ESTAD _ T13. DISTRIB. χ^2, t y F.

CARACTERÍSTICAS.

HACER — ▷ IMPORTANCIA de estas distrib. en la t.ª y p.ª estadística

Relaciones con la normal (absorbida)

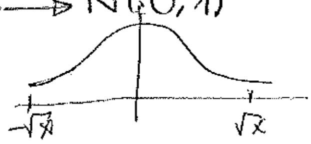
Las tres distrib. que vamos a estudiar en este tema son distribuciones continuas univariantes derivadas de la distrib. normal, puesto que resultan de combinar de $\neq$ forma v.a. normales.

Junto con la distrib. normal, la χ^2 de Pearson, la t-Student y la F de Fisher-Snedecor son ampliamente utilizadas en Inferencia, tanto en el proceso de estimación (distr. de estadísticos en el muestreo) como en la fase de contrastación.

1_ DISTRIB. χ^2 de PEARSON

Debida a Karl Pearson (1900), interesado en la distribución muestral de normales al cuadrado (Test de la bondad del ajuste).

Para llegar a esta distribución consideramos $\xi^* \rightarrow N(0,1)$ y calculamos la distribución de $X = (\xi^*)^2$



Función de distribución:

$$F_X(x) = P(X \leq x) = P(\xi^{*2} \leq x) = P(-\sqrt{x} \leq \xi^* \leq \sqrt{x}) =$$

$$= F_{\xi^*}(\sqrt{x}) - F_{\xi^*}(-\sqrt{x}) \qquad \text{para } x > 0$$

Como ξ^* es simétrica:

$$F_X(x) = F_{\xi^*}(\sqrt{x}) - (1 - F_{\xi^*}(\sqrt{x})) = 2F_{\xi^*}(\sqrt{x}) - 1 \quad , x > 0$$

$$\text{y } F_X(x) = 0 \quad \text{para } x \leq 0$$

Para obtener la función de densidad basta con derivar la función de distribución:

$$f(x) = \cancel{2} f_{\xi^*}(\sqrt{x}) \cdot \underbrace{\frac{1}{\cancel{2}\sqrt{x}}}_{\text{derivada de } \sqrt{x}} = \frac{1}{\sqrt{x}} \cdot \frac{1}{\sqrt{2\pi}} e^{-x/2} =$$

$$= \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{\pi}} \cdot x^{-1/2} \cdot e^{-\frac{x}{2}} = \frac{\left(\frac{1}{2}\right)^{1/2}}{\Gamma(1/2)} \cdot x^{\frac{1}{2}-1} \cdot e^{-\frac{x}{2}} \quad x>0$$

$\Gamma(1/2) = \sqrt{\pi}$

que coincide con la función de densidad de una distrib. $\Gamma(\underbrace{\tfrac{1}{2}}_{p}, \underbrace{\tfrac{1}{2}}_{q}) \rightarrow f(x) = \frac{q^p}{\Gamma(p)} \cdot x^{p-1} \cdot e^{-qx}$, $x>0$

Ésta sería la función de densidad de una distribución χ^2_1, ji-cuadrado con un grado de libertad, pero podemos generalizar la definición anterior a la suma de los cuadrados de n var. normales (0,1) independientes.

~~Propos~~

→ La suma de los cuadrados de n $N(0,1)$ independientes es una $\chi^2_n = \Gamma(\frac{n}{2}, \frac{1}{2})$.

$\xi_i \rightarrow N(0,1)$ v.a.i.i.d $\Rightarrow \chi^2_n = \xi_1^2 + \cdots + \xi_n^2$

Como ξ_i indep $\Rightarrow \xi_i^2 \rightarrow \chi^2_1$ indep $= \Gamma(1/2, 1/2)$.

la distrib. Gamma es reproductiva respecto a p $\Rightarrow$

$$\sum_n \Gamma(1/2, 1/2) = \Gamma\left(\frac{n}{2}, 1/2\right)$$

Aprovechando que $\chi^2_n = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, se deducen fácilmente su función de densidad y sus características:

Función de densidad

$$f(x) = \begin{cases} \dfrac{(1/2)^{n/2}}{\Gamma(n/2)} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{1}{2}x} & x > 0 \\ 0 & x \ngtr 0 \end{cases}$$

Media

$$E[\chi^2_n] = \frac{p}{q} = \frac{n/2}{1/2} = n \quad \leftarrow \text{grados de libertad}$$

Varianza

$$V[\chi^2_n] = \frac{p}{q^2} = \frac{n/2}{(1/2)^2} = 2n \quad \leftarrow \text{doble grados de libertad}$$

Función característica

$$\varphi(t) = \left(1 - \frac{it}{q}\right)^{-p} = (1-2it)^{-n/2}$$

* la función característica no es fácil. $\varphi(t) = \left(1 - \frac{it}{q}\right)^{-p} = \left(1 - \frac{it}{1/2}\right)^{-n/2} = (1-2it)^{-n/2}$

Propiedades

P1 → Si sumamos los cuadrados de n $N(\mu_i, 1)$ indep, se obtiene una χ^2_n no centrada con parámetro de descentralización $\theta = \sum_{i=1}^{n} \mu_i^2$

Si ahora son $N(\mu_i, \sigma)$ el parámetro de descentralización será

$$\delta = \frac{1}{\sigma^2} \sum_{i=1}^{n} \mu_i^2$$

P2 → Propiedad reproductiva. La suma de χ^2 independientes también es χ^2 con grados libertad igual a la suma de los grados libertad de los sumandos.

$$\xi_i \to \chi^2_{n_i} \Rightarrow \xi = \xi_1 + \dots + \xi_k$$
$$\xi \to \chi^2_{n = n_1 + \dots + n_k}$$

Dem: Consecuencia inmediata de la propiedad reproductiva de la Gamma, puesto que ξ pertenece de

$$\xi_i \to \Gamma\left(\frac{n_i}{2}, \frac{1}{2}\right) \Rightarrow \xi \to \Gamma\left(\frac{n_1}{2} + \dots + \frac{n_k}{2}, \frac{1}{2}\right) = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right).$$

P3 → Aproximación a una distrib. normal.

Cuando el nº de grados de libertad es sufic. grande ($n>30$) entonces $\chi^2_n \sim$ Normal.

$$\xi \to \chi^2_n, \; n>30 \Rightarrow \sqrt{2\xi} \approx N(\sqrt{2n-1}, 1).$$

$$\xi_i \to N(0,\sigma) \Rightarrow \frac{\xi_i}{\sigma} \to N(0,1) \Rightarrow \left(\frac{\xi_i}{\sigma}\right)^2 \to \chi^2_1$$
$$\Rightarrow \sum_{i=1}^{n} \left(\frac{\xi_i}{\sigma}\right)^2 \to \chi^2_n.$$

etc.

ESTAD _ T13

2 _ DISTRIB. t de STUDENT

la distribución t de Student fue ideada en 1908 por Gosset ("student" era su pseudónimo) y difundida por Fisher, interesado en la distribución de los estadísticos muestrales cuando las muestras eran pequeñas.

La distribución t-Student se define como el cociente entre una $N(0,1)$ y la raíz cuadrada de una χ^2 entre sus grados de libertad. La t tiene un único parámetro, los grados de libertad, heredado de la χ^2.

$$\boxed{t_n = \frac{N(0,1)}{\sqrt{\chi^2_n / n}}} = \frac{\xi^*}{\sqrt{\frac{\sum \xi_i^2}{n}}} \quad / \quad \begin{array}{l} \xi^* \to N(0,1) \\ \xi_i \to N(0,1) \end{array}$$

Es fácil demostrar que la distrib. t no depende de las varianzas de las distrib. normales, por lo que tb. se puede escribir como

$$t_n = \frac{N(0,\sigma)}{\sqrt{\frac{N(0,\sigma)^2 + \ldots + N(0,\sigma)^2}{n}}} = \frac{\sigma \cdot N(0,1)}{\sqrt{\frac{\sigma^2 N(0,1)^2 + \ldots + \sigma^2 N(0,1)^2}{n}}} =$$

$$= \frac{\cancel{\sigma} \cdot N(0,1)}{\cancel{\sigma}\sqrt{\frac{N(0,1)^2 + \ldots + N(0,1)^2}{n}}} = \frac{\xi^*}{\sqrt{\chi^2_n / n}}$$

Para obtener la función de densidad, partimos de la definición y calculamos la función de densidad conjunta como producto de las f. densidad de la Normal y de la Chi, puesto que son indep.

¡CÓMO?

$$N(0,1) \longrightarrow f(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-u^2/2} \qquad -\infty < u < +\infty$$

$$\chi^2_n \longrightarrow f(v) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} \cdot v^{n/2 - 1} \cdot e^{-v/2} \qquad 0 < v$$

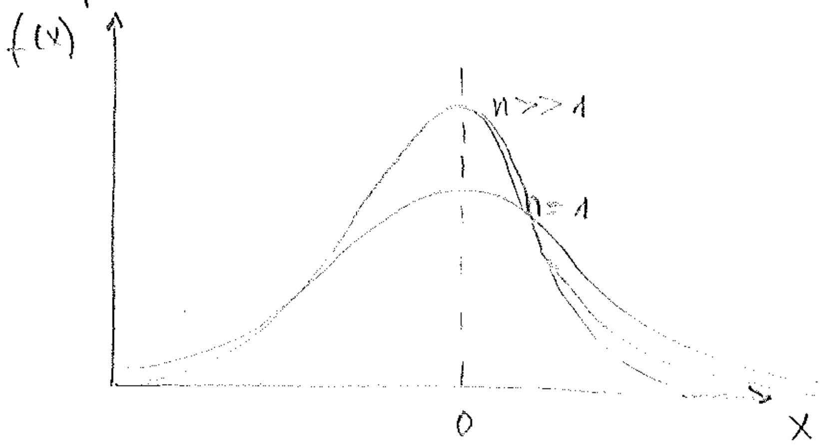
$\downarrow$

$\Gamma(\frac{n}{2}, \frac{1}{2})$

y se llega a:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\ \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty < x < +\infty$$

cuya representación gráfica es:



De la representación gráfica se deduce:

1 _ Es simétrica respecto al origen, $x=0$.
2 _ Tiene una asíntota horizontal en el eje de abscisas.
3 _ Es creciente para $x<0$ y decreciente para $x>0$.
En $x=0$ se alcanza el máximo $\Rightarrow \mu = Me = Mo = 0$.
4 _ Presenta dos puntos de inflexión simétricos $\Rightarrow$ forma de campana como la Normal, pero más achatada.

Observaciones

01 $\longrightarrow t_n \xrightarrow[n \nearrow]{} N(0,1)$. Coinciden prácticamente cuando $n>30$.

Características

Esperanza: $E[\xi] = 0$, por simetría en la f. densidad.

Varianza: $V(\xi) = E[\xi^2] - E(\xi)^2 = \dfrac{n}{n-2}$ para $n>2$.

3_ DISTRIB. F

Casas la llama F de Snedecor
Martín-Pliego la llama F de Fisher-Snedecor } ?

Se utiliza principalmente en problemas relacionados con la varianza, en particular en ANOVA.

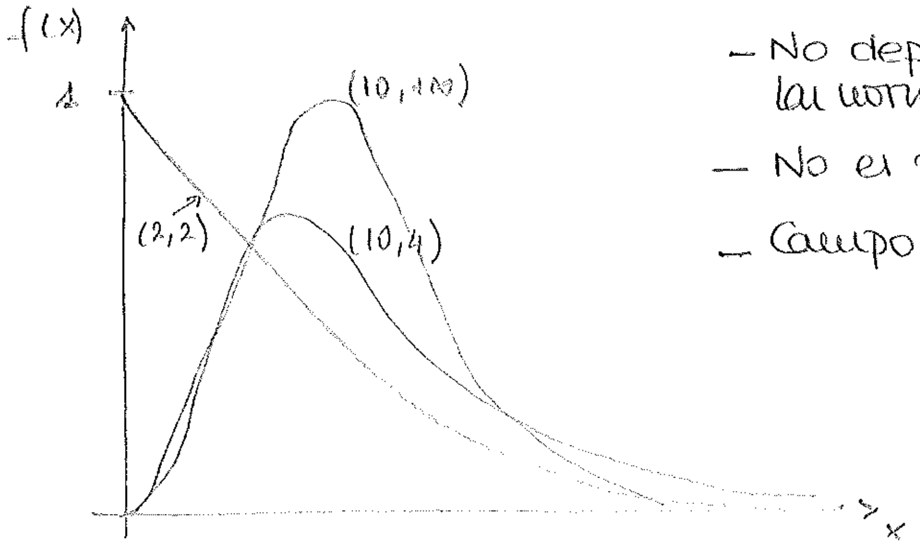
La distribución F se define como cociente de dos distrib. χ^2, y por tanto, como cociente de sumas de cuadrados de normales ($N(0,1)$ ó $N(0,\sigma)$, da igual)

$$F_{m,n} = \frac{\chi^2_m/m}{\chi^2_n/n} = \frac{(\xi_1^{*2}+\dots+\xi_m^{2})/m}{(\xi_1^{*2}+\dots+\xi_n^{*2})/n}$$

Para obtener la función de densidad, se calcula la f. densidad {conjunta} del producto de las χ^2 (son indep), ó sea utilizando la Gamma, y una vez obtenida hacemos la marginal de la 1ª variable.

Función de densidad — no estoy segura, viene ✗

$$f(x) = \frac{m^{m/2} \cdot n^{n/2} \cdot \Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\cdot\Gamma(n/2)}\, x^{\frac{m}{2}-1}(n+mx)^{-\frac{m+n}{2}} \quad , x>0$$



- No depende de la varianza de las normales.
- No es simétrica
- Campo de variación: $[0,+\infty)$

Características

Esperanza : $\mu = E[\xi] = \frac{n}{n-2}$, si $n>2$

Varianza : $\sigma^2 = V[\xi] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, si $n>4$

Propiedades

P1 → Propiedad de reciprocidad : La inversa de una F también es una F, cambiando los grados de libertad.

$$\frac{1}{F(m,n)} = F(n,m)$$

Dem Por definición.

Si $F_{m,n} = \frac{\chi^2_m/m}{\chi^2_n/n}$, entonces $\frac{1}{F_{m,n}} = \frac{\chi^2_n/n}{\chi^2_m/m} = F_{n,m}$

P2 → $F(1,n) = (t_n)^2$

Por def : $F_{1,n} = \frac{\chi^2_1/1}{\chi^2_n/n} = \frac{N(0,1)^2}{\chi^2_n/n}$

$t_n = \frac{N(0,1)}{\sqrt{\chi^2_n/n}} \Rightarrow t_n^2 = F_{1,n}$.

Pearson | test de la χ^2
distrib χ^2
"desviación estándar"

Biometría
Genética

Quién es quién

Karl Pearson
(Londres 1857-Londres 1936)

A principios de la década de 1880, tres famosos matemáticos, Karl Pearson, Francis Galton y Edgeworth dieron lugar a una revolución estadística en Europa. Karl Pearson, uno de estos tres matemáticos, fue considerado, gracias a su ambición y determinación, padre de la ciencia de la Estadística en el siglo XX.

Nacido el 27 de marzo de 1857 en Londres, Inglaterra, a los nueve años, Karl fue enviado a la University College School de Londres. Tras recibir un título de Bachiller con mención especial en matemáticas por el King's College en 1879, Pearson, continuó su formación con estudios de física, metafísica y Darwinismo en Alemania.

Cuando volvió a Londres, Pearson se casó con una joven llamada Maria Sharpe. La joven pareja tuvo tres hijos, Sigrid, Helga y Egon. Este duradero y feliz matrimonio terminó con la muerte de Maria en 1928. Pearson se casaría de nuevo con una mujer que era colaboradora en su departamento, Margaret Victoria Child. Para sustentar su familia, Pearson volvió al University College donde destacaría como profesor y conferenciante y donde continuaría trabajando hasta unos meses antes de su muerte.

La iniciativa y determinación de Karl Pearson puede que pesaran más que su capacidad matemática y podrían haber sido la principal razón de su éxito. De los tres líderes de la revolución estadística, Pearson probablemente no fuera el más erudito, pero reconoció el potencial y la inteligencia de Edgeworth y el trabajo de Galton.

Pearson y Galton tuvieron una especial amistad durante todos los años que se conocieron. Aunque al principio Pearson criticó el trabajo de Galton, después cambió de opinión y se unió a él. Fue Galton quien financió económicamente a Pearson cuando comenzó con su revista estadística Biométrika. Quizás este apoyo y su amistad fue lo que llevó a Pearson a aceptar el encargo de la familia de Galton de escribir su biografía tras su muerte. En 1911, Pearson comenzó a escribir lo que finalmente serían tres volúmenes de "La vida, escritos y trabajos de Francis Galton".

En cuanto a Edgeworth, existe evidencia que sugiere que este fue la figura crucial en el desarrollo intelectual de Pearson. Aunque los dos matemáticos no fueron siempre grandes amigos, mantuvieron un mutuo respeto el uno por el otro. Mientras trabajaba en sus curvas de distribución, Pearson tenía un objetivo en mente: hacerlo mejor de lo que lo había hecho Edgeworth. Al mismo tiempo, Edgeworth quería desarrollar su aproximación mucho antes de que Pearson tuviera oportunidad de hacerlo. Aunque finalmente, se estableció entre los dos una relación cordial.

En Julio de 1900, una de las más importantes contribuciones de Pearson a la Estadística fue presentada en la publicación de un artículo. Esta contribución era el Test de la χ^2.

Pearson usó esta fórmula para obtener la distribución muestral de χ^2 en grandes muestras, las cuales estaba particularmente interesado en estudiar, como una función de k, la cual resultó ser una forma especial de la distribución de Pearson tipo 3, ahora conocida como "distribución χ^2 para K-1 grados de libertad". Además daba una pequeña tabla de la integral de la distribución de χ^2 desde 1 a 70 y para k desde 3 a 20. Este test de la χ^2 de bondad del ajuste es ciertamente una de las mayores y más útiles contribuciones de Pearson a los tests estadísticos.

Además de por esta cuestión, Pearson es conocido por otras importantes contribuciones en diferentes campos incluyendo antropología, biométrica, genética, método científico y teoría estadística.

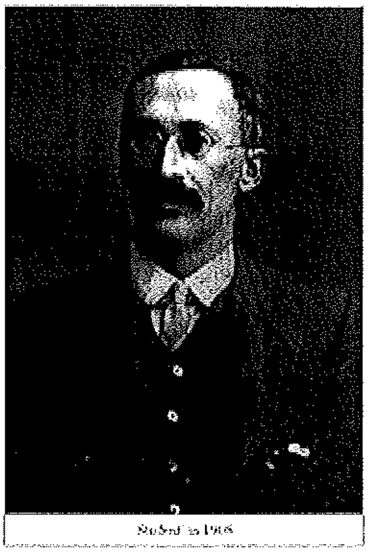
Tras el reconocimiento por su trabajo sobre curvas de distribución, Pearson continuó recibiendo reconocimientos y honores. En 1893, comenzó su serie de 18 artículos titulados "Mathematical Contributions to the Theory of Evolution", que contendrían parte de su trabajo más valioso. El mismo año que empezó estos artículos, Pearson acuñó el término "desviación estándar". Entre 1906 y 1914 Pearson estuvo consagrado al desarrollo de un centro de postgrado para promover el desarrollo de la estadística como una rama de las matemáticas aplicadas. Por último, en el verano de 1933, tras una larga vida consagrada al avance estadístico, Pearson abandonó su trabajo en la Universidad. El hecho de que tras la retirada de Pearson el departamento de estadística aplicada fuera dividido en dos unidades independientes muestra el importante trabajo soportado por Pearson. Incluso después de la muerte de Karl en 1936, su apellido continúa siendo uno de los más destacados en el campo de las matemáticas.

No hay duda de que las contribuciones de Pearson a lo largo de su vida consolidaron la Estadística como una disciplina por derecho propio.

Regresión
Diseño de experimentos

William Sealy Gosset

De Wikipedia, la enciclopedia libre



William Sealy Gosset, 1876-1937

William Sealy Gosset (13 de junio de 1876 – 16 de octubre de 1937) fue un químico y estadístico, mejor conocido por su sobrenombre literario ***Student***. Nacido en Canterbury, hijo de Agnes Sealy Vidal y el coronel Frederic Gosset, asistió a la famosa escuela privada Winchester College, antes de estudiar química y matemática en el New College de Oxford. Tras graduarse en 1899, se incorporó a la destilería de Arthur Guinness e Hijo en Dublín.

Guinness era un negocio agroquímico progresista y Gosset podría aplicar sus conocimientos estadísticos tanto a la destilería como a la granja (para seleccionar las mejores variedades de cebada. Gosset adquirió ese conocimiento mediante estudio, prueba y error así como pasando dos temporadas durante 1906/7 en el laboratorio bioquímico de Karl Pearson. Gosset y Pearson tenían una buena relación y este último ayudó a Gosset con la matemática de sus artículos. Pearson contribuyó a los artículos de 1908, pero no apreció lo suficiente su importancia. Los artículos se referían a la importancia de las pequeñas muestras para la destilería, mientras que el biólogo disponía normalmente de cientos de observaciones y no veía la urgencia en el desarrollo de métodos basados en unas pocas muestras.

Otro investigador de Guinness había publicado anteriormente un artículo que contenía secretos industriales de la destilería. Para evitar futuras exposiciones de información confidencial, Guinness prohibió a sus empleados la publicación de artículos independientemente de la información que contuviesen. Esto significaba que Gosset no podía publicar su trabajo usando su propio nombre. De ahí el uso de su pseudónimo *Student* en sus publicaciones, para evitar que su empleador lo detectara. Por tanto, su logro más famoso se conoce ahora como la distribución t de Student, que de otra manera hubiera sido la distribución t de Gosset.

Usando este pseudónimo, Pearson publicó *El error probable de una media* y casi todos los artículos de Gosset en su publicación *Biometrika*. Sin embargo, fue R.A. Fisher quien apreció la importancia de los trabajos de Gosset sobre muestras pequeñas, tras recibir correspondencia de Gosset en la que le decía 'le envío una copia de las Tablas de Student, ¡ya que es la única persona que probablemente las use jamás!. *Fisher creyó que Gosset había efectuado una "revolución lógica". Irónicamente la estadística* t *por la que Gosset es famoso fue realmente creación de Fisher. La estadística de Gosset era* $z = t/\sqrt{(n - 1)}$. *Fisher introdujo la forma* t *debido a que se ajustaba a su teoría de grados de libertad. Fisher es responsable también de la aplicación de la distribución* t *a la regresión.*

Aunque fueron introducidos por otros, los residuos estudentizados reciben su nombre en honor a Student porque, al igual que con el problema que llevó a la distribución t de Student, la idea de ajustar usando las desviaciones estándar estimadas la base del concepto.

El interés de Gosset en el cultivo de la cebada le llevó a especular que el diseño de experimentos debería dirigirse no sólo a mejorar la producción media, sino también a desarrollar variedades cuya producción no se viese afectada (robusta) por las variaciones en el suelo y el clima. Este principio sólo aparece más adelante en las ideas de Fisher y luego en el trabajo de Genichi Taguchi en los años 1950.

Distribución F

De Wikipedia, la enciclopedia libre

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

Una variable aleatoria de distribución F se construye como cociente de dos variables de distribución Chi-cuadrada:

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

donde

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

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

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

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

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

La función de distribución es

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

donde I es la función beta incompleta regularizada.

Distribuciones relacionadas

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

Enlaces externos

- Tabla de valores críticos de una distribución F (http://www.itl.nist.gov/div898/handbook/eda/section3/eda3673.htm)
- Prueba de significación mediante la distribución F (http://home.clara.net/sisa/signhlp.htm)