Algunas distribuciones de probabilidad útiles en estadística

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Computación
Primer cuatrimestre de 2023

Cambios de variables n-dimensionales

Proposición

Supongamos que X es una vector que se distribuye según una densidad $f(x)$ con soporte en $\overline{U}$ siendo U un abierto $\mathbb{R}^n$, y que $\varphi : U \to V$ es una función de C^1 y biyectiva con inversa C^1, donde V es otro abierto de $\mathbb{R}^n$ entonces, si consideramos el vector aleatorio $Y = \varphi(X)$, Y se distribuye en V según la densidad

$$f(\varphi^{-1}(y))|det(D\varphi^{-1})(y)|$$

Las densidades χ^2

Recordamos de la clase 9 que si $X \sim N(0,1)$ una variable aleatoria con distribución normal estándar $Y = X^2$ se distribuye según la densidad $\chi_1^2 = \Gamma(1/2, 1/2)$

Sean ahora $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con distribución normal estándar, y consideremos la variable aleatoria

$$Y_n = X_1^2 + X_2^2 + \ldots + X_n^2$$

¿cuál es la distribución de Y_n ? Por lo que vimos en una clase antes del parcial, es $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. Es decir, que la densidad de Y_n será

Densidad χ_n^2

$$f_{Y_n}(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{n/2-1} e^{-y/2} \quad (y > 0)$$

Esta densidad se conoce como densidad χ^2 con n grados de libertad. Las fórmulas para la esperanza y variancia de las distribuciones gama nos dan que

$$E[Y_n] = n, \quad \mathrm{Var}[Y_n] = 2n$$

Gráfico de la densidad χ^2_n

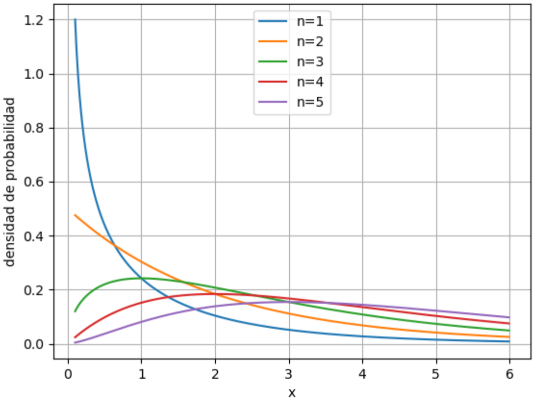
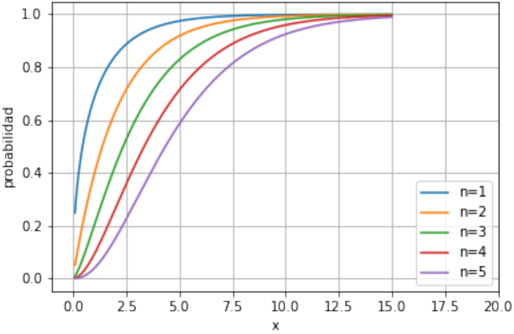


Gráfico de la distribución acumulada de una χ^2_n



Densidad del cociente de dos variables aleatorias independientes

Supongamos que X e Y son variables aleatorias continuas independientes, con densidades f_X y f_Y respectivamente. Supongamos además que Y está concentrada en la semirrecta positiva $(0, +\infty)$. Quremos calcular la densidad del cociente $T = X/Y$.

La densidad conjunta del vector aleatorio (X, Y) será $f_X(x)f_Y(y)$ como consecuencia de independencia de las variables X e Y.

Consideramos ahora el cambio de variable $(T, V) = \varphi(X, Y)$ donde donde

$$(u, v) = \varphi(x, y) = (x/y, y)$$

entonces la función inversa será

$$(x, y) = \varphi^{-1}(t, v) = (tv, v)$$

Y la diferencial de φ^{-1} es

$$D\varphi^{-1}(t, v) = \begin{pmatrix} v & t \\ 0 & 1 \end{pmatrix}$$

de modo que el Jacobiano es v.

Densidad del cociente de dos variables aleatorias independientes (2)

De acuerdo al teorema de cambio de variable, encontramos que el vector (T, V) se distribuye según la densidad conjunta

$$f_X(tv)f_Y(v)v$$

e integrando respecto la variable v podemos recuperar la densidad (marginal) de T que resulta ser:

$$T \simeq \int_0^\infty f_X(tv)f_Y(v)v\, dv \qquad (1)$$

La densidad t de Student

Sea X una variable aleatoria con distribución χ^2 con n grados de libertad, Y una variable aleatoria con distribución normal estándar y supongamos que X e Y son independientes. Queremos calcular la densidad de la variable aleatoria

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

[El porqué esta variable aleatoria es interesante, lo veremos más adelante al desarrollar conceptos de estadística]
Ya vimos que la densidad de X es la χ^2_n. Consideramos $\varphi : (0, +\infty) \to (0, +\infty)$ dada por

$$\varphi(x) = \sqrt{\frac{x}{n}}$$

es una función C^1 biyectiva cuya inversa $\varphi^{-1}(y) = ny^2$ también lo es.

La densidad t de Student (2)

Aplicando la fórmula de cambio de variables, encontramos que la densidad de $U = \sqrt{\frac{X}{n}}$ es

$$\begin{aligned} f_Y(y) &= \frac{(1/2)^{n/2}}{\Gamma(n/2)} \, (ny^2)^{n/2-1} e^{-ny^2/2} 2ny \; I_{(0,+\infty)}(y) \\ &= \frac{2n^{n/2}}{2^{n/2}\Gamma(n/2)} \, y^{n-1} \; e^{-ny^2/2} \; I_{(0,+\infty)}(y) \end{aligned}$$

La densidad t de Student (3)

Utilizando la fórmula para la densidad del cociente de dos variables aleatorias independientes, vemos que T se distribuye según la densidad

$$\begin{aligned} f_T(t) &= \int_0^\infty f_X(tv) f_Y(v) v \, dv \\ &= \frac{2n^{n/2}}{2^{n/2}\Gamma(n/2)\sqrt{2\pi}} \int_0^\infty e^{-t^2v^2/2} \; v^{n-1} \; e^{-nv^2/2} \; v \, dv \\ &= \frac{2^{(1-n)/2} n^{n/2}}{\Gamma(n/2)\sqrt{\pi}} \int_0^\infty e^{-(t^2+n)v^2/2} \; v^n \, dv \quad (t > 0) \end{aligned}$$

La densidad t de Student (4)

Hacemos el cambio de variable $x = \frac{v^2}{2}(t^2 + n)$, entonces esta integral se transforma en

$$\begin{aligned}
f_T(t) &= \frac{2^{(1-n)/2} n^{n/2}}{\Gamma(n/2)\sqrt{\pi}} \frac{1}{n+t^2} \int_0^\infty e^{-x} \left(\frac{2x}{n+t^2} \right)^{(n-1)/2} dx \\
&= \frac{n^{n/2}}{\Gamma(n/2)\sqrt{\pi}} \frac{1}{(n+t^2)^{(n+1)/2}} \int_0^\infty e^{-x} x^{(n-1)/2} dx \\
&= \frac{n^{n/2}}{\Gamma(n/2)\sqrt{\pi}} \Gamma\left(\frac{n+1}{2} \right) \frac{1}{(n+t^2)^{(n+1)/2}} \\
&= \frac{1}{\Gamma(n/2)\sqrt{n\pi}} \Gamma\left(\frac{n+1}{2} \right) \frac{n^{(n+1)/2}}{(n+t^2)^{(n+1)/2}}
\end{aligned}$$

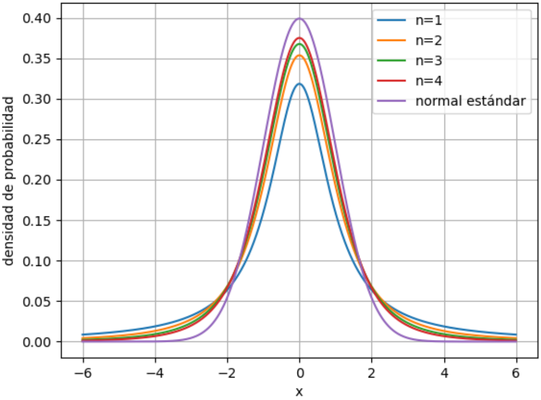
Conclusión

Finalmente obtenemos

Distribución t de Student con n grados de libertad

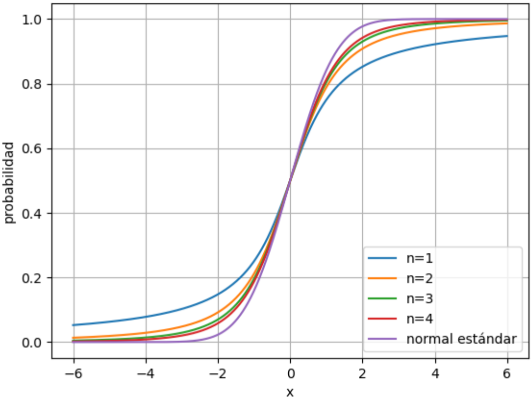
$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \qquad (t > 0)$$

Gráfico de la densidad t de Student



Cuando $n \to +\infty$, estas curvas convergen a la densidad normal estándar [¡ejercicio fácil de límites!].

Gráfico de la distribución acumulada de una t de Studient



Cuando $n \to +\infty$, estas curvas convergen a la distribución acumulada de una normal estándar.

Programando intervalos de confianza en Python

Intervalos de confianza

```
import numpy as np
from scipy.stats import t,chi2

x = np.random.normal(size=100)*2+3
mu = x.mean()
s = x.std(ddof=1)
n = len(x)
alfa = 0.05 # El nivel de error admido,
# 1-alfa es el nivel de confianza.

t_critico = np.abs(t.ppf((1-alfa)/2,n-1))
intervalo_de_confianza_mu= (mu-s*t_critico/np.sqrt(n),
  mu+s*t_critico/np.sqrt(n))

a= chi2.ppf(1-alfa/2,n-1)
b= chi2.ppf(alfa/2,n-1)
intervalo_de_confianza_varianza= ((n-1)*s**2/a, (n-1)*s**2/b)
```

Programando intervalos de confianza en Python

```
Verdadero valor de la media= 3
Verdadero valor de la varianza= 4
Estimación puntual de la media= 3.0918233475192323
Estimación puntual de la varianza= 3.9811561426531097
Intervalo de confianza para la media
(3.0792798078747565, 3.104366887163708)
Intervalo de confianza para varianza
(3.069057427663205, 5.372528009333845)
```