
Distribuciones continuas de probabilidad. Aplicación al calculo de intervalos de confianza en poblaciones normales

por *Guillermo Sánchez*

<http://web.usal.es/guillermo>

Universidad de Salamanca

<http://www.usal.es>

Actualizado : Corresponde a un tutorial desarrollado por el autor en el año 2000. Se ha vuelto a ejecutar con la versión 8 de *Mathematica* sin practicamente modificar el contenido.

Introducción

En esta práctica se describen y dan ejemplos de las distribuciones continuas de mayor empleo en Control Estadístico de Calidad (CEC). Asimismo se muestra el uso de estas para la determinación del intervalo de confianza.

Recordemos que una **distribución de probabilidad** es un modelo matemático que relaciona el valor de la variable con la probabilidad de ocurrencia de dicho valor en la población considerada $P\{x = x\} = p(x)$. Cuando el parametro toma forma de función continua en un determinado intervalo estamos ante una *Distribución Continua*.

Ejemplo 1.- Sea una variable aleatoria X el contenido real de una garrafa de aceite, en Litros. La distribución de probabilidad se asume que es de la forma: $f(x)=1/1.5$ en el intervalo $15.5 \leq x \leq 17.0$ (es decir es continua en dicho intervalo). Esta distribución es llamada *distribución uniforme*. La probabilidad de que el contenido de una deterenada garrafa sea menor o igual a 16.0 L se calcula sencillamente como:

$$\int_{15.5}^{16.0} \frac{1}{1.5} dx$$

0.333333

Representa graficamente la distribución anterior, a calcular su media y varianza [Recuerda que la media corresponde a la esperanza matematica $E[X] = \int_a^b x p(x) dx$ y que $V[X] = E[X^2] - (E[X])^2$].

Las distribución continuas mas utilizadas en CEC son la Normal, Exponencial, T-Student, Gamma y Weibull.

Distribuciones continuas

El Mathematica dispone probablemente mas funciones de distribución incorporada que ningun otro programa:

Continuous Distributions: tutorial/ContinuousDistributions

Distribucion Normal

Es probablemente la distribución mas importante. Entre otras muchas propiedades, en muestreo estadístico se utiliza la siguiente:

Tomados n valores x_i de una población no necesariamente normal la media muestral \bar{x} sigue aproximadamente una distribución de media μ y varianza $\frac{\sigma^2}{n}$, es decir $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

Para obtener información de ella podemos utilizar la ayuda en la forma habitual . Por ejemplo

? NormalDistribution*

NormalDistribution[μ , σ] represents a normal (Gaussian) distribution with mean μ and standard deviation σ .

NormalDistribution[] represents a normal distribution with zero mean and unit standard deviation. >>

Ojo: Observese que aplica el convenio N(media, desviación estándar) en vez del más usual de N(Media, varianza), por eso cuando se emplee algún programa de cálculo por primera vez es conveniente consultar el criterio que aplica en algunos términos

PDF[NormalDistribution[m, σ], x]
 CDF[NormalDistribution[m, σ], x]
 Mean[NormalDistribution[m, σ]]
 Variance[NormalDistribution[m, σ]]

$$\frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$$

$$\frac{1}{2} \operatorname{Erfc}\left[\frac{m-x}{\sqrt{2}\sigma}\right]$$

m

σ^2

Ejemplo 2.-Calcular para una distribución Normal -de media 7 y desviación standard 2 -, a) El valor de la función densidad y de la función de distribución para $x=5$, b) asimetría y c) curtosis. La forma de utilizar estas funciones es similar a la de las distribuciones discretas, se procederá como sigue:

Nota: Para calcular la densidad de probabilidad y la Distribución (acumulada) de probabilidad se utiliza, respectivamente PDF y CDF, cuyo significado es fácil obtener directamente de la ayuda

PDF[NormalDistribution[3, 2], 5] // N
 CDF[NormalDistribution[3, 2], 5] // N
 Skewness[NormalDistribution[3, 2]]
 Kurtosis[NormalDistribution[3, 2]]

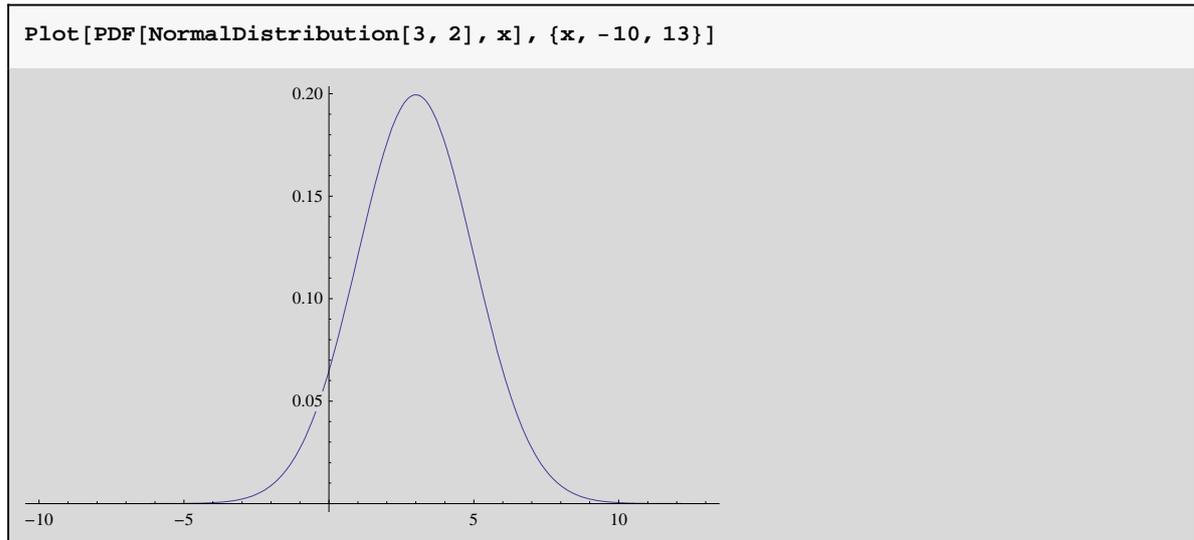
0.120985

0.841345

0

3

La representación gráfica de la función de densidad se puede hacer fácilmente



Aproximación de las distribución Binomial y de Poisson a la Normal

Para valores grandes de n , con $np \geq 10$, $p \geq 1/2$, la distribución binomial se puede aproximar por una normal de media np y varianza $np(1-p)$. Asimismo una distribución de Poisson se puede aproximar a una normal cuando la media $\lambda \geq 15$, siendo mejor mientras mas alto es λ .

Distribucion Exponencial

Esta función se utiliza fundamentalmente para estimar el tiempo de fallo de un equipo o un componente.

? ExponentialDistribution

ExponentialDistribution[λ] represents an exponential distribution with scale inversely proportional to parameter λ . >>

PDF[ExponentialDistribution[λ], x]
 CDF[ExponentialDistribution[λ], x]
 Mean[ExponentialDistribution[λ]]
 Variance[ExponentialDistribution[λ]]

$$\begin{cases} e^{-x\lambda} \lambda & x \geq 0 \\ 0 & \text{True} \end{cases}$$

$$\begin{cases} 1 - e^{-x\lambda} & x \geq 0 \\ 0 & \text{True} \end{cases}$$

$$\frac{1}{\lambda}$$

$$\frac{1}{\lambda^2}$$

Ejemplo 3.- La tasa de fallo de una lampara electrica es de 10^{-4} h^{-1} . ¿Cual es la duración media estimada?.

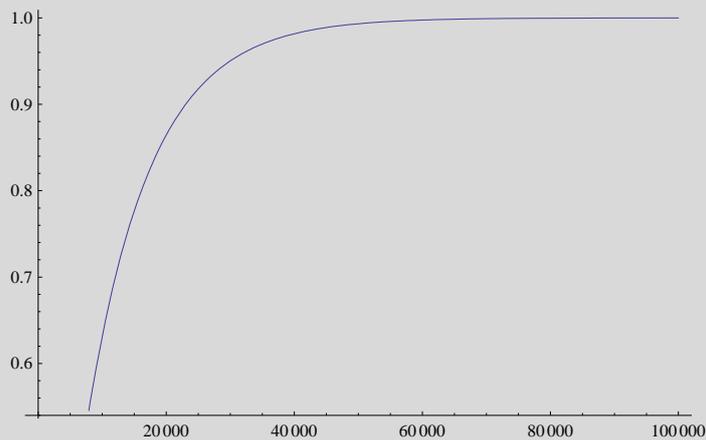
En este caso $\lambda=10^{-4}$ y por tanto

```
Mean[ExponentialDistribution[10-4]] (*en horas*)
```

```
10 000
```

La probabilidad de fallo en función del tiempo, para un periodo [0,T], con T =100000, podemos representarla como sigue:

```
Plot[CDF[ExponentialDistribution[10-4], t], {t, 0, 100 000}]
```



Distribucion Gamma

Esta función se utiliza especialmente para estimar la probabilidad de fallo en sistemas redundantes. las características fundamentales son:

```
Information["GammaDistribution", LongForm -> False]
```

GammaDistribution[α , β] represents a gamma distribution with shape parameter α and scale parameter β .

GammaDistribution[α , β , γ , μ] represents a generalized gamma

distribution with shape parameters α and γ , scale parameter β , and location parameter μ . >>

```
PDF[GammaDistribution[ $\alpha$ ,  $\beta$ ], x]
```

```
CDF[GammaDistribution[ $\alpha$ ,  $\beta$ ], x]
```

```
Mean[GammaDistribution[ $\alpha$ ,  $\beta$ ]]
```

```
Variance[GammaDistribution[ $\alpha$ ,  $\beta$ ]]
```

$$\begin{cases} \frac{e^{-\frac{x}{\beta}} x^{1+\alpha} \beta^{-\alpha}}{\text{Gamma}[\alpha]} & x > 0 \\ 0 & \text{True} \end{cases}$$

$$\begin{cases} \text{GammaRegularized}\left[\alpha, 0, \frac{x}{\beta}\right] & x > 0 \\ 0 & \text{True} \end{cases}$$

```
 $\alpha \beta$ 
```

```
 $\alpha \beta^2$ 
```

Ejemplo 4.-Consideremos un sistema de alimentación de corriente eléctrica esta formado por dos generadores diesel redundantes conectados en paralelo (un equipo está desconectado pero listo para funcionar en caso de fallo del otro). La tasa de fallo de cada

equipo esta dado por una distribución exponencial con una tasa de fallo de 10^{-4} h^{-1} . El conjunto será una distribución gamma con $\alpha=2$, y $\beta=1/\lambda=10^4$. El tiempo medio de fallo esperado para el conjunto es:

```
Mean[GammaDistribution[2, 10^4]]
```

```
20 000
```

Distribucion Weibull

Esta función es muy versatil pues una apropiada elección de los parametros alfa y beta permite ajustar datos experimentales a esta función.

```
Information["WeibullDistribution", LongForm -> False]
```

WeibullDistribution[α , β] represents a Weibull distribution with shape parameter α and scale parameter β .

WeibullDistribution[α , β , μ] represents a Weibull

distribution with shape parameter α , scale parameter β , and location parameter μ . >>

```
PDF[WeibullDistribution[ $\alpha$ ,  $\beta$ ], x]
```

```
CDF[WeibullDistribution[ $\alpha$ ,  $\beta$ ], x]
```

```
Mean[WeibullDistribution[ $\alpha$ ,  $\beta$ ]]
```

```
Variance[WeibullDistribution[ $\alpha$ ,  $\beta$ ]]
```

$$\begin{cases} \frac{e^{-\left(\frac{x}{\beta}\right)^\alpha} \alpha \left(\frac{x}{\beta}\right)^{-1+\alpha}}{\beta} & x > 0 \\ 0 & \text{True} \end{cases}$$

$$\begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha} & x > 0 \\ 0 & \text{True} \end{cases}$$

$$\beta \text{Gamma}\left[1 + \frac{1}{\alpha}\right]$$

$$\beta^2 \left(-\text{Gamma}\left[1 + \frac{1}{\alpha}\right]^2 + \text{Gamma}\left[1 + \frac{2}{\alpha}\right] \right)$$

Ejemplo 5.- El tiempo de fallo de un computador central satisface una función de Weibull con $\alpha=1/2$ y $\beta=1000$ h. El tiempo medio de fallo esperado es:

```
Mean[WeibullDistribution[1 / 2, 1000]]
```

```
2000
```

Distribucion χ^2 (ji-cuadrado)

Esta función es util para estimar σ^2 a partir de una muestra aleatoria formada por n valores x_i , $i=1, \dots, n$ en una población normal verificandose que $\chi^2 \sim (n-1) \frac{S^2}{\sigma^2}$.

```
? ChiSquareDistribution
```

ChiSquareDistribution[ν] represents a χ^2 distribution with ν degrees of freedom. >>

PDF[ChiSquareDistribution[n], x]
 CDF[ChiSquareDistribution[n], x]
 Mean[ChiSquareDistribution[n]]
 Variance[ChiSquareDistribution[n]]

$$\begin{cases} \frac{2^{-n/2} e^{-x/2} x^{-1+\frac{n}{2}}}{\text{Gamma}\left[\frac{n}{2}\right]} & x > 0 \\ 0 & \text{True} \end{cases}$$

$$\begin{cases} \text{GammaRegularized}\left[\frac{n}{2}, 0, \frac{x}{2}\right] & x > 0 \\ 0 & \text{True} \end{cases}$$

n

2 n

Distribucion t-student

Esta función es util en muestreo cuando el número de valores es pequeño.

? StudentTDistribution

StudentTDistribution[ν] represents a Student *t* distribution with ν degrees of freedom.
 StudentTDistribution[μ, σ, ν] represents a Student *t*
 distribution with location parameter μ, scale parameter σ, and ν degrees of freedom. >>

PDF[StudentTDistribution[n], x]
 CDF[StudentTDistribution[n], x]
 Mean[StudentTDistribution[n]]
 Variance[StudentTDistribution[n]]

$$\frac{\left(\frac{n}{n+x^2}\right)^{\frac{1+n}{2}}}{\sqrt{n} \text{Beta}\left[\frac{n}{2}, \frac{1}{2}\right]}$$

$$\left(\frac{1}{2} \sqrt{n} \text{Beta}\left[\frac{n}{2}, \frac{1}{2}\right] + x \text{Hypergeometric2F1}\left[\frac{1}{2}, \frac{1+n}{2}, \frac{3}{2}, -\frac{x^2}{n}\right]\right) / \left(\sqrt{n} \text{Beta}\left[\frac{n}{2}, \frac{1}{2}\right]\right)$$

$$\begin{cases} 0 & n > 1 \\ \text{Indeterminate} & \text{True} \end{cases}$$

$$\begin{cases} \frac{n}{-2+n} & n > 2 \\ \text{Indeterminate} & \text{True} \end{cases}$$

Distribucion F

Esta función se emplea entre otras cosas para comparar las varianzas muestrales S_1^2 y S_2^2 de dos poblaciones

$N(\mu_1, \sigma_1^2)$ $N(\mu_2, \sigma_2^2)$. Tiene la propiedad $F_{n_1-1, n_2-1} \sim \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$

? FRatioDistribution

FRatioDistribution[n, m] represents an F-ratio distribution with n numerator and m denominator degrees of freedom. >>

PDF[FRatioDistribution[n_1, n_2], x]

CDF[FRatioDistribution[n_1, n_2], x]

Mean[FRatioDistribution[n_1, n_2]]

Variance[FRatioDistribution[n_1, n_2]]

$$\begin{cases} \left(x^{-1+\frac{n_1}{2}} n_1^{-\frac{n_1}{2}} n_2^{\frac{n_2}{2}} (x n_1 + n_2)^{\frac{1}{2}(-n_1-n_2)} \right) / \text{Beta}\left[\frac{n_1}{2}, \frac{n_2}{2}\right] & x > 0 \\ 0 & \text{True} \end{cases}$$

$$\begin{cases} \text{BetaRegularized}\left[\frac{x n_1}{x n_1 + n_2}, \frac{n_1}{2}, \frac{n_2}{2}\right] & x > 0 \\ 0 & \text{True} \end{cases}$$

$$\begin{cases} \frac{n_2}{-2+n_2} & n_2 > 2 \\ \text{Indeterminate} & \text{True} \end{cases}$$

$$\begin{cases} \frac{2 n_2^2 (-2+n_1+n_2)}{n_1 (-4+n_2) (-2+n_2)^2} & n_2 > 4 \\ \text{Indeterminate} & \text{True} \end{cases}$$

Determinación de intervalos estadísticos para la media en poblaciones normales

Introducción

Un problema frecuente en medidas de variables estadísticas (inspección, intercomparaciones, validaciones, etc) es el establecimiento de intervalos de confianza, tolerancia y de predicción.

Aquí vamos a mostrar como se pueden estimar los intervalos de confianza en el caso de poblaciones normales cuando se desea garantizar que una proporción de la población este entre determinados límites con un probabilidad determinada.

Vamos a utiliza el programa *Mathematica*, del que vamos a necesitar el siguiente paquete. Ejecuta la celda

```

BeginPackage["Estadistica`Intervalos`"]
(*v1 desarrollado por Guillermo Sanchez *)
Klexacto::usage = "Calcula la K unilateral por un metodo exacto";
g1::usage = "Calcula la K unilateral";
g2::usage = "Calcula la K bilateral";
K1::usage = "Muestra la K unilateral";
K2::usage = "Muestra la K bilateral";
limiteBilateral::usage = "Muestra los limites superior en inferior del intervalo";
limiteUnilateral::usage = "Muestra los limites superior en inferior del intervalo";
z1::usage = "Calcula los intervalos de confianza de la media, varianza conocida, para el
t1::usage = "Calcula los intervalos de confianza de la media , varianza desconocida, par
p1::usage = "Calcula si se acepta o rechaza una muestra ";

Begin["`Private`"]

g1[n_, alfa_, p_] := Module[{zp, za, a, b}, zp = Quantile[NormalDistribution[0, 1], p];
  za = Quantile[NormalDistribution[0, 1], 1 - alfa]; a = 1 - za^2/(2*(n - 1));
  b = zp^2 - za^2/n; K1[n, 1 - alfa, p] -> (zp + Sqrt[zp^2 - a*b])/a;
g2[n_, alfa_, p_] := Module[{z, ji2}, z = Quantile[NormalDistribution[0, 1], (1 + p)/2];
  ji2 = Quantile[ChiSquareDistribution[n - 1], alfa];
  K2[n, 1 - alfa, p] -> z*Sqrt[(n - 1)/ji2]*(1 + 1/(2*n));

limiteBilateral[media_, sd_, n_, alfa_, p_] = {"LS"-> media + g2[n, alfa, p][[2]]*sd,
  "LI"-> media - g2[n, alfa, p][[2]]*sd};

limiteUnilateral[media_, sd_, n_, alfa_, p_] = {"LS"-> media + g1[n, alfa, p][[2]]*sd,

z1[media_, sd_, n_, alfa_] :=
Module[{za, zb}, za = Quantile[NormalDistribution[0, 1], 1 - alfa];
zb = Quantile[NormalDistribution[0, 1], 1 - alfa/2];
{"unilateral:", {"LS" -> media + za*sd/Sqrt[n], "LI" -> media - za*sd/Sqrt[n]}, "bilate
t1[media_,sd_,n_,alfa_] :=
Module[{za,zb},za=Quantile[StudentTDistribution[n-1],1-alfa];
zb=Quantile[StudentTDistribution[n-1],1-alfa/2];
{"unilateral:",{"LS"->media+za*sd/Sqrt[n],"LI"->media-za*sd/Sqrt[n]},
"bilateral:",{"LS"->media+zb*sd/Sqrt[n],"LI"->media-zb*sd/Sqrt[n]}}];

p1[mean_, sigma_, size_, alfa_, p_, uL_, lL_] :=
Module[{pLS, pLI}, pLS = FindRoot[g1[size, alfa, pS][[2]] == (uL - mean)/sigma,
{pS, 0.5, 0.99}, MaxIterations -> 50][[1,2]];
pLI = FindRoot[g1[size, alfa, 1 - pI][[2]] == (mean - lL)/sigma,
{pI, 0.01, 0.2}, MaxIterations -> 50][[1,2]]; pLS - pLI >= p];
End[]

EndPackage[]

```

Estadistica`Intervalos`

Estadistica`Intervalos`Private`

Estadistica`Intervalos`Private`

Lo conveniente es que salve la celda anterior como paquete en un directorio que llamara Estadistica, entonces copie ese directorio en Mathematica Application, entonces podrá llamar al paquete que siempre que lo necesite:

```
Needs["Estadística`Intervalos`"]
```

```
? "Estadística`Intervalos`*"
```

▼ Estadística`Intervalos`

g1	K1	K2	limiteUnilateral	t1
g2	K1exacto	limiteBilateral	p1	z1

Notación

μ , σ = Valores verdaderos de la media y desviación estandar de la población.

\bar{y} , s = Estimadores de la media y desviación estandar muestral.

$z(\alpha)$ = El percentil α de la distribución normal estándar $N(0,1)$. Ejemplo: el valor de z para una normal $N(0,1)$ por debajo del cual se encuentra el 95% de la población es $z(0.95) = 1.64$.

$t(m; \alpha)$ = El percentil α de la t Student con m grados de libertad. Ejemplo: el valor de t para $\alpha = 95\%$ y $m = 5$ es $t(5; 0.95) = 2.01$.

$t'(m, \delta; \alpha)$ = El percentil α de la t Student no central con m grados de libertad y un parametro de descentralización δ . Su interpretación es mas compleja que en los casos anteriores.

T-Student No central

Vamos a necesitar la T-Student No central. La información disponible en al ayuda es la siguiente:

```
? NoncentralStudentTDistribution
```

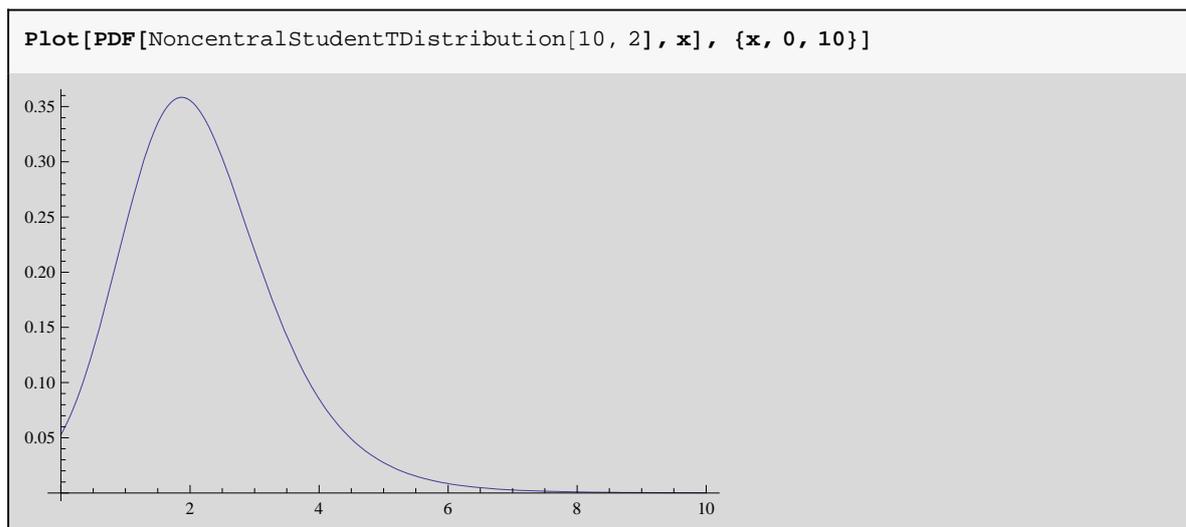
NoncentralStudentTDistribution[ν , δ] represents a noncentral Student t distribution with ν degrees of freedom and noncentrality parameter δ . >>

La función de densidad de esta distribución es

```
PDF[NoncentralStudentTDistribution[n,  $\delta$ ], x]
```

$$\frac{1}{\pi} 2^n e^{-\frac{\delta^2}{2}} n^{1+\frac{n}{2}} (n+x^2)^{\frac{1}{2}(-1-n)} \text{Gamma}\left[\frac{1+n}{2}\right] \text{HermiteH}\left[-1-n, -\frac{x\delta}{\sqrt{2}\sqrt{n+x^2}}\right]$$

Su representación gráfica es para $n = 10$, y $\delta = 2$



Intervalos de confianza

Conceptos

Al dar un estadístico (ej: la media), además de su valor frecuentemente debemos precisar la incertidumbre existente en la estimación.

Sea \bar{y} la media, llamaremos intervalo de confianza para \bar{y} con nivel de confianza $(1 - \alpha)$ aquel que tiene una probabilidad de $(1 - \alpha)100\%$ de contener la verdadera media μ . Los límites generalmente suelen ser:

- Bilateral centrado \bar{y} que consisten en calcular $\bar{y} \pm k \sigma$ tal que la verdadera media está comprendida en el intervalo con una probabilidad $100p\%$.
- Unilateral superior que consisten en calcular $\bar{y} + k \sigma$ tal que la verdadera media μ está por debajo $\bar{y} + k \sigma$ con una probabilidad $100p\%$.
- Unilateral inferior que consisten en calcular $\bar{y} - k \sigma$ tal que la verdadera media μ está por encima de $\bar{y} - k \sigma$ con una probabilidad $100p\%$.

Si tomamos n muestras de una población, y de cada muestra calculamos la media y_i , las expresiones específicas para los límites de los distintos intervalos de \bar{y} , que denotamos por Y_p para un nivel de confianza $(1 - \alpha)100\%$ son las siguientes:

Varianza conocida

Si conocemos la varianza de la población, σ , por el teorema central del límite, sabemos que el error relativo en la determinación de μ mediante la media muestral \bar{y} es: $z = (\bar{y} - \mu) / (\sigma / \sqrt{n})$ es una variable normal estándar de lo que se deduce los siguientes límites para los intervalos

Límite unilateral inferior

$$Y_p = \bar{y} - \frac{z(1 - \alpha)}{\sqrt{n}} \sigma$$

Límite unilateral superior

$$Y_p = \bar{y} + \frac{z(1 - \alpha)}{\sqrt{n}} \sigma$$

Límite bilateral inferior

$$Y_p = \bar{y} - \frac{z(1 - \alpha / 2)}{\sqrt{n}} \sigma$$

Límite bilateral superior

$$Y_p = \bar{y} + \frac{z(1 - \alpha / 2)}{\sqrt{n}} \sigma$$

Varianza desconocida

Si desconocemos la varianza de la población utilizaremos la distribución t que la que se verifica $t = (\bar{y} - \mu)/(s/\sqrt{n})$, donde s es la desviación estándar muestral, con la que se obtiene los siguientes límites para los intervalos

Límite unilateral inferior

$$Y_p = \bar{y} - \frac{t(n-1, 1-\alpha)}{\sqrt{n}} s$$

Límite unilateral superior

$$Y_p = \bar{y} + \frac{t(n-1, 1-\alpha)}{\sqrt{n}} s$$

Límite bilateral inferior

$$Y_p = \bar{y} - \frac{t(n-1, 1-\alpha/2)}{\sqrt{n}} s$$

Límite bilateral superior

$$Y_p = \bar{y} + \frac{t(n-1, 1-\alpha/2)}{\sqrt{n}} s$$

Intervalos de tolerancia

Lo haremos en una sesión nueva por ello salimos de la actual

```
Quit[]
```

Conceptos

Definimos el $100p$ -ésimo percentil al la valor Y_p por debajo del cual está el $100p$ % de la población, o, lo que es equivalente, aquel valor para el que las observaciones de muestra aleatoria de la población cae con una probabilidad p . En ocasiones nos interesa estimar el valor de Y_p asociándole un intervalo de confianza, $(1 - \alpha) 100\%$.

Ejemplo: Deseamos calcular el valor por debajo del cual está el 95% de la población con un nivel de confianza del 95%. Problemas similares consisten en calcular límites centrales dentro de los cuales está el $100p\%$ de la población, y límites inferiores por encima del cual está $100p\%$ de la población. Asimismo un límite superior de confianza $(1 - \alpha) 100\%$ para el percentil $100p$ -ésimo es equivalente al límite superior de tolerancia por debajo del cual está al menos una proporción p de la población con una probabilidad $1 - \alpha$. Una interpretación similar cabe del límite inferior de tolerancia.

Las expresiones específicas para obtener los límites de tolerancia Y_p son las siguientes

Límite unilateral inferior

$$Y_p = \bar{y} - \frac{1}{\sqrt{n}} t'(n-1, -z(p) \sqrt{n}; 1-\alpha) s$$

Límite unilateral superior

$$Y_p = \bar{y} + \frac{1}{\sqrt{n}} t'(n-1, z(p) \sqrt{n}; 1-\alpha) s$$

Límite bilateral inferior

$$Y_p = \bar{y} - \frac{1}{\sqrt{n}} t'(n-1, -z(p) \sqrt{n}; 1 - \frac{\alpha}{2}) s$$

Límite bilateral superior

$$Y_p = \bar{y} + \frac{1}{\sqrt{n}} t'(n-1, z(p) \sqrt{n}; 1 - \frac{\alpha}{2}) s$$

Calculo de $k_{n,\alpha,p}$ (método exacto)

Los factores que en la formulas (9) a (12) multiplican a s se conocen como $k_{n,\alpha,p}$ y están dados por la ecuación

$$k_{n,\alpha,p} = \frac{1}{\sqrt{n}} t'(n-1, z(p) \sqrt{n}; 1-\alpha)$$

De forma exacta puede calcularse aplicando:

```
K1Exacto[n_, α_, p_] := Module[{z}, z = Quantile[NormalDistribution[0, 1], p];
  (1 / Sqrt[n]) * Quantile[NoncentralStudentTDistribution[n - 1, z * Sqrt[n]],
  1 - α]]
```

Ejemplo: La $k(95/95)$ unilateral para una muestra de tamaño 2 a 10 son

```
Table[{r, K1Exacto[r, 0.05, 0.95]}, {r, 2, 10}]
{{2, 26.2597}, {3, 7.6559}, {4, 5.14387}, {5, 4.20268},
 {6, 3.70768}, {7, 3.39947}, {8, 3.18729}, {9, 3.03124}, {10, 2.91096}}
```

Los valores son practicamente identicos a los encontrados en las tablas consultadas

Calculo de $k_{n,\alpha,p}$ (Método aproximado)

El método anterior requiere un consumo eleva de recursos de ordenador que lo hace poco adecuado cuando el valor n empieza a ser elevado ($n > 10$). Una aproximación excelente a (13) puede obtenerse para el caso bilateral, que es el mas complicado aplicando la siguiente aproximación

$$k_{n,\alpha,p} \approx z[(1+p)/2] \left(\frac{(n-1)}{\chi^2(n-1; \alpha)} \right)^{1/2} \left(1 + \frac{1}{2n} \right)$$

Expresado en el lenguaje del *Mathematica* es

```
g2[n_, α_, p_] := Module[{z, ji2}, z = Quantile[NormalDistribution[0, 1],  $\frac{1+p}{2}$ ];
  ji2 = Quantile[ChiSquareDistribution[n - 1], α]; z  $\sqrt{\frac{n-1}{ji2} \left(1 + \frac{1}{2n}\right)}$ 
```

Ejemplo: La k(95/95) bilateral para una muestra de tamaño $n = 50$ y otra para $n = 1000$ son

```
g2[50, 0.05, 0.95]
```

```
2.37889
```

```
g2[1000, 0.05, 0.95]
```

```
2.03608
```

Estos valores muestran una excelente aproximación para los encontrados en tablas

También podemos aplicar una aproximación para el caso unilateral cuando el valor n empieza a ser elevado ($n > 10$).

$$k_{n,\alpha,p} = \frac{1}{\sqrt{n}} t'(n-1, z(p) \sqrt{n}; 1-\alpha) \approx \left(z(p) + [z^2(p) - ab]^{1/2} \right) / a$$

donde

$$a = 1 - [z(1-\alpha)]^2 / [2(n-1)]$$

$$b = [z(p)]^2 - [z(1-\alpha)]^2 / n$$

Expresado en el lenguaje del *Mathematica* es

```
g1[n_, α_, p_] := Module[{zp, za, a, b}, zp = Quantile[NormalDistribution[0, 1], p];
  za = Quantile[NormalDistribution[0, 1], 1 - α];
  a = 1 -  $\frac{za^2}{2(n-1)}$ ; b = zp^2 -  $\frac{za^2}{n}$ ;  $\frac{zp + \sqrt{zp^2 - ab}}{a}$ 
```

```
g1[10, 0.05, 0.95]
```

```
2.8748
```

Los aproximación aumenta con el tamaño de la muestras: Ejemplo: Calcular la k(95/95) unilateral para $n = \{10, 100, \text{de } 10 \text{ en } 10\}$:

```
Table[{n, g1[n, 0.05, 0.95]}, {n, 10, 100, 10}]
```

```
{{10, 2.8748}, {20, 2.37835}, {30, 2.20851}, {40, 2.11721}, {50, 2.05849},
 {60, 2.01681}, {70, 1.98533}, {80, 1.9605}, {90, 1.94029}, {100, 1.92344}}
```

Criterios para determinar si una muestra está dentro del intervalo de tolerancia para un P y alfa dados

Dado un intervalo de tolerancia definido por los límites superior e inferior (LS, LI) pretendemos saber si una muestra de media \bar{x} y desviación estandar s está comprendida en el intervalo de tolerancia para un P

dado

Conocidos los LS e LI de aceptación, y los parámetros muestrales \bar{x} , s y n , se trata de calcular p fijado α y ver si este p es mayor o igual que un P dado. Para ello:

a) Hacemos $Y_p = LS$ en (9) y $Y_p = LI$ en (10) y obtenemos

$$g1[n, \text{alfa}, pS] == \frac{LS - \bar{x}}{s} \text{ y de aquí } pS \left[g1[n, \text{alfa}], \frac{LS - \bar{x}}{s} \right] = p_{LS}$$

$$g1[n, \text{alfa}, 1 - pI] == \frac{\bar{x} - LI}{s} \text{ y de aquí } pI \left[g1[n, \text{alfa}], \frac{\bar{x} - LI}{s} \right] = p_{LI}$$

donde p_{LS} indica la proporción de la población comprendida entre LS y 0 y p_{LI} indica la proporción de la población comprendida entre LI y 0 (estrictamente en vez de 0 es $-\infty$, pero asumimos que la variable x sólo toma valores positivos).

b) La población será aceptada si se verifica que $p_{LS} - p_{LI} = p \geq P$, siendo P la proporción de la población que se considera aceptable.

Creamos una sentencia que nos automatice el anterior proceso (para $P = 0.95$ y $\text{alfa} = 0.05$), donde `mean` es la media de la muestra \bar{x} , `size` su tamaño n , `uL` le LS, `lL` el LI y `sigma` la s de la muestra

```
p[mean_, size_, uL_, lL_, sigma_] := Module[{pLS, pLI},
  pLS = FindRoot[g1[size, 0.05, pS] == (uL - mean)/sigma, {pS, 0.5, 0.99}][[1,2]];
  pLI = FindRoot[g1[size, 0.05, 1 - pI] == (mean - lL)/sigma,
  {pI, 0.01, 0.2}][[1,2]]; {pLS, pLI, pLS - pLI >= 0.95}]
```

Ejemplo

Sea $\bar{x} = 10.1$, $s = 0.3$, $LS = 10.8$, $LI = 9.3$, $n = 50$, ¿Es rechazable la muestra para un $\text{alfa} = 0.05$ y un $P = 0.95$? La solución aplicando la sentencia anterior es que con dicha muestra podríamos aceptar la población

```
p[10.1, 50, 10.8, 9.3, 0.3]
{0.970028, 0.0151466, True}
```

Aclaración.

Este criterio da valores muy próximos al caso bilateral, aunque algo más conservadores como era de esperar. Veámoslo con un ejemplo

Sea $\bar{x} = 10$, $s = 0.3$, $n = 50$, calculemos el caso centrado, pero utilizando en un caso los k unilaterales con α fijo = 0.05, utilizando el criterio bilateral y unilaral

Con el criterio bilateral los límites de tolerancia estarían en

```
10 + g2[50, 0.05, 0.95] 0.3
10.7137
```

```
10 - g2[50, 0.05, 0.95] 0.3
9.28633
```

Con el caso unilateral serían

```
10 + g1[50, 0.05, 0.975] 0.3
10.7276
```

```
10 - g1[50, 0.05, 0.975] 0.3
```

```
9.27242
```

Determinación de intervalos estadísticos para la media en poblaciones normales

Objeto

Vamos a determinar el intervalo de confianza para la desviación estándar y la varianza:

- Para una población normal con media μ y desviación estándar σ de la que se toma una o varias muestras.
- Para varias poblaciones normales que tienen una misma varianza, $N(\bar{x}_1, \sigma)$, $N(\bar{x}_2, \sigma)$, ..., $N(\bar{x}_n, \sigma)$, de cada una de las cuales se toma una muestra.

Intervalo de confianza para la varianza de una población normal

Sean X_1, X_2, \dots, X_n variables aleatorias (v.a.) normales $N(0,1)$ independientes entre sí. Se puede demostrar que la variable:

$$Y = X_1^2 + X_2^2 + \dots + X_n^2$$

es un distribución χ_n^2 con n grados de libertad

La función de densidad (fdd) de χ_n^2 es

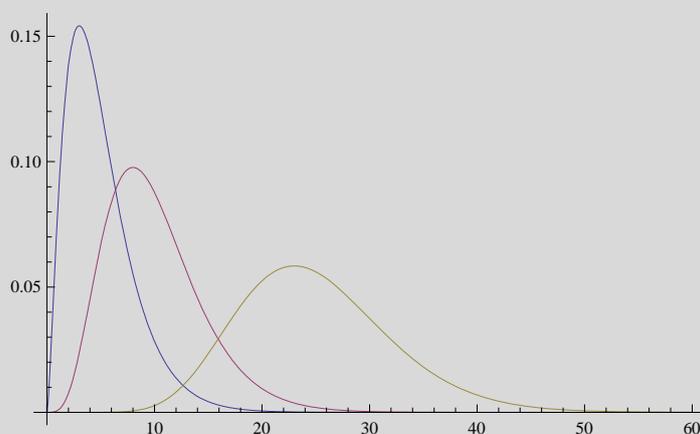
```
PDF[ChiSquareDistribution[n], x]
```

$$\begin{cases} \frac{2^{-n/2} e^{-x/2} x^{-1+\frac{n}{2}}}{\Gamma\left[\frac{n}{2}\right]} & x > 0 \\ 0 & \text{True} \end{cases}$$

La función gamma se define como $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$

En el fig. se representa la fdd de χ_n^2 para distintos valores de n .

```
Plot[{PDF[ChiSquareDistribution[5], x], PDF[ChiSquareDistribution[10], x],
      PDF[ChiSquareDistribution[25], x]}, {x, 0, 60}, PlotRange -> All]
```



De una muestra normal $N(\mu, \sigma)$ tomamos una muestra aleatoria $\{X_1, X_2, \dots, X_n\}$. Si a cada valor X_i le restamos la media μ y lo dividimos entre la varianza σ y lo elevamos al cuadrado obtendremos (2) que es una distribución de

las mismas características que (1), es decir tendremos una distribución χ_n^2

$$Y = \left(\frac{X_1 - \mu}{\sigma}\right)^2 + \left(\frac{X_2 - \mu}{\sigma}\right)^2 + \dots + \left(\frac{X_n - \mu}{\sigma}\right)^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \equiv \chi_n^2$$

Fijado un coef. de confianza γ podemos encontrar dos números a y b de forma que $P(a \leq \chi_n^2 \leq b) = \gamma$, donde $a = \chi_{n, (1-\gamma)/2}^2$ y $b = \chi_{n, (1+\gamma)/2}^2$ para lo que procedemos como sigue:

$$\begin{aligned} \gamma &= P\left(a \leq \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \leq b\right) = \\ &P\left(\frac{1}{b} \sum_{i=1}^n (X_i - \mu)^2 \leq \sigma^2 \leq \frac{1}{a} \sum_{i=1}^n (X_i - \mu)^2\right) \end{aligned}$$

Sustituimos a y b por sus valores y reagrupamos los términos para obtener el intervalo de confianza para σ^2

$$\begin{aligned} \frac{1}{\chi_{n, (1+\gamma)/2}^2} \sum_{i=1}^n (X_i - \mu)^2 \leq \sigma^2 \leq \frac{1}{\chi_{n, (1-\gamma)/2}^2} \sum_{i=1}^n (X_i - \mu)^2 = \\ \frac{n s^2}{\chi_{n, (1+\gamma)/2}^2} \leq \sigma^2 \leq \frac{n s^2}{\chi_{n, (1-\gamma)/2}^2} \end{aligned}$$

En el caso en que la media μ sea desconocida se tiene en cuenta que de acuerdo al teorema de Fisher $(n-1) s_c^2 / \sigma^2$ es una distribución χ_{n-1}^2 y se procede de forma similar al caso anterior obteniéndose

$$\frac{(n-1) s_c^2}{\chi_{n, (1+\gamma)/2}^2} \leq \sigma^2 \leq \frac{(n-1) s_c^2}{\chi_{n, (1-\gamma)/2}^2}$$

Determinación del intervalo de confianza de σ de varias poblaciones normales que tienen en común la misma varianza

Sean varias poblaciones $N(\bar{x}_1, \sigma)$, $N(\bar{x}_2, \sigma)$, ..., $N(\bar{x}_g, \sigma)$, ..., $N(\bar{x}_m, \sigma)$, de cada una de las cuales se toma una muestra de tamaño n con la que se determina la desviación estándar de dicha muestra. Un ejemplo de este tipo es el caso de un proceso productivo del que vamos tomando periódicamente muestras aleatorias independientes del que medimos un parámetro p . Para dicho parámetro de cada muestra tenemos su media y varianza \bar{x}_g , s_g^2 .

Pasado un tiempo tendremos las varianzas $\{s_1^2, \dots, s_m^2\}$ de m muestras que asumimos tomadas con el proceso bajo control. Nuestro objetivo es calcular la dispersión natural σ del proceso, que suponemos desconocida, con un determinado intervalo de confianza. A los límites de dichos intervalos los denominamos Límite Superior de Control (LSC_s) y Límite Inferior de Control (LIC_s). Para cualquier muestra que tomemos con posterioridad mediremos su desviación típica, s_m , si s_m está comprendido en el intervalo (LSC_s , LIC_s) consideramos que dicha muestra se ha tomado con el proceso está bajo control, en caso contrario consideraremos que está desajustado.

Para determinar el intervalo de confianza para s hemos de calcular la esperanza matemática $E[s]$ y su dispersión $D[s]$

Esperanza matemática de s , $E[s]$

Recordemos que la esperanza matemática para funciones continuas se define como

$$E[z] = \int_{-\infty}^{\infty} z f_z(z) dz$$

En nuestro caso z es s . Para determinar la fdd $f_z(z)$ tenemos en cuenta que $n \frac{s^2}{\sigma^2}$ sigue una distribución χ_{n-1}^2 y hacemos la transformación $n \frac{s^2}{\sigma^2} = n \frac{y}{\sigma^2} = x$. De esta forma obtenemos la fdd de s^2

$$f_{s^2}(y) = f_{\chi_{n-1}^2}(x) \frac{dx}{dy} = f_{\chi_{n-1}^2}(x[Y]) \frac{n}{\sigma^2}$$

$$\mathbf{fs2} = \text{PDF}[\text{ChiSquareDistribution}[n-1], x] \frac{n}{\sigma^2} /. x \rightarrow \frac{n}{\sigma^2} y$$

$$\frac{1}{\sigma^2} n \left(\begin{cases} \left(2^{\frac{1-n}{2}} e^{-\frac{ny}{2\sigma^2}} \left(\frac{ny}{\sigma^2} \right)^{-1+\frac{1}{2}(-1+n)} \right) / \text{Gamma}\left[\frac{1}{2}(-1+n)\right] & \frac{ny}{\sigma^2} > 0 \\ 0 & \text{True} \end{cases} \right)$$

Como $s = \sqrt{s^2}$ para obtener la fdd (PDF) de s hacemos el cambio de variable $z = \sqrt{y}$

$$f_s(z) = f_{s^2}(y) \frac{dy}{dz} = f_{s^2}(y[z]) 2z$$

$$\mathbf{f[z_]} = \mathbf{fs2} 2z /. y \rightarrow z^2$$

$$\frac{1}{\sigma^2} 2nz \left(\begin{cases} \left(2^{\frac{1-n}{2}} e^{-\frac{nz^2}{2\sigma^2}} \left(\frac{nz^2}{\sigma^2} \right)^{-1+\frac{1}{2}(-1+n)} \right) / \text{Gamma}\left[\frac{1}{2}(-1+n)\right] & \frac{nz^2}{\sigma^2} > 0 \\ 0 & \text{True} \end{cases} \right)$$

Para calcular la esperanza matemática sustituimos en (6)

$$\text{Integrate}[z f[z], \{z, 0, \infty\}, \text{Assumptions} \rightarrow \text{Re}[n] > 0 \ \&\& \ \text{Re}\left[\frac{n}{\sigma^2}\right] > 0] // \text{FullSimplify}$$

$$\left\{ \begin{array}{l} -\left(\sqrt{2} \sigma \text{Gamma}\left[\frac{n}{2}\right]\right) / \left(\sqrt{n} \text{Gamma}\left[\frac{1}{2}(-1+n)\right]\right) \quad n > 0 \ \&\& \ \sigma < 0 \\ \left(\sqrt{2} \sigma \text{Gamma}\left[\frac{n}{2}\right]\right) / \left(\sqrt{n} \text{Gamma}\left[\frac{1}{2}(-1+n)\right]\right) \quad n > 0 \ \&\& \ \sigma > 0 \\ 0 \quad \text{True} \end{array} \right.$$

por tanto

$$E[s] = c_2 \sigma$$

con

$$c_2[n_] = \frac{\sqrt{2} \text{Gamma}\left[\frac{n}{2}\right]}{\sqrt{n} \text{Gamma}\left[\frac{1}{2}(-1+n)\right]}$$

Dispersión de s , $D[s]$

La dispersión esta dada por

$$D[s] = \sqrt{E[s^2] - (E[s])^2}$$

como

$$E[s^2] = \frac{n-1}{n} \sigma^2 \text{ y } E[s] = c_2 \sigma$$

resulta

$$D[s] = c_3 \sigma$$

con

$$c3[n_] = \sqrt{\frac{n-1}{n} - c2[n]^2} // FullSimplify$$

$$\sqrt{\left(1 - \frac{1}{n} - \frac{2 \text{Gamma}\left[\frac{n}{2}\right]^2}{n \text{Gamma}\left[\frac{1}{2}(-1+n)\right]^2}\right)}$$

Determinación del intervalo de confianza

El intervalo de confianza para s se puede determinar a) Utilizando el criterio habitual que se emplea en la construcción de cartas de control que consiste en calcularlo para 3σ o b) fijando el nivel de confianza α

Intervalo de confianza para 3σ

Se calcula como sigue

$$E[s] - k D[s] \leq \sigma \leq E[s] + k D[s]$$

a) Para σ conocida

$$LSC_s = B_2 \sigma \text{ con } B_2 = c_2 + 3c_3$$

$$B2[n_] = c2[n] + 3 c3[n] // FullSimplify$$

$$\frac{\sqrt{2} \text{Gamma}\left[\frac{n}{2}\right]}{\sqrt{n} \text{Gamma}\left[\frac{1}{2}(-1+n)\right]} + 3 \sqrt{\left(1 - \frac{1}{n} - \left(2 \text{Gamma}\left[\frac{n}{2}\right]^2\right) / \left(n \text{Gamma}\left[\frac{1}{2}(-1+n)\right]^2\right)\right)}$$

$$LIC_s = B_1 \sigma \text{ con } B_1 = c_2 - 3c_3$$

$$B1[n_] = c2[n] - 3 c3[n] // FullSimplify$$

$$\frac{\sqrt{2} \text{Gamma}\left[\frac{n}{2}\right]}{\sqrt{n} \text{Gamma}\left[\frac{1}{2}(-1+n)\right]} - 3 \sqrt{\left(1 - \frac{1}{n} - \left(2 \text{Gamma}\left[\frac{n}{2}\right]^2\right) / \left(n \text{Gamma}\left[\frac{1}{2}(-1+n)\right]^2\right)\right)}$$

a) Para σ desconocida

$$LSC_s = B_4 \bar{s} \text{ con } B_4 = \frac{c_2 + 3c_3}{c_2}$$

$$LIC_s = B_3 \bar{s} \text{ con } B_3 = \frac{c_2 - 3c_3}{c_2}$$

$$B4[n_] = \frac{c2[n] + 3 c3[n]}{c2[n]} // FullSimplify$$

$$1 + \left(3 \sqrt{n} \text{Gamma}\left[\frac{1}{2}(-1+n)\right] \sqrt{\left(1 - \frac{1}{n} - \left(2 \text{Gamma}\left[\frac{n}{2}\right]^2\right) / \left(n \text{Gamma}\left[\frac{1}{2}(-1+n)\right]^2\right)\right)}\right) / \left(\sqrt{2} \text{Gamma}\left[\frac{n}{2}\right]\right)$$

$$\mathbf{B3[n_]} = \frac{\mathbf{c2[n]} - 3 \mathbf{c3[n]}}{\mathbf{c2[n]}} // \mathbf{FullSimplify}$$

$$1 - \left(3 \sqrt{n} \operatorname{Gamma}\left[\frac{1}{2}(-1+n)\right] \sqrt{\left(1 - \frac{1}{n} - \left(2 \operatorname{Gamma}\left[\frac{n}{2}\right]^2\right) / \left(n \operatorname{Gamma}\left[\frac{1}{2}(-1+n)\right]^2\right)\right)} \right) / \left(\sqrt{2} \operatorname{Gamma}\left[\frac{n}{2}\right]\right)$$

Ejemplo: Calcular B4 para $n = 5, 10, 15, 20, 25$

```
Table[{i, B4[i]}, {i, 5, 25, 5}] // N
```

```
{ {5., 2.089}, {10., 1.71629}, {15., 1.5718}, {20., 1.48977}, {25., 1.43521} }
```

Intervalo de confianza dado α

Se trata de calcular: $P(s \leq \text{LSC}_s / \sigma = \sigma_0) = \alpha$, para ello se realiza la siguiente transformación:

$$P\left(\frac{ns^2}{\sigma^2} \leq \frac{n\text{LSC}_s^2}{\sigma^2} / \sigma = \sigma_0\right) = \alpha$$

como $\frac{ns^2}{\sigma^2} = \frac{n\text{LSC}_s^2}{\sigma^2} = \chi^2_{n-1}$ se trata de calcular

$$\text{LSC}_s = \sigma \sqrt{\frac{\chi^2_{\alpha, n-1}}{n}}$$

Si σ es desconocida, que es lo normal, se tiene en cuenta que $\sigma = \frac{\bar{s}}{c_2}$ resultando

$$\text{LSC}_s = \frac{\bar{s}}{c_2} \sqrt{\frac{\chi^2_{\alpha, n-1}}{n}}$$

$$\text{Llamamos } c\text{Alfa} = \frac{1}{c_2} \sqrt{\frac{\chi^2_{\alpha, n-1}}{n}}$$

$$\mathbf{cAlfa}[\alpha_, n_]:= \frac{1}{\mathbf{c2[n]}} \mathbf{Sqrt}\left[\frac{1}{n} \mathbf{Quantile}[\mathbf{ChiSquareDistribution}[n-1], 1-\alpha]\right]$$

Ejemplo: El coeficiente a aplicar para calcular el Límite superior de control para un nivel de confianza bilateral de $\alpha = 0.05$ y un tamaño de muestra de 25 uds será

```
cAlfa[0.025, 25]
```

```
1.2941
```

Ejemplo: Supongamos que con las características anteriores hemos obtenido con el proceso bajo control que $\bar{s} = 2.0$. Con posterioridad hemos tomado una muestra m obteniendo $s_m = 2.8$. Para saber si el proceso sigue bajo control comprobamos si $s_m \leq \text{LSC}_s$

```
2 cAlfa[0.025, 25]
```

```
2.5882
```

Como 2.8 es mayor que este valor el proceso estaría fuera de control

Es importante aclarar que usualmente se suele aplicar el criterio 3σ pues si se elijen niveles de confianza muy altos ($\alpha > 0.01$) el riesgo de falsos rechazos puede ser inaceptable.