Universidad Nacional del Litoral
Facultad de Ingeniería y Ciencias Hídricas

ESTADÍSTICA

Ingeniería Informática

TEORÍA

Mg.Ing. Susana Vanlesberg
Profesor Titular

UNIDAD 4

MODELOS PROBABILÍSTICOS

MODELOS PARA VARIABLES ALEATORIAS CONTINUAS

MODELO EXPONENCIAL

Este modelo surge al considerar el ***tiempo hasta*** la primera ocurrencia de un evento que pueda ser considerado como proceso de Poisson.

Si la variable aleatoria es ahora el tiempo transcurrido hasta que se verifica la primera ocurrencia, entonces será una variable continua, la probabilidad que T exceda algún valor t es lo mismo que decir que **no** se verificaron ocurrencias en ese intervalo de longitud t, lo que es equivalente a decir que la variable aleatoria **Nº de ocurrencias** de tipo Poisson toma el valor 0:

$$P\,(T > t) = P(x = O)$$

$$P(x = 0) = \frac{e^{-\lambda t}\lambda^{0}}{0!} = e^{-\lambda t}$$

Esto permite obtener la función de distribución de T variable continua:

$$F(t) = P(T \leq t) = 1 - P(T > t\,) = 1 - e^{-\lambda t} \qquad t \geq 0 \qquad (1)$$

La función de densidad se obtiene derivando la expresión anterior:

$$f(t) = \lambda e^{-\lambda t} \qquad (2)$$

Cumpliendo con las propiedades de estacionariedad e independencia de los procesos de Poisson, $e^{-\lambda t}$ da la probabilidad de que no se verifiquen eventos en algún intervalo de tiempo de longitud t, estando o no el origen en el tiempo 0.
Por ejemplo si el origen estuviese en el momento de la enésima ocurrencia, $e^{-\lambda t}$ representaría la probabilidad de que el tiempo hasta la próxima ocurrencia (n+1) sea mayor que t.

Esto quiere demostrar sencillamente que los tiempos entre ocurrencias de eventos de tipo Poisson son independientes y se distribuyen en forma exponencial.

Características

$$E(T) = \int_0^{\infty} t\lambda e^{-\lambda t} dt$$

$$si\, u = t\lambda \qquad du = \lambda dt \qquad \frac{du}{\lambda} = dt$$

$$\frac{1}{\lambda}\int_0^{\infty} ue^{-u} du = \frac{1}{\lambda}\left[e^{-u}(-u-1)\right]_0^{\infty}$$

$$E(T) = \frac{1}{\lambda} \qquad (3)$$

Recordar que λ en los procesos de Poisson, representa el número promedio de ocurrencias, aquí, 1/λ representa el tiempo promedio entre ocurrencias.

$$Var(T) = E(T^2) - E^2(T)$$

$$E(T^2) = \int_0^{\infty} t^2 \lambda e^{-\lambda t} dt$$

$$si\, v = t^2 \qquad dv = 2tdt$$

$$\lambda e^{-\lambda t} dt = du \qquad u = \int_0^{\infty} \lambda e^{-\lambda t} dt = -e^{-\lambda t}$$

$$E(T^2) = \left[-e^{-\lambda t} t^2\right]_0^{\infty} - \int_0^{\infty} -e^{-\lambda t}\, 2t\, dt =$$

$$= 0 + 2\int_0^{\infty} e^{-\lambda t} tdt \frac{\lambda}{\lambda} = \frac{2}{\lambda}\left(\frac{1}{\lambda}\right)$$

$$E(T^2) = \frac{2}{\lambda^2}$$

$$Var(T) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \qquad (4)$$

Una característica de los procesos de Poisson es que no tienen memoria. Esto significa que el comportamiento futuro es independiente de lo registrado en el presente o en el pasado.

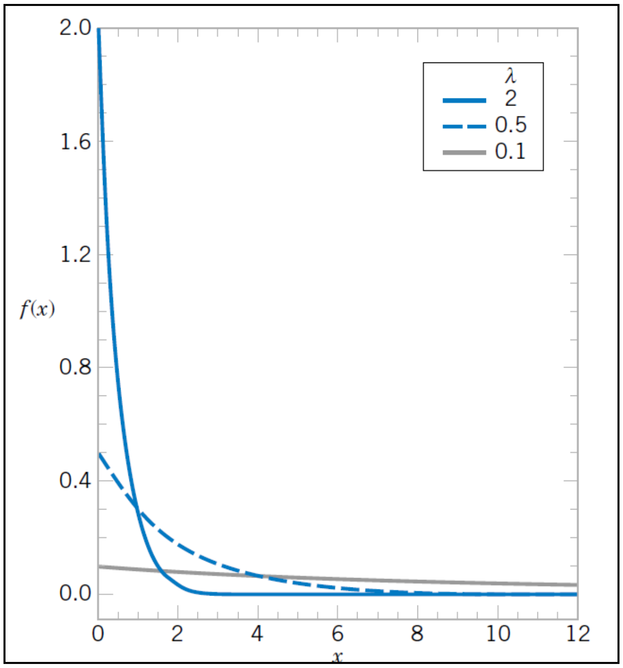


Figura Nº 1- Distribución exponencial para distintos valores del parámetro

Ejemplo

El personal de la compañía Onda S.L. usa una Terminal para realizar sus pedidos internacionales. Si el tiempo que cada comercial gasta en una sesión en la Terminal tiene una distribución exponencial de media 36 minutos, encontrar:

a) Probabilidad de que un comercial utilice la Terminal 30 minutos o menos.

b) Si un comercial ha estado por lo menos 30 minutos en la Terminal, ¿Cuál es la probabilidad de que pase al menos una hora más en la Terminal?

$1/\lambda = 36 \qquad \lambda = 1/36$

$$a) \quad P(X \leq 30) = \int_0^{30} 1/36 * e^{-1/36*x}\,dx = 0.565$$

$$b) \quad P((X \geq 60)/(X \geq 30)) = P(X \geq 60) = \int_{60}^{+\infty} 1/36 * e^{1/36*x}\,dx = 0.188$$

MODELO NORMAL

Este modelo suele conocerse como **Modelo de las sumas**, ya que la incertidumbre en algunas variables puede ser el resultado de efectos combinados de algunas causas que contribuyen, siendo difícil de separar y observar a cada una. En algunas situaciones si se conoce el mecanismo por el cual las causas individuales afectan a la variable de interés se puede determinar un modelo para la variable resultante sin estudiar en detalle los efectos individuales, particularmente no es necesario conocer la distribución de las causas.

El modelo de la variable resultado, presenta una función de densidad doble exponencial cuya forma es la de una campana. Este hecho está contenido en el Teorema del Límite Central, el cual es uno de los resultados más importantes de la Teoría de Probabilidad. Este teorema será desarrollado más adelante pero se puede hacer referencia a su enunciado: *bajo condiciones generales, cuando el número de variables que intervienen en la suma que origina una variable, es cada vez más grande, la distribución de esta suma tiende a aproximarse al modelo Normal*.

La inmensa importancia práctica del modelo Normal reside entre otras razones en que lo que se plantea en el teorema del Límite Central, puede hacerse sin el conocimiento exacto de:
- las distribuciones marginales de cada variable que interviene en la suma, - de su número, - de su distribución conjunta. Ya que la variación aleatoria en algunos fenómenos naturales se origina de un número de variaciones aditivas, no debe sorprender el hecho que gráficos que pueden aproximarse a este modelo, se observen con frecuencia en la naturaleza.

La función de densidad que caracteriza esta distribución es una doble exponencial con la siguiente forma:

$$f(x) = ke^{-c(x-m)^2} \text{ , } -\infty \leq x \leq \infty \qquad (5)$$

siendo **m** la distancia al centro de la distribución que por la simetría de la distribución es igual a su media μ.

Las constantes k y c se pueden determinar: k por considerar que f(x) para ser función de densidad deberá cumplir la condición de ser igual a uno para todos los valores que la variable puede tomar:

$$\int_{-\infty}^{\infty} k\, e^{-c(x-\mu)^2}\, dx = 1$$

$$k \int_{-\infty}^{\infty} e^{-c(x-\mu)^2}\, dx = 1$$

$$llamando\ a\,(x-\mu) = y,\quad dx = dy, luego$$

$$k \int_{-\infty}^{\infty} e^{-c\,y^2}\, dy = 1$$

$$siendo \int_{0}^{\infty} e^{-a^2x^2}\, dx = \frac{\sqrt{\pi}}{2a} \qquad a^2 = c \longrightarrow a = \sqrt{c}$$

$$por\ simetría\ se\ tiene:$$

$$k\, 2\, \frac{\sqrt{\pi}}{2\sqrt{c}} = 1$$

$$k = \frac{\sqrt{c}}{\sqrt{\pi}}$$

Para completar la función de densidad es necesario evaluar la otra constante, c. Para esto se utiliza la varianza:

$$Var(x) = \frac{\sqrt{c}}{\sqrt{\pi}} \int_{-\infty}^{\infty} (X-\mu)^2\, e^{-c(X-\mu)^2}\, dx$$

Para resolver esta integral se usará un resultado proveniente de la resolución de integrales impropias dependientes de un parámetro:

$$\int_0^{\infty} e^{-c\,x^2}\,dx = \frac{1}{2}\frac{\sqrt{\pi}}{\sqrt{c}} \text{ para c} > 0$$

$$\text{luego} \quad \text{llamando} \quad \mathrm{F(c)} = \int_0^{\infty} e^{-c\,x^2}\,dx$$

y aplicando la siguiente propiedad

$$\mathrm{F(x)} = \int_a^{\infty} f(t,x)dt \longrightarrow F'(x) = \int_a^{\infty} \frac{\partial f}{\partial x}(t,x)dt$$

$$\mathrm{F(c)} = \int_0^{\infty} f(t,c)dt \longrightarrow F'(c) = \int_0^{\infty} \frac{\partial f}{\partial c}(t,c)dt$$

en este caso origina :

$$\mathrm{F(c)} = \int_0^{\infty} e^{-c\,x^2} dx \quad \text{donde}$$

$$\mathrm{f(t,c)} = e^{-c\,t^2} \quad \text{luego} \quad \frac{\partial f}{\partial c} = -t^2 e^{-c\,t^2}$$

$$\textit{con} \quad \text{lo} \quad \text{cual } \mathrm{F'(c)} = \int_0^{\infty} -t^2 e^{-c\,t^2} dt$$

$$\textit{pero} \quad \text{como} \quad \mathrm{F'(c)} = \frac{\mathrm{dF}}{\mathrm{dc}} = \frac{\mathrm{d}}{\mathrm{dc}}\left(\frac{1}{2}\frac{\sqrt{\pi}}{\sqrt{c}}\right) = -\frac{\sqrt{\pi}}{4}\mathrm{c}^{-\frac{3}{2}}$$

$$\text{resulta:}$$

$$\int_0^{\infty} -t^2 e^{-c\,t^2} dt = -\frac{\sqrt{\pi}}{4} c^{-\frac{3}{2}}$$

$$\textit{luego} \quad \int_{-\infty}^{\infty} t^2 e^{-c\,t^2} dt = \frac{\sqrt{\pi}}{2} c^{-\frac{3}{2}}$$

$$y \quad \frac{\sqrt{\mathrm{c}}}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-c\,t^2} dt = \frac{\sqrt{c}}{\sqrt{\pi}}\frac{\sqrt{\pi}}{2} c^{-\frac{3}{2}} = \frac{c^{-1}}{2}$$

$$\mathrm{Var(X)} = \frac{1}{2\,\mathrm{c}} \quad \text{de} \quad \text{donde}$$

$$\mathrm{c} = \frac{1}{2\,\sigma^2}$$

Esto permite completar la expresión de la función de densidad:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \le x \le \infty \qquad (6)$$

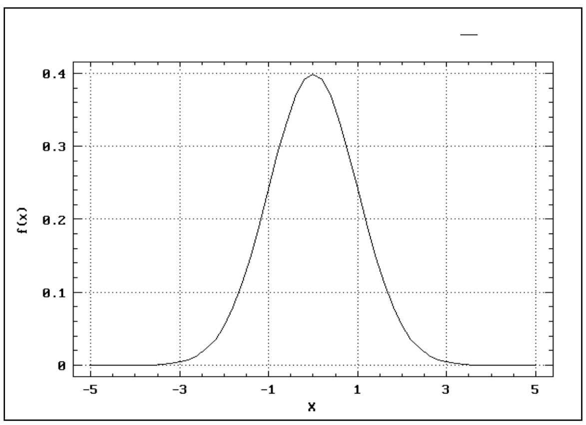
La función acumulativa será a partir de esta función anterior, la siguiente:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \qquad (7)$$

μ y σ son sus parámetros. σ desvío estándar y μ la media. Así es que la función queda completamente definida si se conocen estos dos valores.

- $[(x-\mu)/\sigma]^2$: es la parte práctica de la función ya que contiene un valor determinado de la variable y a los parámetros. El hecho de estar elevado al cuadrado hace que dos valores distintos de la variable, que tengan la misma desviación absoluta de la media μ, van a tener igual valor de la función de densidad. Esto muestra claramente la característica de simétrica alrededor de μ, que posee la distribución Normal. Al ser este exponente negativo, cuanto mayor es la desviación de x respecto de μ, será menor la densidad de probabilidad de x, esto es lo que se observa en las colas de la distribución. Ahora cuando x coincide con μ, el exponente es cero y el valor de la función de densidad es máxima en ese punto e igual a $1/\sigma \pi^{1/2}$, o sea que $x=\mu$ es el valor del modo.

- El valor de μ hace desplazar al modelo hacia la izquierda o la derecha, mientras que el valor de σ cambia su forma sin desplazarlo. Este efecto se observa en las siguientes figuras:



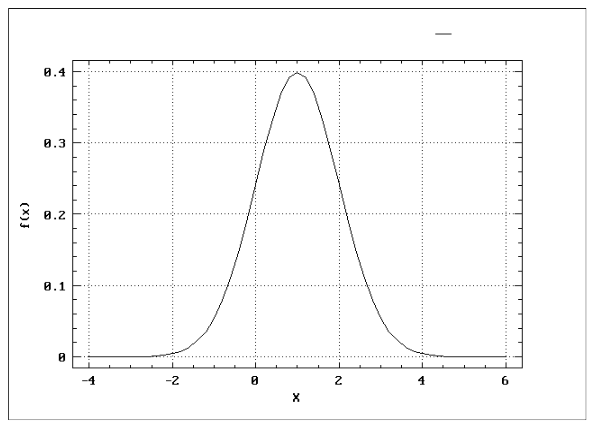


Fig. Nº 2 - Dist. Normal con $\mu=0$ y $\sigma=1$ **Fig. Nº 3** – Dist.Normal con $\mu=1$ y $\sigma=1$

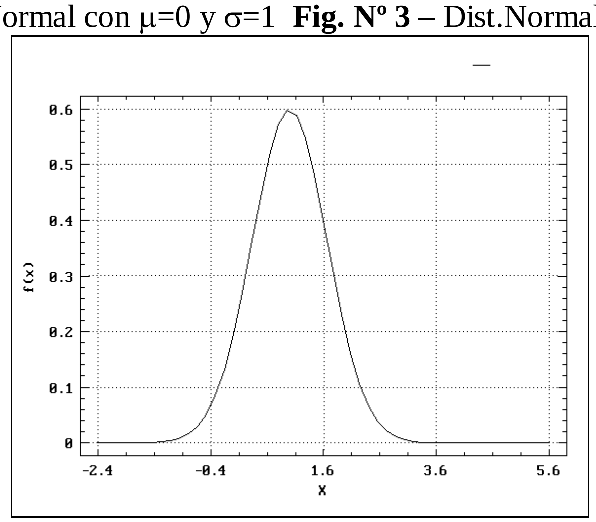


Fig. Nº 4 – Distribución Normal con $\mu=1$ y $\sigma=2/3$

- Debido a que la distribución Normal tiene amplitud infinita, su curva nunca toca el eje x, la probabilidad de un intervalo muy alejado de μ será prácticamente despreciable y el 99% de su área está encerrada por μ±3σ.

En las siguientes figuras se observa el efecto de tomar distintos valores de μ y σ:

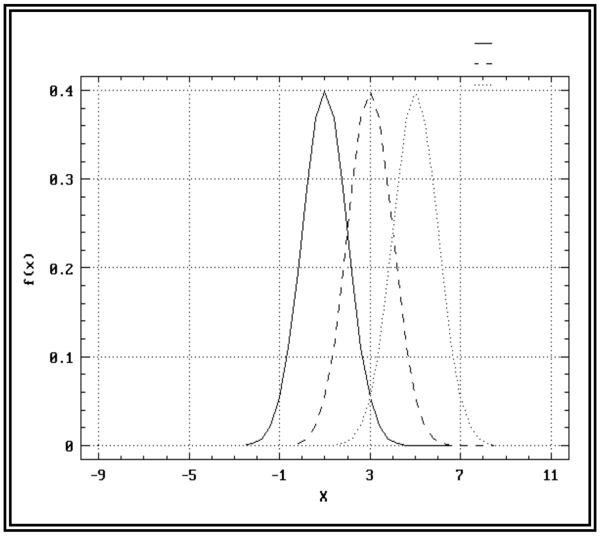


Fig. Nº 5 - Curvas Normales: distintos μ e igual σ

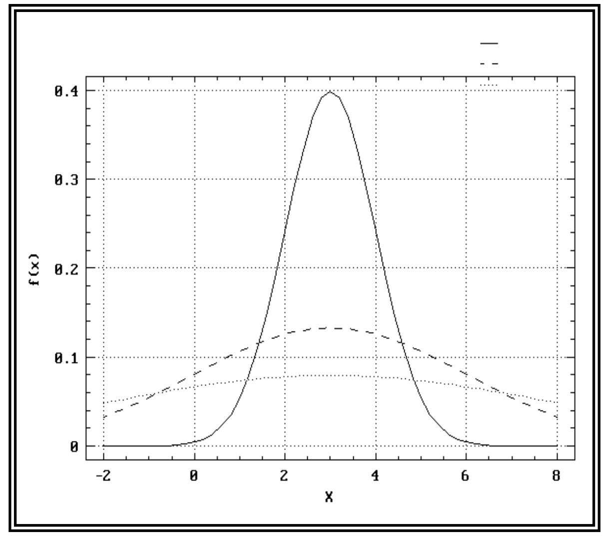


Fig. Nº 6 - Curvas Normales: igual μ y distintos σ

Características

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx =$$

$$= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} dx =$$

$$siendo \ \frac{X-\mu}{\sigma} = y \quad x = \sigma\, y + \mu \quad dx = \sigma\, dy$$

$$E(X) = \int_{-\infty}^{\infty} (\sigma\, y + \mu) \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{1}{2}y^2} \sigma\, dy =$$

Integrando por partes,

$$= \frac{\sigma}{\sqrt{2\pi}\ \sigma} \int_{-\infty}^{\infty} (\sigma\ \mathrm{y} + \mu) e^{-\frac{y^2}{2}} dy =$$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} (\sigma\ \mathrm{y}\ \mathrm{e}^{-\frac{y^2}{2}}\ dy + \int_{-\infty}^{\infty} \mu\ e^{-\frac{y^2}{2}}\ dy \right] =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{y}\ \mathrm{e}^{-\frac{y^2}{2}}\ dy + \mu \ \ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{y^2}{2}} d\left(\frac{y}{2}\right)^2 + \mu =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \left[-e^{-\frac{y^2}{2}} \right]_{-\infty}^{\infty} + \mu = 0 + \mu$$

$$E(X) = \mu \tag{8}$$

$$Var(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$\frac{x-\mu}{\sigma} = y \qquad \mathrm{dy} = \frac{\mathrm{dx}}{\sigma}$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 y^2 e^{-\frac{1}{2}y^2}\ \sigma\ dy =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} d\left(\frac{y^2}{2}\right)$$

$$u = y \qquad \mathrm{dv} = e^{-\frac{1}{2}y^2} d\left(\frac{y^2}{2}\right) \Rightarrow v = -e^{\frac{y^2}{2}}$$

$$\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sigma^2 \left\{ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty}\ e^{-\frac{1}{2}y^2} dy \right\}$$

$$Var(X) = \sigma^2$$

(9)

Otras características del modelo normal que merecen destacarse:

-Muchas variables aleatorias continuas se distribuyen en forma aproximadamente normal. Es de destacar que los errores de mediciones repetidas de alguna dimensión particular, se dice que tienen distribución normal. Esto se debe a que cualquier medida se supone que es igual a un valor verdadero más un error; este error puede considerarse como el resultado de un gran conjunto de factores que están presentes en ese momento, y cada factor ejerce un pequeño efecto en la magnitud y sentido del error. Los errores actúan independientemente y con igual fuerza para aumentar o disminuir el valor de la medición observada, y a largo plazo se anulan. Esto hace considerar a los errores de medida como distribuidos normalmente, con valor medio cero, y se los suele denominar "errores al azar".

-Este modelo sirve como buena aproximación de modelos discretos, como el Binomial o el de Poisson, bajo circunstancias especiales.

- El supuesto de normalidad de las poblaciones permite obtener buenos resultados de métodos elaborados bajo este supuesto, aunque en realidad no se cumpla en forma estricta.

-Muchos estadísticos calculados a partir de grandes muestras se distribuyen en forma aproximadamente normal, lo cual facilita el trabajo de inferencia estadística.

Modelo Normal estándar

La distribución normal se encuentra tabulada y existen, además, rutinas computacionales que permiten trabajar con ella.

Para fines prácticos y para ganar en eficiencia y rapidez, el concepto de distribución estándar es muy importante.

Se denomina modelo estándar a aquel cuya media es 0 y su desvío es uno. La variable $z=(x-\mu)/\sigma$ se denomina *variable aleatoria estandarizada*, cuya media es 0 y su desvío es unitario.

La función de densidad del modelo estandarizado es:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \qquad -\infty < z < +\infty \qquad (10)$$

y su función acumulativa es la siguiente:

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt \qquad (11)$$

Características de este modelo:

$$E(z) = E\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}E(x-\mu) = \frac{1}{\sigma}\left[E(x) - E(\mu)\right] = \frac{1}{\sigma}(\mu - \mu) \qquad (12)$$

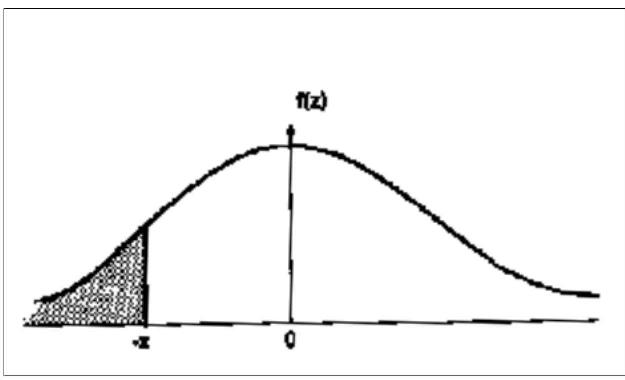
$$E(z) = 0$$

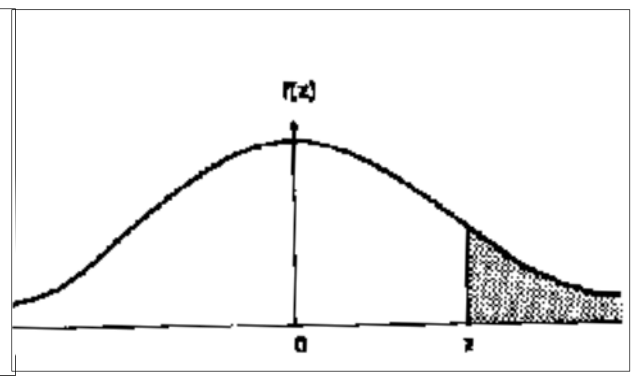
$$Var(z) = Var\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(x-\mu) = \frac{1}{\sigma^2}Var(x) - 0 = \frac{\sigma^2}{\sigma^2}$$

$$Var(\mu) = E\left[\mu - E(\mu)\right]^2 = E(\mu-\mu)^2 = 0$$

$$Var(z) = 1 \qquad (13)$$

Para el manejo de las tablas, a veces es necesario tener en cuenta el carácter simétrico de la distribución, ya que algunas están elaboradas solamente para valores positivos. Para obtener probabilidades o densidades de valores negativos de la variable se procede de la siguiente manera:





$$P(Z \le -z) = P(Z \ge z) = 1 - P(Z \le z)$$

Esta simetría del modelo normal respecto a su media implica que los momentos centrados de orden impar sean cero. Los de orden par pueden considerarse a partir de la siguiente expresión:

$$\mu_n = E(x-\mu)^n = \frac{n!}{2^{\frac{n}{2}}\left(\frac{n}{2}\right)!}\sigma^n \qquad n = 2, 4, \ldots \qquad (14)$$

La asimetría será, por lo tanto, igual a cero:

$$As = \frac{\mu_3}{\sigma^3} = 0 \qquad (15)$$

y la kurtosis, en función de la expresión general de momentos centrados de orden par, será:

$$K = \frac{\mu_4}{\sigma^4}$$

$$\mu_4 = \frac{4!}{2^{\frac{4}{2}}\left(\frac{4}{2}\right)!}\sigma^4 = 3\sigma^4$$

$$K = 3 \qquad (16)$$

Es por esta característica que suelen compararse los coeficientes de kurtosis de distintos modelos con el de la normal, permitiendo clasificarlos según sean mayores, menores o iguales que 3.

Ejemplo

Una máquina fabrica tornillos cuyas longitudes se distribuyen normalmente con media 20 mm y varianza 0.25 mm. Un tornillo se considera defectuoso si su longitud difiere de la media más de 1 mm. Los tornillos se fabrican de forma independiente. ¿Cuál es la probabilidad de fabricar un tornillo defectuoso?

$$\sigma = 0.5$$

$$\sigma^2 = 0.25$$

$$x \longrightarrow N(20,0.5)$$

$$P(def.) = P(x > 21) + P(x < 19) = P\left(z > \frac{21-20}{0.5}\right) + P\left(z < \frac{19-20}{0.5}\right) = P(z > 2) + P(z < -2) =$$
$$[1 - P(z \le 2)] + [1 - P(z \le 2)] = [1 - 0.9773] + [1 - 0.9773] = 0.0452$$

Distribución de la suma de variables aleatorias Normales

El Teorema del Límite Central (que se expondrá más detalladamente más adelante), expresa que la suma de un número dado de variables aleatorias tiende a distribuirse en forma Normal. Es de esperar que si se suma un número de variables aleatorias independientes y Normales cada una, la suma será distribuida en forma Normal.

$$\begin{aligned}
&x_1 \sim N(\mu_1, \sigma_1)\\
&x_2 \sim N(\mu_2, \sigma_2)\\
&\quad .\\
&.x_n \sim N(\mu_n, \sigma_n)\\
&U = \sum_{i=1}^{n} x_i\\
&\text{con parámetros}\\
&\mathrm{E(U)} = \mathrm{E}\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} \mu_i\\
&\sigma^2(U) = \sigma^2\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \sigma^2(x_i) = \sum_{i=1}^{n} \sigma_i^2\\
&\mathit{entonces} \quad \mathrm{U} \sim \mathrm{N}(\sum \mu_i ; \sqrt{\sum_{i=1}^{n} \sigma_i^2})
\end{aligned}$$

Aproximación del modelo Binomial al Normal

En el esquema que corresponde a una variable aleatoria con distribución Binomial, cuando n aumenta y p no varía, la variable aleatoria

$$\frac{x - np}{\sqrt{npq}}$$

se distribuye aproximadamente Normal:

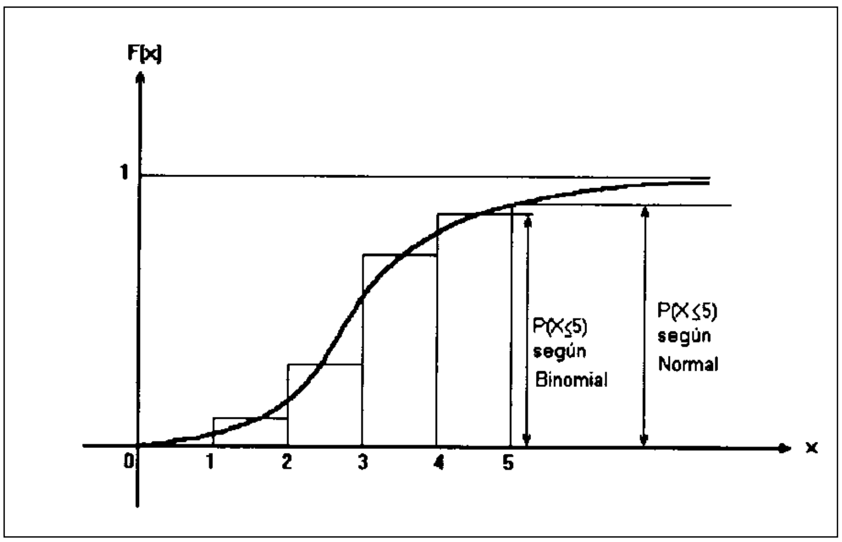
$$P(X \le x_0) = F_N\left(\frac{x_0 - np}{\sqrt{npq}}\right)$$

siendo x_0, un valor cualquiera de x.

El primer miembro de la última ecuación pertenece a la distribución Binomial, cuya función de distribución es escalonada. El segundo miembro pertenece a la distribución Normal que es continua. A pesar de esta diferencia, para grandes valores de **n** ambas pueden considerarse iguales para cada valor, ya que la función continua pasa por la mitad de cada escalón de la función discreta. Esto es una buena aproximación para valores de n*p>5, cuando p≤1/2 y n*q>5 con p>1/2. Estas aproximaciones son válidas si n es pequeño introduciendo un factor de corrección por continuidad cuyo valor es igual a 1/2.

En la figura siguiente se observa que una mejor aproximación a P(X≤5) se obtiene por tomar la ordenada de la distribución continua 1/2 unidad a la derecha de 5. En general:

$$P(X \leq x_0) = F_N\left(\frac{x_0 + \frac{1}{2} - np}{\sqrt{npq}}\right)$$



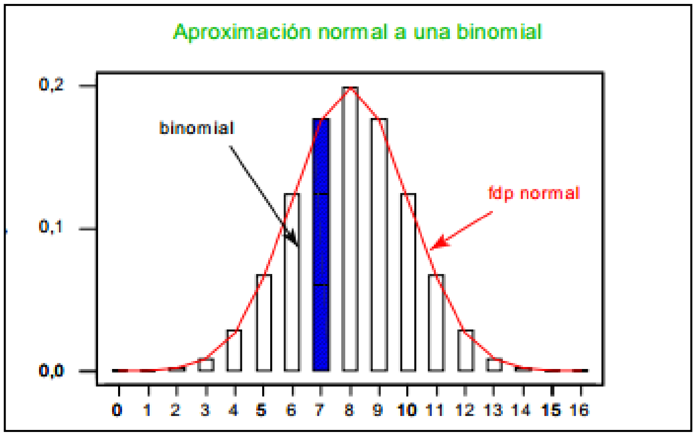


Figura Nº 7 – Aproximación Binomial con la Normal

DISTRIBUCIONES RELACIONADAS CON LA DISTRIBUCIÓN NORMAL

Distribución Chi-cuadrado

Este modelo describe la distribución de la suma de los cuadrados de ν variables aleatorias independientes, con distribución N(0,1):

$$V = \sum_{i=1}^{\upsilon} X_i^2 \sim \chi_{\upsilon}^2 \qquad (17)$$

Si las variables que intervienen en la suma no fuesen N(0,1), se las debería estandarizar, con lo cual se obtendría:

$$V = \sum_{i=1}^{\upsilon} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi_{\upsilon}^2 \quad (18)$$

V es también una variable aleatoria, ya que es el resultado de la suma de variables aleatorias. Por ser una suma de cuadrados, varía desde 0 a infinito.

La función de densidad surge de considerar la función N(0,1) de cada componente al cuadrado, y es la siguiente:

$$f(\chi^2)=\begin{cases}\dfrac{1}{2^{\frac{\upsilon}{2}}\Gamma\left(\dfrac{\upsilon}{2}\right)}\chi^{2\left(\frac{\upsilon-2}{2}\right)}e^{-\frac{\chi^2}{2}}, para\ \chi^2>0 \\ \\ 0,\ en\ otro\ caso\end{cases} \qquad (19)$$

siendo Γ la función gamma y ν el número de variables al cuadrado que intervienen en la suma, y se denominan "grados de libertad".

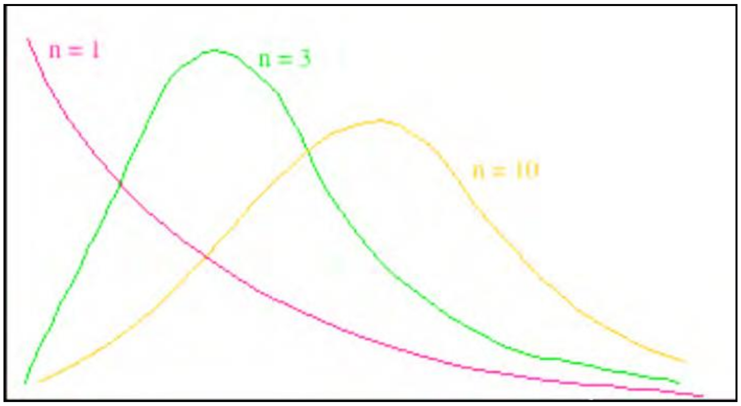
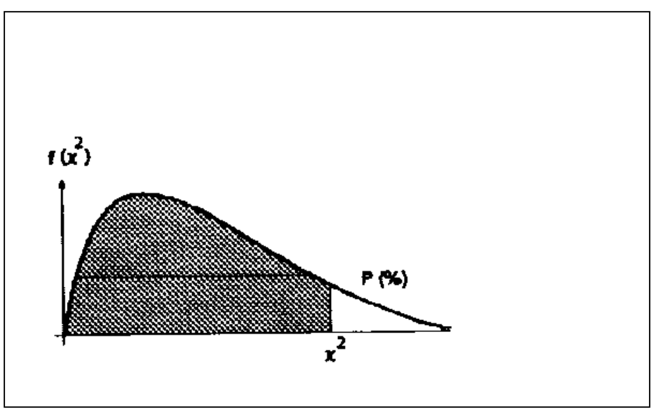


Figura Nº 8 – Distribución chi- cuadrado

Las características de esta distribución son:

$$E(\xi^2)=\nu$$
$$\sigma^2(\xi^2)=2\nu \qquad (20)$$

Las tablas permiten evaluar probabilidades del siguiente tipo:



Propiedades de la distribución chi-cuadrado

1.- La variable solo puede tomar valores positivos.
2.- Es asimétrica.
3.- Depende del parámetro **υ** (grados de libertad).
4.- Su esperanza matemática es υ, y su varianza, 2υ.
5.- *Propiedad aditiva o reproductiva* : Si χ^{2n} y χ^{2m} son dos variables Chi cuadrado con n y m grados de libertad respectivamente, independientes entre sí, entonces la suma de las dos variables es una variable Chi-cuadrado con n+m grados de libertad. Esto se puede generalizar a la suma de cualquier número de variables Chi-cuadrado, independientes.
6.- Al aumentar el número de grados de libertad, la distribución Chicuadrado se aproxima asintóticamente a una distribución normal.

Distribución t de Student

Este modelo surge como el cociente entre una variable N(0,1) y la raíz cuadrada de una variable aleatoria distribuida χ^2 dividida por sus grados de libertad, siendo estas dos variables independientes:

$$t_\upsilon = \frac{x}{\sqrt{\frac{V}{\upsilon}}}, \quad \begin{cases} x \sim N(0,1) \\ V \sim \chi^2_\upsilon \end{cases} \qquad (21)$$

El rango de la variable t varía entre -∞ y +∞, y su función de densidad surge de las distribuciones de las variables aleatorias componentes:

$$f(t) = \frac{1}{\sqrt{\pi\upsilon}} \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)} \left(1+\frac{t^2}{\upsilon}\right)^{-\frac{\upsilon+1}{2}}, \quad para \quad -\infty < t < \infty \qquad (22)$$

siendo ν los grados de libertad asociados a la componente χ^2

Sus características son:

$$E(t) = 0$$
$$\sigma^2(t) = \frac{\upsilon}{\upsilon-2}, \; \upsilon > 2 \qquad (23)$$

La gráfica es parecida a la de la distribución Normal, y es simétrica respecto del cero. Para grandes valores de ν se la puede aproximar a la Normal.

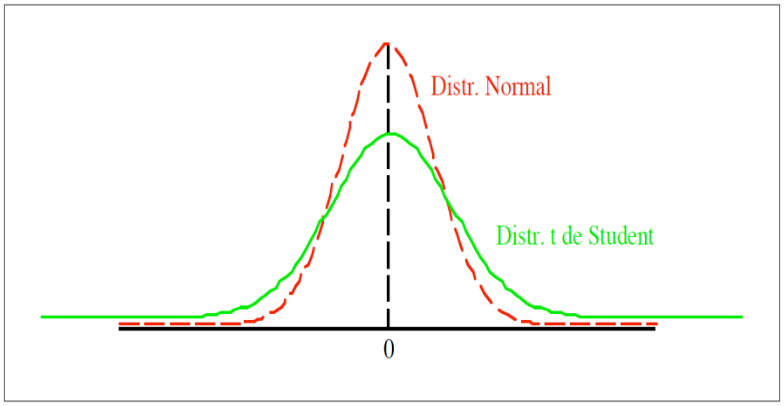


Figura Nº 9 - Distribución de Student

Existen tablas que brindan probabilidades para distintos valores de la variable y algoritmos computacionales que permiten obtener probabilidades con igual o mayor grado de precisión que las tablas.

Propiedades de la distribución "t"

1.- Depende de un único parámetro, el número de grados de libertad.
2.- El rango de la variable es todo el eje real $(-\infty, +\infty)$.
3.- Su gráfica es simétrica respecto al eje de ordenadas OY.
4.- El valor $x = 0$ es la media, mediana y moda de la distribución.
5.- Al aumentar n, se va haciendo cada vez más apuntada la gráfica de su función de densidad, siendo el límite para $n \rightarrow \infty$ la curva normal tipificada.

Distribución F de Snedecor

Si X e Y son dos variables aleatorias independientes cada una con distribución ξ^2, con ν_1 y ν_2 grados de libertad; luego la distribución F está definida por el siguiente cociente:

$$F_{\upsilon_1,\upsilon_2} = \frac{\frac{X}{\upsilon_1}}{\frac{Y}{\upsilon_2}} = \frac{\frac{\chi_1^2}{\upsilon_1}}{\frac{\chi_2^2}{\upsilon_2}} \qquad (24)$$

Como es el cociente entre dos variables aleatorias χ^2, que son positivas, también será positiva, y su gráfica similar a la de χ^2

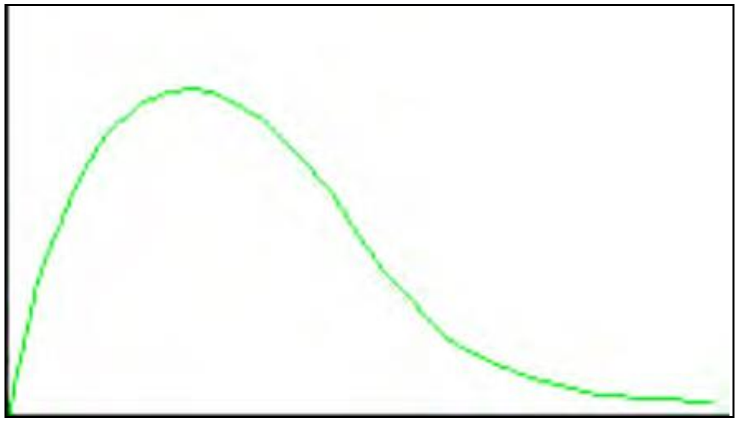


Figura Nº 10 - Distribución de Snedecor

Su función de densidad surge del conocimiento de las variables aleatorias ξ^2 que la integran, y es la siguiente:

$$f(F)=\frac{\Gamma\left(\frac{\upsilon_1+\upsilon_2}{2}\right)\upsilon_1^{\frac{\upsilon_1}{2}}\upsilon_2^{\frac{\upsilon_2}{2}}F^{\frac{\upsilon_1}{2}-1}}{\Gamma\left(\frac{\upsilon_1}{2}\right)\Gamma\left(\frac{\upsilon_2}{2}\right)\left(\upsilon_2+\upsilon_1 F\right)^{\frac{\upsilon_1+\upsilon_2}{2}}}\ ,\quad para\ F>0 \qquad (25)$$

Se utiliza fundamentalmente en la parte de inferencia estadística. Existen tablas que brindan valores de probabilidad del siguiente tipo: $P(F>F_p,\ \nu_1,\ \nu_2)=P(\%)$. Debido a que la función depende de los grados de libertad, se necesita una tabla a triple entrada para obtener valores tabulados de F que corresponden a distintas probabilidades y a distintos valores de los grados de libertad.

A veces es necesario tener en cuenta la siguiente relación:

$$F_{1-\alpha,\upsilon_2,\upsilon_1}=\frac{1}{F_{\alpha,\upsilon_1,\upsilon_2}} \qquad (26)$$

MODELOS DE VALORES EXTREMOS

En muchas situaciones prácticas especialmente en ingeniería, es de interés trabajar con el mayor o el menor valor de un número de variables aleatorias.

Si la variable y es considerada el máximo de una serie de n variables aleatorias $x_1 \ldots x_n$ es posible obtener series y_t que estarán compuestas por los máximos. Es posible así obtener una expresión para la probabilidad de que el máximo sea menor o igual que un valor dado:

$P(Y \leq y)=F(y)=P(para\ todo\ n\ de\ x_i \leq y)$

Si los valores de x son independientes, entonces:

$$F(y) = P(X_1 \leq y)\, P(X_2 \leq y) \ldots P(X_n \leq y) = \\ = F_{x_1}(y) F_{x_2}(y) \ldots F_{x_n}(y)$$

Si los x_i son idénticamente distribuidos, F(x), luego:

$F(y)=[F_x(y)]^n$

y, cuando n → : se deriva el modelo buscado.

De acuerdo a las características de la distribución inicial se originan tres tipos de modelos asintóticos de valores extremos:

Modelo Tipo I:

Surge de aquellas distribuciones iniciales que no tienen límite superior. El extremo de la curva correspondiente a la función de densidad debe decrecer tan rápidamente como una función exponencial; entonces, valores extremos provenientes de una distribución normal, log-normal, gamma, pueden ser ajustados por un Modelo Tipo I de máximos. En cambio, si la distribución inicial no es limitada en la dirección de los mínimos, se origina un Modelo Tipo I de mínimos.

Modelo Tipo II:

Este modelo surge de aquellas distribuciones iniciales ilimitadas, y que poseen un número finito de momentos.

Modelo Tipo III:

Este modelo surge cuando la distribución inicial está limitada en la dirección del valor extremo, es así que la distribución de valores mínimos provenientes de distribuciones log-normal, gamma y beta pueden ajustarse por un Modelo Tipo III.

Existen una serie de condiciones que se deben establecer en el desarrollo del modelo:

-Las observaciones de las que se extraen los valores extremos deben ser independientes.
-Las observaciones deben ser hechas bajo las mismas condiciones, es decir, las distribuciones iniciales y los parámetros que contienen deben ser los mismos.
-El número de observaciones n de las cuales se extraen los valores extremos debe ser grande; en algunas situaciones, día y años son unidades naturales de periodicidad.

Modelo Tipo III -Weibull

Este modelo se origina con las mismas consideraciones hechas para el anterior, a las cuales se les agrega las siguientes:

-La distribución de los X_i está limitada superiormente por un valor ω.

-La distribución de X_i es del siguiente tipo:

$$P(X \leq x) = 1 - c(\omega - x)^k$$
$$con\ x \leq \omega, k > 0$$

De esto se obtienen las funciones para el mayor valor de los X_i:

$$F(x) = e^{-\left(\frac{\omega - x}{\omega - \mu_0}\right)^k},\ x < \omega$$
$$f(x) = \frac{k}{\omega - \mu_0}\left(\frac{\omega - x}{\omega - \mu_0}\right)^{k-1} e^{-\left(\frac{\omega - x}{\omega - \mu_0}\right)^k} \qquad (37)$$

Esta forma del modelo no es muy aplicada en muchas ramas de las ciencias, quizá por el límite superior que presenta. Una transformación de la variable resulta en el conocido modelo de Weibull. Si se trabaja con el valor negativo de la variable z = -x, estando x definida para valores menores que ω, z estará definida para valores mayores que -ω. Llamando -ω = ε, y observando que las probabilidades de no excedencias serán ahora de excedencias, la función de probabilidad estará dada por:

$$P(Z \geq z) = e^{-\left(\frac{z - \varepsilon}{\mu_0 - \varepsilon}\right)^k}$$

Por lo tanto, la función de distribución será:

$$P(Z \leq z) = 1 - e^{-\left(\frac{z-\varepsilon}{\mu_0 - \varepsilon}\right)^k} \qquad (38)$$

$$f(z) = \frac{k}{\mu_{0-}\varepsilon}\left(\frac{z-\varepsilon}{\mu_{0-}\varepsilon}\right)^{k-1} e^{-\left(\frac{z-\varepsilon}{\mu_{0-}\varepsilon}\right)^k}$$

Al trabajar con los valores negativos de una serie y maximizarlos, se estará obteniendo como extremo un valor mínimo, ya que el mayor valor de una serie negativa es el menor valor en valor absoluto.

Los parámetros de este modelo son μ_0, k y ε, siendo este último el límite inferior de los valores mínimos.

Las características son las siguientes:

$$E(x) = \mu = \varepsilon + (\mu_0 - \varepsilon)\Gamma\left(1 + \frac{1}{k}\right) \qquad (39)$$

$$Var(x) = \sigma^2 = (\mu_0 - \varepsilon)^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right] \qquad (40)$$

Generalmente es posible adoptar ε = 0, lo cual simplificaría las expresiones del modelo:

$$F(z) = 1 - e^{-\left(\frac{z}{\mu_0}\right)^k} \qquad (41)$$

$$f(z) = \frac{k}{\mu_0}\left(\frac{z}{\mu_0}\right)^{k-1} e^{-\left(\frac{z}{\mu_0}\right)^k}$$

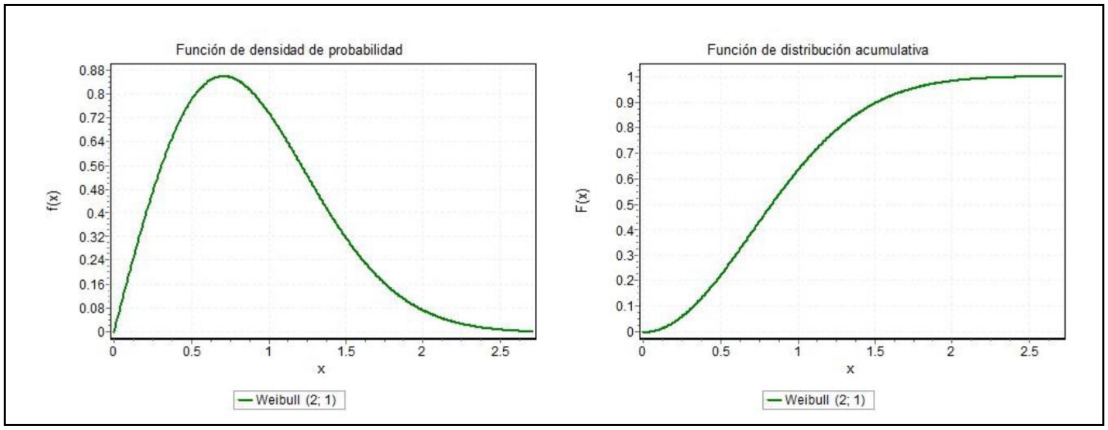


Figura Nº 16 – Funciones de la distribución de Weibull para distintos valores de sus parámetros

La primera distribución representada es la que tiene como parámetro de forma igual a 2 y de escala igual a 1; al no haber parámetro de localización, este se supone igual a 0, con lo cual la densidad y la distribución existen para valores mayores que ese valor. A la vista de la gráfica de la función de densidad, se puede deducir que en este caso la distribución es asimétrica positiva, mientras que de la representación de la función de distribución se puede deducir que el crecimiento suele ser constante hasta que x= 1.5, a partir de donde empieza a decaer ligeramente para crecer cada vez menos.

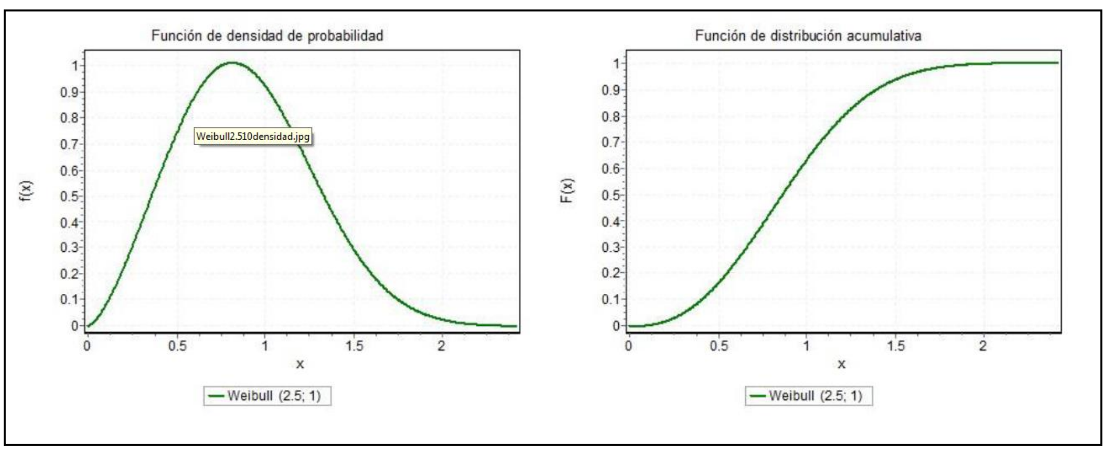


Figura Nº 17 – Funciones de la distribución de Weibull para distintos valores de sus parámetros

En este caso, el parámetro de forma vale 2.5, mientras que el de escala es igual a 1. La distribución es asimétrica positiva al estar la cola a la derecha y los valores con mayor probabilidad más a la izquierda.

Se observa que se tienen gráficas de forma flexible a medida que se cambian los parámetros con lo cual se adaptan a una buena cantidad de fenómenos tanto el modelo tipo I como el tipo III.