Inferencia Estadística

Foto: José Juan Castro

Los Del DGIIM, `losdeldgiim.github.io`

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada

Inferencia Estadística

Los Del DGIIM, `losdeldgiim.github.io`

Granada, 2025

Índice general

1. Ejercicios de clase

Esta sección tiene el propósito de recoger todos los ejercicios propuestos en clase por parte de la profesora y que fueron resueltos por los alumnos en pizarra.

1.1. Estadísticos muestrales

Ejercicio 1.1.1. Obtener la función masa de probabilidad conjunta de una m.a.s. de $X \rightsquigarrow B(k_0, p)$ y la función de densidad de una m.a.s. de $X \rightsquigarrow U(a, b)$.

Recordamos que si $X \rightsquigarrow B(k_0, p)$, entonces:

$$P[X = x] = \binom{k_0}{x} p^x (1-p)^{n-x} \qquad \forall x \in \{0, \dots, k_0\}$$

Por lo que si tenemos una m.a.s. de n variables independientes e idénticamente distribuidas a X, $(X_1, \dots, X_n)$, su función de densidad vendrá dada por:

$$\begin{aligned} P[X_1 = x_1, \dots, X_n = x_n] &\overset{\text{indep.}}{=} \prod_{i=1}^{n} P[X_i = x_i] \overset{\text{id. d.}}{=} \prod_{i=1}^{n} P[X = x_i] \\ &= \prod_{i=1}^{n} \binom{k_0}{x_i} p^{x_i} (1-p)^{k_0 - x_i} = p^{\sum\limits_{i=1}^{n} x_i} (1-p)^{nk_0 - \sum\limits_{i=1}^{n} x_i} \prod_{i=1}^{n} \binom{k_0}{x_i} \\ &\forall x_i \in \{0, \dots, k_0\} \end{aligned}$$

Si ahora $X \rightsquigarrow U(a, b)$ para ciertos $a, b \in \mathbb{R}$ con $a < b$, entonces:

$$f_X(x) = \frac{1}{b-a} \qquad \forall x \in [a, b]$$

de donde:

$$f_{(X_1,\dots,X_n)}(x_1, \dots, x_n) \overset{\text{indep.}}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \overset{\text{id. d.}}{=} \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{b-a} = \frac{1}{(b-a)^n} \qquad \forall x \in [a, b]$$

Ejercicio 1.1.2. Para cada realización muestral, $(x_1, \dots, x_n) \in \mathcal{X}^n$, $F^*_{x_1,\dots,x_n}$ es una función de distribución en $\mathbb{R}$. En particular es una función a saltos, con saltos de amplitud $1/n$ en los sucesivos valores muestrales ordenados de menor a mayor, supuestos que sean distintos, y de saltos múltiplos en el caso de que varios valores muestrales coincidieran.

En las condiciones del enunciado, es decir, suponiendo que $x_1, \ldots, x_n$ están ordenados de menor a mayor y son distintos, entonces es fácil ver que:

$$F^*_{x_1,\ldots,x_n}(x) = \begin{cases} 0 & \text{si } x < x_1 \\ 1/n & \text{si } x_1 \leqslant x < x_2 \\ \vdots & \\ 1 & \text{si } x > x_n \end{cases} \qquad \forall x \in \mathbb{R}$$

Por lo que es claro que $F^*_{x_1,\ldots,x_n}$ es no decreciente, continua por la derecha, con límite 0 en $-\infty$ y con límite 1 en $+\infty$.

Ejercicio 1.1.3. $\forall x \in \mathbb{R}$, $F^*_{X_1,\ldots,X_n}(x)$ es una variable aleatoria tal que $nF^*_{X_1,\ldots,X_n}(x) \rightsquigarrow B(n, F(x))$ y:

$$E[F^*_{X_1,\ldots,X_n}(x)] = F(x), \qquad Var[F^*_{X_1,\ldots,X_n}(x)] = \frac{F(x)(1-F(x))}{n}$$

donde $F(x)$ es la función de distribución de X.

Recordamos que:

$$F^*_{X_1,\ldots,X_n}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{]-\infty,x]}(X_i) \qquad \forall x \in \mathbb{R}$$

Fijado $x \in \mathbb{R}$, tenemos que $I_{]-\infty,x]}(X) \rightsquigarrow B(1, P[X \leqslant x]) \equiv B(1, F(x))$, por lo que por la propiedad reproductiva de la binomial tenemos que:

$$nF^*_{X_1,\ldots,X_n}(x) \rightsquigarrow B(n, F(x))$$

Por lo que:

$$nE[F^*_{X_1,\ldots,X_n}(x)] = E[nF^*_{X_1,\ldots,X_n}(x)] = nF(x)$$

de donde:

$$E[F^*_{X_1,\ldots,X_n}(x)] = F(x)$$

Para la varianza:

$$n^2 Var[F^*_{X_1,\ldots,X_n}(x)] = Var[nF^*_{X_1,\ldots,X_n}(x)] = nF(x)(1-F(x))$$

de donde:

$$Var[F^*_{X_1,\ldots,X_n}(x)] = \frac{F(x)(1-F(x))}{n}$$

Ejercicio 1.1.4. Para valores grandes de n, en virtual del Teorema Central del Límite:

$$F^*_{X_1,\ldots,X_n}(x) \rightsquigarrow \mathcal{N}\left(F(x), \frac{F(x)(1-F(x))}{n}\right)$$

Sea $(X_1, \ldots, X_n)$ una m.a.s. de n muestras, sea:

$$S_n = \sum_{i=1}^{n} I_{]-\infty,x]}(X_i) \qquad \forall n \in \mathbb{N}$$

Por el Teorema Central del Límite tenemos que:

$$\frac{S_n - E[S_n]}{\sqrt{Var[S_n]}} \overset{n\to\infty}{\rightsquigarrow} \mathcal{N}(0,1) \Longrightarrow S_n \overset{n\to\infty}{\rightsquigarrow} \mathcal{N}\left(F(x), \frac{F(x)(1-F(x))}{n}\right)$$

Como $S_n \rightsquigarrow B(n, F(x))$, entonces tenemos que:

$$\begin{aligned} E[S_n] &= nF(x) \\ Var[S_n] &= nF(x)(1-F(x)) \end{aligned}$$

Por lo que:

$$F^*_{X_1,\dots,X_n}(x) = \frac{1}{n}S_n \overset{n\to\infty}{\rightsquigarrow} \mathcal{N}\left(F(x), \frac{F(x)(1-F(x))}{n}\right)$$

Ejercicio 1.1.5. Dada una muestra aleatoria simple formada por las observaciones $(3, 8, 5, 4, 5)$, obtener su función de distribución muestral y realizar la representación gráfica.

Aplicando la definición de la función de distribución muestral obtenemos que:

$$F^*_{(3,8,5,4,5)}(x) = \begin{cases} 0 & \text{si } x < 3 \\ 1 & \text{si } 3 \leqslant x < 4 \\ 2 & \text{si } 4 \leqslant x < 5 \\ 4 & \text{si } 5 \leqslant x < 8 \\ 5 & \text{si } x \geqslant 8 \end{cases}$$

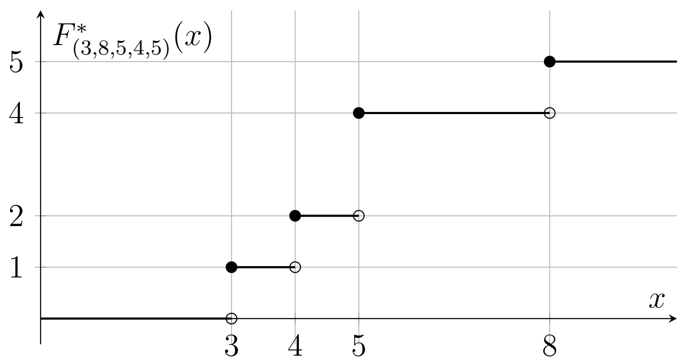

Figura 1.1: Representación gráfica de $F^*_{(3,8,5,4,5)}(x)$.

Ejercicio 1.1.6. Sea X una variable aleatoria con distribución $B(1, p)$ con $p \in (0, 1)$. Se toma una muestra de tamaño 5, $(X_1, X_2, X_3, X_4, X_5)$, y se obtiene la siguiente observación $(0, 1, 1, 0, 0)$. Determinar el valor de los estadísticos estudiados en la observación.

Aplicando las fórmulas vistas en clase obtenemos:

- Media: 0,4.
- Varianza: 0,24.

- Cuasivarianza: 0,3.
- $x_{(1)} = 0$, $x_{(2)} = 0$, $x_{(3)} = 0$, $x_{(4)} = 1$, $x_{(5)} = 1$.

Ejercicio 1.1.7. Sea $(X_1, \ldots, X_n)$ una m.a.s. y $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$, entonces:

$$M_{\overline{X}}(t) = \left(M_X(t/n)\right)^n$$

$$M_{\overline{X}}(t) = E\left[e^{t\overline{X}}\right] = E\left[e^{\frac{t}{n}\sum\limits_{i=1}^{n} X_i}\right] = M_{\sum\limits_{i=1}^{n} X_i}\left(\frac{t}{n}\right) \overset{\text{indep.}}{=} \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) \overset{\text{id. d.}}{=} \left(M_X\left(\frac{t}{n}\right)\right)^n$$

Ejercicio 1.1.8. Obtener la distribución muestral de $\overline{X}$ para $(X_1, \ldots, X_n)$ una m.a.s. de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

$$M_{\overline{X}}(t) = \left(M_X\left(\frac{t}{n}\right)\right)^n = \left(e^{\mu t + \frac{\sigma^2 t^2}{2n^2}}\right)^n = e^{\mu t + \frac{\sigma^2 t^2}{2n}}$$

Luego $\overline{X} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, ya que la función generatriz de momentos caracteriza la distribución.

Proposición 1.1. *Si tenemos una m.a.s.* $(X_1, \ldots, X_n)$*, entonces:*

$$F_{X_{(n)}}(x) = \left(F_X(x)\right)^n \qquad \forall x \in \mathbb{R}$$
$$F_{X_{(1)}}(x) = 1 - \left(1 - F_X(x)\right)^n$$

Demostración. Para la distribución del máximo:

$$\begin{aligned} F_{X_{(n)}}(x) &= P[X_{(n)} \leqslant x] = P[X_1 \leqslant x, \ldots, X_n \leqslant x] \overset{\text{indep.}}{=} \prod_{i=1}^{n} P[X_i \leqslant x] \\ &\overset{\text{id. d.}}{=} \prod_{i=1}^{n} P[X \leqslant x] = \left(F_X(x)\right)^n \end{aligned}$$

Para la del mínimo:

$$\begin{aligned} F_{X_{(1)}}(x) &= P[X_{(1)} \leqslant x] = 1 - P[X_{(1)} > x] = 1 - P[X_1 > x, \ldots, X_n > x] \\ &\overset{\text{indep.}}{=} 1 - \prod_{i=1}^{n} P[X_i > x] \overset{\text{id. d.}}{=} 1 - \left(P[X > x]\right)^n = 1 - \left(1 - F_X(x)\right)^n \end{aligned}$$

□

Ejercicio 1.1.9. Obtener las distribuciones muestrales de $X_{(1)}$ y $X_{(n)}$ para $X \rightsquigarrow U(a, b)$.

Si $X \rightsquigarrow U(a, b)$, entonces:

$$F_X(x) = \frac{x - a}{b - a} \qquad \forall x \in [a, b]$$

Por lo que aplicando la Proposición superior:

$$F_{X_{(n)}}(x) = (F_X(x))^n = \left(\frac{x-a}{b-a}\right)^n \qquad \forall x \in [a,b]$$

$$F_{X_{(1)}}(x) = 1-(1-F_X(x))^n = 1-(1-F_X(x))^n = 1-\left(1-\frac{x-a}{b-a}\right)^n$$
$$= 1-\left(\frac{b-x}{b-a}\right)^n \qquad \forall x \in [a,b]$$

1.2. Distribuciones en el muestreo de poblaciones normales

Proposición 1.2. *Sea* $X \rightsquigarrow \mathcal{N}(0,1)$*, entonces* $X^2 \rightsquigarrow \chi^2(1)$*.*

Demostración. Sea $Y = X^2 = h(X)$, entonces $X = \pm\sqrt{Y} = h^{-1}(y)$, por lo que:

$$f_Y(y) = f_X(h_1^{-1}(y))\left|\frac{dh_1^{-1}(y)}{dy}\right| + f_X(h_2^{-1}(y)) + \left|\frac{dh_2^{-1}(y)}{dy}\right|$$

Como $X \rightsquigarrow \mathcal{N}(0,1)$, entonces:

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \qquad \forall x \in \mathbb{R}$$

De donde:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{\frac{-(\sqrt{y})^2}{2}}\left|\frac{1}{2\sqrt{y}}\right| + \frac{1}{\sqrt{2\pi}}e^{\frac{-(-\sqrt{y})^2}{2}}\left|\frac{-1}{2\sqrt{y}}\right| = \frac{1}{\sqrt{2\pi y}}e^{\frac{-y}{2}} \qquad \forall y > 0$$

Por lo que $Y \rightsquigarrow \chi^2(1)$. □

Ejercicio 1.2.1. Calcula el valor de k o la probabilidad inducida:

a) $P[\chi^2(10) \geqslant k] = 0{,}005$.
   $k = 25{,}1881$.

b) $P[\chi^2(45) \leqslant k] = 0{,}005$.

$$P[\chi^2(45) \geqslant k] = 0{,}995 \Longrightarrow k = 24{,}3110$$

c) $P[\chi^2(14) \geqslant 21{,}06]$
   0,1

d) $P[\chi^2(20) \leqslant 12{,}44]$

$$P[\chi^2(20) \leqslant 12{,}44] = 1 - P[\chi^2(20) \geqslant 12{,}44] = 1 - 0{,}9 = 0{,}1$$

Ejercicio 1.2.2. Calcula el valor de k o la probabilidad inducida:

a) $P[t(26) \geqslant k] = 0{,}05$
   $k = 1{,}7056$

b) $P[t(20) \leqslant k] = 0{,}25$
   $k = -0{,}6870$

c) $P[t(26) \geqslant k] = 0{,}9$
   $k = -1{,}3150$

d) $P[t(21) \geqslant 1{,}721]$
   0,05

e) $P[t(11) \leqslant 0{,}697]$
   0,75

f) $P[t(8) \leqslant -2{,}306]$
   0,025

Ejercicio 1.2.3. Calcula el valor de k o la probabilidad inducida:

a) $P[F(7,3) \leqslant k] = 0{,}95$
   $k = 8{,}89$

b) $P[F(8,4) \geqslant k] = 0{,}01$

$$0{,}01 = 1 - P[F(8,4) \leqslant k] \Longrightarrow P[F(8,4) \leqslant k] = 0{,}99 \Longrightarrow k = 14{,}8$$

c) $P[F(2,2) \leqslant 19]$
   0,95

d) $P[F(3,5) \geqslant 12{,}1]$

$$P[F(3,5) \geqslant 12{,}1] = 1 - P[F(3,5) \leqslant 12{,}1] = 1 - 0{,}99 = 0{,}01$$

e) $P[F(60,40) \leqslant k] = 0{,}05$
   $k = 0{,}627$

1.2.1. Varias demostraciones

Tenemos una $(X_1, \dots, X_n)$ m.a.s. con $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, si tomamos:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Demostraciones importantes que pueden caer.

Proposición 1.3. *En dichas condiciones, veamos que:*

$$\overline{X}, (X_1 - \overline{X}, \dots, X_n - \overline{X}) \text{ son independientes}$$

Demostración. Para ellos, usaremos la caracterización por la función generatriz de momentos conjunta:

$$M_{\overline{X},X_1-\overline{X},\dots,X_n-\overline{X}}(t,t_1,\dots,t_n) \overset{?}{=} M_{\overline{X}}(t)M_{X_1-\overline{X},\dots,X_n-\overline{X}}(t_1,\dots,t_n)$$

$$\begin{aligned}
M_{\overline{X},X_1-\overline{X},\dots,X_n-\overline{X}}(t,t_1,\dots,t_n) &= E[e^{(t,t_1,\dots,t_n)\cdot(\overline{X},X_1-\overline{X},\dots,X_n-\overline{X})}] \\
&= E\left[e^{t\overline{X}+\sum\limits_{i=1}^{n}(X_i-\overline{X})t_i}\right] \\
&= E\left[e^{\frac{t}{n}\sum\limits_{i=1}^{n}X_i+\sum\limits_{i=1}^{n}X_it_i-\sum\limits_{i=1}^{n}\overline{X}t_i}\right] \\
&= E\left[e^{\frac{t}{n}\sum\limits_{i=1}^{n}X_i+\sum\limits_{i=1}^{n}X_it_i-\overline{X}\sum\limits_{i=1}^{n}t_i}\right] \\
&= E\left[e^{\frac{t}{n}\sum\limits_{i=1}^{n}X_i+\sum\limits_{i=1}^{n}X_it_i-\frac{1}{n}\sum\limits_{i=1}^{n}X_i\sum\limits_{i=1}^{n}t_i}\right] \\
&= E\left[e^{\frac{t}{n}\sum\limits_{i=1}^{n}X_i+\sum\limits_{i=1}^{n}X_it_i-\sum\limits_{i=1}^{n}X_i\frac{1}{n}\sum\limits_{i=1}^{n}t_i}\right] \\
&= E\left[e^{\frac{t}{n}\sum\limits_{i=1}^{n}X_i+\sum\limits_{i=1}^{n}X_it_i-\sum\limits_{i=1}^{n}X_i\bar{t}}\right] \\
&= E\left[e^{\sum\limits_{i=1}^{n}X_i\frac{t}{n}+\sum\limits_{i=1}^{n}X_it_i-\sum\limits_{i=1}^{n}X_i\bar{t}}\right] \\
&= E\left[e^{\sum\limits_{i=1}^{n}X_i\left(\frac{t}{n}+t_i-\bar{t}\right)}\right] \\
&= E\left[\prod_{i=1}^{n}e^{X_i\left(\frac{t}{n}+t_i-\bar{t}\right)}\right] \\
&\overset{\text{indep.}}{=} \prod_{i=1}^{n}E\left[e^{X_i\left(\frac{t}{n}+t_i-\bar{t}\right)}\right] = \prod_{i=1}^{n}M_{X_i}\left(\frac{t}{n}+t_i-\bar{t}\right) \\
&\overset{(*)}{=} \prod_{i=1}^{n}e^{\left(\frac{t}{n}+t_i-\bar{t}\right)\mu+\left(\frac{t}{n}+t_i-t\right)^2\frac{\sigma^2}{2}} \\
&= e^{\sum\limits_{i=1}^{n}\left[\left(\frac{t}{n}+t_i-\bar{t}\right)\mu+\frac{\sigma^2}{2}\left(\frac{t^2}{n^2}+\left(t_i-\bar{t}\right)^2+2\frac{t}{n}(t_i-\bar{t})\right)\right]} \\
&= e^{\sum\limits_{i=1}^{n}\frac{\mu t}{n}+\sum\limits_{i=1}^{n}\mu(t_i-\bar{t})+\frac{\sigma^2}{2}\left(\sum\limits_{i=1}^{n}\frac{t^2}{n^2}+\sum\limits_{i=1}^{n}(t_i-\bar{t})^2+2\sum\limits_{i=1}^{n}\frac{t}{n}(t_i-\bar{t})\right)} \\
&= e^{\frac{\cancel{n}\mu t}{\cancel{n}}\cancel{\mu\sum t_i-\bar{t}}+\frac{\sigma^2}{n^2}\left(\frac{nt^2}{n^2}+\sum\limits_{i=1}^{n}(t_i-\bar{t})^2+2\frac{t}{n}\sum\limits_{i=1}^{n}(t_i-\bar{t})\right)} \\
&= e^{\mu t+\frac{\sigma^2t^2}{2n}+\frac{\sigma^2}{2}\sum\limits_{i=1}^{n}(t_i-\bar{t})^2}
\end{aligned}$$

Sabemos que:

$$\begin{aligned}
M_{\overline{X}}(t) &= M_{\left(\overline{X}, X_1-\overline{X}, \ldots, X_n-\overline{X}\right)}(t, 0, \ldots, 0) = e^{\mu t+\frac{\sigma^2 t^2}{2n}} \\
M_{(X_1-\overline{X}, \ldots, X_n-\overline{X})}(t_1, \ldots, t_n) &= M_{(\overline{X}, X_1-\overline{X}, \ldots, X_n-\overline{X})}(0, t_1, \ldots, t_n) \\
&= e^{\frac{\sigma^2}{n}\sum\limits_{i=1}^{n}(t_i-\bar{t})^2}
\end{aligned}$$

Por lo que es cierto que el producto de las funciones generatrices de momentos es la generatriz de mmoentos conjunta, luego las variables son independientes. □

Corolario 1.3.1. *Como corolario de la Proposición anterior, tenemos que:*

- *Se vio ya, y se saca de la demostración de arriba.*
- *Lema de Fisher:* $\overline{X}$ *y* S^2 *son independientes.*

  Como S^2 *es función del vector de la Proposición anterior, tenemos que es independiente con* $\overline{X}$*, ya que las funciones de variables independientes son independientes.*
- $\dfrac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$

  Para demostrarlo:

  $$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \rightsquigarrow \chi^2(n)$$

  Ahora, queremos ver que:

  $$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$$

  Para ello:

  $$\begin{aligned}
  \frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} &= \frac{\sum\limits_{i=1}^{n}(X_i-\overline{X}+\overline{X}-\mu)^2}{\sigma^2} \\
  &= \frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2 + \sum\limits_{i=1}^{n}(\overline{X}-\mu)^2 + 2\sum\limits_{i=1}^{n}(X_i-\overline{X})(\overline{X}-\mu)}{\sigma^2} \\
  &= \frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} + n\frac{(\overline{X}-\mu)^2}{\sigma^2}
  \end{aligned}$$

  Y como:

  $$n\frac{(\overline{X}-\mu)^2}{\sigma^2} \rightsquigarrow \chi^2(1)$$

  Buscamos ver lo que sigue lo de la derecha ($A = B + C$)*. Para ello, usaremos la función generatriz de momentos. tenemos que* $B = f(S^2)$ *y* $C = f(\overline{X})$*, luego* B *y* C *son independientes, por lo que:*

  $$M_{A=B+C}(t) \overset{indep.}{=} M_B(t)M_C(t) = M_B(t)\frac{1}{(1-2t)^{\frac{1}{2}}} \qquad t < \frac{1}{2}$$

Y sabemos que:

$$M_A(t) = \frac{1}{(1-2t)^{\frac{n}{2}}}$$

De donde:

$$M_B(t) = \frac{M_A(t)}{M_C(t)} = \frac{\frac{1}{(1-2t)^{\frac{n}{2}}}}{\frac{1}{(1-2t)^{\frac{1}{2}}}} = \frac{1}{(1-2t)^{\frac{n-1}{2}}} \qquad t < \frac{1}{2}$$

Por lo que $B \rightsquigarrow \chi^2(n-1)$

- $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \rightsquigarrow t(n-1)$

  Para ello, al igual que la χ^2*, lo más sencillo es ir a la construcción de* t*:*

$$\left.\begin{array}{l} X \rightsquigarrow \mathcal{N}(0,1) \\ Y \rightsquigarrow \chi^2(n) \\ indep \end{array}\right\} \frac{X}{\sqrt{Y/n}} \rightsquigarrow t(n)$$

  Como:

$$\overline{X} \rightsquigarrow \mathcal{N}(\mu, \sigma^2) \Longrightarrow \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$$
$$\frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$$

  que son independientes por el Lema de Fisher. Si aplicamos la construcción:

$$\frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}} = \frac{\overline{X}-\mu}{S/\sqrt{n}} \rightsquigarrow t(n-1)$$

Este último corolario ayuda a inferir parámetros de las distribuciones. Ver punto 2.3.1. Inferencia sobre x significa que queremos averiguar el valor de x. El corolario de ayer nos sirve para usar otros estadísticos en lugar de otros que deberíamos usar, pero con 1 parámetro desconocido en lugar de 2. Aprenderse fórmulas de 2.3.1.

1.2.2. Dos poblaciones normales

Teorema 1.4 (extensión del Lema de Fisher)**.** *Los vectores* $(\overline{X}, \overline{Y})$*,* (S_1^2, S_2^2) *son independientes.*

Demostración. Usemos la función generatriz de momentos, ya que son independientes si y solo si:

$$M_{(\overline{X},\overline{Y},S_1^2,S_2^2)}(t_1,t_2,s_1,s_2) \stackrel{?}{=} M_{(\overline{X},\overline{Y})}(t_1,t_2) M_{(S_1^2,S_2^2)}(s_1,s_2)$$

Como las X y las Y son independientes:

$$M_{\left(\overline{X},\overline{Y},S_1^2,S_2^2\right)}(t_1,t_2,s_1,s_2) = M_{\left(\overline{X},S_1^2\right)}(t_1,s_2)M_{\left(\overline{Y},S_2^2\right)}(s_1,s_2) \stackrel{\text{Lema Fisher}}{=} M_{\overline{X}}(t_1)M_{S_1^2}(s_1)M_{\overline{Y}}(t_2)M_{S_2^2}(s_2)$$

Ahora, como X e Y son independientes:

$$M_{\overline{X}}(t_1)M_{S_1^2}(s_1)M_{\overline{Y}}(t_2)M_{S_2^2}(s_2) = M_{\left(\overline{X},\overline{Y}\right)}(t_1,t_2)M_{\left(S_1^2,S_2^2\right)}(s_1,s_2)$$

$\square$

Corolario 1.4.1. *A partir de entonces (aunque el 3o es el único que necesita el Teorema anterior):*

1. *Tenemos (aunque no sea corolario de Fisher):*

$$\frac{n_2}{n_1}\frac{\sigma_2^2}{\sigma_1^2}\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} \rightsquigarrow F(n_1,n_2)$$

Que es equivalente a que:

$$\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2/n_1\sigma_1^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2\sigma_2^2} \rightsquigarrow F(n_1,n_2)$$

Por construcción de $F(n_1,n_2)$:

$$\left.\begin{array}{l} X \rightsquigarrow \chi^2(m) \\ \qquad \textit{independientes} \\ Y \rightsquigarrow \chi^2(n) \end{array}\right\} \Longrightarrow \frac{X/m}{Y/n} \rightsquigarrow F(n_1,n_2)$$

Como ayer vimos que:

$$\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sigma_1^2} \rightsquigarrow \chi^2(n_1)$$
$$\frac{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}{\sigma_2^2} \rightsquigarrow \chi^2(n_2)$$

Como X e Y son independientes, tenemos funciones en función de X e Y, luego estas dos variables son independientes. Ahora:

$$\frac{\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{n_1\sigma_1^2}}{\dfrac{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}{n_2\sigma_2^2}} \rightsquigarrow F(n_1,n_2) \rightsquigarrow F(n_1,n_2)$$

2. *Tenemos:*
$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow F(n_1-1, n_2-1)$$
Seguimos buscando la construcciónde F, por lo que buscamos aplicar lo que sabemos:
$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \rightsquigarrow \chi^2(n_1-1)$$
$$\frac{(n_2-1)S_2^2}{\sigma_2^2} \rightsquigarrow \chi^2(n_2-1)$$
Que son independientes por ser funciones de X e Y, que son independientes. Dividimos:
$$\frac{\dfrac{\cancel{(n_1-1)}S_1^2}{\sigma_1^2\cancel{(n_1-1)}}}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2(n_2-1)}} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow F(n_1-1, n_2-1)$$
En particular, si $\sigma_1 = \sigma_2$, se tiene:
$$\frac{S_1^2}{S_2^2} \rightsquigarrow F(n_1-1, n_2-1)$$

3. *La construcción de la t-Student era:*
$$\left.\begin{array}{l} X \rightsquigarrow \mathcal{N}(0,1) \\ \qquad \textit{independientes} \\ Y \rightsquigarrow \chi^2(n) \end{array}\right\} \Longrightarrow \frac{X}{\sqrt{Y/n}} \rightsquigarrow t(n)$$
Queremos probar:
$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2}{\sigma_1^2} + \dfrac{(n_2-1)S_2^2}{\sigma_2^2}}} \sqrt{\frac{n_1+n_2-2}{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2}{n_2}}} \rightsquigarrow t(n_1+n_2-2)$$
Ahora:
$$\left.\begin{array}{l} \overline{X} \rightsquigarrow \mathcal{N}\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \\ \overline{Y} \rightsquigarrow \mathcal{N}\left(\mu_2, \frac{\sigma_2^2}{n_2}\right) \end{array}\right\} \Longrightarrow \overline{X} - \overline{Y} \rightsquigarrow \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$
Tipificamos:
$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1)$$
Para el denominador ahora, de la S^2 sabemos:
$$\frac{(n_1-1)S_1^2}{n_1^2} \rightsquigarrow \chi^2(n_1-1)$$
$$\frac{(n_2-1)S_2^2}{n_2^2} \rightsquigarrow \chi^2(n_2-1)$$

Como X e Y son independientes, estas dos son independientes, y al sumar dos χ^2 independientes tenemos la propiedad reproductiva:

$$\frac{(n_1-1)S_1^2}{n_1^2}+\frac{(n_2-1)S_2^2}{n_2^2} \rightsquigarrow \chi^2(n_1+n_2-2)$$

Por la extensión del Lema de Fisher tenemos que las dos variables aleatorias que hemos calculado son independientes. Procedemos ahora a aplicar la construcción de la t:

$$\frac{\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}}{\sqrt{\dfrac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}+\dfrac{(n_2-1)S_2^2}{\sigma_2^2}}{n_1+n_2-2}}} \rightsquigarrow t(n_1+n_2-2)$$

Las del punto 2.4.1. hay que aprenderlas de memoria también

2. Relaciones de Ejercicios

2.1. Estadísticos muestrales

Ejercicio 2.1.1. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. Dar el espacio muestral y calcular la función masa de probabilidad de $(X_1, \ldots, X_n)$ en cada uno de los siguientes casos:

a) $X \rightsquigarrow \{B(k_0, p) : p \in (0, 1)\}$ Binomial.

El espacio muestral en este caso es $\mathcal{X}^n$, donde:

$$\mathcal{X} = \{0, 1, ..., k_0\}$$

Recordamos que si $X \rightsquigarrow B(k_0, p)$, entonces:

$$P[X = x] = \binom{k_0}{x} p^x (1-p)^{k_0 - x} \qquad \forall x \in \mathcal{X}$$

Por tanto, para nuestra m.a.s. tendremos la función masa de probabilidad:

$$\begin{aligned} P[X_1 = x_1, \ldots, X_n = x_n] &\overset{\text{indep.}}{=} \prod_{i=1}^n P[X_i = x_i] \overset{\text{id. d.}}{=} \prod_{i=1}^n P[X = x_i] \\ &= \prod_{i=1}^n \binom{k_0}{x_i} p^{x_i} (1-p)^{k_0 - x_i} = p^{\sum\limits_{i=1}^n x_i} (1-p)^{nk_0 - \sum\limits_{i=1}^n x_i} \prod_{i=1}^n \binom{k_0}{x_i} \\ &\forall (x_1, \ldots, x_n) \in \mathcal{X}^n \end{aligned}$$

b) $X \rightsquigarrow \{\mathcal{P}(\lambda) : \lambda \in \mathbb{R}^+\}$ Poisson.

El espacio muestral de X es:

$$\mathcal{X} = \mathbb{N} \cup \{0\}$$

Recordamos que si $X \rightsquigarrow \mathcal{P}(\lambda)$, entonces:

$$P[X = x] = e^{-\lambda} \frac{\lambda^x}{x!} \qquad \forall x \in \mathcal{X}$$

Por tanto:

$$P[X_1 = x_1, \ldots, X_n = x_n] \overset{\text{indep.}}{=} \prod_{i=1}^n P[X_i = x_i] \overset{\text{id. d.}}{=} \prod_{i=1}^n P[X = x_i]$$

$$= \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \cdot \frac{\lambda^{\sum\limits_{i=1}^n x_i}}{\prod\limits_{i=1}^n x_i} \qquad \forall (x_1, \ldots, x_n) \in \mathcal{X}^n$$

c) $X \rightsquigarrow \{BN(k_0, p) : p \in (0, 1)\}$ Binomial Negativa.

El espacio muestral de X es:

$$\mathcal{X} = \mathbb{N} \cup \{0\}$$

Recordamos que si $X \rightsquigarrow BN(k_0, p)$, entonces:

$$P[X = x] = \binom{x + k_0 - 1}{x} (1-p)^x p^{k_0} \qquad \forall x \in \mathcal{X}$$

Por tanto:

$$P[X_1 = x_1, \dots, X_n = x_n] \overset{\text{indep.}}{=} \prod_{i=1}^{n} P[X_i = x_i] \overset{\text{id. d.}}{=} \prod_{i=1}^{n} P[X = x_i]$$
$$= \prod_{i=1}^{n} \binom{x_i + k_0 - 1}{x_i} (1-p)^{x_i} p^{k_0} = p^{nk_0} (1-p)^{\sum\limits_{i=1}^{n} x_i} \prod_{i=1}^{n} \binom{x_i + k_0 - 1}{x_i}$$
$$\forall (x_1, \dots, x_n) \in \mathcal{X}^n$$

d) $X \rightsquigarrow \{G(p) : p \in (0, 1)\}$ Geométrica.

El espacio muestral de X es:

$$\mathcal{X} = \mathbb{N} \cup \{0\}$$

Recordamos que $G(p) \equiv BN(1, p)$, por lo que si sustituimos en la fórmula obtenida en la Binomial Negativa $k_0 = 1$:

$$P[X_1 = x_1, \dots, X_n = x_n] = p^n (1-p)^{\sum\limits_{i=1}^{n} x_i} \qquad \forall (x_1, \dots, x_n) \in \mathcal{X}^n$$

e) $X \rightsquigarrow \{P_N : N \in \mathbb{N}\}, \quad P_N(X = x) = \dfrac{1}{N}, \quad x = 1, \dots, N.$

El espacio muestral ya nos lo dan: $\mathcal{X} = \{1, \dots, N\}$. Calculemos la masa de probabilidad:

$$P[X_1 = x_1, \dots, X_n = x_n] \overset{\text{indep.}}{=} \prod_{i=1}^{n} P[X_i = x_i] \overset{\text{id. d.}}{=} \prod_{i=1}^{n} P[X = x_i]$$
$$= \prod_{i=1}^{n} \frac{1}{N} = \left(\frac{1}{N}\right)^n \qquad \forall (x_1, \dots, x_n) \in \mathcal{X}^n$$

Ejercicio 2.1.2. Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. Dar el espacio muestral y calcular la función de densidad de $(X_1, \dots, X_n)$ en cada uno de los siguientes casos:

a) $X \rightsquigarrow \{U(a, b) : a, b \in \mathbb{R}, a < b\}$ Uniforme.

El espcio muestral en este caso es $\mathcal{X}^n$, donde:

$$\mathcal{X} = [a, b]$$

Recordamos que si $X \rightsquigarrow U(a,b)$, entonces:

$$f_X(x) = \frac{1}{b-a} \qquad \forall x \in [a,b]$$

Por lo que:

$$\begin{aligned} f_{(X_1,\dots,X_n)}(x_1,\dots,x_n) &\stackrel{\text{indep.}}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \stackrel{\text{id. d.}}{=} \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{b-a} \\ &= \left(\frac{1}{b-a}\right)^n \qquad \forall (x_1,\dots,x_n) \in \mathcal{X}^n \end{aligned}$$

b) $X \rightsquigarrow \{\mathcal{N}(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$ Normal.

El espacio muestral de X es $\mathcal{X} = \mathbb{R}$. Recordamos que si $X \rightsquigarrow \mathcal{N}(\mu,\sigma^2)$, entonces:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \qquad \forall x \in \mathbb{R}$$

Por lo que:

$$\begin{aligned} f_{(X_1,\dots,X_n)}(x_1,\dots,x_n) &\stackrel{\text{indep.}}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \stackrel{\text{id. d.}}{=} \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x_i-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \prod_{i=1}^{n} e^{-\dfrac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\sum\limits_{i=1}^{n} \dfrac{(x_i-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{\frac{-1}{2\sigma^2} \sum\limits_{i=1}^{n} (x_i-\mu)^2} \qquad \forall (x_1,\dots,x_n) \in \mathbb{R}^n \end{aligned}$$

c) $X \rightsquigarrow \{\Gamma(p,a) : p, a \in \mathbb{R}^+\}$ Gamma.

El espacio muestral de X es $\mathcal{X} = \mathbb{R}_0^+$. Recordamos que si $X \rightsquigarrow \Gamma(p,a)$, entonces:

$$f_X(x) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} \qquad \forall x \in \mathbb{R}_0^+$$

Por lo que:

$$\begin{aligned} f_{(X_1,\dots,X_n)}(x_1,\dots,x_n) &\stackrel{\text{indep.}}{=} \prod_{i=1}^{n} f_{X_i}(x_i) \stackrel{\text{id. d.}}{=} \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{a^p}{\Gamma(p)} x_i^{p-1} e^{-ax_i} \\ &= \left(\frac{a^p}{\Gamma(p)}\right)^n \cdot e^{-a\sum\limits_{i=1}^{n} x_i} \cdot \prod_{i=1}^{n} x_i^{p-1} \qquad \forall (x_1,\dots,x_n) \in \mathcal{X}^n \end{aligned}$$

d) $X \rightsquigarrow \{\beta(p,q) : p, q \in \mathbb{R}^+\}$ Beta.

El espacio muestral de X es $\mathcal{X} = [0,1]$. Recordamos que si $X \rightsquigarrow \beta(p,q)$, entonces:

$$f_X(x) = \frac{1}{\beta(p,q)} x^{p-1} (1-x)^{q-1} \qquad \forall x \in [0,1]$$

Donde:

$$\beta(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

Por tanto:

$$\begin{aligned} f_{(X_1,\dots,X_n)}(x_1,\dots,x_n) &\stackrel{\text{indep.}}{=} \prod_{i=1}^n f_{X_i}(x_i) \stackrel{\text{id. d.}}{=} \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \frac{1}{\beta(p,q)} x_i^{p-1}(1-x_i)^{q-1} \\ &= \frac{1}{\beta(p,q)^n} \prod_{i=1}^n x_i^{p-1}(1-x_i)^{q-1} \qquad \forall (x_1,\dots,x_n) \in \mathcal{X}^n \end{aligned}$$

e) $X \rightsquigarrow \{P_\theta : \theta \in \mathbb{R}^+\}, \quad f_\theta(x) = \dfrac{1}{2\sqrt{x\theta}}, \quad 0 < x < \theta.$

Se nos dice que $\mathcal{X} =]0,\theta[$. Calculamos la función de densidad conjunta:

$$\begin{aligned} f_{(X_1,\dots,X_n)}(x_1,\dots,x_n) &\stackrel{\text{indep.}}{=} \prod_{i=1}^n f_{X_i}(x_i) \stackrel{\text{id. d.}}{=} \prod_{i=1}^n f_X(x_i) = \prod_{i=1}^n \frac{1}{2\sqrt{x_i\theta}} \\ &= \frac{1}{\left(2\sqrt{\theta}\right)^n} \prod_{i=1}^n \frac{1}{\sqrt{x_i}} \qquad \forall (x_1,\dots,x_n) \in \mathcal{X}^n \end{aligned}$$

Ejercicio 2.1.3. Se miden los tiempos de sedimentación de una muestra de partículas flotando en un líquido. Los tiempos observados son:

11,5; 1,8; 7,3; 12,1; 1,8; 21,3; 7,3; 15,2; 7,3; 12,1; 15,2;
7,3; 12,1; 1,8; 10,5; 15,2; 21,3; 10,5; 15,2; 11,5

- Construir la función de distribución muestral asociada a a dichas observaciones.

  Si aplicamos la definición de función de distribución muestral obtenemos que esta viene dada por:

$$F_n^*(x) = \begin{cases} 0 & \text{si } x < 1{,}8 \\ 3/20 & \text{si } 1{,}8 \leqslant x < 7{,}3 \\ 7/20 & \text{si } 7{,}3 \leqslant x < 10{,}5 \\ 9/20 & \text{si } 10{,}5 \leqslant x < 11{,}5 \\ 11/20 & \text{si } 11{,}5 \leqslant x < 12{,}1 \\ 14/20 & \text{si } 12{,}1 \leqslant x < 15{,}2 \\ 18/20 & \text{si } 15{,}2 \leqslant x < 21{,}3 \\ 20/20 & \text{si } x \geqslant 21{,}3 \end{cases}$$

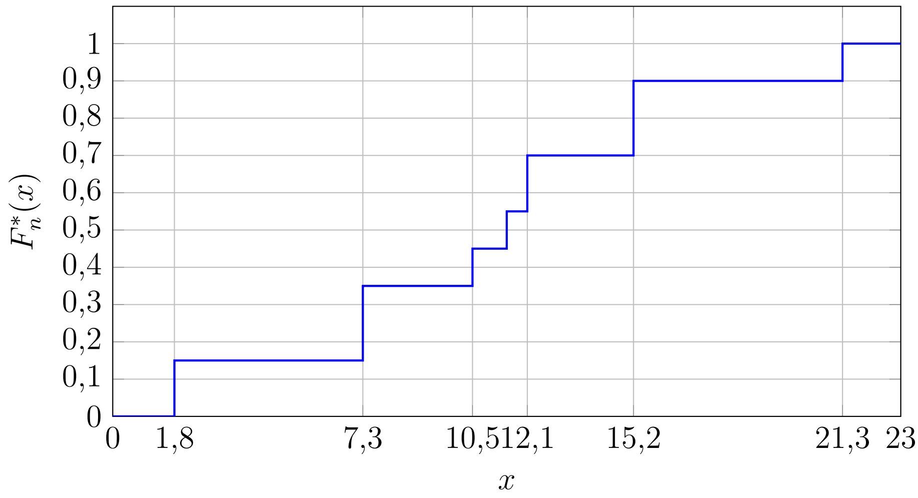

Figura 2.1: Gráfica de la función de distribución muestral.

- Hallar los valores de los tres primeros momentos muestrales respecto al origen y respecto a la media.

  Calculamos primero los tres primeros momentos respecto al origen para luego calcular los centrados respecto a la media a partir de ellos:

$$a_1 = \sum_{i=1}^{n} f_i x_i = 10{,}915 \qquad a_2 = \sum_{i=1}^{n} f_i x_i^2 = 148{,}9325$$

$$a_3 = \sum_{i=1}^{n} f_i x_i^3 = 2280{,}98365$$

$$b_1 = \sum_{i=1}^{n} f_i (x_i - \overline{x}) = 0$$

$$\begin{aligned} b_2 &= \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i^2 - 2x_i\overline{x} + \overline{x}^2) \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{2\overline{x}}{n}\sum_{i=1}^{n} x_i + \overline{x}^2 = a_2 - 2a_1^2 + a_1^2 = a_2 - a_1^2 \\ &= 148{,}9325 - 10{,}915 = 29{,}795275 \end{aligned}$$

$$\begin{aligned} b_3 &= \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^3 = \frac{1}{n}\sum_{i=1}^{n} \left(x_i^3 - 3x_i^2\overline{x} + 3x_i\overline{x}^2 - \overline{x}^3\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i^3 - \frac{3\overline{x}}{n}\sum_{i=1}^{n} x_i^2 + \frac{3\overline{x}^2}{n}\sum_{i=1}^{n} x_i - \overline{x}^3 = a_3 - 3a_1a_2 + 3a_1^3 - a_1^3 \\ &= a_3 - 3a_1a_2 + 2a_1^3 = 4{,}95455925 \end{aligned}$$

- Determinar los valores de los cuartiles muestrales y el percentil 70.

  Para ello, primero ordenamos los datos de menor a mayor y los agrupamos en grupos de $^{20}/_4 = 5$ en 5:

$$1{,}8;\ 1{,}8;\ 1{,}8;\ 7{,}3;\ 7{,}3;\ 7{,}3;\ 7{,}3;\ 10{,}5;\ 10{,}5;\ 11{,}5;\ 11{,}5;\ 12{,}1;\ 12{,}1;$$
$$12{,}1;\ 15{,}2;\ 15{,}2;\ 15{,}2;\ 15{,}2;\ 21{,}3;\ 21{,}3$$

Como en los cambios de agrupaciones de números estos se repiten, hemos obtenido el valor de los cuartiles:

$$q_1 = 7{,}3 \qquad q_2 = 11{,}5 \qquad q_3 = 15{,}2 \qquad q_4 = 21{,}3$$

Para el percentil 70, calculamos:

$$0{,}7 \cdot 20 = 14$$

Como hemos obtenido un número entero, el percentil 70 será:

$$c_{70} = \frac{X_{(14)} + X_{(15)}}{2} = \frac{12{,}1 + 15{,}2}{2} = 13{,}65$$

En el caso de haber obtenido un número no entero (por ejemplo, 14,2), sería $X_{(15)}$.

Ejercicio 2.1.4. Se dispone de una muestra aleatoria simple de tamaño 40 de una distribución exponencial de media 3, ¿cuál es la probabilidad de que los valores de la función de distribución muestral y la teórica, en $x = 1$, difieran menos de 0,01? Aproximadamente, ¿cuál debe ser el tamaño muestral para que dicha probabilidad sea como mínimo 0,98?

Como dice el enunciado, tenemos una m.a.s. $(X_1, \dots, X_n)$ con $n = 40$, todas ellas idénticamente distribuidas a $X \rightsquigarrow exp(\lambda)$. Sabemos de la asignatura de Probabilidad que:

$$E[X] = \frac{1}{\lambda} = 3 \Longrightarrow \lambda = \frac{1}{3}$$

Por lo que $X \rightsquigarrow exp\left(\frac{1}{3}\right)$. Denotaremos por comodidad:

$$F_n^*(x) = F^*_{(X_1,\dots,X_n)}(x)$$

Y el enunciado nos pregunta por:

$$P[|F_n^*(1) - F_X(1)| < 0{,}01]$$

Para ello, primero calculamos $F_X(1)$:

$$F_X(1) = 1 - e^{-\lambda \cdot 1} = 1 - e^{-\lambda} = 1 - e^{-1/3} = \alpha$$

Por lo que nos disponemos ya a calcular la probabilidad:

$$\begin{aligned} P[|F_n^*(1) - F_X(1)| < 0{,}01] &= P[|F_n^*(1) - \alpha| < 0{,}01] = P[-0{,}01 < F_n^*(1) - \alpha < 0{,}01] \\ &= P[-0{,}01 + \alpha < F_n^*(1) < 0{,}01 + \alpha] \\ &= P[40(-0{,}01 + \alpha) < 40F_n^*(1) < 40(0{,}01 + \alpha)] \end{aligned}$$

Y como sabemos que $Y = 40F_n^*(1) \rightsquigarrow B(40, F_X(1)) \equiv B(40, \alpha)$:

$$P[|F_n^*(1) - F_X(1)| < 0{,}01] = P[40(-0{,}01 + \alpha) < Y < 40(0{,}01 + \alpha)]$$

Si ahora tomamos:

$$\alpha = 1 - e^{-1/3} \approx 0{,}283469$$

Entonces:

$$\begin{aligned} 40(0{,}01 + \alpha) &\approx 40(0{,}01 + 0{,}283469) = 11{,}73876 \\ 40(-0{,}01 + \alpha) &\approx 40(-0{,}01 + 0{,}283469) = 10{,}93876 \end{aligned}$$

Por lo que:

$$P[|F_n^*(1) - F_X(1)| < 0{,}01] \approx P[10{,}93876 < Y < 11{,}73876] = P[Y = 11]$$

De donde usando la masa de probabilidad de la Binomial:

$$P[Y = 11] = \binom{40}{11}(0{,}283469)^{11}(1 - 0{,}283469)^{40-11} \approx 0{,}139$$

Para el segundo apartado, como para $n = 40$ obtenemos una probabilidad de 0,139, podemos intuir que para que dicha probabilidad sea como mínimo 0,98, nos es necesario un valor de n grande, por lo que podemos suponer que:

$$F_n^*(1) \rightsquigarrow \mathcal{N}\left(\alpha, \frac{\alpha(1-\alpha)}{n}\right)$$

De donde:

$$Z = \frac{\sqrt{n}(F_n^*(1) - \alpha)}{\sqrt{\alpha(1-\alpha)}} \rightsquigarrow \mathcal{N}(0, 1)$$

Buscamos el valor de n que verifica:

$$0{,}98 \leqslant P[|F_n^*(1) - F_X(1)| < 0{,}01] = P\left[|Z| < \frac{\sqrt{n}0{,}01}{\sqrt{\alpha(1-\alpha)}}\right]$$

Si aplicamos propiedades conocidas de la Normal, si $a \in \mathbb{R}$, entonces:

$$P[|Z| < a] = P[-a < Z < a] = P[Z < a] - P[Z < -a]$$

Pero:

$$P[Z < -a] = P[Z > a] = 1 - P[Z - a]$$

Por lo que:

$$P[|Z| < a] = P[Z < a] - P[Z < -a] = 2P[Z < a] - 1$$

Volviendo al caso que nos interesa:

$$0{,}98 \leqslant P\left[|Z| < \frac{\sqrt{n}0{,}01}{\sqrt{\alpha(1-\alpha)}}\right] = 2P\left[Z < \frac{\sqrt{n}0{,}01}{\sqrt{\alpha(1-\alpha)}}\right] - 1$$

Luego:

$$0{,}99 = \frac{0{,}98+1}{2} \leqslant P\left[Z < \frac{\sqrt{n}0{,}01}{\sqrt{\alpha(1-\alpha)}}\right]$$

Si consultamos la tabla de la normal $\mathcal{N}(0,1)$, observamos que el primer valor que supera la probbailidad de 0,99 es 2,33, por lo que:

$$2{,}33 = \frac{\sqrt{n}0{,}01}{\sqrt{\alpha(1-\alpha)}} = \frac{\sqrt{n}0{,}01}{\sqrt{0{,}283469(1-0{,}283469)}} \approx 0{,}0221886\sqrt{n}$$

De donde:

$$5{,}4289 = (2{,}33)^2 = (0{,}0221886\sqrt{n})^2 = 0{,}00049233n \Longrightarrow n = \frac{5{,}4289}{0{,}00049233} = 11026{,}95347$$

Por lo que para $n \geqslant 11027$ podemos asegurar que la probabilidad es como mínimo 0,98.

Ejercicio 2.1.5. Se dispone de una muestra aleatoria simple de tamaño 50 de una distribución de Poisson de media 2, ¿cuál es la probabilidad de que los valores de la función de distribución muestral y la teórica, en $x = 2$, difieran menos de 0,02? Aproximadamente, ¿qué tamaño muestral hay que tomar para que dicha probabilidad sea como mínimo 0,99?

Tenemos una m.a.s. $(X_1, \ldots, X_n)$ con $n = 50$ idénticamente distribuidas a $X \rightsquigarrow \mathcal{P}(2)$. Notamos por comodidad:

$$F^*_{(X_1,\ldots,X_n)}(x) = F^*_n(x)$$

Nos preguntan por:

$$P[|F^*_n(2) - F_X(2)| < 0{,}02]$$

Para ello primero calculamos:

$$F_X(2) = \sum_{k=0}^{2} e^{-2}\frac{2^k}{k!} = e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!}\right) = e^{-2}(1+2+2) \approx 0{,}6767$$

Por lo que:

$$P[|F^*_n(2) - 0{,}6767| < 0{,}02] = P[-0{,}02 < F^*_n(2) - 0{,}6767 < 0{,}02] = P[0{,}6567 < F^*_n(2) < 0{,}6967]$$

Como sabemos por lo visto en teoría que:

$$Y = 50F^*_n(2) \rightsquigarrow B(50, F_X(2)) \equiv B(50, \quad 0{,}6767)$$

Multiplicamos por 50 la última expresión:

$$\begin{aligned} P[|F^*_n(2) - 0{,}6767| < 0{,}02] = P[0{,}6567 < F^*_n(2) < 0{,}6967] &= P[32{,}835 < Y < 34{,}835] \\ &= P[Y = 33] + P[Y = 34] \end{aligned}$$

Y calculamos estas dos probabilidades:

$$P[Y=33] = \binom{50}{33}(0{,}6767)^{33}(1-0{,}6767)^{50-33} \approx 0{,}114734$$
$$P[Y=34] = \binom{50}{34}(0{,}6767)^{34}(1-0{,}6767)^{50-34} \approx 0{,}120075$$

Por lo que:

$$P[|F_n^*(2) - 0{,}6767| < 0{,}02] \approx 0{,}114734 + 0{,}120075 = 0{,}234809$$

Para el segundo apartado, como para $n = 50$ obtenemos una probabilidad de 0,234809, podemos intuir que para que dicha probabilidad sea como mínimo 0,99, nos es necesario un valor de n grande, por lo que podemos suponer que:

$$F_n^*(2) \rightsquigarrow \mathcal{N}\left(0{,}6767, \frac{0{,}6767(1-0{,}6767)}{n}\right) \equiv \mathcal{N}\left(0{,}6767, \frac{0{,}218777}{n}\right)$$

Por lo que:

$$Z = \frac{\sqrt{n}(F_n^*(2) - 0{,}6767)}{\sqrt{0{,}218777}} \rightsquigarrow \mathcal{N}(0,1)$$

En dicho caso, buscamos n de forma que:

$$0{,}99 \leqslant P[|F_n^*(2) - F_X(2)| < 0{,}02] = P\left[|Z| < \frac{\sqrt{n}0{,}02}{\sqrt{0{,}218777}}\right]$$

De forma análoga al ejercicio anterior:

$$P\left[|Z| < \frac{\sqrt{n}0{,}02}{\sqrt{0{,}218777}}\right] = 2P\left[Z < \frac{\sqrt{n}0{,}02}{\sqrt{0{,}218777}}\right] - 1$$

Luego:

$$0{,}995 = \frac{0{,}99+1}{2} \leqslant P\left[Z < \frac{\sqrt{n}0{,}02}{\sqrt{0{,}218777}}\right]$$

Y si miramos la tabla de la Normal observamos que el primer valor que supera la probabilidad de 0,995 es 2,58, luego:

$$2{,}58 = \frac{\sqrt{n}0{,}02}{\sqrt{0{,}218777}} = 0{,}042759\sqrt{n}$$

Por lo que:

$$6{,}6564 = (2{,}58)^2 = (0{,}042759\sqrt{n})^2 = 0{,}00182833n$$

Luego:

$$n = \frac{6{,}6564}{0{,}00182833} \approx 3640{,}7$$

Por lo que para $n \geqslant 3641$ podemos asegurar que la probabilidad es como mínimo 0,99.

Ejercicio 2.1.6. Sea $X \rightsquigarrow B(1, p)$ y (X_1, X_2, X_3) una muestra aleatoria simple de X. Calcular la función masa de probabilidad de los estadísticos $\overline{X}$, S^2, $\min X_i$ y $\max X_i$.

Para resolver este ejercicio, como X sigue una distribución discreta, buscamos aplicar el teorema de cambio de variable de discreta a discreta. Para ello, la forma más cómoda será analizar cada uno de los valores que puede tomar la muestra aleatoria simple (X_1, X_2, X_3) y determinar en consecuencia cada uno de los valores que toman $\overline{X}$, S^2, $\min X_i$ y $\max X_i$. Acompañaremos la tabla junto con la probabilidad de que la muestra tome dicho valor, es decir, en la fila correpondiente a (x_1, x_2, x_3) incluiremos $P[X_1 = x_1, X_2 = x_2, X_3 = x_3]$:

P	(X_1, X_2, X_3)	$\overline{X}$	S^2	$\min X_i$	$\max X_i$
$(1-p)^3$	$(0, 0, 0)$	0	0	0	0
$p(1-p)^2$	$(0, 0, 1)$	$1/3$	$1/3$	0	1
$p(1-p)^2$	$(0, 1, 0)$	$1/3$	$1/3$	0	1
$p^2(1-p)$	$(0, 1, 1)$	$2/3$	$1/3$	0	1
$p(1-p)^2$	$(1, 0, 0)$	$1/3$	$1/3$	0	1
$p^2(1-p)$	$(1, 0, 1)$	$2/3$	$1/3$	0	1
$p^2(1-p)$	$(1, 1, 0)$	$2/3$	$1/3$	0	1
p^3	$(1, 1, 1)$	1	0	1	1

Podemos ya calcular la función masa de probabilidad de cada uno de los estadísticos, simplemente sumando las probabilidades de la tabla que corresponden a cada valor del espacio muestral de cada estadístico:

- Para $\overline{X}$:

$$\begin{aligned}
P[\overline{X} = 0] &= P[X_1 = 0, X_2 = 0, X_3 = 0] = (1-p)^3 \\
P[\overline{X} = 1/3] &= \sum_{i=1}^{3} p(1-p)^2 = 3p(1-p)^2 \\
P[\overline{X} = 2/3] &= \sum_{i=1}^{3} p^2(1-p) = 3p^2(1-p) \\
P[\overline{X} = 1] &= P[X_1 = 1, X_2 = 1, X_3 = 1] = p^3
\end{aligned}$$

- Para S^2:

$$\begin{aligned}
P[S^2 = 0] &= P[X_1 = 0, X_2 = 0, X_3 = 0] + P[X_1 = 1, X_2 = 1, X_3 = 1] = p^3 + (1-p)^3 \\
P[S^2 = 1/3] &= \sum_{i=1}^{3} p(1-p)^2 + \sum_{i=1}^{3} p^2(1-p) = 3p(1-p)(p + 1 - p) = 3p(1-p)
\end{aligned}$$

- Para $\min X_i$:

$$\begin{aligned}
P[\min X_i = 1] &= P[X_1 = 1, X_2 = 1, X_3 = 1] = p^3 \\
P[\min X_i = 0] &= 1 - P[\min X_i = 1] = 1 - p^3
\end{aligned}$$

- Para $\max X_i$:

$$P[\max X_i = 0] = P[X_1 = 0, X_2 = 0, X_3 = 0] = (1-p)^3$$
$$P[\max X_i = 1] = 1 - P[\max X_i = 0] = 1 - (1-p)^3$$

Ejercicio 2.1.7. Obtener la función masa de probabilidad o función de densidad de $\overline{X}$ en el muestreo de una variable de Bernoulli, de una Poisson y de una exponencial.

Calculamos la masa de probabilidad o función de densidad en cada caso, suponiendo que tenemos $(X_1, \ldots, X_n)$ una muestra aleatoria simple con variables aleatorias idénticamente distribuidas a X, que sigue una distribución distinta en cada caso y estaremos interesados en calcular la masa de:

$$\overline{X} = \sum_{i=1}^{n}$$

Bernoulli. Supuesto que $X \rightsquigarrow B(1,p)$ para cierto $p \in]0,1[$, si tomamos:

$$Y = \sum_{i=1}^{n} X_i$$

Por la propiedad reproductiva de la Bernoulli, tenemos que $Y \rightsquigarrow B(n,p)$. En dicho caso:

$$P[Y = k] = \binom{n}{k} p^k (1-p)^{n-k} \qquad \forall k \in \{0, \ldots, n\}$$

Por tanto, tendremos que:

$$P\left[\overline{X} = \frac{k}{n}\right] = P[Y = k] = \binom{n}{k} p^k (1-p)^{n-k} \qquad \forall k \in \{0, \ldots, n\}$$

Poisson. Supuesto que $X \rightsquigarrow \mathcal{P}(\lambda)$ para cierto $\lambda \in \mathbb{R}^+$, si tomamos:

$$Y = \sum_{i=1}^{n} X_i$$

Por la propiedad reproductiva de la Poisson, tendremos que:

$$Y \rightsquigarrow \mathcal{P}\left(\sum_{i=1}^{n} \lambda\right) \equiv \mathcal{P}(n\lambda)$$

En dicho caso:

$$P[Y = x] = e^{-n\lambda} \frac{(n\lambda)^x}{x!} \qquad \forall x \in \mathbb{N}$$

Por lo que:

$$P[\overline{X} = {}^{x}/_{n}] = P[Y = x] = e^{-n\lambda} \frac{(n\lambda)^x}{x!} \qquad \forall x \in \mathbb{N}$$

Exponencial. Supuesto ahora que $X \rightsquigarrow exp(\lambda)$ para cierto $\lambda \in \mathbb{R}^+$, tendremos entonces que:

$$M_X(t) = \frac{\lambda}{\lambda - t} \qquad t < \lambda$$

Si aplicamos la igualdad $(*)$ vista en teoría:

$$M_{\overline{X}}(t) \overset{(*)}{=} (M_X(t/n))^n = \left(\frac{\lambda}{\lambda - t/n}\right)^n = \left(\frac{n\lambda}{n\lambda - t}\right)^n$$

Observamos que obtenemos una función generatriz de momentos para $\overline{X}$ igual que para una variable aleatoria de distribución $\Gamma(n, n\lambda)$. Como la función generatriz de momentos de una variable aleatoria caracteriza su distribución, concluimos que $\overline{X} \rightsquigarrow \Gamma(n, n\lambda)$.

Ejercicio 2.1.8. Calcular las funciones de densidad de los estadísticos $\text{máx}\, X_i$ y $\text{mín}\, X_i$ en el muestreo de una variable X con funcion de densidad:

$$f_\theta(x) = e^{\theta - x}, \qquad x > \theta.$$

Calculamos primero la función de distribución, para calcular con mayor comodidad las funciones de distribución de $X_{(n)}$ y $X_{(1)}$:

$$F_\theta(x) = \int_\theta^x f_\theta(t)\, dt = \int_\theta^x e^{\theta - t}\, dt = \left[-e^{\theta - t}\right]_\theta^x = 1 - e^{\theta - x} \qquad \forall x \in \mathbb{R}^+$$

Supuesto ahora que disponemos de una m.a.s. $(X_1, \ldots, X_n)$ idénticamente distribuidas a X cuya función de densidad es la anteriormente dicha, podemos aplicar las fórmulas obtenidas en teoría para calcular las funciones de distribución del mínimo y del máximo. Para el máximo:

$$F_{X_{(n)}}(x) = (F_X(x))^n = (1 - e^{\theta - x})^n \Longrightarrow f_{X_{(n)}} = n(1 - e^{\theta - x})^{n-1} e^{\theta - x} \qquad \forall x \in \mathbb{R}^+$$

Para el mínimo:

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(n))^n = 1 - (1 - 1 + e^{\theta - x})^n = 1 - e^{n(\theta - x)} \qquad \forall x \in \mathbb{R}^+$$

de donde:

$$f_{X_{(1)}}(x) = ne^{n(\theta - x) - 1} \qquad \forall x \in \mathbb{R}^+$$

Ejercicio 2.1.9. El número de pacientes que visitan diariamente una determinada consulta médica es una variable aleatoria con varianza de 16 personas. Se supone que el número de visitas de cada día es independiente de cualquier otro. Si se observa el número de visitas diarias durante 64 días, calcular aproximadamente la probabilidad de que la media muestral no difiera en más de una persona del valor medio verdadero de visitas diarias.

Sea X una variable aleatoria que indica el número de pacientes que visitan diariamente dicha consulta médica, por cómo nos definen X sabemos que $X \rightsquigarrow \mathcal{P}(\lambda)$. Como además nos dicen que la varianza de dicha variable aleatoria es 16, tenemos que $Var(X) = \lambda = 16$. Si tenemos ahora una muestra aleatoria simple $(X_1, \dots, X_n)$ con $n = 64$, nos preguntan por:

$$P[|\overline{X} - E[X]| < 1]$$

Donde $E[X] = \lambda = 16$, ya que $X \rightsquigarrow \mathcal{P}(16)$. Calculamos:

$$P[|\overline{X} - E[X]| < 1] = P[-1 < \overline{X} - 16 < 1] = P[15 < \overline{X} < 17]$$

Aplicamos ahora lo visto en el ejercicio 7, ya que si $X \rightsquigarrow \mathcal{P}(\lambda)$, entonces tendremos que $n\overline{X} \rightsquigarrow \mathcal{P}(n\lambda)$, gracias a la propiedad reproductiva de la Poisson:

$$\begin{aligned} P[|\overline{X} - E[X]| < 1] &= P[15 < \overline{X} < 17] = P[64 \cdot 15 < 64\overline{X} < 64 \cdot 17] \\ &= P[960 < 64\overline{X} < 1088] \end{aligned}$$

Donde $64\overline{X} \rightsquigarrow \mathcal{P}(64 \cdot 16) \equiv \mathcal{P}(1024)$. Para calcular dicha probabilidad, aproximaremos la Poisson a una distribución normal:

$$\mathcal{P}(1024) \approx \mathcal{N}(1024, 1024)$$

Por lo que:

$$\begin{aligned} P[|\overline{X} - E[X]| < 1] = P[960 < 64\overline{X} < 1088] &\approx P\left[\frac{960 - 1024}{\sqrt{1024}} < Z < \frac{1088 - 1024}{\sqrt{1024}}\right] \\ &= P[-2 < Z < 2] = 2P[Z < 2] - 1 \\ &= 2 \cdot 0{,}97725 - 1 = 0{,}9545 \end{aligned}$$

Ejercicio 2.1.10. Una máquina de refrescos está arreglada para que la cantidad de bebida que sirve sea una variable aleatoria con media 200 ml. y desviación típica 15 ml. Calcular de forma aproximada la probabilidad de que la cantidad media servida en una muestra aleatoria de tamaño 36 sea al menos 204 ml.