TEMA 2: Distribuciones en el muestreo de poblaciones normales

2.1. Distribuciones χ^2 de Pearson, t de Student y F de Snedecor.

2.2. Muestreo en una población normal unidimensional.

2.3. Muestreo en dos poblaciones normales unidimensionales.

2.1.1. DISTRIBUCIÓN χ^2 DE PEARSON

Caso particular de la distribución gamma[1] :

$$X \to \chi^2(n),\ n \in \mathbb{N} \ \Leftrightarrow\ X \to \Gamma(n/2, 1/2).$$

Función de densidad: $f(x) = \dfrac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2}, \quad x > 0.$

Función generatriz de momentos: $M_X(t) = \dfrac{1}{(1-2t)^{n/2}}, \quad t < 1/2.$

Momentos: $E[X^k] = 2^k \dfrac{\Gamma(n/2+k)}{\Gamma(n/2)}, \quad k \in \mathbb{N}.$

- *Media*: $E[X] = n.$
- *Varianza*: $Var[X] = 2n.$

Reproductividad

$X_1, \dots, X_n$ independientes y $X_i \to \chi^2(k_i),\ i = 1, \dots, n \ \Rightarrow \sum_{i=1}^{n} X_i \to \chi^2\left(\sum_{i=1}^{n} k_i\right)$

Relación con la distribución normal

a) $X \to \mathcal{N}(0,1) \ \Rightarrow\ X^2 \to \chi^2(1)$

b) $X_1, \dots, X_n$ independientes y $X_i \to \mathcal{N}(0,1),\ i = 1, \dots, n \ \Rightarrow \sum_{i=1}^{n} X_i^2 \to \chi^2(n)$ (*)

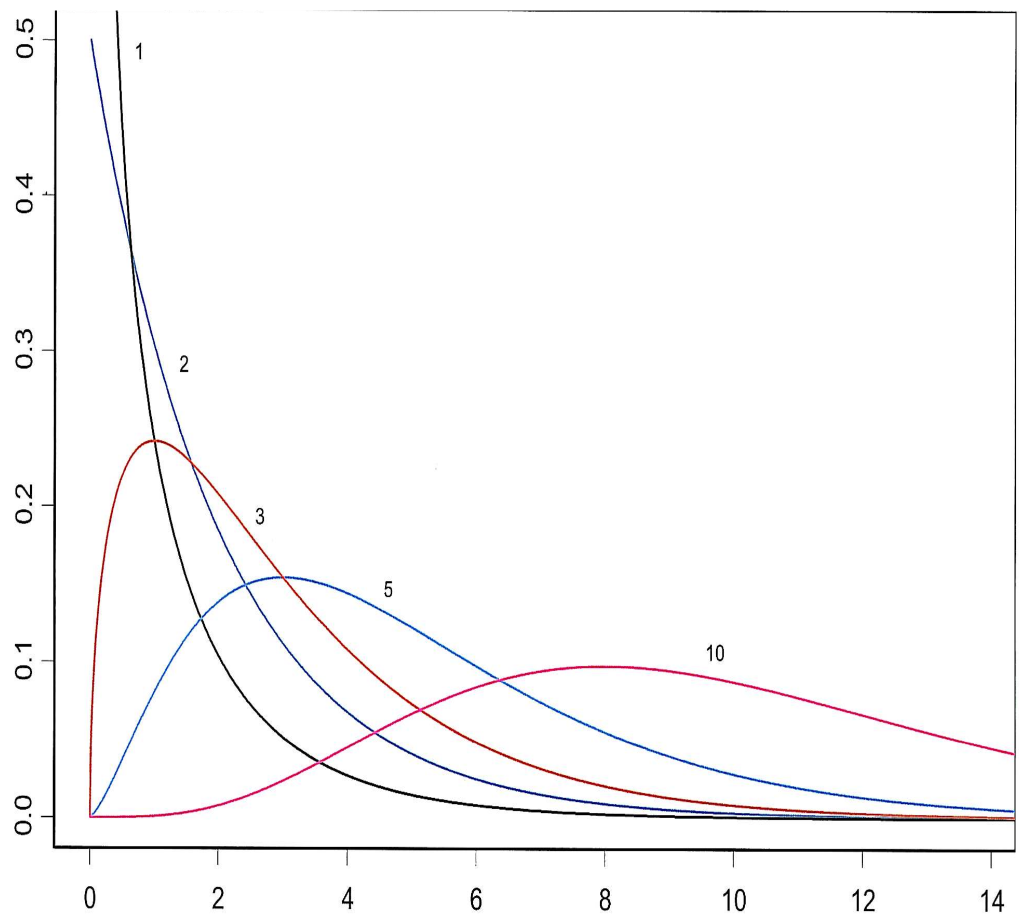
Tablas y aproximaciones: Está tabulada para valores de n pequeños. Para n grande, la expresión (*) como suma de variables independientes e idénticamente distribuidas, con media y varianza finitas, permite usar la siguiente aproximación (*teorema central del límite de Lèvy*):

$$\chi^2(n) \approx \mathcal{N}(n, 2n).$$

[1] $X \to \Gamma(p, a)\ (p, a \in \mathbb{R}^+) \ \Leftrightarrow f(x) = \dfrac{a^p}{\Gamma(p)} x^{p-1} e^{-ax},\ x > 0 \quad \left(\Gamma(p) = \int_0^{+\infty} x^{p-1} e^{-x} dx\right)$

Gráfica de la función de densidad de $\chi^2(n)$:

- Asimétrica a la derecha y unimodal.
- Para $n = 1$, $\lim_{x \to 0} f(x) = +\infty$ y la función de densidad es estrictamente decreciente.
- Para $n = 2$, $f(0) = 1/2$ y la función de densidad es estrictamente decreciente.
- Para $n \geq 3$, $f(0) = 0$, crece hasta la moda y luego decrece.



2.1.2. DISTRIBUCIÓN t DE STUDENT

Es la distribución del cociente entre una variable con distribución $\mathcal{N}(0,1)$ y la raíz cuadrada de una con distribución χ^2 dividida por sus grados de libertad, ambas independientes:

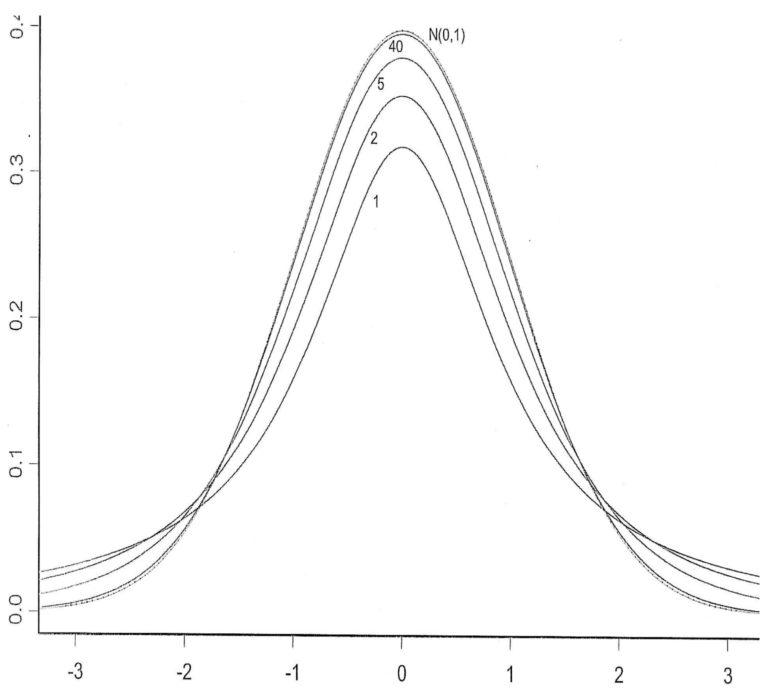
$$X \text{ e } Y \text{ independientes}, \; X \to \mathcal{N}(0,1), \; Y \to \chi^2(n) \;\Rightarrow\; T = \frac{X}{\sqrt{Y/n}} \to t(n)$$

Función de densidad: $f(t) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \dfrac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}.$

Momentos: $\exists E[T^k] \Leftrightarrow k < n$.

- $n > 1 \Rightarrow \exists E[T] = 0.$
- $n > 2 \Rightarrow \exists Var[T] = \dfrac{n}{n-2}.$

Gráfica de la función de densidad de $t(n)$: Es similar a la de la $\mathcal{N}(0,1)$ (simétrica alrededor del cero y unimodal) y, de hecho, se aproxima a ella cuando $n \to +\infty$. Ya que la varianza es mayor que uno, las colas son más gruesas que las de la normal y la gráfica es más aplastada (*distribución platicúrtica*).



Tablas: Está tabulada para valores de n pequeños. Para n grande se aproxima por la $\mathcal{N}(0,1)$.

2.1.3. DISTRIBUCIÓN F DE SNEDECOR

Es la distribución del cociente entre dos variables independientes con distribución χ^2, cada una dividida por sus grados de libertad:

$$X \text{ e } Y \text{ independientes}, \; X \to \chi^2(m), \; Y \to \chi^2(n) \;\Rightarrow\; F = \frac{X/m}{Y/n} \to F(m,n)$$

Función de densidad: $g(f) = \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\dfrac{m}{n}\right)^{\frac{m}{2}} f^{m/2-1}\left(1+\dfrac{m}{n}f\right)^{-\frac{m+n}{2}}, \quad f>0.$

Momentos: $\exists E[F^k] \Leftrightarrow k < n/2$.

- $n > 2 \Rightarrow \exists E[F] = \dfrac{n}{n-2}$.
- $n > 4 \Rightarrow \exists Var[F] = \dfrac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$.

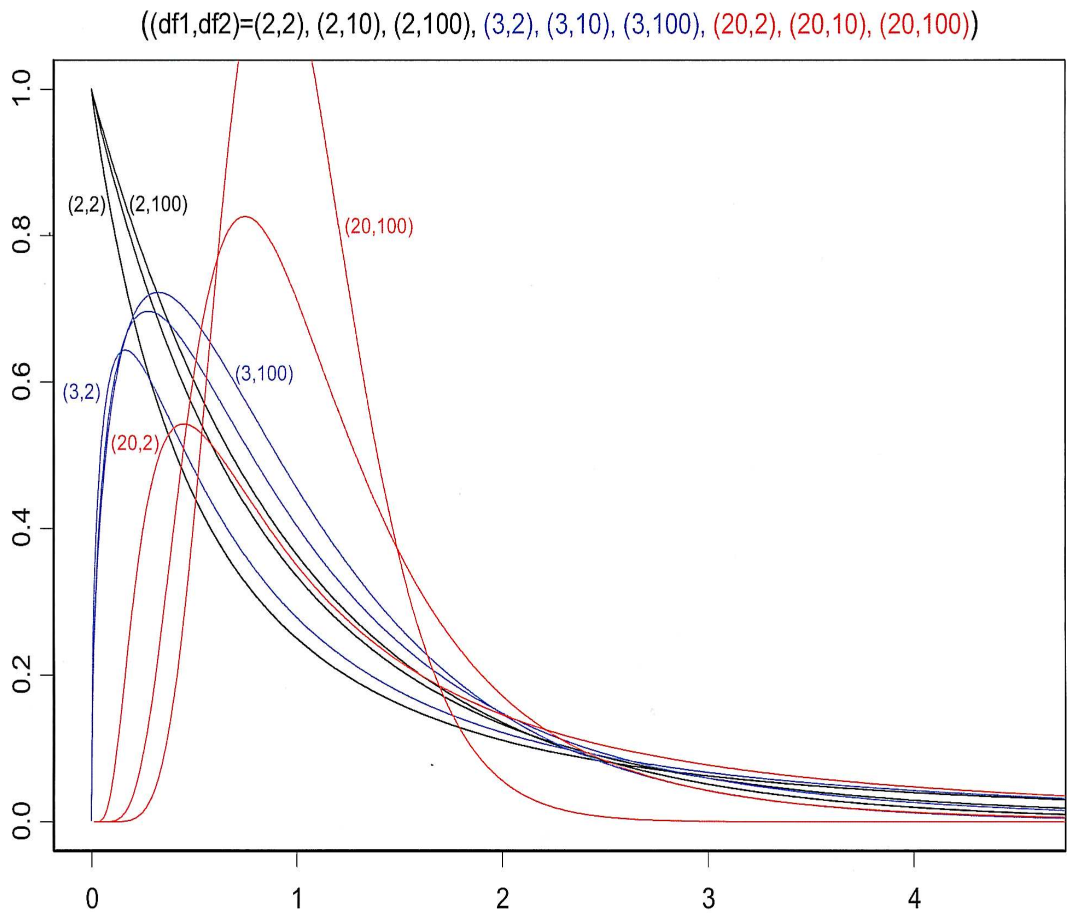
Propiedades:

- $F \to F(m,n) \Leftrightarrow F^{-1} \to F(n,m)$.
- $T \to t(n) \Leftrightarrow T^2 \to F(1,n)$.

Tablas: Como las anteriores, esta distribución está tabulada y, usualmente, las tablas incluyen aproximaciones para valores grandes de m y n.

Gráfica de la función de densidad de $F(m, n)$

- Asimétrica a la derecha y unimodal.
- $F(1, n)$: $\lim_{f \to 0} g(f) = +\infty$ y la función de densidad es estrictamente decreciente.
- $F(2, n)$: $g(0) = 1$ y la función de densidad es estrictamente decreciente.
- $F(m, n),\ m > 2$: $g(0) = 0$, crece hasta el valor modal y luego decrece.



2.2. MUESTREO EN UNA NORMAL UNIDIMENSIONAL

$$(X_1, \ldots, X_n) \text{ m.a.s. de } X \rightarrow \mathcal{N}(\mu, \sigma^2), \quad \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \quad S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

LEMA DE FISHER

Los estadísticos $\overline{X}$ y S^2 son independientes

DISTRIBUCIONES ASOCIADAS AL MUESTREO

Variable	Distribución	Uso
$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$	$\mathcal{N}(0, 1)$	Inferencia sobre μ cuando σ^2 es conocida
$\dfrac{\overline{X} - \mu}{S/\sqrt{n}}$	$t(n-1)$	Inferencia sobre μ cuando σ^2 es desconocida
$\dfrac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2}$	$\chi^2(n)$	Inferencia sobre σ^2 cuando μ es conocida
$\dfrac{(n-1)S^2}{\sigma^2}$	$\chi^2(n-1)$	Inferencia sobre σ^2 cuando μ es desconocida

2.3. MUESTREO EN DOS NORMALES UNIDIMENSIONALES

- $(X_1,\ldots,X_{n_1})$ m.a.s de $X \to \mathcal{N}(\mu_1,\sigma_1^2)$, $\quad \overline{X} = \dfrac{\sum_{i=1}^{n_1} X_i}{n_1}$, $\quad S_1^2 = \dfrac{\sum_{i=1}^{n_1} (X_i - \overline{X})^2}{n_1 - 1}$.
- $(Y_1,\ldots,Y_{n_2})$ m.a.s de $Y \to \mathcal{N}(\mu_2,\sigma_2^2)$, $\quad \overline{Y} = \dfrac{\sum_{j=1}^{n_2} Y_j}{n_2}$, $\quad S_2^2 = \dfrac{\sum_{j=1}^{n_2} (Y_j - \overline{Y})^2}{n_2 - 1}$.
- $(X_1,\ldots,X_{n_1})$, $(Y_1,\ldots,Y_{n_2})$ independientes.

EXTENSIÓN DEL LEMA DE FISHER

Los vectores $(\overline{X},\overline{Y})$ y (S_1^2,S_2^2) son independientes

DISTRIBUCIONES ASOCIADAS AL MUESTREO

Variable	Distribución	Uso
$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$	$\mathcal{N}(0,1)$	Inferencia sobre $\mu_1-\mu_2$ (σ_1^2,σ_2^2 conocidas)
$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}+\dfrac{(n_2-1)S_2^2}{\sigma_2^2}}\sqrt{\dfrac{\sigma_1^2/n_1+\sigma_2^2/n_2}{n_1+n_2-2}}}$	$t(n_1+n_2-2)$	
$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$	$t(n_1+n_2-2)$	Inferencia sobre $\mu_1-\mu_2$ ($\sigma_1^2=\sigma_2^2$ desconocidas)
$\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2\Big/n_1\sigma_1^2}{\sum_{j=1}^{n_2}(Y_j-\mu_2)^2\Big/n_2\sigma_2^2}$	$F(n_1,n_2)$	Inferencia sobre $\dfrac{\sigma_2^2}{\sigma_1^2}$ (μ_1,μ_2 conocidas)
$\dfrac{S_1^2\Big/\sigma_1^2}{S_2^2\Big/\sigma_2^2}$	$F(n_1-1,n_2-1)$	Inferencia sobre $\dfrac{\sigma_2^2}{\sigma_1^2}$ (μ_1,μ_2 desconocidas)