Probabildiad II 2020-2

Lecture : Distribuciones de Funciones de Vectores Aleatorios

Lecturer: F. Baltazar-Larios *Scribe:*

1. ESTADÍSTICAS DE ORDEN

Consideremos que en el espacio de probabilidad $(\Omega, \mathscr{F}, \mathbb{P})$ estarán definidas todas las variables aleatorias consideradas en este documento.

Definición 1.1. *Una colección finita de v.a.i.i.d. $X_1, \ldots, X_n$ es llamada muestra aleatoria (m.a.) (n es tamaño de la muestra).*

Nota 1.1. *Para cada $\omega \in \Omega$ tenemos que $X_i(\omega) := x_i \in \mathbb{R}$ con $i = 1, \ldots, n$.* ***Notación:*** *X denota a la v.a. y x a la evalución de un punto muestral ω en X.*

Ejemplo 1.1. *Supongamos que se lanzan de forma independiente tres dados identificables. Sea $X_i =$ el resultado obtenido en el i-ésimo dado.*

Tenemos que la colección finita de v.a.i.i.d. X_1, X_2, X_3 es un m.a. de tamaño 3 con distribución Uniforme en $\{1, \ldots, 6\}$. Si se realiza el experimeto aleatorio y obtenemos $\omega_1 = 4$, $\omega_2 = 1$ y $\omega_1 = 2$ entonces tenemos que $X_1(\omega_1) = x_1 = 4$, $X_2(\omega_2) = x_2 = 1$ y $X_3(\omega_3) = x_3 = 2$.

Definición 1.2. *Sea $X_1, \ldots, X_n$ una m.a. y $f : \mathbb{R}^n \to \mathbb{R}$ una función medible, entonces decimos que la función $f(X_1, \ldots, X_n)$ es una* ***estadística*** *de la m.a.*

Ejemplo 1.2. *Si en el Ejemplo 1.1 definimos a $f(X_1, X_2, X_3) = \sum_{i=1}^{3} \frac{X_i}{3}$ entonces tenemos que $f(4,, 1, 2) = \frac{7}{3}$.*

Si tenemos una m.a. $X_1, \ldots, X_n$ y evaluamos cada v.a. en un elemento del espacio muestral (ω) obtenemos una colección un vector en $\mathbb{R}^n$. En particular, esos reales pueden ser ordenados de menor a mayor (con repetición), si $X_{(i)}(\omega)$ representa el i-ésimo real una vez que fueron ordenados, tenemos $X_{(1)}(\omega) \leq \ldots X_{(n)}(\omega)$.En el Ejemplo 1.1 tenemos que $x_{(1)} = 1, x_{(2)} = 2$ y $x_{(3)} = 4$.

Si variamos el argumento (ω) obtenermos una la siguiente sucesión de variables aleatorias.

Definición 1.3. *Sea $X_1,\ldots,X_n$ una m.a., definimos a sus* ***estadísticas de orden*** *como*

$$\begin{aligned} X_{(1)} &= \text{mín}\{X_1,\ldots,X_n\} \\ X_{(2)} &= \text{mín}\{X_1,\ldots,X_n\}-\{X_{(1)}\} \\ X_{(3)} &= \text{mín}\{X_1,\ldots,X_n\}-\{X_{(1)},X_{(2)}\} \\ &\vdots \\ X_{(n)} &= \text{máx}\{X_1,\ldots,X_n\}. \end{aligned}$$

A $X_{(i)}$ la llamaremos i-ésima estadística de orden de la m.a., $i=1,\ldots,n$.

Nota 1.2. *Una observación importante es que $X_{(1)},\ldots,X_{(n)}$ no es una colección de v.a. independientes.*

Proposición 1.1. *Sea $X_1,\ldots,X_n$ una m.a. con función de densidad f y función de distribución F. La función de densidad de la i-ésima estadística de orden está dada por*

$$f_{X_{(i)}}(x)=\binom{n}{i} i f(x)[F(x)]^{i-1}[1-F(x)]^{n-i}.$$

Proof. Para cada $i\in 1,\ldots,n$ definamos la v.a.

$$Y_i=1_{(-\infty,x]}(X_i)=\begin{cases}1 & \text{si} \quad X_i\le x\\ 0 & \text{si} \quad X_i>x.\end{cases}$$

Tenemos que $Y_i\sim\text{Bernoulli}(F(x))$, entonces $S_n:=\sum_{i=1}^n Y_i\sim\text{Bin}(n,F(x))$. Entonces

$$\begin{aligned} F_{X_{(i)}}(x) &= \mathbb{P}(X_{(i)}\le x) \\ &= \mathbb{P}(S_n\ge i) \\ &= \sum_{k=i}^{n}\binom{n}{k}[F(x)]^k[1-F(x)]^{n-k}, \end{aligned}$$

derivando con respecto a x se obtiene la función de densidad. **Ejercicio Extra (EE) 1:** Derivar y obtener la función de densidad.

□

Corolario 1.1. *Sea $X_1,\ldots,X_n$ una m.a. con función de densidad f y función de distribución F. La función de densidad de la mínima y máxima estadística de orden están dadas por*

$$f_{X_{(1)}}(x)=nf(x)[1-F(x)]^{n-1}$$

y

$$f_{X_{(n)}}(x)=nf(x)[F(x)]^{n-1}.$$

En las Figuras 1 y 2 se ilustra la función de densidad de una m.a. para variables aleatorias como las del Ejemplo 1.1 para las mínima y máxima estadísticas de orden.

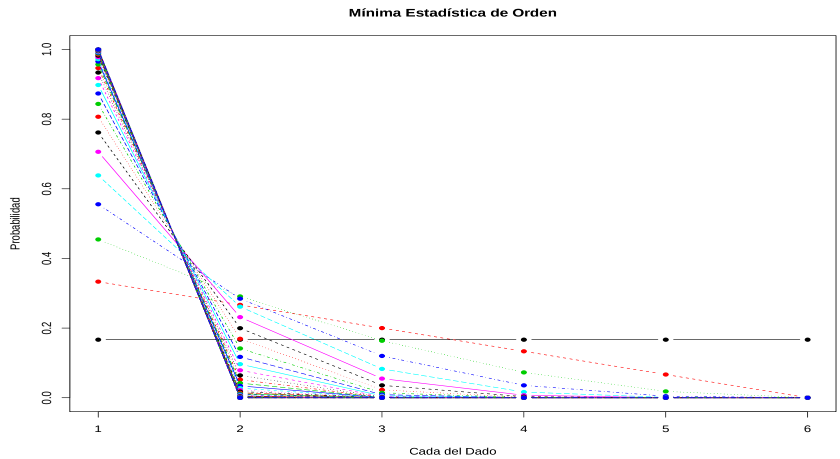

FIGURE 1. Función de densidad para n=1:100

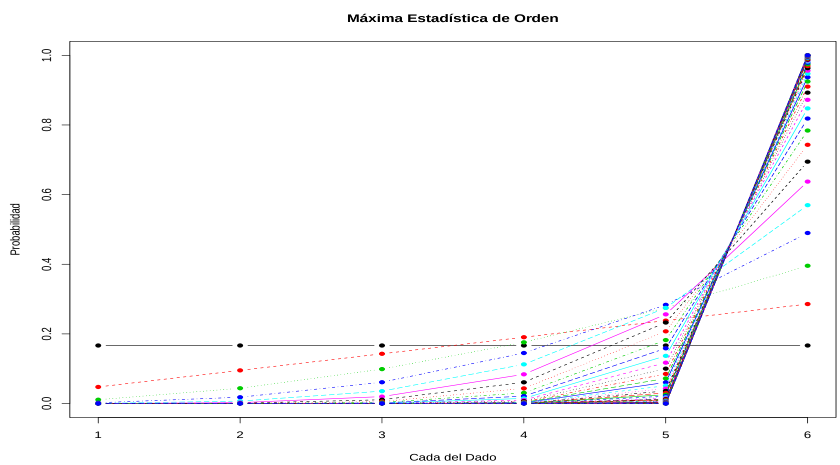

FIGURE 2. Función de densidad para n=1:100

Dada la m. a. $X_1,\ldots,X_n$ una estadística de interés es la medida de ”distancia” entre la mínima y máxima estadísticas de orden conocida como rango y definida por $R_n := X_{(n)} - X_{(1)}$.

Proposición 1.2. *Sea $X_1,\ldots,X_n$ una m.a. con función de densidad f y función de distribución F. La función de densidad de su rango está dada por*

$$f_{R_n}(x) = n(n-1)\int_{-\infty}^{\infty} f(y)f(x+y)[F(x+y)-F(y)]dy \quad y \geq 0.$$

Proof. Primero calcularemos la función de distribución del vector aleatorio $(X_{(1)}, X_{(n)})$ y con ésta su correspondiente función de densidad.

$$\begin{aligned} F_{X_{(1)},X_{(n)}}(x_1,x_2) &= \mathbb{P}(X_{(1)} \leq x_1, X_{(n)} \leq x_2) \\ &= \mathbb{P}(X_{(n)} \leq x_2) - \mathbb{P}(X_{(n)} \leq x_2, X_{(1)} > x_1) \\ &= [F(x_2)]^n - \mathbb{P}(x_1 < X_1 \leq x_2, \ldots, x_1 < X_n \leq x_2) \\ &= [F(x_2)]^n - [F(x_2) - F(x_1)]^n. \end{aligned}$$

Derivando tenemos que

$$f_{X_{(1)},X_{(n)}}(x_1,x_2) = n(n-1)f(x_1)f(x_2)[F(x_2)-F(x_1)]^{n-2}.$$

Por el Ejercicio 1.1 tenemos que

$$f_{R_n}(x) = n(n-1)\int_{\mathbb{R}} f(u)f(x+u)[F(x+u)-F(u)]^{n-2}du.$$

EE2: Derivar y obtener esta función de densidad. □

Ejercicio 1.1. *Dado que el vector aleatorio (X,Y) tiene función de densidad $f_{(X,Y)}$, probar que la función de densidad de la estadística $Z = Y - X$ está dada por*

$$f_Z(z) = \int_{\mathbb{R}} f_{X,Y}(u, z+u)du.$$

Ejercicio 1.2. *Sea $X_1,\ldots,X_n$ una m.a. con función de densidad f y función de distribución F. Calcular la función de densidad conjunta de las n estadísticas de orden. Para este objetivo deben obtener, primero, obetener la expresión genérica para una muestra de tamaño n y después probar que es válida.*

2. Transformaciones de Vectores Aleatorios

Consideremos el vector aleatorio $\boldsymbol{X} = (X_1,\ldots,X_n)$ con función de densidad $f_{\boldsymbol{X}}$ y a la función medible $g:\mathbb{R}^n \to \mathbb{R}^k$, entonces $\boldsymbol{Y} = g(\boldsymbol{X})$ es un k vector aleatorio, si $k=1$ entonces $\boldsymbol{Y}$ es una variable aleatoria.

Teorema 2.1. *Sea $\boldsymbol{X}$ un vector aleatorio continio con soporte $S_{\boldsymbol{X}} \subset \mathbb{R}^n$ y con función de densidad $f_{\boldsymbol{X}}$. Sea $g: S_{\boldsymbol{X}} \to \mathbb{R}^n$ una función continua con inversa g^{-1} diferenciable. Entonces el vector $\boldsymbol{Y} = g(\boldsymbol{X})$ tiene soporte en $g(S_{\boldsymbol{X}})$ y función de densidad*

$$f_{\boldsymbol{Y}}(y_1,\ldots,y_n) = \begin{cases} f_{\boldsymbol{X}}(g^{-1}(y_1,\ldots,y_n))|J(y_1,\ldots,y_n)| & si \quad (y_1,\ldots,y_n) \in g(S_{\boldsymbol{X}}) \\ 0 \quad en\ otro\ caso, \end{cases}$$

donde

$$J(y_1,\ldots,y_n) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n}. \end{vmatrix}$$

Ejemplo 2.1. *Sea $\boldsymbol{X}$ un vector aleatorio con función de densidad dada por*

$$f_{\boldsymbol{X}}(x_1,x_2) = \begin{cases} \frac{1}{2} & si \quad -1 \leq x_1+x_2, \leq 1, -1 \leq x_1-x_2, \leq 1 \\ 0 & en\ otro\ caso, \end{cases}$$

y definimos a $\boldsymbol{Y} = (X_1+X_2, -X_1+X_2)$.

Método 1: Podemos escribir

$$\boldsymbol{Y}^t = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \boldsymbol{X}^t = A\boldsymbol{X}^t.$$

Entonces la transformación invesa está dada por

$$\boldsymbol{X}^t = A^{-1}\boldsymbol{Y}^t = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} \boldsymbol{Y}^t.$$

Entonces tenemos que

$$|J(y_1,y_2)| = |\det(A^{-1})| = 1/2,$$

y

$$f_{\boldsymbol{Y}}(y_1,y_2) = f_{\boldsymbol{X}}((A^{-1}(y_1,y_2)^t)^t)(1/2) = \begin{cases} \frac{1}{4} & \text{si} \quad -1 \leq y_1,y_2 \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Método 2: Tenemos que $\boldsymbol{Y} = (y_1,y_2) = g(\boldsymbol{X}) = (x_1+x_2, x_2-x_1)$, entonces $g^{-1}(\frac{y_1-y_2}{2}, \frac{y_1+y_2}{2})$,

$$f_{\boldsymbol{X}}(g^{-1}(y_1,y_2)) = \frac{1}{2} \quad -1 \leq y_1,y_2 \leq 1$$

y

$$J = \begin{vmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{vmatrix} = 1/2.$$

Por lo tanto

$$f_{\boldsymbol{Y}}(y_1,y_2) = f_{\boldsymbol{X}}(g^{-1}(y_1,\ldots,y_n))|J(y_1,\ldots,y_n)| = \begin{cases} \frac{1}{4} & \text{si} \quad -1 \leq y_1,y_2 \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Ejercicio 2.1. *Sean $X_1,X_2 \sim U(0,1)$ (idependientes), encontrar la función de densidad de X_1+X_2 y X_1X_2.*

Ejercicio 2.2. *Sean $\boldsymbol{X}$ un vector aleatorio en $\mathbb{R}^2$ con función de densiad $f_{\boldsymbol{X}}$. Utilizando el Teorema 2.1, probar que la función de densiad de X_1+X_2 está dada por*

$$f_{X_1+X_2}(u) = \int_{\mathbb{R}} f_{\boldsymbol{X}}(u-v,v)dv.$$

Ejemplo 2.2. *Sea (X_1,X_2) un vector aleatorio absolutamente continuo con función de densidad $f_{X_1,X_2}(x_1,x_2)$ y $\Pr(X_2=0)=0$, entonces la función de densidad de la variable aleatoria $Z = \frac{X_1}{X_2}$ es*

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_1,X_2}(zy,y)dy.$$

Tomamos $\boldsymbol{Y} = (y_1, y_2) = g(\boldsymbol{X}) = (x_1/x_2, x_2)$, entonces $g^{-1}(y_1 y_2, y_2)$ y

$$J = \begin{vmatrix} y_2 & y_1 \\ 0 & 1 \end{vmatrix} = y_2.$$

Entonces

$$f_{Z,X_2}(z,y) = f_{X_1,X_2}(zy,y)|y|$$

y

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_1,X_2}(zy,y)|y|dy.$$

3. Estadísticas de Prueba

3.1. Distribución Ji (chi)-Cuadrada χ^2. El estadístico ji-cuadrado es utilizado para someter a prueba hipótesis referidas a distribuciones de frecuencias, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula (bondad de ajuste). Nos dice en qué medida se ajusta la distribución de frecuencias obtenida con los datos de una muestra.

Definición 3.1. *La variable aleatoria X tiene distribución ji-cuadrada con $n > 0$ grados de libertad ($X \sim \chi^2(n)$) si su función de densidad está dada por*

$$f_X(x) = \begin{cases} \left(\frac{1}{2}\right)^{n/2} \frac{x^{n/2-1} e^{-x/2}}{\Gamma(n/2)} & si \quad x > 0 \\ 0 & si \quad x \leq 0. \end{cases}$$

Proposición 3.1. *Si $X \sim N(0,1)$, entonces $X^2 \sim \chi^2(1)$.*

Proof. Tenemos que

$$\begin{aligned} F_{X^2}(x) &= \mathbb{P}(X^2 \leq x) \\ &= \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{x}) \\ &= F_X(\sqrt{x}) - F_X(-\sqrt{x}). \end{aligned}$$

Entonces

$$\begin{aligned} f_{X^2}(x) &= \frac{f_X(\sqrt{x})}{2\sqrt{x}} + \frac{f_X(-\sqrt{x})}{2\sqrt{x}} \\ &= \frac{f_X(\sqrt{x})}{\sqrt{x}} \\ &= \frac{e^{-x/2}}{\sqrt{2\pi x}} \\ &= \frac{1}{\Gamma(1/2)} \left(\frac{1}{2}\right)^{1/2} e^{-x/2} x^{1/2-1}. \end{aligned}$$

□

Ejercicio 3.1. *Sea $X_1, \ldots, X_n$ una m.a. normal estándar. Probar que $S_n = \sum_{i=1}^n X_i^2 \sim \chi^2(n)$.*

3.2. **Distribución F de Fisher-Snedecor.** La distribución **F** se utliza como estadística de prueba para checar si dos muestras aleatorias provienen de poblaciones de la misma varianza. Además se utiliza en anális de varianza que consiste en la comparación simultanea de medias poblacionales.

Definición 3.2. *La variable aleatoria X tiene distribución **F** con parámetros $n > 0$ y $m > 0$ ($X \sim F(n,m)$) si su función de densidad es*

$$f_X(x) = \begin{cases} \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} x^{n/2-1} \left(1+\frac{n}{m}x\right)^{-(n+m)/2} & si \quad x > 0 \\ 0 & si \quad x \leq 0. \end{cases}$$

Proposición 3.2. *Si $X \sim F(n,m)$, entonces*

(1) $\mathbb{E}[X] = \frac{m}{m-2}$ *con $m > 2$,*
(2) $Var[X] = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}$ *para $m > 4$.*

Proof. Tenemos que

$$\mathbb{E}[X] = \int_0^\infty \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} x^{n/2} \left(1+\frac{n}{m}x\right)^{-(n+m)/2} dx,$$

haciendo el cambio de variable $y = \left(1+\frac{n}{m}x\right)^{-1}$, entonces $x = \frac{m}{n}\left(\frac{1-y}{y}\right)$ y $dy = -\frac{n}{m}y^2 dx$, de esta forma

$$\begin{aligned} \mathbb{E}[X] &= \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} \int_0^1 \left(\frac{m}{n}\left(\frac{1-y}{y}\right)\right)^{n/2} y^{(n+m)/2} \frac{m}{n} y^{-2} dy \\ &= \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{m}{n}\right) \int_0^1 (1-y)^{(n/2+1)-1} y^{(m/2-1)-1} dy \\ &= \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{m}{n}\right) \frac{\Gamma(m/2-1)\Gamma(n/2+1)}{\Gamma((n+m)/2)} \\ &= \frac{\Gamma(m/2-1)(n/2)\Gamma(n/2)}{\Gamma(n/2)(m/2-1)\Gamma(m/2-1)} \left(\frac{m}{n}\right) \\ &= \frac{(n/2)}{(m/2-1)} \left(\frac{m}{n}\right) \\ &= \frac{m}{m-2}. \end{aligned}$$

EE3 Encontrar la varianza.

□

Proposición 3.3. *Si $X_1 \sim \chi^2(n_1)$ y $X_2 \sim \chi^2(n_2)$ independientes, entonces*

$$\frac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2).$$

Ejercicio 3.2. *Utilizando el Ejemplo 2.2 probar la Proposición 3.3.*

3.3. **Distribución T de Student.** La distribución de probabilidad **T** de Student se ultiza para estimar la media de una población con distribución normal cuando el tamaño de la muestra es pequeño (menor a 30).

Definición 3.3. *La variable aleatoria X tiene distribución $\boldsymbol{T}$ con parámetros $n > 0$ (grados de libertad) ($X \sim t(n)$) si su función de densidad es*

$$f_X(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}.$$

Proposición 3.4. *Si $X \sim t(n)$, entonces*

(1) $\mathbb{E}[X] = 0$,
(2) $Var[X] = \frac{n}{n-2}$ *para* $n > 2$.

Proof. Tenemos que

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} x(1+x^2/n)^{-(n+1)/2} dx = \int_{-\infty}^{\infty} h(x) dx,$$

desde que $h(-x) = -h(x)$ para toda $x \in \mathbb{R}$, es decir, h es una función impar entonces $\mathbb{E}[X] = 0$.
Para el cálculo de la varianza tenemos que

$$\begin{aligned} Var(X) &= \mathbb{E}[X^2] \\ &= \int_{-\infty}^{\infty} \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} x^2(1+x^2/n)^{-(n+1)/2} dx \\ &= \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \int_{-\infty}^{\infty} x^2(1+x^2/n)^{-(n+1)/2} dx \\ &= 2\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \int_{0}^{\infty} x^2(1+x^2/n)^{-(n+1)/2} dx \end{aligned}$$

haciendo el cambio de variabl e $y = \left(1 + x^2/n\right)^{-1}$, entonces $x^2 = \frac{n(1-y)}{y}$ y $dx = -\frac{\sqrt{n}}{2y^{3/2}(1-y)^{1/2}}dy$, de esta forma tenemos que

$$\begin{aligned}
Var(X) &= 2\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}\int_0^1 \frac{n(1-y)}{y} y^{(n+1)/2} \frac{\sqrt{n}}{2y^{3/2}(1-y)^{1/2}} dy \\
&= \frac{n}{\sqrt{\pi}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \int_0^1 (1-y)^{3/2-1} y^{(n/2-1)-1} dy \\
&= \frac{n}{\sqrt{\pi}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \frac{\Gamma((n-2)/2)\Gamma(3/2)}{\Gamma((n+1)/2)} \\
&= \frac{n}{\sqrt{\pi}} \frac{\Gamma((n-2)/2)\Gamma(3/2)}{\Gamma(n/2)} \\
&= \frac{n}{\sqrt{\pi}} \frac{\Gamma((n-2)/2)(1/2)\Gamma(1/2)}{(n/2-1)\Gamma(n/2-1)} \\
&= \frac{n(1/2)}{(n/2-1)} \\
&= \frac{n}{n-2}
\end{aligned}$$

□

Proposición 3.5. *Si $X_1 \sim N(0,1)$ y $X_2 \sim \chi^2(n)$ independientes, entonces*

$$\frac{X_1}{\sqrt{X_2/n}} \sim t(n).$$

Ejercicio 3.3. *Utilizando el Ejemplo 2.2 probar la Proposición 3.5.*

Proposición 3.6. *Si $X \sim t(n)$ entonces $X^2 \sim F(1,n)$.*

Proof. Tenemos que

$$\begin{aligned}
F_{X^2}(x) &= \mathbb{P}(X^2 \le x) \\
&= \mathbb{P}(-\sqrt{x} \le X \le \sqrt{x}) \\
&= F_X(\sqrt{x}) - F_X(-\sqrt{x}).
\end{aligned}$$

Entonces

$$\begin{aligned}
f_{X^2}(x) &= \frac{f_X(\sqrt{x})}{2\sqrt{x}} + \frac{f_X(-\sqrt{x})}{2\sqrt{x}} \\
&= \frac{f_X(\sqrt{x})}{\sqrt{x}} \\
&= \frac{\Gamma((n+1)/2)}{\sqrt{x}\sqrt{n\pi}\Gamma(n/2)} (1 + x/n)^{-(n+1)/2}.
\end{aligned}$$

□