DISTRIBUCIÓN NORMAL MULTIVARIANTE
(DERIVADAS DE LA NORMAL)

INTRODUCCIÓN
PROPIEDADES
LINEALIDAD

DISTRIBUCIONES DERIVADAS DE LA NORMAL
χ^2 DE PEARSON
t DE STUDENT
F DE SNEDECOR

DISTRIBUCIÓN NORMAL MULTIVARIANTE

Un vector aleatorio X=, sigue una distribución normal k-variante o k-dimensional

de vector de medias ,

$$\mathbf{M} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ \mu_k \end{pmatrix}$$

y de matriz de varianzas,

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & . & . & \sigma_{1,k} \\ \sigma_{1,2} & \sigma_2^2 & . & . & \sigma_{2,k} \\ . & . & . & . & . \\ . & . & . & . & . \\ \sigma_{1,k} & \sigma_{2,k} & . & . & \sigma_k^2 \end{pmatrix}$$

VNIVERSITAT DE VALÈNCIA OpenCourseWare

Estadística
Curso 2010-2011
Juan Mtnez. de Lejarza
Ignacio Mtnez. de Lejarza

lo que se expresa como: X~ $N_k(M;V)$ si su función de densidad conjunta obedece a la expresión:

$$f(X) = f(x_1, x_2, \ldots, x_k) = \frac{e^{-\frac{1}{2}\left[(X-M)' V^{-1} (X-M)\right]}}{\sqrt{(2\pi)^k \det V}}$$

PROPIEDADES DE LA DISTRIBUCIÓN NORMAL MULTIVARIANTE:

1.-Si un vector aleatorio sigue una distribución normal multivariante, puede demostrarse que todas las distribuciones marginales son normales, de forma que cada $x_i \sim N(m_i ; s_i)$. Igualmente el resultado recíproco se cumple también: dadas k variables aleatorias normales, su distribución conjunta es una normal k-dimensional.

2.- Puede demostrarse que las distribuciones condicionadas de cualquier dimensión son también normales y que, en particular, las de dimensión uno son normales unidimensionales que tienen por media el valor esperado por la regresión lineal, y por varianza la varianza residual, de esa regresión.

3.- Un caso importante de distribución normal k-dimensional es aquél, en el que todas las variables son independientes. En este caso todas las covarianzas serán nulas y la matriz de varianzas será diagonal:

$$\mathbf{V=}\begin{pmatrix} \sigma_1^2 & 0 & . & . & 0 \\ 0 & \sigma_2^2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \sigma_k^2 \end{pmatrix}$$

4.- Un resultado aún más importante es que en el caso de que tengamos un conjunto de variables normales la incorrelación implica independencia estocástica, cosa, que recordemos que, en general, no es cierta, pero sí en el caso de normalidad de las variables.

TRANSFORMACIONES LINEALES DE NORMALES MULTIVARIANTES:

Dado un vector aleatorio X tal que, $X \rightarrow N_n[M,V]$

y dada la matriz de transformación: $A = \begin{pmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ . & . & . & . \\ a_{k1} & a_{k2} & . & a_{kn} \end{pmatrix}$

el nuevo vector aleatorio k-dimensional $Y = A\ X$

será tal que : $Y \rightarrow N_k[AM\ ,\ AVA']$

DISTRIBUCIONES DERIVADAS DE LA DISTRIBUCIÓN NORMAL:

DISTRIBUCIÓN χ^2 de PEARSON

Esta distribución, junto con la t de Student y la F de Snedercor (además de la normal) son de fundamental importancia para el desarrollo de la inferencia estadística. Como las otras dos , es la distribución de una cierta característica de los datos obtenidos aleatoriamente a partir de una distribución normal . En consecuencia puede hacerse derivar de un proceso experimental de selección aleatoria, aunque también puede ponerse en relación con las distribuciones Eulerianas.

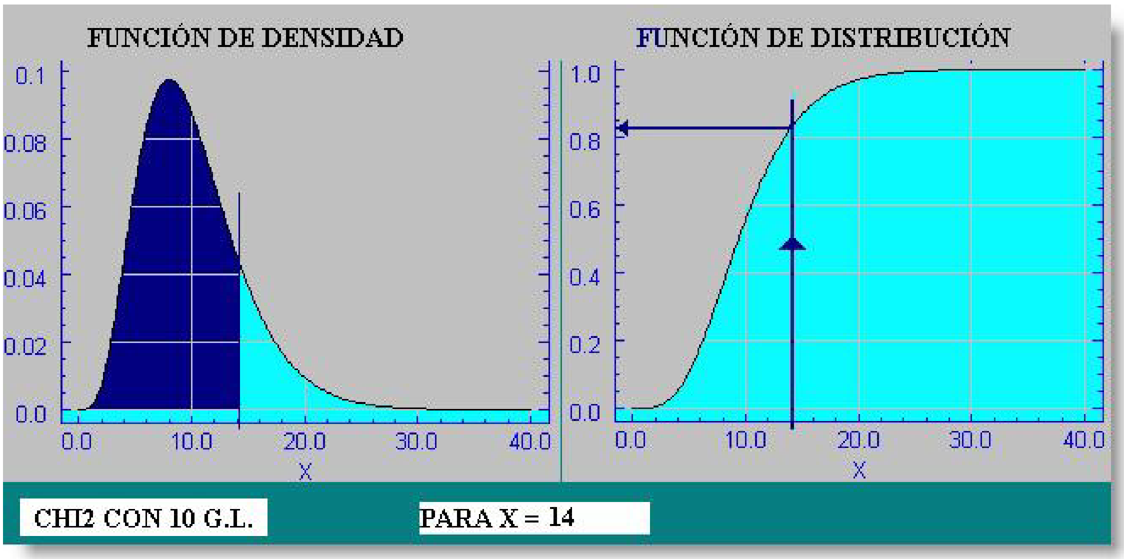
Se define la χ^2 con n grados de libertad, como la distribución que sigue la variable suma de los cuadrados de n variables normales (tipificadas) [0;1] independientes.

Así, dadas n variables aleatorias independientes $Z_i \quad (i = 1, 2, 3..n)$

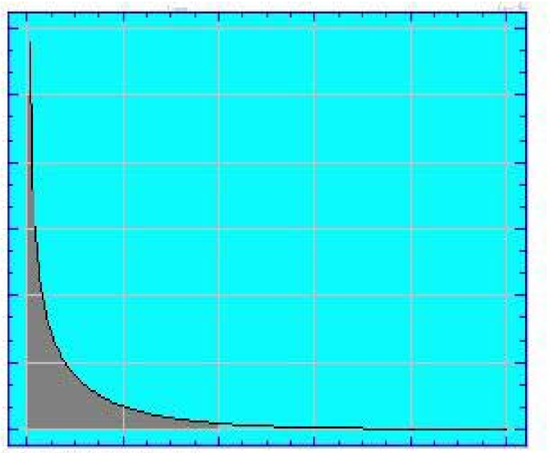
tales que : $Z_i \Rightarrow N[0;1] \ \ \forall i$

la variable $\chi = \sum_{i=1}^{n} z_i$ seguirá una χ_n^2 con n grados de libertad.

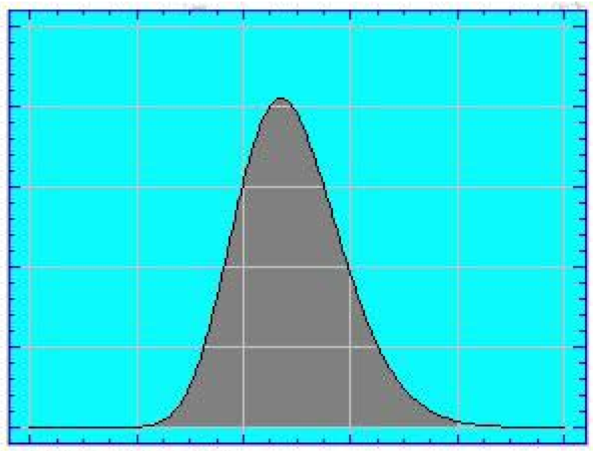
Dado que la χ_n^2 es la suma de n normales tipificadas al cuadrado podemos afirmar, por el motivo de ser precisamente cuadrados, que el campo de actuación-variación de la variable así distribuída será siempre positivo .La forma de la función de densidad y distribución variarán según los grados de libertad, siendo su forma habitual la campaniforme, así



El cálculo de las diversas probabilidades para los diferentes valores de la variable X, se explicita normalmente en tablas desarrolladas para los diversos valores de grados de libertad; nosotros planteamos un script (programa) para su cálculo directo. (Ir a script de la chi2)



chi 2 con 1 g.l



chi 2 con 49 g.l.

Como hemos dicho ,la representación gráfica de la distribución (su forma) varía según los valores que tome su parámetro n (grados de libertad); así y como puede observarse en el gráfico animado , para grados de libertad bajos (sobre todo para un g.l.) el conjunto de la probabilidad queda muy próxima al valor cero de la variable ; de esta característica surge , para algunos tipos de contrastes que utilizan la Chi2 , la corrección de Yates

La función de densidad de la χ_n^2 vendrá dada por:

$$f(x;n) = \frac{e^{-x/2} \cdot x^{\left(n/2 - 1\right)}}{2^{n/2} \cdot \Gamma\left(n/2\right)}$$:

Siendo $\tau(n/2)$ una distribución gamma de Euler de parámetro $(n/2)$
y siendo n (número de grados de libertad) el único parámetro de la distribución χ^2_n

La función generatriz de momentos vendrá dada por

$$\alpha(t) = (1-2t)^{-n/2}$$

partiendo de ella podemos establecer por aplicación del teorema de los momentos que la media vendrá dada $\mu = n$ siendo la varianza $\sigma = 2n$

La distribución JHI-2 (chi 2) cumple el teorema de adición para su parámetro n (grados de libertad) , así la suma de dos chi2 con n y m grados de libertad respectivamente , no será otra cosa que una chi2 con (n+m) grados de libertad . Es lógico pensar que si una chi2 con n grados de libertad es la suma de n normales tipificadas(independientes) al cuadrado ; y que una chi2 con m grados de libertad es la suma de m normales tipificadas(independientes) al cuadrado , la suma de ambas será , obviamente, la suma de (m+n) normales tipificadas (independientes) al cuadrado : es decir, una chi2 con (n+m) grados de libertad.

DISTRIBUCIÓN t DE STUDENT

La distribución t de student (desarrollada por Gosset) es, con la chi2, la F de Snedecor, y, por supuesto, la normal, transcendental para aplicaciones inferenciales, en especial para aquellas en las que se desconoce la varianza; dado que no depende de las varianzas de las variables que la integran.

Su expresión formal parte de dos variables X e Y tales que :

$$X \to N[0;1] \quad e \quad Y \to \chi_n^n$$

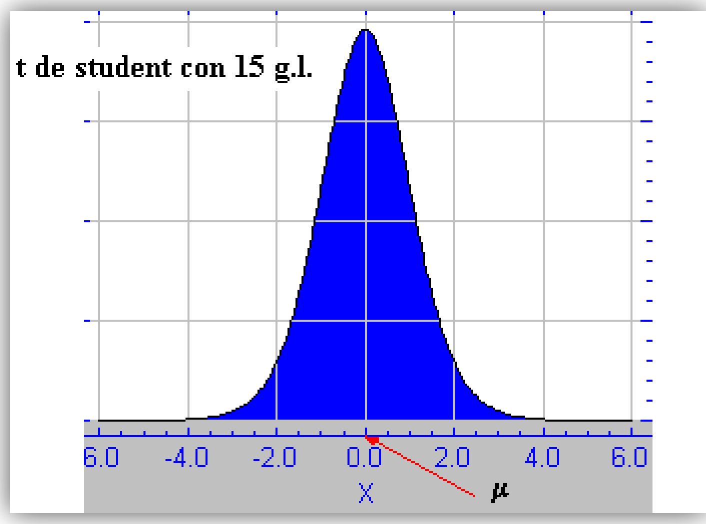
de manera que

$$t = \frac{X}{\sqrt{\frac{Y}{n}}}$$

siendo t una nueva distribución conocida como t de student con n grados de libertad.

La distribución t de student. admite, también , una definición alternativa , tomada como la distribución marginal de la primera variable de una distribución "normal-gamma" ; en este sentido la expresión de su función de densidad vendría dada por :

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}}\left[1+\frac{x^{2}}{n}\right.$$



siendo n los grados de libertad que actuán de parámetro y la función τ gamma de Euler

La distribución t de student con n grados de libertad tiene siempre como media el valor "0" , sea cuales fueren los grados de libertad ; es simétrica ,asintóticamente tiende a $\pm\ \infty$,de forma campaniforme, al igual que la distribución normal, teniendo como varianza el valor $\frac{n}{n-2}$

Los valores de la función de probabilidad para los diversos x_i están tabulados atendiendo al único parámetro de la distribución, es decir, el número de grados de libertad; evidentemente, es más recomendable para su cálculo la utilización del programa que presentamos.

En otro orden de cosas la distribucción t de student con n grados de libertad congerge en ley a una normal tipificada (estandarizada) cuando el número de grados de libertad tiende a infinito

DISTRIBUCIÓN F DE SNEDECOR.

Diremos que una variable aleatoria F tiene una distribución F de Snedecor con m grados de libertad en el numerador y n grados de libertad en el denominador

($F \rightarrow F_{m,n}$) cuando la variable aleatoria F sea:

$$F=\frac{\frac{U}{m}}{\frac{V}{m}}$$

donde U y V son dos variables aleatorias independientes que tienen por distribución:

$U \rightarrow \chi^2_m$ y $W \rightarrow \chi^2_n$ respectivamente

La determinación de probabilidades y valores críticos resulta sencilla mediante la Tabla de la F de Snedecor

---