154	2.048	1.701	1.480	1.313	0.855	0.530	0.256
29	3.396	3.127	2.965	2.848	2.756	2.462	2.282	2.150	2.045	1.699	1.479	1.311	0.854	0.530	0.256
30	3.385	3.118	2.957	2.841	2.750	2.457	2.278	2.147	2.042	1.697	1.477	1.310	0.854	0.530	0.256
31	3.375	3.109	2.950	2.835	2.744	2.453	2.275	2.144	2.040	1.696	1.476	1.309	0.853	0.530	0.256
32	3.365	3.102	2.943	2.829	2.738	2.449	2.271	2.141	2.037	1.694	1.475	1.309	0.853	0.530	0.255
33	3.356	3.094	2.937	2.823	2.733	2.445	2.268	2.138	2.035	1.692	1.474	1.308	0.853	0.530	0.255
34	3.348	3.088	2.931	2.818	2.728	2.441	2.265	2.136	2.032	1.691	1.473	1.307	0.852	0.529	0.255
35	3.340	3.081	2.926	2.813	2.724	2.438	2.262	2.133	2.030	1.690	1.472	1.306	0.852	0.529	0.255
40	3.307	3.055	2.902	2.792	2.704	2.423	2.250	2.123	2.021	1.684	1.468	1.303	0.851	0.529	0.255
45	3.281	3.034	2.884	2.776	2.690	2.412	2.241	2.115	2.014	1.679	1.465	1.301	0.850	0.528	0.255
50	3.261	3.018	2.870	2.763	2.678	2.403	2.234	2.109	2.009	1.676	1.462	1.299	0.849	0.528	0.255
55	3.245	3.004	2.859	2.752	2.668	2.396	2.228	2.104	2.004	1.673	1.460	1.297	0.848	0.527	0.255
60	3.232	2.994	2.849	2.744	2.660	2.390	2.223	2.099	2.000	1.671	1.458	1.296	0.848	0.527	0.254
65	3.220	2.984	2.841	2.736	2.654	2.385	2.219	2.096	1.997	1.669	1.457	1.295	0.847	0.527	0.254
70	3.211	2.977	2.834	2.730	2.648	2.381	2.215	2.093	1.994	1.667	1.456	1.294	0.847	0.527	0.254
75	3.202	2.970	2.828	2.725	2.643	2.377	2.212	2.090	1.992	1.665	1.454	1.293	0.846	0.527	0.254
80	3.195	2.964	2.823	2.720	2.639	2.374	2.209	2.088	1.990	1.664	1.453	1.292	0.846	0.526	0.254
90	3.183	2.954	2.815	2.713	2.632	2.368	2.205	2.084	1.987	1.662	1.452	1.291	0.846	0.526	0.254
100	3.174	2.946	2.808	2.706	2.626	2.364	2.201	2.081	1.984	1.660	1.451	1.290	0.845	0.526	0.254
150	3.145	2.923	2.787	2.688	2.609	2.351	2.191	2.072	1.976	1.655	1.447	1.287	0.844	0.526	0.254
∞	3.090	2.878	2.748	2.652	2.576	2.326	2.170	2.054	1.960	1.645	1.440	1.282	0.842	0.524	0.253

Distribución F

Usada en teoría de probabilidad y estadística, la **distribución F** es una distribución de probabilidad continua. También se le conoce como **distribución F de Snedecor** (por George Snedecor) o como **distribución F de Fisher-Snedecor** (por Ronald Fisher).

Una variable aleatoria de distribución *F* se construye como el siguiente cociente:

$$F = \frac{U_1/d_1}{U_2/d_2}$$

donde

- U_1 y U_2 siguen una distribución chi-cuadrado con d_1 y d_2 grados de libertad respectivamente, y
- U_1 y U_2 son estadísticamente independientes.

La distribución *F* aparece frecuentemente como la *distribución nula* de una prueba estadística, especialmente en el análisis de varianza. Véase el test F.

La función de densidad de una $F(d_1, d_2)$ viene dada por

$$g(x) = \frac{1}{\mathrm{B}(d_1/2, d_2/2)} \left(\frac{d_1\,x}{d_1\,x + d_2}\right)^{d_1/2} \left(1 - \frac{d_1\,x}{d_1\,x + d_2}\right)^{d_2/2} x^{-1}$$

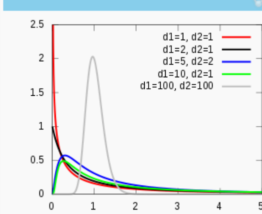
para todo número real $x \geq 0$, donde d_1 y d_2 son enteros positivos, y B es la función beta.

La función de distribución es

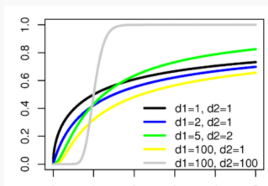
$$G(x) = I_{\frac{d_1 x}{d_1 x + d_2}}(d_1/2, d_2/2)$$

donde *I* es la función beta incompleta regularizada.

Fisher-Snedecor



Función de densidad de probabilidad



Distribuciones relacionadas [editar]

- $Y \sim \chi^2_{\nu_1}$ es una distribución ji-cuadrada cuando $Y = \lim_{\nu_2\to\infty} \nu_1 X$ para $X \sim \mathrm{F}(\nu_1, \nu_2)$.

Aplicaciones [editar]

- Pruebas de homocedasticidad

Enlaces externos [editar]

- Tablas de la distribución F de Fisher-Snedecor
- Distribution Calculator Calcula las probabilidades y valores críticos para las distribuciones normal, *t*, ji-cuadrada y *F*
- [1] Calcular la probabilidad de una distribución F-Snedecor con R (lenguaje de programación)

Función de distribución de probabilidad

Parámetros	$d_1 > 0,\ d_2 > 0$ grados de libertad
Dominio	$x \in [0; +\infty)$
Función de densidad (pdf)	$\dfrac{\sqrt{\frac{(d_1\,x)^{d_1}\,d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\left(\frac{d_1}{2},\frac{d_2}{2}\right)}$
Función de distribución (cdf)	$I_{\frac{d_1 x}{d_1 x+d_2}}(d_1/2, d_2/2)$
Media	$\dfrac{d_2}{d_2-2}$ para $d_2 > 2$
Moda	$\dfrac{d_1-2}{d_1}\ \dfrac{d_2}{d_2+2}$ para $d_1 > 2$
Varianza	$\dfrac{2\,d_2^2\,(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ para $d_2 > 4$
Coeficiente de simetría	$\dfrac{(2d_1+d_2-2)\sqrt{8(d_2-4)}}{(d_2-6)\sqrt{d_1(d_1+d_2-2)}}$ para $d_2 > 6$

[editar datos en Wikidata]

Ejercicios distribucion f fisher

832 palabras | 4 páginas

"DISTRIBUCION F DE FISHER"

1.- Para una distribución F encuentra:

a) f0.05 con gl=7 y gl=15; = 2.71

b) f0.05 con gl=15 y gl=7; = 3.51

c) f0.01 con gl=24 y gl=19; = 2.92

d) f0.95 con gl=19 y gl=24; = 1/F = 1/(24,19,0.05) = 1/ 2.11 = 0.4339

e) f0.99 con gl=28 y gl=12; = 1/F = 1/(12,28,0.01) = 1/ 2.9 = 0.3448

2.- Encuentra el valor:

a) F(10,20,0.95) y un extremo derecho.

1F20,10,0.05=12.77=0.3610

b) F(10,15,0.99) y extremo izquierdo

F(15,10,0.01) = 4.56

c) F(16,18,0.02) y ambos extremos. No existe valor en tablas d) F(8,20,0.01) y un extremo izquierdo.

F(8,20,0.01)= 3.56

e) F(8,20,0.05) y un extremo derecho.

1F20,8,0.05=15.36=0.1865

Se requiere que la temperatura permanezca constante durante la operación de horneado. Se hizo un estudio para medir la varianza en la temperatura de los dos hornos en funcionamiento. Antes de que el termostato reestableciera la flama, la variancia en la temperatura del horno A fue iguala 2.4, resultante de 16 medidas. La variancia del horno B fue 3.2, resultante de 12 mediciones.. Proporciona esta información evidencia suficiente para concluir que existe una diferencia en las variancias para los dos hornos. Utiliza α=0.01.

Horno 1 | Horno 2 | N1= 16 | N2= 12 | σ2= 2.4 | σ2= 3.2 | F=16-13.212-12.4(1)=1.81

F(11,15,0.05) No existe en tablas

9.- Se realizó un estudio para decidir si hay o no la misma variabilidad en la presión sanguínea sistólica entre hombres y mujeres. Se utilizaron muestras aleatorias de 16 hombres y 13 mujeres para contrastar la afirmación de los investigadores en el sentido de que las variancias eran diferentes. Utiliza α=0.05 y los siguientes datos:

Hombres: 120 120 118 112 120 114 130 114 124 125 130 100 120 108 112 122
Mujeres: 122 102 118 126 108 130 104 116 102 122 120 118 130

S12=62.19
S22=98.34

F=16-198.3413-162.34=1.97

6.10 Distribución de Weibull (opcional)

La tecnología actual permite que los ingenieros diseñen muchos sistemas complicados cuya operación y seguridad dependen de la confiabilidad de los diversos componentes que conforman los sistemas. Por ejemplo, un fusible se puede quemar, una columna de acero se puede torcer o un dispositivo sensor de calor puede fallar. Componentes idénticos, sujetos a idénticas condiciones ambientales, fallarán en momentos diferentes e impredecibles. Ya examinamos el papel que desempeñan las distribuciones gamma y exponencial en estos tipos de problemas. Otra distribución que se ha utilizado ampliamente en años recientes para tratar con tales problemas es la **distribución de Weibull**, introducida por el físico sueco Waloddi Weibull en 1939.

Distribución de Weibull La variable aleatoria continua X tiene una **distribución de Weibull**, con parámetros α y β, si su función de densidad es dada por

$$f(x; \alpha, \beta) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}, & x > 0, \\ 0, & \text{en otro caso,} \end{cases}$$

donde $\alpha > 0$ y $\beta > 0$.

En la figura 6.30 se ilustran las gráficas de la distribución de Weibull para $\alpha = 1$ y diversos valores del parámetro β. Vemos que las curvas cambian de manera considerable para diferentes valores del parámetro β. Si permitimos que $\beta = 1$, la distribución de Weibull se reduce a la distribución exponencial. Para valores de $\beta > 1$ las curvas adoptan ligeramente la forma de campana y se asemejan a las curvas normales, pero muestran algo de asimetría.

La media y la varianza de la distribución de Weibull se establecen en el siguiente teorema. Se solicita al lector que haga la demostración en el ejercicio 6.52 de la página 206.

Teorema 6.8: La media y la varianza de la distribución de Weibull son

$$\mu = \alpha^{-1/\beta}\Gamma\left(1 + \frac{1}{\beta}\right) \text{ y } \sigma^2 = \alpha^{-2/\beta}\left\{\Gamma\left(1 + \frac{2}{\beta}\right) - \left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2\right\}.$$

Al igual que la distribución gamma y la exponencial, la distribución de Weibull se aplica a problemas de confiabilidad y de prueba de vida como los de **tiempo de operación**

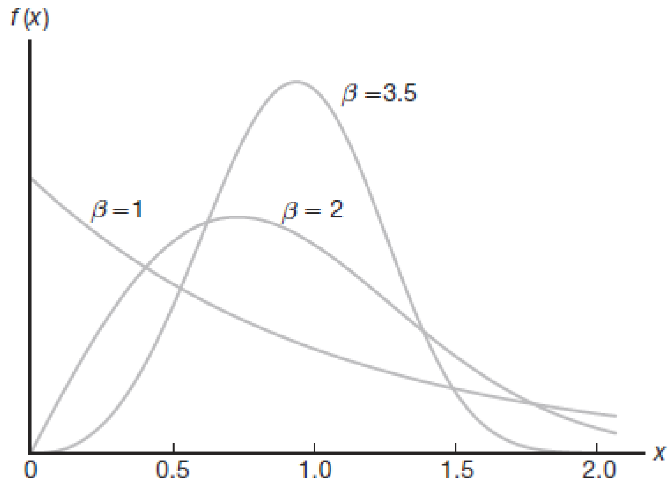


Figura 6.30: Distribuciones de Weibull ($\alpha = 1$).

antes de la falla o la **duración de la vida** de un componente, que se miden desde algún tiempo específico hasta que falla. Representemos este tiempo de operación antes de la falla mediante la variable aleatoria continua T, con función de densidad de probabilidad $f(t)$, donde $f(t)$ es la distribución de Weibull. Ésta tiene la flexibilidad inherente de no requerir la propiedad de falta de memoria de la distribución exponencial. La función de distribución acumulativa (fda) para la distribución de Weibull se puede escribir en forma cerrada y realmente es muy útil para calcular probabilidades.

Fda para la distribución de Weibull

La función de distribución acumulativa para la distribución de Weibull es dada por

$$F(x) = 1 - e^{-\alpha x^{\beta}}, \quad \text{para} \quad x \geq 0,$$

para $\alpha > 0$ y $\beta > 0$.

Ejemplo 6.24: El tiempo de vida X, en horas, de un artículo en el taller mecánico tiene una distribución de Weibull con $\alpha = 0.01$ y $\beta = 2$. ¿Cuál es la probabilidad de que falle antes de 8 horas de uso?

Solución: $P(X < 8) = F(8) = 1 - e^{-(0.01)8^2} = 1 - 0.527 = 0.473.$

La duración de un componente eléctrico sigue una *distribución exponencial* con media 10000 horas. Se pide:

a) Calcular la probabilidad de que si el componente ha durado más de 20000 horas, dure más de 21000 horas. Comparar esta probabilidad con la probabilidad de que dure entre 0 y 1000 horas. Comentar razonadamente el resultado.

b) Si se instalan 4 de esos componentes en serie en un aparato, calcular la probabilidad de que el aparato siga funcionando al cabo de 10000 horas.

SOLUCIÓN:

a) $\Pr(T > 21000|T > 20000) = \frac{e^{-2.1}}{e^{-2}} = 0.905 = e^{-0.1} = \Pr(T > 1000)$;

b) $\Pr(Funcione) = \Pr(T > 10000)^4 = 0.018$.

Sea X una variable aleatoria de Weibull de parámetro $\beta > 1$ que representa la duración de un componente hasta que se averíe.
Para montar un circuito, buscamos componentes que nos duren al menos 500 unidades de tiempo.
Para seleccionar esos componentes nos dan a elegir entre 2 *tipos*.
Los componentes del *Tipo 1* están sin estrenar, mientras que los componentes del *Tipo 2* no son nuevos.
¿Que tipo de componentes es el más adecuado para nosotros?

SOLUCIÓN:

Si X es una Weibull de parámetro $\beta > 1$, el componente envejece.
Cada vez le resulta mas difícil sobrevivir; es decir,

$$\Pr(X > t_0 + t|X > t_0) < \Pr(X > t).$$

Por tanto nos interesa el componente sin estrenar.

Ejercicio Fiabilidad

El tiempo T en segundos que tarda en conectarse a un servidor durante un día laborable sigue una distribución de Weibull de parámetros $\alpha = 0.6$ y $\beta = 1/4$, mientras que un fin de semana es una Weibull de parámetros $\alpha = 0.24$ y $\beta = 1$, donde la densidad de la Weibull se escribe como

$$f(t) = \frac{\beta}{\alpha^\beta} t^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}.$$

Se quiere saber:

(a) Tiempo medio que tardaremos en conectarnos en ambos tipos de día;

(b) Calcula, para ambos tipos de día, la probabilidad de tardar más de 10 segundos en realizar la conexión

(c) Si llevamos ya 5 segundos esperando a que se efectué la conexión, ¿cual es la probabilidad de que la conexión se demore aun 10 segundos más?

(d) ¿Era de esperar el resultado que se obtiene en (c)?.

SOLUCIÓN:

L = laborable; F= fin de semana;

(a) $\mu_L = 14.4$ seg., $\mu_F = 0.24$ seg. ;

(b) $\Pr(T_L > 10) = 0.133$, $\Pr(T_F > 10) \approx 0$;

(c) $\Pr(T_L > 15|T_L > 5) = 0.5845$, $\Pr(T_F > 15|T_F > 5) = \Pr(T_F > 10)$;

(d) El resultado era previsible al ser $\beta_L < 1$;

Análisis Weibull: Ejemplos Básicos de como usarlo para los Análisis de Confiabilidad

Introducción

El análisis de Weibull es la técnica mayormente escogida para estimar una probabilidad basada en datos medidos o asumidos. La distribución de Weibull, descubierta por el sueco Walodi Weibull, fue anunciada por primera vez en un escrito en 1951.

El modelo Weibull tiene una interesante propiedad ligada a que según sean los valores de, puede presentar tasas de fallo crecientes, decrecientes o constantes. Así, cuando β=1 el modelo Weibull se convierte en exponencial y presenta tasa de fallos constante. El modelo exponencial es por tanto un caso particular del modelo Weibull, cuando β>1 el modelo tiene tasa de falla creciente y cuando β<1 presenta tasa de falla decreciente. El modelo Weibull es muy versátil y en la práctica es uno de los más utilizados.

El objetivo de este artículo es presentar una serie de ejemplos prácticos basada en mi concepto de compartir el conocimiento, que permitan dar una visión general del uso del análisis Weibull.

1. Función de Densidad de Probabilidad

No está contemplado dar un curso de estadística, sin embargo, hay conceptos básicos que tenemos que tener claros para el desarrollo que cualquier análisis probabilístico. Como por ejemplo, qué es la Función de Densidad de Probabilidad; esta "caracteriza del comportamiento de una población[1] en tanto especifica la posibilidad relativa de que una variable aleatoria continua X tome un valor cercano a x. ".

La función de Densidad (f (t) es la Función de Densidad de Probabilidad (PDF)) y la Función de Distribución, estas funciones sirven para estudiar los datos de duración. La PDF es a menudo calculada a partir de, en nuestro caso, datos de fallas reales.

F(t) es la función de distribución acumulada (CDF). Es el área bajo la curva f(t) de 0 a t. (algunas veces llamada la **no confiabilidad o probabilidad de falla acumulada**).

- **F(t)=∫f(t)dt**
- **F(t) = Pr(T ≤ t) =** $\int_0^t f(x)dx$

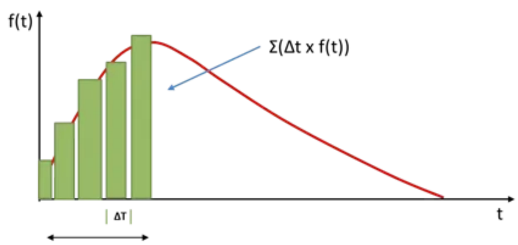


Figura 1. F(t) es la función de distribución acumulada(CDF).
Fuente: El autor.

Para el caso de análisis Weibull las funciones son las siguientes:

• Distribución de densidad de probabilidad (PDF)[3]

$$f(t) = \left(\frac{\beta t^{\beta-1}}{\eta^{\beta}}\right) e^{\left(-\frac{t}{\eta}\right)^{\beta}} \quad \textbf{(1)}$$

•Distribución de probabilidad acumulada (CDF)

$$F(t) = 1 - e^{\left(-\frac{t}{\eta}\right)^{\beta}} \quad \textbf{(2)}$$

• Parámetro α

$$\hat{\alpha} = \left(\frac{\sum_{i=1}^{n} X_i^{\beta}}{n}\right)^{1/\beta} \quad \textbf{(3)}$$

• Parámetro β

$$\frac{\sum_{i=1}^{n}\left[X_i^{\beta}\ln(t_i)\right]}{\sum_{i=1}^{n}X_i^{\beta}}-\frac{1}{\beta}=\frac{1}{n}\sum_{i=1}^{n}\ln(X_i) \quad \textbf{(4)}$$

Donde:

- t – es la variable de tiempo.
- β – es el parámetro de forma.
- η – es el parámetro de escala o característica de vida.

La distribución de Weibull es útil por su habilidad para simular un amplio rango de distribuciones como la Normal, la Exponencial, etc. Las técnicas discutidas en la distribución de Weibull son similares a las usadas con las distribuciones Normal y Log-Normal[2].

Supongamos que usted quiere entrar al negocio de ventas en mercado electrónico, realizó una investigación de los 10 productos que más se venden y escogió el producto numero dos (2), "Accesorios para teléfonos móviles", según los "Blogs de Shopify", para iniciarse en el mundo de las ventas online. Ahora la pregunta del millón: ¿Cuánto tengo que comprar o invertir para revender y poder generar utilidad? En vista de su nulo conocimiento de este negocio, se consigue un socio capitalista que está empapado del mercado y este le dice, una vez que ya calculó el precio de venta que quiere, que usted compre ***solamente los suficientes "Accesorios para teléfonos móviles" de tal manera que, una vez terminadas las ventas, solo le queden sin vender aproximadamente el 10% de estos accesorios***. Usted, como buen ingeniero de mantenimiento y conocedor de estadística, decide hacer un **ANÁLISIS DE WEIBULL** (dado que a nosotros los ingenieros nos gusta complicarnos la vida y "si lo puedes hacer difícil, ¿Por qué hacerlo fácil?"). Además, la cantidad de "Accesorios" a vender es un número aleatoriamente variable.

Nuevamente realizaron un análisis de mercado con sus posibles competidores, con la ayuda del socio y apoyados por un software o simplemente a mano, tabularon los datos de venta promedio de diez posibles competidores y procedieron a calcular lo que se conoce como **RANGO MEDIO** (se puede encontrar en los libros de estadística), para cada uno de ellos. El Rango Medio es un número entre 0 a 1 que refleja en orden ascendente la fracción del valor del dato que es menor que el mismo dato. Así, se obtiene la Tabla No. 1, una vez ordenados los datos:

ORDEN	DATO	RANGO MEDIO
1	1,500	6.73%
2	2,000	16.35%
3	2,300	25.96%
4	2,700	35.58%
5	3,100	45.19%
6	3,350	54.81%
7	3,700	64.42%
8	3,983	74.04%
9	4,283	83.65%
10	4,583	93.27%

Tabla 1. Resumen de ventas promedios de sus competidores.
Fuente: El autor.

Como resultado inicial de acuerdo a lo indicado en la tabla No.1, hay un 93.26% en el que la cantidad de accesorios a vender será menor al 4,583 ya que el rango medio es 0.93269. Pero, ¿Cuánto comprar para revender y poder cumplir con el requerimiento del socio capitalista?

Continuando con el desarrollo del ANÁLISIS DE WEIBULL y con una hoja en Excel, dibujamos en un gráfico el doble logaritmo de los datos en cuestión, y obtuvimos el Gráfico de Weibull (ver Figura No. 2).

Utilizando el "Excel", calculamos a la ecuación de la recta de la hoja logarítmica y obtuvimos la siguiente:

- **$Y = 5E\text{-}09x^{2.2725}$ (5)**
- **$R^2 = 0.9815$ (6)**

La pendiente de la línea recta que intercepta la mayoría de los puntos en el Gráfico de Weibull, es también el Factor de Forma β (2.2725) e indica a qué tipo de distribución de probabilidad se aproxima (normal, lognormal, exponencial, etc.). La vida característica, η, es el momento en que se espera que sea el 63,2% del Rango Medio de la línea recta. Este 63.2% es cierto para todas las distribuciones de Weibull, independientemente del parámetro de forma β, que corresponde cuando t = η (aproximadamente 3,800) en la ecuación 1, como lo indica la IEC 61649-2008 – Weibull analysis.

Volviendo a nuestro ejemplo para un 10% de ventas no satisfechas, se localiza en la línea recta con una probabilidad de 0.90 (90%). Para este dato, el punto ocurre en aproximadamente los 4,250 accesorios. (Y justo en ese momento nos dimos cuenta para qué estudiamos probabilidad y estadísticas en la universidad).

Con el Gráfico de Weibull, usted puede hacer estimaciones de probabilidades utilizando la línea recta, o simplemente leyendo la probabilidad en la escala vertical, para un dato (en este caso, un número de accesorios).

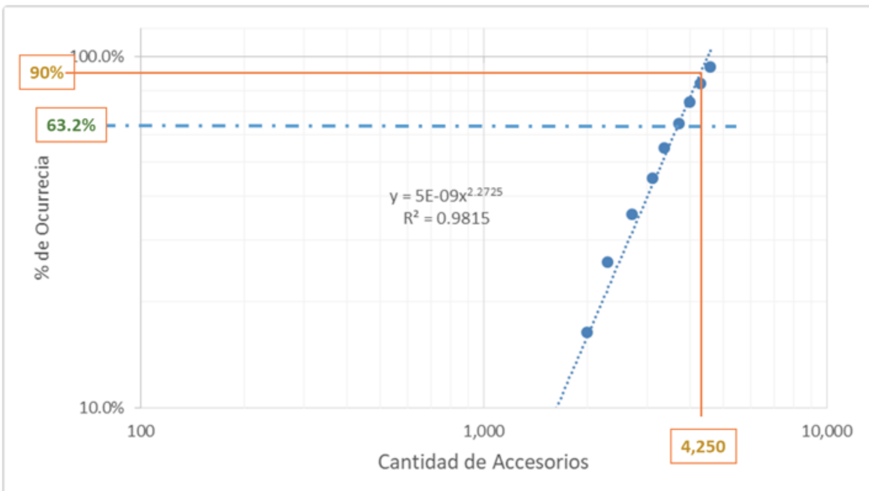


Figura 2. Gráfico de Weibull.
Fuente: El autor.

De acuerdo a la gráfica de Weibull, el número estimado de "Accesorios para teléfonos móviles" a comprar, para un 10% de accesorios no vendidos, se localiza en la línea recta con una probabilidad de 0.90 (90%). Para este dato, el punto ocurre en aproximadamente los 4,250 accesorios.

Utilizando el "Excel", calculamos a la ecuación de la recta de la hoja logarítmica y obtuvimos la siguiente:

- $Y = 5E\text{-}09x^{2.2725}$ (5)
- $R^2 = 0.9815$ (6)

La pendiente de la línea recta que intercepta la mayoría de los puntos en el Gráfico de Weibull, es también el Factor de Forma β (2.2725) e indica a qué tipo de distribución de probabilidad se aproxima (normal, lognormal, exponencial, etc.). La vida característica, η, es el momento en que se espera que sea el 63,2% del Rango Medio de la línea recta. Este 63.2% es cierto para todas las distribuciones de Weibull, independientemente del parámetro de forma β, que corresponde cuando t = η (aproximadamente 3,800) en la ecuación 1, como lo indica la IEC 61649-2008 – Weibull analysis.

Volviendo a nuestro ejemplo para un 10% de ventas no satisfechas, se localiza en la línea recta con una probabilidad de 0.90 (90%). Para este dato, el punto ocurre en aproximadamente los 4,250 accesorios. (Y justo en ese momento nos dimos cuenta para qué estudiamos probabilidad y estadísticas en la universidad).

Con el Gráfico de Weibull, usted puede hacer estimaciones de probabilidades utilizando la línea recta, o simplemente leyendo la probabilidad en la escala vertical, para un dato (en este caso, un número de accesorios).

Ejemplo de Análisis Weibull aplicado a Fallas de Equipos

La técnica de Análisis Weibull puede ser usada para estimar probabilidad de muchos casos, por lo que continuando con nuestros ejemplos, nos enfocaremos en un análisis de confiabilidad asociado a una estadística de fallas de equipos de una planta Fraccionadora de Gas, para lo cual tenemos la siguiente estadística de Tiempo Para Falla (TPF).

No. Orden (i)	TPF	Rango Medio
1	240.00	4.02%
2	264.00	9.77%
3	312.00	15.52%
4	408.00	21.26%
5	528.00	27.01%
6	576.00	32.76%
7	600.00	38.51%
8	672.00	44.25%
9	744.00	50.00%
10	840.00	55.75%
11	864.00	61.49%
12	936.00	67.24%
13	1,152.00	72.99%
14	1,992.00	78.74%
15	2,472.00	84.48%
16	5,832.00	90.23%
17	13,776.00	95.98%

Tabla 2. Datos estadísticos de TPF.
Fuente: El autor.

Analizando los datos de la tabla No. 2, como primera observación tenemos que hay un 95.98% que los TPF sean menores a 95.98%, por supuesto que es una información, pero no nos ayuda mucho ya que tenemos datos desde 240 horas a 13,776 horas, por lo que debemos analizar la gráfica de Weibull (Fig.3).

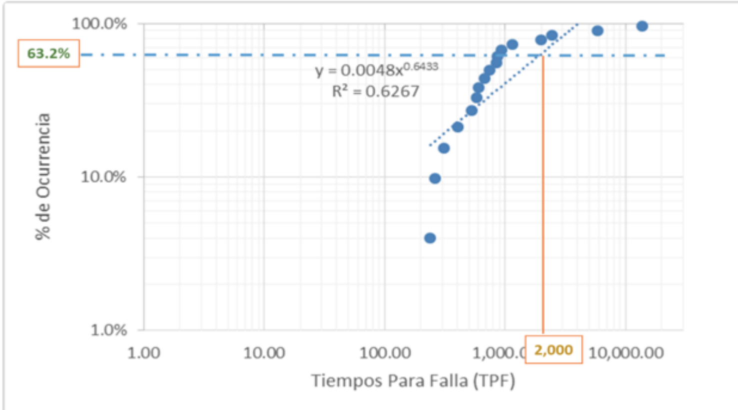


Figura 3. Gráfico de Weibull (TPF).
Fuente: El autor.

Del análisis tenemos que el Factor de Forma β es 0.6433 lo que indica que tenemos una tasa de falla descendente, al igual que una t = η (aproximadamente 2,000).

Ya con los datos de forma y escala podemos realizar cualquier estimado de confiabilidad utilizando la ecuación (7) o el gráfico de Weibull.

R(t) = 1-F(t) (7)

Por ejemplo, si queremos estimar cuál es la confiabilidad o probabilidad para que los equipos no fallen a t1=500 horas o t2=3000 horas, obtendremos lo siguiente:

- **R(t1) = 80%**
- **R(t2) = 10%**

Referencias

1. Wikipedia
2. FUNDAMENTOS DEL ANÁLISIS DE WEIBULL – Por Robert B. Abernethy, FL, USA
3. IEC 61649-2008 – Weibull analysis

Autor: Arquímedes Ferrera
Venezuela
Socio fundador de E&M Solutions Group., y Gerente General de E&M Solutions, en México
Empresa: E&M Solutions Group
Correo: arquimedes.ferrera@eymsolutions.com

DISTRIBUCIÓN LOG-NORMAL

1. INTRODUCCIÓN

Se trata de la densidad de probabilidad de una variable log x distribuida según una función normal:

$$X = N(\mu,\sigma) \qquad Y = e^X$$

Con este cambio de variable quedará:

Función de distribución: $G(y) = P(Y \le y) = P(e^X \le y) = P(X \le \log y) = F(\log y)$
Función de densidad : $g(y) = G'(y) = F'(\log y) * (1/y)$

$$\boxed{g(y) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{y}\exp\left(-\frac{(\log y-\mu)^2}{2\sigma^2}\right); \quad y \ge 0}$$

También es conocida como Ley de Galton-Mac. Aliester o ley del efecto proporcional, según Calot (1988).

Los parámetros principales que la caracterizan son:

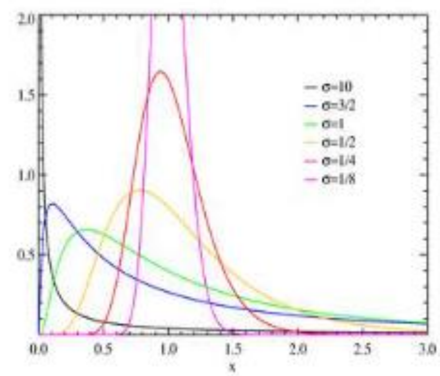
Parámetros	$s \ge 0$ $-\infty \le \mu \le \infty$
Soporte	$x \in [0; +\infty)$
pdf	$\frac{e^{-[\frac{\ln(x)-\mu}{\sigma}]^2/2]}}{x\sigma\sqrt{2\pi}}$
cdf	$\frac{1}{2}+\frac{1}{2}\mathrm{Erf}\left[\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right]$
Media	$e^{\mu+\sigma^2/2}$
Mediana	e^{μ}

1

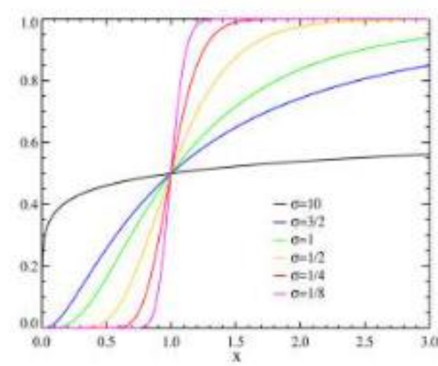
Moda	$e^{\mu-\sigma^2}$
Varianza	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$
Asimetría	$e^{-\mu-\sigma^2/2}(e^{\sigma^2}+2)\sqrt{e^{\sigma^2}-1}$
Curtosis	$e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-6$
Entropía	$\frac{1}{2}+\frac{1}{2}\ln(2\pi\sigma^2)$

A continuación se muestran unos gráficos de la función de distribución y de la función de densidad:

Función de densidad



Función de distribución



2

Puede comprobarse que la mediana está comprendida entre la moda[1] y la media y más cerca de la media que la moda, en particular, puede comprobarse que la mediana[2] está casi dos veces más cerca de la media que de la moda.

La distribución lognormal es una probabilidad frecuentemente utilizada para expresar el comportamiento de observaciones con asimetría positiva, en donde la mayoría de los valores ocurren en las proximidades de un valor mínimo.

Según Cabrera (1998), una condición para la validez de que una variable se distribuya Lognormal es que x sea la resultante de un número elevado de causas independientes con efectos positivos, que se componen de manera multiplicativa y cada una de estas causas tiene un efecto despreciable frente al global.

Esta distribución es característica en conjuntos de datos donde existe mayor frecuencia de valores pequeños, por lo cual la media se desplaza hacia la derecha y esto hace que el mejor estadígrafo de posición sea la moda y no la media aritmética (Conferencia UNACH, 1995). Esta consideración se valora, pero no se comparte en lo referente a la valoración del **centro** de los datos por considerarse que el mismo puede hallarse con más exactitud en el valor de la mediana, la cual se conoce no es influida por valores extremos, lo cual no ocurre con la moda. También se considera que otra medida de posición válida para esta distribución es la media geométrica (Peña, 1994).

ÍNDICE DE EVENTOS QUE PRESENTAN UNA DISTRIBUCIÓN LOG-NORMAL

1. Patrones de abundancia de especies.
2. Distribución log-normal de las concentraciones ambientales.
3. Modelo log-normal del precio de las acciones.
4. Análisis de la comunidad de una laguna costera en la costa sur occidental de México.
5. Cuantificación de la vitamina B2.
6. Distribución del peso molecular de los polímeros.
7. Predicción de sismos una ojeada al futuro.
8. Factores que afectan las tasas de captura de langostino amarillo en la zona norte de chile.
9. Comportamiento de las precipitaciones en el sector del lago Titicaca (Bolivia) durante el fenómeno "El Niño".
10. Producción de nanopartículas de Cobre.

[1] Es el valor para el cual le distribución toma su máximo absoluto.
[2] $F(x_{med})=P(X\leq x)=P(X\geq x)=0.5$

1. Patrones de abundancia de especies

El desarrollo de Fisher y colaboradores llamó la atención de Frank Preston, un ingeniero inglés. Motivado por los trabajos de Fisher, Preston publicó en 1948 un trabajo sobre la abundancia y la rareza de las especies biológicas que marcó el desarrollo de la teoría ecológica por varias décadas. El trabajo de Preston demostró que, aunque el modelo de Fisher era esencialmente correcto, el supuesto que la abundancia media es una característica fija en cada especie, era innecesariamente restrictivo. En efecto, el supuesto de Fisher implica que la cantidad de recursos que conquista inicialmente una especie en una comunidad permanece constante a lo largo del tiempo evolutivo, aunque nuevos competidores le disputen su nicho ecológico. En contraposición, Preston partió de un supuesto mucho menos restrictivo, basado en un razonamiento estrictamente demográfico, que puede demostrarse fácilmente mediante el cálculo. **Así logaritmo de la abundancia de una especie en una comunidad depende del logaritmo de la abundancia inicial y de las variaciones en su tasa real de crecimiento a lo largo del tiempo**. Preston infirió correctamente que si una especie se encuentra en equilibrio con su medio, las tasas reales de crecimiento oscilarán aleatoriamente, a veces incrementando la población cuando ocurren periodos favorables y a veces disminuyéndola cuando ocurren periodos desfavorables. **De acuerdo con el Teorema del Límite central (Teorema que demuestra que todas las variables estadísticas aditivas tienen una distribución normal de Gauss), el logaritmo del número de individuos tendrá una variación Gaussiana a lo largo del tiempo. El corolario que sacó Preston de esta demostración fue que, si la distribución del logaritmo de las abundancias de una especie varía de forma Gaussiana a lo largo del tiempo, entonces, en un tiempo dado, la distribución del logaritmo de las abundancias de varias especies en un sólo tiempo y lugar también debe variar normalmente. Es decir, que en un instante dado habrá algunas especies que se presenten en grandes abundancias y otras que lo hagan en cantidades mucho más bajas, pero el logaritmo de sus abundancias tendrá una distribución normal.**

En resumen, el modelo de Preston, conocido como modelo de log-normal predice que la cantidad de especies presenten en una comunidad tendrá una relación Gaussiana o normal en el logaritmo de sus abundancias.

Es probable que las distribuciones de las abundancias de las especies sean log-normales. A nivel de comunidad es de esperar que los factores que gobiernan la distribución y la abundancia de las especies sean independientes entre sí y que afecten de manera multiplicativa a las variables que de ellos dependen. En resumen, la distribución log-normal, es predecible cuando un conjunto de datos depende del producto de variables aleatorias. Los factores que definen la abundancia de las especies tienden a actuar de esta manera.

Una de las consecuencias más importante del modelo de Preston es que, al igual que el de Fisher, es capaz de describir la cantidad de especies a hallar en un área como una función de la cantidad de individuos en el total de la muestra. Pero a diferencia del modelo de Fisher, que predice un rango muy amplio de curvas, el modelo de Preston predice una relación del tipo: numero de especies igual a k por A elevado a z, donde A y z son coeficientes derivados del modelo. Para el caso

4

de comunidades con distribución de abundancias de tipo log-normal, Preston demostró que el valor del exponente z debería estar cerca de 0,23 en muestras relativamente grandes y un valor algo mayor en muestras pequeñas (May, 1975), pero la ecuación anterior puede linealizarse sacando logaritmos, de forma que:

$$\log(s)= \log(k) + z\log(A)$$

Es decir, que la relación entre el logaritmo del área y el logaritmo del número de especies para muestras de distintos tamaños debería dar una recta con pendiente aproximada de 0,23 y una pendiente algo mayor en datos de comunidades muy pobres en especies. Si así no ocurriera, deberíamos rechazar el modelo de log-normal de Preston como modelo subyacente en los patrones de abundancia y rareza de las especies biológicas.

www.geocities.com/CollegePark/ Classroom/7370/pagina5b.htm

2. La distribución log-normal de las concentraciones ambientales

La valoración higiénica clásica de un puesto de trabajo se efectúa comparando la exposición a contaminantes que sufre el trabajador que lo ocupa con las correspondientes "exposiciones máximas permisibles" contempladas en el criterio de valoración elegido. En general es la concentración media ponderada en el tiempo el parámetro básico a través del cual se cuantiza la exposición y su medición se realiza mediante un procedimiento de toma de muestras/análisis. Un puesto de trabajo queda caracterizado cuando se ha determinado su ciclo de trabajo, es decir, el mínimo conjunto ordenado de tareas que se repite idéntica y sucesivamente; entre dos ciclos cualesquiera no deben existir diferencias macroscópicamente observables. A efectos de valoración higiénica la exposición a contaminantes quedará caracterizada por la duración del ciclo de trabajo y las concentraciones medias existentes durante el mismo; en consecuencia las mediciones que se efectúen para determinar dichas concentraciones deberán cubrir uno o varios (pero siempre un número entero) de ciclos de trabajo.

Se ha demostrado experimentalmente que la concentración medida durante un determinado ciclo de trabajo es una variable aleatoria que sigue una distribución de probabilidad lognormal (es decir, que los logaritmos de dicha variable siguen una ley normal). Ello significa que las concentraciones pueden variar teóricamente entre cero e infinito, y que la probabilidad de que la concentración medida esté más o menos alejada de la concentración media real depende de la mayor o menor desviación típica (dispersión) de la distribución o, lo que es lo mismo, de la mayor o menor variabilidad de los factores aleatorios que influyen sobre la concentración.

La variabilidad de las concentraciones medidas suele ser, en la práctica, considerable. Como parámetro indicador de la misma acostumbra a emplearse la llamada desviación standard geométrica (GSD) de las concentraciones; la GSD es el antilogaritmo de la desviación standard de la distribución de los logaritmos de las concentraciones. La desviación standard geométrica puede variar

5

teóricamente desde 1 (concentración constante) hasta cualquier valor positivo superior a la unidad, aunque en la práctica los valores encontrados suelen hallarse en el intervalo de 1,25 a 2,5. (Figura 1).

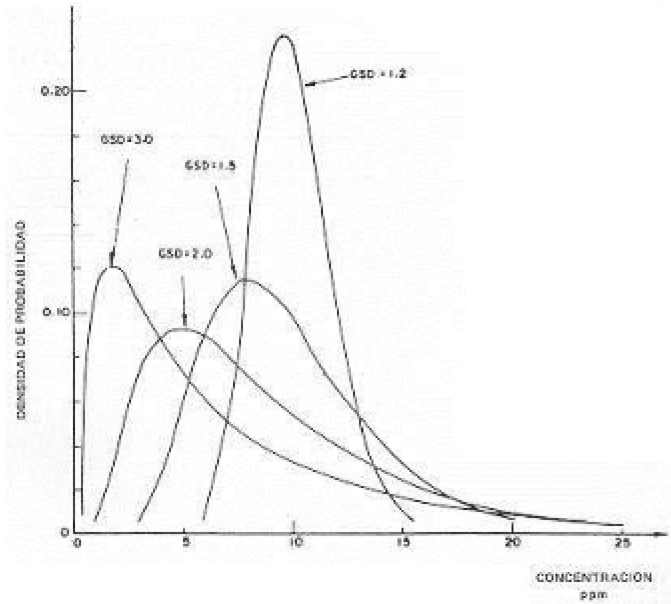


Fig. 1: Distribuciones lognormales de igual media aritmética (10 ppm) y distintos G.S.D.

Distribuciones lognormales de igual media aritmética (10 ppm) y distintos G.S.D.

En la tabla que se inserta a continuación se indica, en el supuesto de que la concentración media real fuese 10 ppm, la amplitud del intervalo en el que se encontrarían el 50% de las muestras obtenidas para distintos valores de GSD; ello implica por tanto que el 50% restante se encontrarían fuera de dicho intervalo.

GSD	INTERVALO, ppm
1,25	8,6 - 11,6
1,50	7,6 - 13,2
1,75	6,8 - 14,6
2,00	6,3 - 16,0
2,25	5,8 - 17,3
2,5	5,4 - 18,6

Conociendo el tipo de distribución que siguen las concentraciones ambientales es posible tratar los resultados obtenidos en una serie de muestras para llegar a una estimación de la concentración media real. Dicha estimación puede expresarse, a través de límites de confianza, en la forma: "la concentración media está comprendida entre A y B, con un α % de probabilidad."

www.mtas.es/insht/ntp/ntp_140.htm

6

3. Modelo log-normal del precio de las acciones

Abstract

El modelo de Black-Schooles se basa en el supuesto de que los precios de las acciones siguen lo que se conoce como distribución log-normal.
Mientras que una variable con distribución normal puede tomar valor positivo o negativo, una variable distribuida lognormalmente sólo puede ser positiva, con media, moda y mediana todas diferentes.

En las siguientes páginas trataremos de demostrar la hipótesis de log-normalidad de los precios de las acciones.

Introducción

¿Es razonable pensar que los precios de un activo subyacente se distribuyen de manera normal? Más allá de la exacta distribución de los precios en el mundo real, el supuesto de la distribución normal tiene serios defectos. Una curva de distribución normal es simétrica, por lo cual, bajo el supuesto de la normalidad, para todo posible incremento abrupto de precios en el activo subyacente existe la posibilidad de una caída en los mismos de igual magnitud. Es decir que, si por ejemplo permitimos la posibilidad de que cuando el activo subyacente vale \$50 éste pueda incrementarse en \$75 a \$125, también tendríamos que permitir la posibilidad de que los precios cayeran en igual magnitud a - \$25. Como todos sabemos es imposible que un activo adquiera un valor negativo, por lo que suponer que los mismos se distribuyen normalmente es una grave falencia1.

¿Qué podríamos hacer entonces al respecto? Si definimos volatilidad como el porcentaje de cambio en los precios de un activo subyacente, tasa de interés y volatilidad pueden considerarse similares en términos de que ambos representan tasas de retorno. La primera diferencia entre la tasa de interés y la volatilidad es que el interés generalmente acumula una tasa positiva mientras que la volatilidad representa una combinación de retornos positivos y negativos. Si invertimos una suma de dinero a una tasa fija, el valor del principal siempre se incrementará, pero si invertimos en un activo subyacente con una volatilidad distinta de cero, el precio del instrumento puede subir o bajar. La volatilidad, definida como desvío estándar, no nos dice nada acerca de la dirección que tomarán los movimientos de precios. De la misma manera que el interés, la volatilidad puede calcularse a diferentes intervalos. Con el propósito de valuar teóricamente las opciones, se asume que la volatilidad se calcula de manera continua (de la misma manera en que se dan los cambios en el precio del activo subyacente).

Cuando se asume que los cambios en los precios se distribuyen normalmente, el cálculo continuo de estos cambios causan que **los precios** al vencimiento **se distribuyan log -normalmente**. Tal distribución es simétrica a la derecha, debido a que los incrementos de precios resultan de tasas de retornos positivas cada vez más grandes.

La log-normalidad en el precio de las acciones

"Una variable tiene distribución lognormal si el logaritmo natural de la variable se distribuye normalmente".

Los parámetros claves que describen el comportamiento del precio de las acciones cuando se hace una hipótesis lognormal son:

1. el rendimiento esperado de las acciones
2. la volatilidad del precio de las acciones

La rentabilidad esperada es la rentabilidad media anual obtenida por los inversores en un período de tiempo corto. Llamaremos a ésta μ. La volatilidad es la medida de nuestra incertidumbre sobre los movimientos futuros del precio de las acciones, es decir, es la medida de nuestra incertidumbre sobre los cambios proporcionales del precio de las acciones.
Llamaremos a la volatilidad σ.
La hipótesis lognormal para los precios de las acciones implica, por lo tanto, que lnST es normal, donde ST es el precio de las acciones en un tiempo futuro T. Puede demostrarse que la media y la desviación estándar de ln ST son:

$$\ln ST \approx \phi\left[\ln S + \left(\mu - \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right] \qquad (1)$$

Donde:

S_T = el precio de la acción en un momento futuro T

S = el precio de la acción al momento 0

$\phi(\mu,\sigma)$ = denota una distribución normal con media μ y desviación estándar σ.

La ecuación (1) muestra que LnST está normalmente distribuido, por lo cual ST tiene distribución lognormal.
Una variable que tiene una distribución lognormal puede tomar cualquier valor entre cero e infinito. De la ecuación (1) y de las propiedades de la distribución lognormal puede obtenerse que el valor esperado de ST, E(ST) viene dado por:

$$E(S_T) = Se^{\mu T}$$

Donde:

$E(S_T)$ = es el valor esperado

S_T = valor de la acción al momento T

μ = es la media de la distribución o rentabilidad esperada

T = período de tiempo

Esto corresponde con la definición de μ como la tasa de rentabilidad esperada. La varianza de ST, var(ST), puede demostrarse que viene dada por:

8

$$\operatorname{var}(S_T) = S^2 e^{2\mu T}(e^{\sigma^2 T} - 1)$$

A partir de la ecuación (1) puede demostrarse que:

$$\ln\frac{S_T}{S} \approx \phi\left[\left(\mu + \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right]$$

La expresión ln(ST/S) es la rentabilidad compuesta continua proporcionada por las acciones en un tiempo T. La ecuación precedente muestra que está normalmente distribuida.

Conclusión

El modelo de Black - Scholes es un modelo de tiempo continuo. Asume que la volatilidad de un determinado instrumento subyacente es constante durante toda la vida de la opción, pero esta volatilidad se calcula continuamente. Estos dos supuestos implican que los posibles precios del subyacente al vencimiento de la opción se distribuyen log - normalmente. El supuesto de log – normalidad incluido en el modelo de Black - Scholes supera el problema inicialmente presentado. Una distribución log – normal permite incremento de precios ilimitados (el logaritmo de +∞ es +∞) mientras que permite caídas pero sólo hasta cero (el logaritmo de -∞ es 0). Ésta es una representación más realista de cómo se distribuyen los precios en el mundo real.

Podemos resumir los supuestos más importantes que gobiernan los movimientos de precios en el modelo de Black - Scholes:

- Los cambios en los precios de un instrumento subyacente son aleatorios y no pueden ser manipulados artificialmente, no es posible predecir de antemano la dirección en la que se moverán los mismos.
- El porcentaje de cambio en el precio de un instrumento subyacente está distribuido normalmente.
- Debido a que se asume que el porcentaje de cambio en el precio del subyacente se calculó de manera continua, los precios del subyacente al vencimiento se distribuirán log – normalmente.

"*El punto más importante a tener en cuenta es que el modelo de BS supone que los cambios en los precios son aleatorios y que la dirección de dichos cambios no puede ser prevista. Este supuesto puede crear cierta resistencia en los operadores de opciones*".

www.bcr.com.ar/.../images/pdf/
Modelo%20de%20valuación%20de%20opciones%20sobre%20acciones%202005_DIC.pdf

9

4. Análisis de la comunidad de una laguna costera en la costa sur occidental de México.

ABUNDANCIA RELATIVA DE LAS ESPECIES Y MODELOS DE AJUSTE

La comunidad analizada está compuesta por 97 especies de las cuales 72 son peces, 24 son invertebrados (15 crustáceos, seis moluscos, dos medusas y un poliqueto) y además se registró una tortuga marina. Las especies más abundantes resultaron ser el camarón café (*Penaeus californiensis*), el camarón blanco (*P. vannamei*), una sardina (*Lile stolifera*), anchovetas (*Anchoa spp*) y varias especies de mojarras (*Gerreidae*).

La abundancia relativa de las especies se describió mediante la distribución log-normal.

Con el interés de lograr entender mejor los mecanismos que determinan la estructura de la comunidad, el estudio se ha enfocado, desde hace algún tiempo, hacia el conocimiento de la abundancia relativa de las especies analizado desde un punto de vista estadístico, pretendiendo con ello tomar en cuenta no sólo la manera en la que los recursos se distribuyen entre los distintos elementos que la constituyen sino también los posibles mecanismos de interacción de unas especies con otras y con su medio ambiente, que por otra parte, se traduzcan en situaciones tangibles que determinen el grado de dominancia, estabilidad, madurez y sucesión de una comunidad. Como resultado de tal interés se han propuesto varios modelos, los que desafortunadamente tienen más valor heurístico que analítico, pues simplemente describen con mayor o menor precisión la abundancia relativa de las especies que componen la comunidad, pero dicen muy poco con respecto a su estructura. Los modelos mejor conocidos y ampliamente discutidos en la literatura son la serie logarítmica, propuesta por Fisher y el modelo de la llamada distribución log normal de Preston (1948)[3]. Los datos recopilados en el presente estudio se utilizaron con la intención de hacer el ajuste de los mismos al modelo más adecuado; por lo tanto, en las líneas que siguen a continuación se hace una breve exposición de los postulados en que se basa cada uno y los resultados del ajuste: [4]

La serie log normal. Esta distribución se basa en la hipótesis según la cual el nicho de cada especie se considera dependiente de una multitud de factores distintos que determinan la amplitud del tal nicho y consecuentemente los recursos de que la comunidad dispone se deben repartir entre las especies de ésta de una manera equivalente a una curva normal, de modo que tanto las especies abundantes como las raras se dispondrán hacia los extremos de la distribución,

[3] Ver el primer documento.

[4] Yo sólo he recogido los resultados de la distribución log-normal, los resultados de la serie logarítmica los he omitido, ya que no son objeto de este trabajo.

mientras que la mayor parte del inventario, que está representado por especies de frecuencia intermedia, ocupará la parte central de la curva. Este modelo tiene la peculiaridad de que los intervalos de la distribución son sus logaritmos de base dos[5] y de acuerdo con lo antes indicado, su valor modal se localiza en el intervalo que corresponde al mayor número de individuos. Además, debido a que suele haber muchas especies raras que no aparecen representadas en la muestra la distribución log normal suele aparecer truncada hacia la izquierda y en una comunidad no alterada el valor modal generalmente se localiza en el primer intervalo. En un trabajo posterior a aquel en el que se postuló por primera vez este modelo, Williams, (1964) hace un análisis bastante ambicioso sobre la aplicación de esta serie, así como de la logarítmica a distintos grupos de colecciones faunísticas. La función que describe la serie log normal es la siguiente:

$$S = S_0 e^{-(aR)^2} \qquad \text{donde:}$$

S = Es el número de especies en el R-ésimo intervalo a la izquierda y derecha del valor modal.
So = Es el número de especies en el intervalo modal.
a = Parámetro estimado a partir de los datos

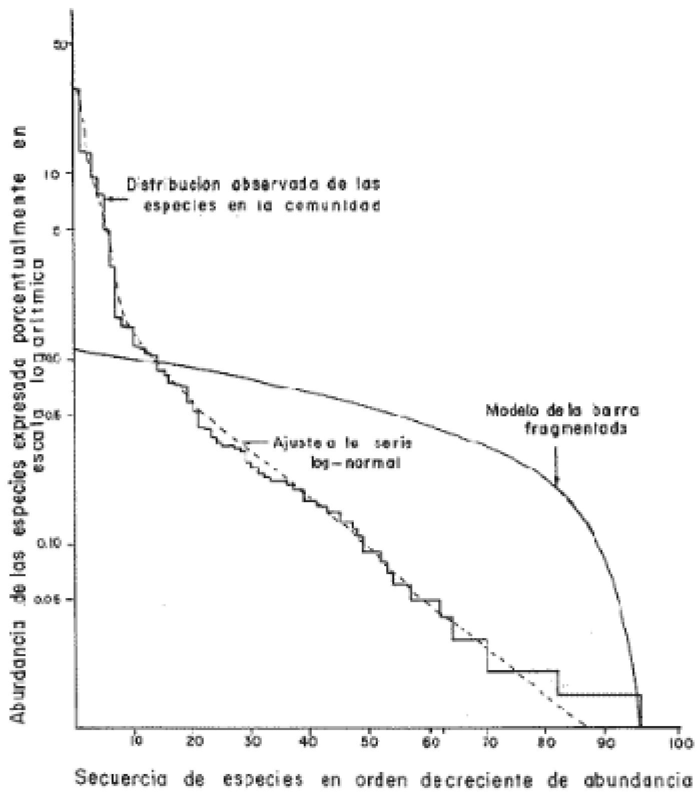


Fig. 11. Número y abundancia relativa de las especies de la comunidad y las distribuciones teóricas esperadas de acuerdo con el modelo de MacArtur y con la log normal.

[5] Aunque en la introducción se ha definido la distribución log-normal con logaritmos neperianos, la base no tiene demasiada importancia.

11

Como resultado de la aplicación de los datos a esta serie se encontró un ajuste más aceptable que en los dos primeros casos, según puede observarse en las figuras 11 y 12.

En atención a los patrones que determinan la abundancia y diversidad de las especies, May (1975) hace una revisión analítica del tema, con énfasis en la intención de separar de los modelos de distribución a los aspectos que reflejen algún detalle de la estructura de las comunidades los que tengan un significado simplemente estadístico gobernado por leyes de grandes números. Una de las características por él señaladas y que vale la pena hacer notar, se refiere al hecho de que en los modelos que describen la abundancia de las especies, la distribución log normal refleja el teorema del límite central, mientras que en aquellos casos donde es posible hacer el ajuste de los modelos tales como el de la barra fragmentada, o la serie logarítmica, está implícito algún aspecto de la biología de la comunidad.

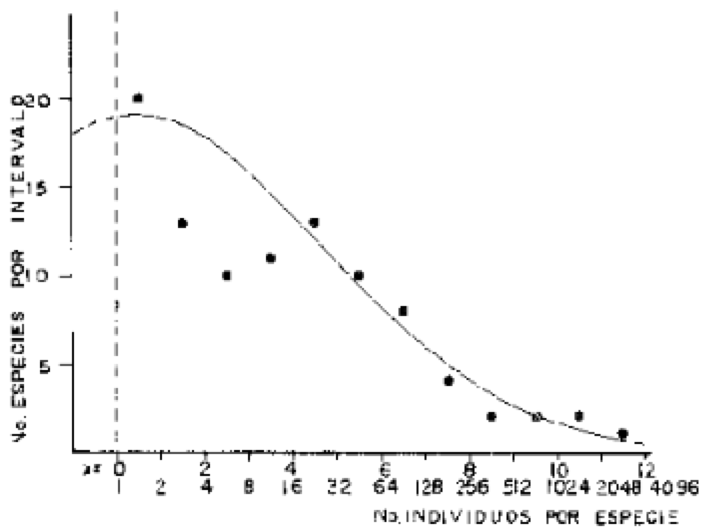


Fig. 12. Relación que describe la abundancia relativa de las especies y su ajuste a la serie log-normal.

May (1975), menciona que las distribuciones de abundancia del tipo log normal en comunidades con organismos "oportunistas", como posiblemente sea el caso de las que son objeto del presente estudio, reflejan poco detalle sobre la estructura de la comunidad y que las especies dominantes son simplemente aquellas que recientemente han disfrutado de un valor grande de r, o sea su potencial biótico y por lo mismo, en momentos distintos habrá diferentes dominantes. Por lo que respecta a la peculiaridad observada en la serie log

12

normal de que el valor de a usualmente vale 0.2 y el de y 29 1, es decir, la relación existente entre el intervalo donde el número de especies es máximo, dividido por el intervalo donde se localiza la especie más abundante, May sugiere que en su opinión estas características no tienen mayor significado que el de ser propiedades matemáticas de la distribución log normal. Al respecto de los modelos de la barra fragmentada y la serie logarítmica o geométrica, él opina que son distribuciones características de comunidades relativamente simples cuya dinámica está dominada por algún factor individual, pues en un extremo está el citado en primer término, que es una expresión estadística realista de una distribución intrínsecamente uniformes; y en el otro está la serie logarítmica que con frecuencia expresa estadísticamente el proceso desigual del "niche preemption", o sea, el derecho de ocupar un nicho disponible antes de que otra especie lo haga, del cual la forma ideal es la serie geométrica. Por otra parte, la serie log normal, que probablemente es la más importante, se puede considerar en muchos aspectos como intermedia entre las otras dos.

http://biblioweb.dgsca.unam.mx/cienciasdelmar/centro/1979-2/articulo68.html

5. Cuantificación de la vitamina B2

Los valores de referencia para la cuantificación de riboflavina (vitamina B2) sérica fueron establecidos en el presente trabajo a partir de una muestra del personal que recibe asistencia en los servicios del Instituto Superior de Medicina Militar "Dr. Luis Díaz Soto". Se estudiaron 88 sujetos de uno y otro sexos, con edades entre 17 y 50 a. Todos cumplían la condición de supuestamente sanos según criterios aplicados. La vitamina B2 fue cuantificada por un proceder analítico de tipo fluorimétrico desarrollado por Natelson y otros.

Mediante el control de calidad de esta técnica se pudieron obtener resultados aceptables en la precisión con coeficientes de variación de 5,27 y 3,74 % para la reproducibilidad y repetibilidad respectivamente, y en la exactitud con coeficiente de correlación de 0,99 y un índice de recuperación de 96,96 %. Según el análisis estadístico de los datos de la muestra **se estableció que la variable vitamina B2 tenía una distribución logarítmica normal**. No se encontraron diferencias significativas según el sexo. El recorrido de valores de referencia cuantificados se enmarcó en un límite mínimo de 5,01 µg/100 mL y máximo de 9,57 µg/100 mL, con una media de 7,29 mg/100 mL, expresado como X ± 2DE[6]. Estos valores se insertan razonablemente en el entorno de los recorridos informados por otros autores, obtenidos por iguales o diferentes métodos.

www.bvs.sld.cu/revistas/mil/vol29_3_00/mil05300.pdf

[6] Desviación estándar

13

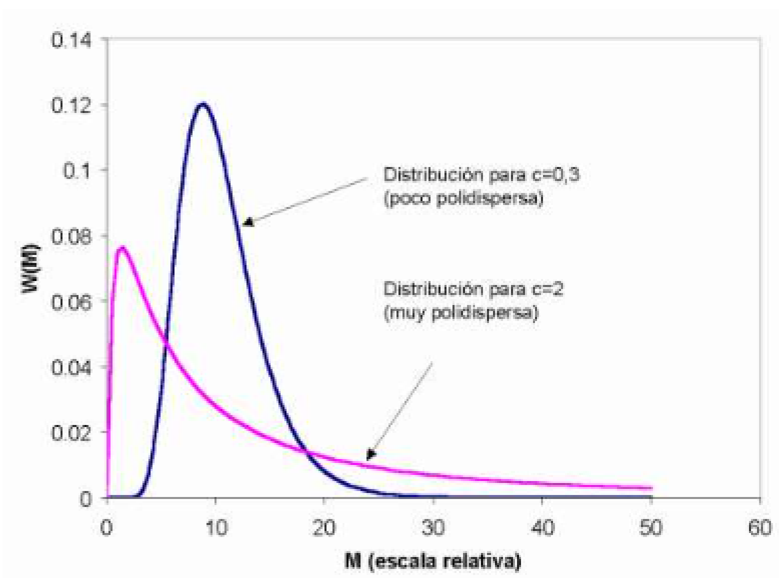
6. Distribución del peso molecular de los polímeros

El **comportamiento de cualquier polímero** ***depende de la distribución de masas en este polímero.*** Esta distribución de masas afectará tanto a las propiedades termomecánicas (capacidad de elongación, punto de rotura...) como físico-químicas (solubilidad, estabilidad...). No se comportará igual un polietileno de cadenas cortas que otro de cadenas largas, el segundo tendrá propiedades mecánicas más fuertes pero sin embargo costará más de trabajar ya que para su procesado se requerirán temperaturas más altas.

Afortunadamente **los polímeros presentan una** ***distribución*** **de pesos moleculares** que sigue una ley fija: la **distribución log-normal** dada por la ecuación:

$$W(M) = \frac{1}{c.\sqrt{\pi}.M}.\exp\left(-\left(\frac{\ln(M/M_0)}{c}\right)^2\right)$$

donde W(M) corresponde a la fracción másica de una cadena de peso "M", M_0 corresponde a la masa para la cual el producto M.W(M) es máxima y "c" es un parámetro que mide la polidispersidad de la mezcla (la anchura de su distribución).



Conviene además definir un par de pesos moleculares: M_n y M_W. Su definición es la siguiente:

14

$$M_n = \frac{1}{\int_0^{\infty} \frac{W(M)}{M} . dM}$$

$$M_w = \int_0^{\infty} M.W(M).dM$$

Obsérvese que a partir de estas tres definiciones (M_n, M_w y $W(M)$) y teniendo en cuenta las correspondientes a $H(\tau)$[7] y τ[8] y admitiendo una distribución log-normal para una mezcla de polímeros, es posible, obtener una función tipo:

$$G'(\omega)+G''(\omega) = f(\omega,c,M_0)$$

Por tanto, en principio sólo habría que probar diferentes valores de c y M_0 hasta hacer que los resultados de la función $f(\omega,c,M_0)$ coincidieran con los experimentales $G'(\omega)+G''(\omega)$.

Desgraciadamente esto no es totalmente posible, ya que analizando $f(\omega,c,M_0)$ puede demostrarse que esta función es independiente del parámetro M_0. De modo que a partir de los ensayos oscilatorios sólo es posible obtener uno de los parámetros de la mezcla. El segundo parámetro (M_0), deberá hallarse por métodos distintos...

www.angel.qui.ub.es/~curco/Reologia/viscoel.html

7. Predicción de Sismos: Una ojeada al futuro[9]

El caso de los sismos característicos

Veamos ahora el caso de la posible existencia de "sismos característicos", definidos en la siguiente forma: "Un evento característico es un sismo que rompe repetidamente el mismo segmento de falla y cuyas dimensiones definen tal segmento" (Nishenko y Buland, 1987). En la tabla 1 del trabajo citado encontramos 14 segmentos definidos de esta manera, entre ellos, por cierto, el de Parkfield. Estos segmentos están definidos con base en 62 temblores "característicos", o sea 48 intervalos entre temblores, lo que hace un promedio de 3.43 intervalos para cada segmento. No es una muestra gigantesca. Un ejemplo típico es el siguiente:

---

7 $H(\tau_i) = \frac{G_N^0}{\beta} M_i . W(M_i)$

8 $\tau_1 = \frac{\eta_0}{G_N^0} \frac{M_n^{1,25} M_1^{\beta}}{M_w^{3,4}}$

9 Este documento no es demasiado preciso y probablemente parte de su contenido podría haberse omitido de este trabajo.

15

Región: san Marcos

Eventos	Intervalo T	Promedio T ave
1907-1845	62	56.0
1957-1907	50	

Esto nos dice que el segmento de San Marcos (cerca de Acapulco) está definido por tres sismos "característicos", en 1845, 1907 y en 1957 (que fue llamado "Sismo del Ángel"). Los intervalos respectivos fueron de 62 y 50 años, y el intervalo promedio fue de 56.0 años.

Ahora bien, en el mismo número del Bulletin of the Seismological Society of America aparece otro artículo (Nishenko y Singh, 1987) que casualmente habla del segmento de San Marcos. Dice lo siguiente: "Los sismos de 1937 y de 1950 y 1957 representan cada uno una ruptura parcial de la zona de 1907... Por lo tanto, los intervalos observados de recurrencia para la región de Acapulco-Ometepec durante el presente siglo varían entre más de 30 a 50 años (o sea, 1937 a 1907 y 1957 a 1907)". Nótese que el primer autor de ambos artículos es la misma persona.

Resulta que uno de los sismos mencionados en el primer artículo, el de 1907, no era definitorio solamente del segmento de San Marcos, sino también de un segmento más grande que lo incluye y que ahora se llama la "región" de Acapulco-Ometepec. Esta "región" se rompió parcialmente en el sismo de San Marcos de 1957, y también parcialmente en otros sismos (1937, 1950) que no se mencionan en el primer artículo. Todos ellos, sin embargo siguen siendo sismos "característicos".

¿No que los sismos característicos "definen" los segmentos en que ocurren? ¿Cómo puede decirse entonces que tanto el sismo de 1907 como el de 1957 "define" el segmento de San Marcos, y que al mismo tiempo el de 1907 "define" el segmento de Ometepec, y además la región de Acapulco-Ometepec, que no es la misma?.

En cuanto al intervalo promedio, ya no sabemos si es 56.0 años como afirma el primer artículo, o menos de 50 como dice el segundo. Quién sabe a qué sismos "característicos" se refiere cada uno de los artículos. Si los sismos "característicos" rompen repetidamente el mismo segmento de falla no debería admitirse traslapes ni rupturas parciales. Una de dos: o bien algunos sismos citados no son eventos característicos (lo que arrojaría dudas sobre el autor común de ambos trabajos), o bien los datos de la Tabla 1 eran incompletos. Aceptaremos esta última hipótesis, porque es la más compleja y por lo tanto probablemente más real; por lo demás, el co-autor del segundo trabajo es el más distinguido conocedor de la sismicidad de México, especialista en la zona de Guerrero y Oaxaca.

Ahora bien, si la muestra de San Marcos era incompleta, ello debería modificar el promedio Tave cuyo papel en el primer trabajo es muy interesante. En efecto, Nishenko y Buland (1987) normalizaron los intervalos T mediante subdivisión por Tave (digamos, dividen 62 y 50 por 56), y hacen lo propio con todas las 14 regiones. Luego juntan todos los datos y los grafican en un mismo

16

histograma, al que calzan una **distribución logarítmico-normal**. Reproduzco la conclusión de este procedimiento que suena a brujería: "Por lo tanto, **la distribución de los intervalos de recurrencia para cada segmento de falla también es logarítmico-normal y In(T) obedece a una distribución normal**" (Nishenko y Buland, 1987).

Sobra decir que la distribución de muestras combinadas de 14 procesos logarítmicos-normales no tiene por qué ser también logarítmico-normal. Pero eso no es todo. La normalización de las muestras no se justifica por nada. Los promedios Tave aún en el caso de que las muestras fueran completas, tienen una enorme varianza ya que el tamaño de las muestras es apenas de 3 a 4. No existe razón alguna para que la muestra combinada siguiera alguna distribución en particular. Por lo demás, los autores nunca efectúan una prueba de normalidad de log(T).

Podría argumentarse que **la distribución logarítmico-normal** posee unas propiedades interesantes y que yo mismo he especulado con ella en mi libro (1974) por **ser apta para representar la distribución de magnitudes de los temblores. Este resultado se basa en la idea de autosimilitud de las fracturas en la Tierra, cuyo mecanismo fuera discutido por primera vez por Kolmogorov (1941) y que ahora se ha hecho famoso con el nombre de fractalidad.**

En tal caso, sin embargo, ¿cómo explicar el hecho de que los intervalos medios, en dos "segmentos" tan cercanos como Parkfield y Pallett Creek, ambos en la falla de San Andrés, sean tan diferentes? El intervalo promedio de Parkfield (ya lo mencionamos) es de 21.8 años. El de Pallett Creek, de 194.3 años. Ambos supuestamente definidos con base en sismos "característicos". Ambos sobre la misma falla. El corrimiento anual de la falla es el mismo en ambos lugares. Si hay auto-similitud el mecanismo de fractura debe ser homogéneo (Kolmogorov, 1941). Pero no lo es, puesto que hay diferencias tan enormes en el intervalo promedio de temblores.

Sin embargo, los autores explícitamente declaran que todos los sismos característicos son generados por un solo proceso común. Esto significa que debería poder predecirse un sismo de Pallett Creek mediante observaciones hechas en Parkfield, lo que es absurdo, puesto que los intervalos son completamente diversos. En conclusión, los sismos "característicos" no caracterizan nada, a no ser un gran deseo de predecir fenómenos que aún no entendemos suficientemente bien.

www.ssn.unam.mx/SSN/Doc/Prediccion/cinna.htm

17