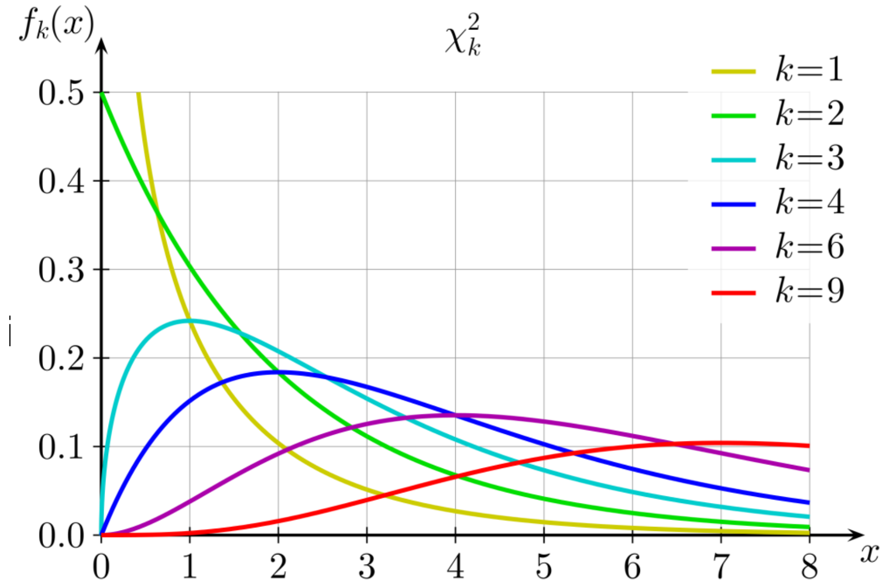
Tema 4 Distribuciones de probabilidad (II)



1. La distribución binomial.
2. La distribución multinomial
3. La distribución de Poisson
4. La distribución uniforme.
5. La distribución exponencial.
6. La distribución normal.
7. La distribución binormal.
8. La distribución multinormal
9. **Distribuciones de Muestreo**
    1. **La distribución de chi-cuadrado**
    2. **La distribución t de Student**
    3. **La distribución F**

9.1 La distribución de chi-cuadrado

Sea un conjunto de n variables aleatorias independientes $x_1, x_2, \ldots, x_n$ distribuidas normalmente con medias $\mu_1, \mu_2, \ldots, \mu_n$ y varianzas $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ respectivamente. La pdf conjunta viene dada por:

$$f(\bar{x}, \bar{\mu}, \bar{\sigma}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] = \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i}$$

Se define la variable aleatoria χ^2 como:

$$\chi^2(n) = \sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$$

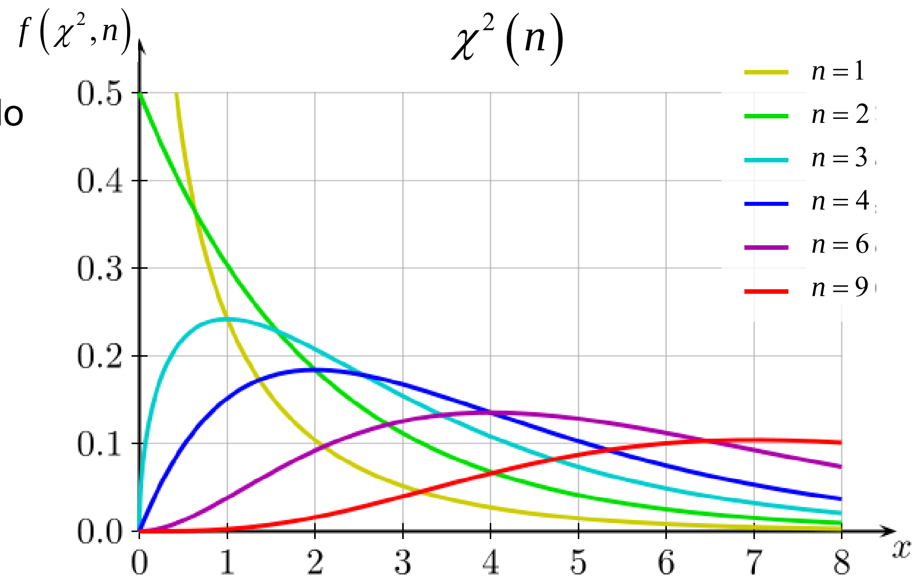
Donde el parámetro n se denomina número de grados de libertad pues cada variable aleatoria puede variar independientemente del resto.

La nueva variable así definida se distribuye de acuerdo

Con la función:

$$f(\chi^2, n) = \frac{(\chi^2)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

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



9.1 La distribución de chi-cuadrado

Función chi-cuadrado con 1 grado de libertad

Para $n = 1$ tenemos solo una variable gaussiana. Hacemos el cambio de variable $z = \left(\dfrac{x-\mu}{\sigma}\right)$
Con lo que la pdf queda como:

$$f(z) = f(x)\left|\frac{dx}{dz}\right| = \frac{f(x)}{\left|\frac{dz}{dx}\right|} = \sigma f(x) = \sigma \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)$$

Para pasar a una variable χ^2 hacemos un nuevo cambio de variable: $q = z^2$
Con lo que la pdf de la nueva variable q queda como:

¡Cuidado!
Función bivaluada

$$f(q)dq = f(z)\left(|J_+| + |J_-|\right)dq = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\left(\frac{dq}{2\sqrt{q}} + \frac{dq}{2\sqrt{q}}\right) = \frac{1}{\sqrt{2\pi q}}\exp\left(-\frac{q}{2}\right)dq$$

$$J_{\pm} = \frac{d(\pm z)}{dq} = \pm\frac{1}{2\sqrt{q}}$$

$$f(\chi^2, n) = \frac{\left(\chi^2\right)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$$

Distribución de χ^2
con 1 grado de libertad

9.1 La distribución de chi-cuadrado

Función chi-cuadrado con 2 grados de libertad

$$\chi^2 = \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2$$

La pdf es: $$f(x_1,x_2) = f_1(x_1)\,f_2(x_2) = \left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]\right)\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left[-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right)$$

Hacemos el cambio de variables a coordenadas polares:

$$\left.\begin{aligned} \frac{x_1-\mu_1}{\sigma_1} &= \rho\cos\phi \\ \frac{x_2-\mu_2}{\sigma_2} &= \rho\sin\phi \end{aligned}\right\} \Rightarrow \quad J = \begin{vmatrix} \frac{\partial x_1}{\partial \rho} & \frac{\partial x_1}{\partial \phi} \\ \frac{\partial x_2}{\partial \rho} & \frac{\partial x_2}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \sigma_1\cos\phi & -\rho\sigma_1\sin\phi \\ \sigma_2\sin\phi & \rho\sigma_2\cos\phi \end{vmatrix} = \rho\sigma_1\sigma_2 \Rightarrow$$

La nueva pdf en coordenadas polares será:

$$f(\rho,\phi) = f_1(x_1)\,f_2(x_2)\,|J| = \left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]\right)\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left[-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right)\rho\sigma_1\sigma_2 = \frac{1}{2\pi}e^{-\frac{1}{2}\rho^2}\rho$$

Integrando para ϕ: $\Rightarrow$ $f(\rho) = \rho e^{-\frac{1}{2}\rho^2}$ Y haciendo el cambio de variable: $\chi^2 = \rho^2$ $\quad \rho = \sqrt{\chi^2}$

$$f(\chi^2)d\chi^2 = f(\rho)|J|d\chi^2 = \rho e^{-\frac{1}{2}\rho^2}\frac{1}{2\rho}d\chi^2 = \frac{1}{2}e^{-\frac{1}{2}\chi^2}d\chi^2$$

$$|J| = \left|\frac{\partial\rho}{\partial\chi^2}\right| = \left|\frac{1}{2\sqrt{\chi^2}}\right| = \left|\frac{1}{2\rho}\right|$$

Distribución de χ^2 con 2 grados de libertad

$$f(\chi^2,n) = \frac{(\chi^2)^{\frac{n}{2}-1}e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$$

9.1 La distribución de chi-cuadrado

Función chi-cuadrado con 3 grados de libertad

Definimos: $z_i = \left(\frac{x_i - \mu_i}{\sigma_i}\right)$ → $R^2 = \chi^2 = z_1^2 + z_2^2 + z_3^2$ La nueva pdf viene dada por::

$$f(z_1, z_2, z_3)dz_1dz_2dz_3 = f(x_1, x_2, x_3)\left|J\right|dz_1dz_2dz_3 = \frac{1}{(2\pi)^{3/2}}\exp\left(-\frac{R^2}{2}\right)dz_1dz_2dz_3$$

$\left|J\right| = \sigma_1\sigma_2\sigma_3$

R viene a ser el radio de una esfera tridimensional

El cambio de variables a coordenadas esféricas equivale a:

$$dz_1dz_2dz_3 = R^2dRd\cos\theta d\phi$$

Integrando para cos(θ) y ϕ :

$$f(R)dR = \frac{2}{\sqrt{2\pi}}R^2\exp\left(-\frac{R^2}{2}\right)dR$$

Probabilidad de que R se encuentre entre R y R+dR

Por último haciendo el cambio de variable:

$$\chi^2 = R^2 \rightarrow R = \sqrt{\chi^2} \rightarrow \left|J\right| = \left|\frac{dR}{d\chi^2}\right| = \frac{1}{2\sqrt{\chi^2}}$$

$$f(\chi^2)d\chi^2 = \frac{2}{\sqrt{2\pi}}R^2\exp\left(-\frac{R^2}{2}\right)\left|J\right|d\chi^2 = \frac{(\chi^2)^{1/2}}{\sqrt{2\pi}}\exp\left(-\frac{\chi^2}{2}\right)d\chi^2$$

$$f(\chi^2, n) = \frac{(\chi^2)^{\frac{n}{2}-1}e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$$

Distribución de χ^2 con 3 grados de libertad

9.1 La distribución de chi-cuadrado

Función chi-cuadrado con *n* grados de libertad

La densidad de probabilidad será: $f(x_1,x_2,\cdots,x_n)=\dfrac{1}{(2\pi)^{n/2}\,\sigma_1\sigma_2\cdots\sigma_n}e^{-\frac{\chi^2}{2}}$ con $\chi^2=\sum_{i=1}^{n}\left(\dfrac{x_i-\mu_i}{\sigma_i}\right)^2$

Pasando a coordenadas tipificadas: $z_i=\left(\dfrac{x_i-\mu_i}{\sigma_i}\right)$; $dz_i=\dfrac{dx_i}{\sigma_i}$ la pdf se transforma en:

$$f(z_1,z_2,\cdots,z_n)dz_1dz_2\cdots dz_n=\frac{1}{(2\pi)^{n/2}\,\sigma_1\sigma_2\cdots\sigma_n}e^{-\frac{\chi^2}{2}}\left|J\right|dz_1dz_2\cdots dz_n=\frac{e^{-\frac{\chi^2}{2}}}{(2\pi)^{n/2}}dz_1dz_2\cdots dz_n$$

$\left|J\right|=\sigma_1\sigma_2\cdots\sigma_n$

$\chi^2=z_1^2+z_1^2+\cdots+z_n^2$

χ^2 es una distancia al cuadrado entre el origen y el punto $(z_1,z_2,\cdots,z_n)$ en un espacio de n dimensiones

La probabilidad de que χ esté entre χ y $\chi+d\chi$ es proporcional a:

$$f(\chi)d\chi\propto\exp\left(-\frac{\chi^2}{2}\right)\chi^{n-1}d\chi$$

$d\chi^2=2\chi d\chi$

En función de χ^2: $f(\chi^2)d\chi^2=K\exp\left(-\dfrac{\chi^2}{2}\right)\left(\chi^2\right)^{\frac{n}{2}-1}d\chi^2$

Cambio de variable

$y=\dfrac{\chi^2}{2};\quad dy=\dfrac{d\chi^2}{2}$

Normalizando: $\displaystyle\int_0^\infty f(\chi^2;n)d\chi^2=\int_0^\infty Ke^{-\frac{\chi^2}{2}}\left(\chi^2\right)^{\frac{n}{2}-1}d\chi^2=K2^{n/2}\int_0^\infty e^{-y}y^{\frac{n}{2}-1}dy=K2^{n/2}\Gamma\left(\frac{n}{2}\right)=1$ $\rightarrow$ $K=\dfrac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}$

Distribución de χ^2 con n grados de libertad

$$f(\chi^2,n)=\frac{\left(\chi^2\right)^{\frac{n}{2}-1}e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$$

9.1 La distribución de chi-cuadrado. Propiedades

Al número de grados de libertad se acostumbra llamarlo por ν

$$f(\chi^2,\nu)=\frac{\left(\chi^2\right)^{\frac{\nu}{2}-1}e^{-\frac{\chi^2}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}$$

Función característica

$$\Phi(t)=E\left[e^{it\chi^2}\right]=\int_0^\infty e^{it\chi^2}f\left(\chi^2,\nu\right)d\chi^2=\left(1-2it\right)^{-\frac{\nu}{2}}$$

A partir de la función característica es fácil obtener los momentos:

Valor medio

$$E\left[\chi^2\right]=\left.\frac{\partial\Phi}{\partial(it)}\right|_{t=0}=\nu$$

$$E\left[\chi^2\right]=\nu$$

Varianza

$$V\left[\chi^2\right]=E\left[\left(\chi^2\right)^2\right]-E\left[\chi^2\right]^2=\left.\frac{\partial^2\Phi}{\partial(it)^2}\right|_{t=0}-\nu^2=\nu(\nu+2)-\nu^2=2\nu$$

$$V\left[\chi^2\right]=2\nu$$

Skewness

$$\gamma_1=\left(\frac{8}{\nu}\right)^{\frac{1}{2}}$$

Kurtosis

$$\gamma_2=\frac{12}{\nu}$$

La distribución de χ^2 tiende a la distribución normal en el límite de $\nu\to\infty$

Visualmente para $\nu\geq 20$ ya se puede considerar gaussiana

9.1 La distribución de chi-cuadrado. Propiedades

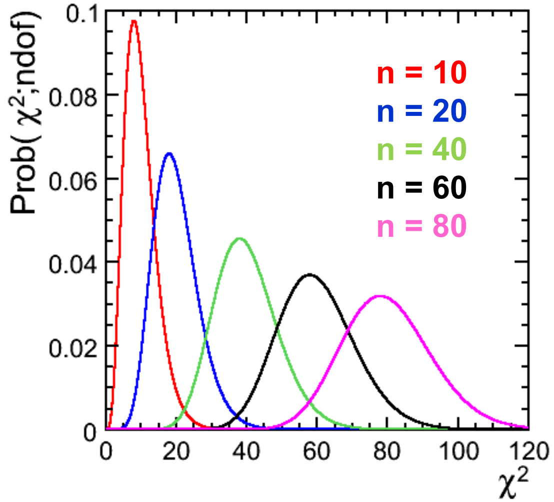
Comportamiento asintótico gaussiano

La distribución de χ^2 tiende a la distribución normal en el límite de $\nu \to \infty$

Definimos la variable estandarizada $y = \dfrac{\chi^2 - \nu}{\sqrt{2\nu}}$ y calculamos su función característica

$$\Phi_y(t) = E\left[e^{ity}\right] = E\left[\exp\left(it\left(\frac{\chi^2-\nu}{\sqrt{2\nu}}\right)\right)\right] = \exp\left(-\frac{it\nu}{\sqrt{2\nu}}\right)E\left[\exp\left(\frac{it\chi^2}{\sqrt{2\nu}}\right)\right] =$$

$$= \exp\left(-\frac{it\nu}{\sqrt{2\nu}}\right)\Phi_{\chi^2}\left(\frac{t}{\sqrt{2\nu}}\right) = \exp\left(-\frac{it\nu}{\sqrt{2\nu}}\right)\left(1-\frac{2it}{\sqrt{2\nu}}\right)^{-\frac{\nu}{2}}$$



$$\ln(1-x) = -\left[x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots\right]$$

Tomando logaritmos y expandiendo

$$\ln\Phi_y(t) = -\frac{it\nu}{\sqrt{2\nu}} - \frac{\nu}{2}\ln\left(1-\frac{2it}{\sqrt{2\nu}}\right) = -\frac{it\nu}{\sqrt{2\nu}} + \frac{\nu}{2}\left[\left(\frac{2it}{\sqrt{2\nu}}\right) + \frac{1}{2}\left(\frac{2it}{\sqrt{2\nu}}\right)^2 + \frac{1}{3}\left(\frac{2it}{\sqrt{2\nu}}\right)^3 + \cdots\right] = -\frac{t^2}{2} + \vartheta\left(\frac{1}{\sqrt{\nu}}\right)$$

$$\ln\Phi_y(t) = -\frac{t^2}{2} + \vartheta\left(\frac{1}{\sqrt{\nu}}\right) \xrightarrow[\nu \to \infty]{} \Phi_z(t) = e^{-\frac{t^2}{2}}$$

Función característica de una distribución normal con media 0 y varianza unidad

En el límite de infinitos grados de libertad la variable y se distribuye como una variable $N(0,1)$

En el límite de infinitos grados de libertad la variable χ^2 se distribuye como una variable $N(\nu, 2\nu)$

9.1 La distribución de chi-cuadrado. Propiedades

Teorema de adición de variables distribuidas chi-cuadrado

La suma de variables distribuidas según χ^2 es una variable χ^2

"Sea $\chi_1^2, \chi_2^2, \ldots, \chi_k^2$ un conjunto de k variables independientes distribuidas χ^2 con $\nu_1, \nu_2, \ldots, \nu_k$ grados de libertad respectivamente. La suma $\chi_s^2 = \chi_1^2 + \chi_2^2 + \ldots + \chi_k^2$ se distribuye también como una χ^2 distribución de con $\nu = \nu_1 + \nu_2 + \ldots + \nu_k$ grados de libertad."

Demostración

Basta con demostrar que la función característica de la variable χ_s^2 tiene la misma forma que la función característica de una variable chi-cuadrado

$$\Phi_{\chi_s^2}(t) = \Phi_{\chi_1^2}(t)\Phi_{\chi_2^2}(t)\cdots\Phi_{\chi_k^2}(t) = (1-2it)^{\frac{\nu_1}{2}}(1-2it)^{\frac{\nu_2}{2}}\cdots(1-2it)^{\frac{\nu_k}{2}} = (1-2it)^{\frac{(\nu_1+\nu_2+\cdots+\nu_k)}{2}}$$

Variables independientes

Función característica de una función de chi-cuadrado con $\nu_1 + \nu_2 + \ldots + \nu_k$ grados de libertad

$$\chi_s^2 = \chi^2(\nu_1 + \nu_2 + \ldots + \nu_k)$$

9.1 La distribución de chi-cuadrado. Propiedades

Contenido de probabilidad

La función acumulativa de χ^2 resulta muy útil en el cálculo de intervalos de confianza y en los test de hipótesis

$$F\left(\chi^2,\nu\right)=\int_0^{\chi^2_\alpha} f\left(\chi^2,\nu\right)d\chi^2=1-\alpha$$

Routine functions ROOT

TMath::Prob(chi2,ndf)

Double_t Prob(Double_t chi2, Int_t ndf)

Computation of the probability for a certain Chi-squared (chi2) and number of degrees of freedom (ndf).

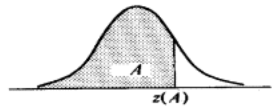
Calculations are based on the incomplete gamma function P(a,x), where a=ndf/2 and x=chi2/2. P(a,x) represents the probability that the observed Chi-squared for a correct model should be less than the value chi2.

The returned probability corresponds to 1-P(a,x), which **denotes the probability that an observed Chi-squared exceeds the value chi2** by chance, even for a correct model.

ν

Valores tabulados

Entry is area A under the standard normal curve from $-\infty$ to $z(A)$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

9.1 La distribución de chi-cuadrado. Propiedades

Relación con la distribución de Poisson

Existe una relación muy útil entre las distribuciones acumulativas de Poisson y de chi-cuadrado

Distribución de Poisson $P(x,\mu) = \frac{\mu^x}{x!} e^{-\mu}$ y función acumulativa de χ^2

$$\sum_{x=0}^{n} P(x,\mu) = \sum_{x=0}^{n} \frac{\mu^x}{x!} e^{-\mu} = 1 - \int_0^{2\mu} f\left(\chi^2, \nu = 2(n+1)\right) d\chi^2 = \int_{2\mu}^{\infty} f\left(\chi^2, \nu = 2(n+1)\right) d\chi^2$$

$$\sum_{x=0}^{n} P(x,\mu) = \int_{2\mu}^{\infty} f\left(\chi^2, \nu = 2(n+1)\right) d\chi^2$$

9.1 La distribución de chi-cuadrado. Propiedades

Relación con el estimador de la varianza

$$\frac{(n-1)s^2}{\sigma^2}$$

Sea $x_1, x_2, \ldots, x_n$ una muestra de medidas distribuidas como $N(\mu, \sigma^2)$

Los estimadores de la media y la varianza vienen dados por:

$$\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Helmert Transformation

$$\left.\begin{aligned}
y_1 &= \frac{x_1 - x_2}{\sqrt{2}} \\
y_2 &= \frac{x_1 + x_2 - 2x_3}{\sqrt{6}} \\
&\vdots \\
y_{n-1} &= \frac{x_1 + x_2 + \cdots + x_{n-1} - (n-1)x_n}{\sqrt{n(n-1)}} \\
y_n &= \frac{x_1 + x_2 + \cdots + x_n}{\sqrt{n}} = \sqrt{n}\,\overline{x}
\end{aligned}\right\}$$

$$y_i = \sum_{j=1}^{n} a_{ij} x_j \quad ; \quad \sum_{i=1}^{n} a_{ij} a_{ik} = \delta_{jk}$$

Queremos demostrar que la variable $\frac{(n-1)s^2}{\sigma^2}$ se distribuye como $\chi^2(n-1)$

$$(n-1)s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 = \sum_{i=1}^{n} y_i^2 - n\overline{x}^2 = \sum_{i=1}^{n} y_i^2 - y_n^2 = \sum_{i=1}^{n-1} y_i^2$$

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij} x_j\right)\left(\sum_{k=1}^{n} a_{ik} x_k\right) = \sum_{j=1}^{n}\sum_{k=1}^{n}\left(\sum_{i=1}^{n} a_{ij} a_{ik}\right) x_k x_j = \sum_{j=1}^{n}\sum_{k=1}^{n} \delta_{jk} x_k x_j = \sum_{j=1}^{n} x_j^2$$

El número de términos independientes es n-1
Luego se distribuye como $\chi^2(n-1)$

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n-1}\left(\frac{y_i}{\sigma_i}\right)^2$$

9.2 La distribución t de Student

Sea x una variable aleatoria normal estándar: $N(0,1)$ y sea χ^2 una variable aleatoria de chi-cuadrado con v grados de libertad. Supongamos que son independientes entre si. Entonces la nueva variable:

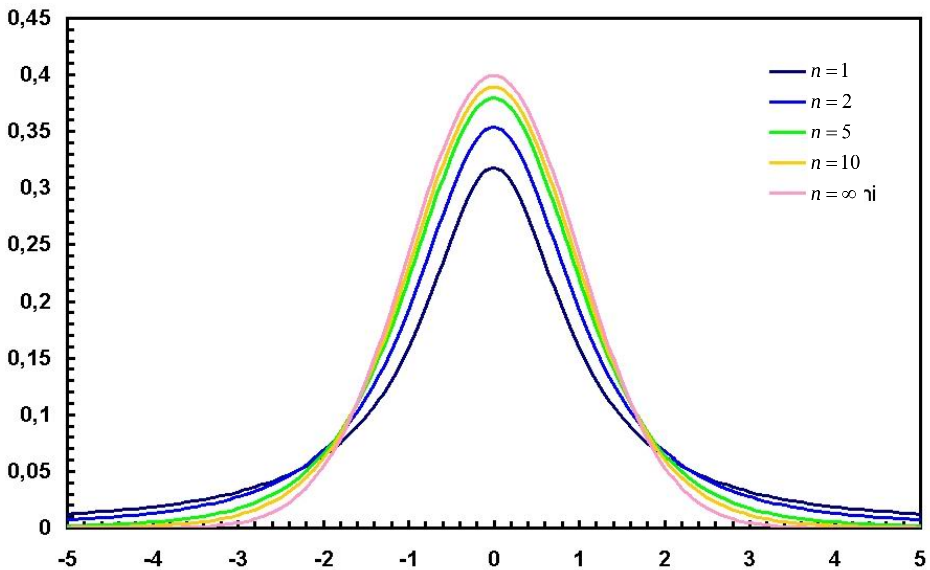
$$t = \frac{x}{\sqrt{\chi^2/\nu}} \quad ; \quad -\infty \le t \le \infty \quad ; \quad \nu > 0$$

Cociente entre una variable normal y la raíz de una variable de chi-cuadrado reducida, independientes entre sí

se distribuye de acuerdo a la distribución **t de Student** que viene dada por:

$$f(t,\nu) = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}$$

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



El parámetro v número de grados de libertad.

Tiende a la distribución normal para $\nu \to \infty$

Las colas de la distribución son más pronunciadas

William Sealy Gosset

9.2 La distribución t de Student

Demostración

Como x e χ^2 son independientes entre si, la pdf conjunta viene dada por: $f(x;\chi^2,\nu)=\dfrac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\dfrac{(\chi^2)^{\frac{\nu}{2}-1}e^{-\frac{\chi^2}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}$

Hacemos el cambio de variables a las nuevas coordenadas:

$$\left.\begin{aligned} t&=\frac{x}{\sqrt{\chi^2/\nu}} \\ \rho&=\chi^2 \end{aligned}\right\} \Rightarrow J=\begin{vmatrix} \dfrac{\partial x}{\partial t} & \dfrac{\partial x}{\partial \rho} \\ \dfrac{\partial \chi^2}{\partial t} & \dfrac{\partial \chi^2}{\partial \rho} \end{vmatrix}=\begin{vmatrix} \sqrt{\chi^2/\nu} & 0 \\ 0 & 1 \end{vmatrix}=\sqrt{\chi^2/\nu} \Rightarrow f(t,\rho,\nu)=\frac{1}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)2^{\frac{1}{2}(\nu+1)}}\rho^{\frac{1}{2}(\nu+1)-1}e^{-\frac{1}{2}\rho\left(1+\frac{t^2}{\nu}\right)}$$

$$\int_0^\infty x^m e^{-ax}dx=\frac{\Gamma(m+1)}{a^{m+1}}$$

Integrando para ρ:

$$f(t,\nu)=\int_0^\infty f(t,\rho,\nu)d\rho=\frac{1}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)2^{\frac{1}{2}(\nu+1)}}\int_0^\infty \rho^{\frac{1}{2}(\nu+1)-1}e^{-\frac{1}{2}\rho\left(1+\frac{t^2}{\nu}\right)}d\rho=\frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}$$

Propiedades

Valor esperado $E[t]=0$

Skewness y Kurtosis

$$\gamma_1=0\ ;\quad \gamma_2=\frac{6}{\nu-4}\qquad \nu>4$$

Distribución simétrica en torno a t = 0

Varianza

$$V[t]=\frac{\nu}{\nu-2}\qquad \nu>2$$

Similaritud entre la distribución t de Student y la distribución normal estándar

9.2 La distribución t de Student. Aplicaciones

El contenido de probabilidad de la distribución t de Student permite:

- Calcular intervalos de confianza cuando la varianza es desconocida
- Realizar test de hipótesis cuando el estadístico se distribuye como una t de Student

En muchas ocasiones no conocemos la resolución de nuestras medidas y la dispersión procede de las propias medidas.

Consideremos una variable aleatoria x distribuida normalmente con media μ y desviación estándar σ entonces:

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

Si no conocemos σ tenemos que estimarla: $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$

Pero entonces $t=\frac{x-\mu}{\hat{\sigma}}$ ya no es una distribución normal estándar !!!!

Solución:

$$t=\frac{x-\mu}{\hat{\sigma}}=\frac{(x-\mu)/\sigma}{\hat{\sigma}/\sigma}=\frac{(x-\mu)/\sigma}{\sqrt{\frac{(n-1)\hat{\sigma}^2}{\sigma^2}\frac{1}{(n-1)}}}=\frac{z}{\sqrt{\chi^2/(n-1)}} \Rightarrow t(n-1)$$

es una variable aleatoria distribuida según la distribución t de Student con n-1 grados de libertad

9.2 La distribución t de Student. Aplicaciones

Extensión a la media. Si tenemos una muestra de medidas $x_1, x_2, \ldots, x_n$ de una distribución: $N\left(\mu, \sigma^2\right)$ con valor medio $\overline{x}$ y s^2

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \Rightarrow \quad \text{es una variable} \quad N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \quad \Rightarrow \quad \frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n}\frac{\left(x_i - \overline{x}\right)^2}{\sigma^2} \quad \text{es una variable} \quad \chi^2(n-1)$$

Además $\overline{x}$ y s^2 son variables independientes entre si, por tanto:

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \quad \text{y} \quad \frac{(n-1)s^2}{\sigma^2} \quad \text{también lo son} \quad \Rightarrow \quad t = \frac{\dfrac{\overline{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)s^2}{\sigma^2}\Big/(n-1)}} = \frac{\overline{x}-\mu}{s/\sqrt{n}}$$

$$\boxed{t = \frac{\overline{x}-\mu}{s/\sqrt{n}}}$$

es una variable aleatoria distribuida según la distribución t de Student con n-1 grados de libertad

Nuestra ignorancia sobre σ en el numerador cancela nuestra ignorancia sobre σ en el denominador de manera que t solo contienen cantidades observadas $\overline{x}, s$, y además sabemos como se distribuye t

9.2 La distribución t de Student. Aplicaciones

INTERVALOS DE CONFIANZA

La probabilidad de que $t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$ se encuentre en un intervalo [a,b] de confianza con un (1-γ) CL:

Intervalo simétrico a = -b

$$P\left[a \le \frac{\overline{x} - \mu}{s/\sqrt{n}} \le b\right] = \int_a^b f(t, n-1)\,dt = 1 - \gamma \quad \Rightarrow \quad P\left[-b \le \frac{\overline{x} - \mu}{s/\sqrt{n}} \le b\right] = \int_{-b}^b f(t, n-1)\,dt = 1 - \gamma$$

Para obtener b basta con llamar a la función de ROOT

b = TMath::StudentQuantile (p, ndf)

$p = 1 - \gamma/2$

$$P\left[\overline{x} - \frac{bs}{\sqrt{n}} \le \mu \le \overline{x} + \frac{bs}{\sqrt{n}}\right] = 1 - \gamma$$

EJEMPLO

$x_1 = 2.2$ $\qquad \overline{x} = \frac{1}{4}\sum_{i=1}^{4} x_i = 3.70$

$x_2 = 4.3$ $\qquad s^2 = \frac{1}{4-1}\sum_{i=1}^{4}(x_i - \overline{x})^2 = 5.00\widehat{6}$

$x_3 = 1.7$

$x_4 = 6.6$ $\qquad s = 2.237$

Para un intervalo de confianza del (1-γ) = 68.27 % (one sigma)

$p = 1 - \gamma/2 = 0.84135$

b = TMath::StudentQuantile (0.84135, 3)=1.197

$$\left[\overline{x} - \frac{bs}{\sqrt{n}}, \overline{x} + \frac{bs}{\sqrt{n}}\right] = [3.70 - 1.34, 3.70 + 1.34] = [2.36, 5.04]$$

9.3 La distribución F

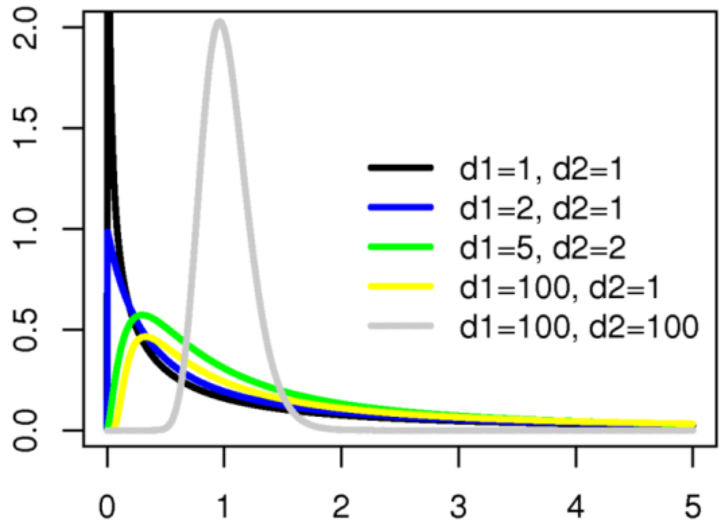
Sean χ_1^2 y χ_2^2 dos variables aleatorias de chi-cuadrado con ν_1 *y* ν_2 grados de libertad respectivamente, se define la variable F como:

$$F = \frac{\chi_1^2 / \nu_1}{\chi_1^2 / \nu_2} \quad ; \quad 0 \le F \le \infty \quad ; \quad \nu_1, \nu_2 > 0$$

Cociente entre dos variables de chi-cuadrado

se distribuye de acuerdo a la distribución ***F*** :

$$f(F;\nu_1,\nu_2) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{F^{\frac{\nu_1}{2}-1}}{\left(1+\frac{\nu_1 F}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}}$$



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

Se denomina función distribución F con (ν_1, ν_2) grados de libertad

Tiende a la distribución normal para $\nu \to \infty$

Las colas de la distribución son más pronunciadas