Inferencia Estadística - Guia 1

Nicolas Ferrer
e-mail: nicolas.ferrer.747@gmail.com

Junio 2021

2 Ejercicio 2

2.1 Inciso a (p. 111 C&B)

Recordemos que para una **familia exponencial** de m parámetros $\{f(x;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m\}$:

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{m} w_i(\boldsymbol{\theta})t_i(x)\right) \tag{1}$$

Donde:

- $h(x) \geq 0$ y $t_1(x), \ldots, t_m(x)$ son funciones valuadas en los reales que *no* dependen de $\boldsymbol{\theta}$.
- $c(\boldsymbol{\theta}) \geq 0$ y $w_1(\boldsymbol{\theta}), \ldots, w_m(\boldsymbol{\theta})$ son funciones valuadas en los reales que *no* dependen de x.

Dado que $c(\boldsymbol{\theta}) \geq 0$ y $\exp(\cdot)$ siempre evalúa a un valor positivo, el conjunto de valores de x para los cuales $f(x|\boldsymbol{\theta})$ es positivo no depende de $\boldsymbol{\theta}$. Recordar que la definición de una familia exponencial debe incluir la especificación de Θ, el conjunto de valores posibles de $\boldsymbol{\theta}$. Por lo tanto, si Θ incluye valores de $\boldsymbol{\theta}$ para los cuales $c(\boldsymbol{\theta}) = 0$, estaríamos ante un modelo que no pertenece a una familia exponencial. Podemos ver un ejemplo de una familia exponencial satisfaciendo esta condición en el punto 3.1.

2.2 Inciso b

A partir de este punto haremos uso frecuente del logaritmo natural, la función exponencial y sus propiedades. Algunas de las más importantes son:

- $x = \ln e^x = e^{\ln x}$
- $\ln(a * b) = \ln a + \ln b$
- $\ln a^b = b * \ln a$

Supongamos sin pérdida de generalidad que tenemos una familia exponencial de un único parámetro tal que:

$$f(x|\theta) = h(x)c(\theta) \exp\big(w(\theta)t(x)\big)$$

Entonces, tomando logaritmo natural de ambos lados y utilizando sus propiedades:

$$\begin{aligned}
\ln f(x|\theta) &= \ln\Big(h(x)c(\theta) \exp\big(w(\theta)t(x)\big)\Big) \\
\ln f(x|\theta) &= \ln h(x) + \ln c(\theta) + \ln\Big(\exp\big(w(\theta)t(x)\big)\Big) \\
\ln f(x|\theta) &= \underbrace{\ln h(x)}_{=m(x)} + \underbrace{\ln c(\theta)}_{=-d(\theta)} + \underbrace{w(\theta)t(x)}_{a(x)b(\theta)} \\
\ln f(x|\theta) &= a(x)b(\theta) + m(x) - d(\theta)
\end{aligned}$$

Exponenciando de ambos lados y reordenando términos:

$$f(x|\theta) = \exp\big(a(x)b(\theta) + m(x) - d(\theta)\big) \tag{2}$$

3 Ejercicio 3

3.1 Inciso a: Poisson

Operando y aplicando propiedades de logaritmo natural:

$$\begin{aligned}
f(x,\lambda) &= \frac{e^{-\lambda}\lambda^x}{x!} \\
\ln f(x,\lambda) &= -\lambda + \ln\frac{1}{x!} + \ln\lambda^x \\
\ln f(x,\lambda) &= -\lambda - \ln x! + x\ln\lambda \\
f(x,\lambda) &= \exp\big(x\ln\lambda - \lambda - \ln x!\big)
\end{aligned}$$

Esta última expresión tiene la forma de la ecuación (2) del inciso 2.b donde: $a(x) = x$, $b(\lambda) = \ln\lambda$, $c(\lambda) = -\lambda$, $d(x) = -\ln x!$, con lo cual comprobamos que un modelo Poisson pertenece a la familia exponencial.

Una manera mas directa de llegar a una solución es notar que:

$$\lambda^x = \exp(x\ln\lambda)$$

Por lo cual:

$$f(x,\lambda) = \frac{1}{x!}e^{-\lambda}\exp\left(x\ln\lambda\right)$$

Tiene la forma de familia exponencial con $h(x) = x!^{-1}$, $c(\lambda) = e^{-\lambda}$, $w(\lambda) = \ln\lambda$, $t(x) = x$.

Volviendo al punto 1.a, puede verificarse que, para cualquier valor de $\lambda \in (0,\infty)$ $f(x,\lambda) > 0$ para todo $x \geq 0$.[1]

3.2 Inciso b: Exponencial

Aplicamos una herramienta que aparecerá en forma frecuente a lo largo de la materia: la función indicadora $\mathbb{1}(x)$. En este caso, la definimos como:

$$\mathbb{1}(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Notar que $\mathbb{1}(x)$ es sólo función de x, con lo cual puede cumplir el rol de $h(x)$. Entonces, dada la definición provista para el modelo exponencial, podemos reescribir su función de densidad como:

$$f(x,\lambda) = \mathbb{1}(x)\frac{1}{\lambda}\exp\left(-\frac{1}{\lambda}x\right)$$

La cual tiene forma de familia exponencial con $h(x) = \mathbb{1}(x)$, $c(\lambda) = \lambda^{-1}$, $w(\lambda) = -\lambda^{-1}$, $t(x) = x$.

3.3 Inciso c: Truncada en θ

Este modelo estadístico no pertenece a una familia exponencial. θ establece una cota inferior al conjunto de valores de x para los cuales la función está definida, y por lo tanto, el soporte de la función $f(x,\theta)$ depende de θ.

[1] Recordar que $0! = 1$.

3.4 Inciso d: Laplace

$$f_X(x;\mu,\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x-\mu|}{\sigma}\right)$$

La distribución de Laplace sólo puede expresarse como una familia exponencial si el parámetro de locación $\mu \in \mathbb{R}$ es conocido. En este caso, estaríamos hablando de un modelo estadístico con un único parámetro de escala $\theta = \sigma$. De ser así, resulta fácil ver que:

- $h(x) = 1$
- $c(\sigma) = (2\sigma)^{-1}$
- $t(x) = |x-\mu|$
- $w(\sigma) = -\sigma^{-1}$

Por lo tanto, la distribución de Laplace con parámetros desconocidos $\Theta = \{\mu, \sigma\}$ no pertenece a una familia exponencial.

3.5 Inciso e: Loc-escala Cauchy

$$f(x;\mu,\sigma) = \frac{1}{\pi}\left[\frac{\sigma}{(x-\mu)^2+\sigma^2}\right]$$

Al intentar factorizar esta expresión aplicando logaritmo de ambos lados y luego exponenciando nos encontramos ante un problema:

$$\begin{aligned}\ln f(x;\mu,\sigma) &= -\ln\pi + \ln\sigma - \ln\left((x-\mu)^2+\sigma^2\right)\\ f(x;\mu,\sigma) &= \frac{1}{\pi} + \sigma + \exp\left(\frac{1}{(x-\mu)^2+\sigma^2}\right)\end{aligned}$$

La expresión dentro de la función exponencial del último término no puede factorizarse en una expresión de la forma $\sum_i^m w_i(\boldsymbol{\theta})t_i(x)$.

3.6 Inciso f: Gamma

Notar que un modelo estadístico Gamma posee dos parámetros, por lo cual tenemos $\boldsymbol{\theta} = \{\lambda, k\}$.

$$\begin{aligned}f(x;\lambda,k) &= \lambda e^{-\lambda x}\frac{(\lambda x)^{k-1}}{\Gamma(k)}\\ f(x;\lambda,k) &= \frac{\lambda^k}{\Gamma(k)}\frac{x^k}{x}e^{-\lambda x}\end{aligned}$$

Dado que $x^k = \exp(k\ln x)$. Por lo tanto:

$$f(x;\lambda,k) = \frac{\lambda^k}{\Gamma(k)}\frac{1}{x}\exp\left(k\ln x - \lambda x\right)$$

Esta expresión tiene la forma de una familia exponencial de $m = 2$ parámetros (ver ecuación (1)) con:

- $h(x) = x^{-1}$
- $c(\boldsymbol{\theta}) = \frac{\lambda^k}{\Gamma(k)}$
- $w_1(\boldsymbol{\theta}) = k$, $t_1 = \ln x$
- $w_2(\boldsymbol{\theta}) = -\lambda$, $t_2 = x$

3.7 Inciso g: Beta

Al igual que en el caso del modelo Gamma, tenemos dos parámetros, con lo cual $\boldsymbol{\theta} = \{\beta, \alpha\}$.

$$f(x; \beta, \alpha) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Una vez más, usamos que:

$$x^{\alpha} = \exp\big(\alpha \ln x\big)$$
$$(1-x)^{\beta-1} = \exp\big((\beta-1)\ln(1-x)\big)$$

Por lo tanto:

$$f(x; \beta, \alpha) = \frac{1}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \exp\big(\alpha \ln x + (\beta-1)\ln(1-x)\big)$$

Esta expresión tiene la forma de una familia exponencial de $m = 2$ parámetros (ver ecuación (1)) con:

- $h(x) = x^{-1}$
- $c(\boldsymbol{\theta}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$
- $w_1(\boldsymbol{\theta}) = \alpha$, $t_1 = \ln x$
- $w_2(\boldsymbol{\theta}) = (\beta - 1)$, $t_2 = \ln(1-x)$

4 Ejercicio 4

Por la definición de familia de locación-escala vista en clase[2] (y la ayuda provista), sabemos que:

$$F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right)$$

Esto vale para cualquier valor de x. Sabemos por la definición de x_p que:

$$F_X(x_p) = F_Z\left(\frac{x_p-\mu}{\sigma}\right) = p$$

Pero $F_Z(z_p) = p$, con lo cual:

$$\frac{x_p - \mu}{\sigma} = z_p$$
$$x_p = \sigma z_p + \mu$$

5 Ejercicio 5

Considerar a las $X = x$ como *fijas* implica que el investigador establece un valor de referencia para dichas variables (pensar en cualquier característica de los individuos bajo estudio, edad, peso, nacionalidad, etc.).

Dado que dichas variables están *fijas*, $h(X)$ sólo desplaza la media de la distribución de Y respecto a la de ε, de la misma manera que β_0. Por lo tanto, tenemos un modelo de locación-escala con $\mu = \beta_0 + h(X)$.

$$Y = \beta_0 + h(X) + \sigma\varepsilon \quad \text{con } \varepsilon \sim N(0,1) \Rightarrow \frac{Y - \beta_0 - h(X)}{\sigma} \sim N(0,1)$$

En el caso que asumimos que $h(X) = \beta X$, es decir que existe una relación *lineal* entre Y y X, estaremos ante el modelo de *regresión lineal.*

[2]Para una demostración de la existencia de una variable aleatoria $Z \sim f(z)$ tal que $X = \sigma Z + \mu$, ver el teorema 3.5.6 en la página 120 del libro de Casella y Berger.

6 Ejercicio 6

6.1 Inciso a

Sea $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ la serie variables aleatorias que representa la condición de empleo de los n habitantes de un país, podemos representar cada una de dichas variables aleatorias como un modelo Bernoulli tal que:

$$X_i \sim Ber(p)$$

Dicho modelo posee un parámetro p que representa la probabilidad de "éxito" (en nuestro caso, desempleo) en la población, y el soporte de X_i se define como:

$$\text{Soporte}(Ber(x; p)) = \{0, 1\}$$

Donde $X_i = 0$ representa un individuo empleado y $X_i = 1$ un individuo desempleado[3].

6.2 Inciso b

El estadístico $T = \sum_i^{1000} X_i/1000$ representa la proporción de desempleados ("éxitos") en la muestra.

6.3 Inciso c

Sea Z el número de desempleados en nuestra muestra de n personas, podemos pensar a Z como el número de éxitos de una secuencia de n experimentos Bernoulli. Z entonces sigue una distribución Binomial con parámetros (n, p), donde n es el número de experimentos y p la probabilidad de "éxito".

En nuestro caso particular, tenemos $n = 1000$ y $p = 0.072$. La esperanza de una variable binomial está dada por:

$$E(Z) = np = 72$$

6.4 Inciso d

Queremos encontrar $\text{Var}(T) = \text{Var}\left(\sum_{i=1}^{n} X_i/n\right)$. Utilizando que las variables aleatorias $\{X_i\}_{i=1}^{n}$ son independientemente distribuidas:

$$\text{Var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) = \frac{\text{Var}(X)}{n}$$

Dado que $X_i \sim Ber(p) \Rightarrow \text{Var}(X_i) = p(1-p)$. Entonces,

$$\text{Var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{n}{n^2}p(1-p) = \frac{p(1-p)}{n}$$

Entonces $\text{Var}(T) = p(1-p)/n$. Para $n = 1000, p = 0.072$:

$$\text{Var}(T) = 0.000066816.$$
$$\text{Desv. Est}(T) \approx 0.00817$$

[3] Podríamos definir la variable de manera inversa y que p represente la probabilidad de empleo.

6.5 Inciso e

Queremos calcular $P(Z=k)$ con $Z \sim B(n,p)$:

$$P(Z=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Reemplazando $k = 60, n = 1000$:

$$P(Z=60) = \binom{1000}{60} p^{60} (1-p)^{940}$$

Se puede ver que el máximo de esta expresión se alcanza para $p = 0.06$, es decir $p = k/n$. Cualquier valor más alto o más bajo reduce la probabilidad de ocurrencia del evento $Z = 60$. Si tomamos $p = 0.072 \Rightarrow P(Z = 60) = 0.017$. Una herramienta útil para realizar este tipo de cuentas es stattrek.com

6.6 Inciso f

Para una muestra de 1000 personas, el hecho de que haya al menos 40 desempleados es equivalente a $T = \bar{X} = 0.04$. Por lo tanto, queremos usar el TLC para aproximar $P(T \geq 0.04)$.

Por TCL:

$$T^* = \left(\frac{T - \mu_X}{\sigma_X / \sqrt{n}} \right) \xrightarrow{D} N(0,1)$$

Entonces, queremos conocer la probabilidad de que T^* (una variable normal estándar) sea mayor a $\frac{0.04 - \mu_X}{\sigma_X / \sqrt{n}}$.
Para $X \sim Ber(0.072)$:

$$\mu_X = p = 0.072 \qquad \sigma_X = \sqrt{p(1-p)} \approx 0.258$$

Entonces, si $n = 1000$:

$$\frac{0.04 - \mu_X}{\sigma_X / \sqrt{n}} \approx -\frac{0.032}{0.00815} \approx -3.92$$

Usando una tabla normal, calculamos que:

$$P(T^* \geq -3.92) = 1 - P(T^* \leq -3.92) \approx 1$$

Notar que la probabilidad de que la proporción de desempleados sea mayor al 4% crece a medida que aumenta el tamaño de la muestra (partiendo del supuesto de que la proporción en la población es 7.2%).

7 Ejercicio 7

7.1 Inciso a

Dado nuestro conocimiento empírico sobre la distribución del ingreso, querríamos elegir algún modelo paramétrico que asigne mayor densidad a niveles bajos y medios de ingreso. Algunas opciones razonables serían los modelos Exponencial, Log-Normal o Pareto. Por su simplicidad trabajaremos con el modelo exponencial, el cual posee un único parámetro, λ.

7.2 Inciso b

La elección del modelo exponencial trae una ventaja particular: se trata de una familia exponencial determinada por un único parámetro: λ.

De lo visto en clase, sabemos que para cualquier familia exponencial el estadístico:

$$T(\mathbf{X}) = \sum_{i=1}^{n} t(X_i)$$

Posee varias propiedades atractivas:

- Es *suficiente*: resume toda la información relevante en la muestra para estimar θ.
- Es *minimal*: representa la manera mas "compacta" o "eficiente" de representar la información sobre θ en la muestra.
- Es *completo*: lo cual posee relevancia para varios teoremas que veremos más adelante.

Por lo visto en el inciso 3.2, sabemos que para el modelo Exponencial $t(x) = x$ es un estadístico completo y minimal suficiente para λ. Por lo tanto, podríamos proponer

$$\bar{X} = n^{-1} \sum_{i=1}^{n} x_i = n^{-1} T(\mathbf{X})$$

Como estimador de λ. Este estimador hereda las propiedades de $T(\mathbf{X})$ y posee la ventaja de ser más fácil de interpretar, al representar al ingreso medio de los hogares de la muestra.

7.3 Inciso c

Llamemos Q_1, Q_2 y Q_3 a los cuartiles 1, 2 y 3, los cuales están definidos como:

$$\begin{aligned} Q_1 &= \{x : F(X \leq x) = 0.25\} \\ Q_2 &= \{x : F(X \leq x) = 0.50\} \\ Q_3 &= \{x : F(X \leq x) = 0.75\} \end{aligned}$$

Utilizando que para el modelo exponencial que $F(X \leq x) = 1 - e^{-\frac{x}{\lambda}}$ y nuestra estimación $\hat{\lambda} = \bar{X} = 35$ tenemos:

$$\begin{aligned} Q_1 &\approx 10.06 \\ Q_2 &\approx 24.26 \\ Q_3 &\approx 48.52 \end{aligned}$$

8 Ejercicio 8

8.1 Inciso a

$$\begin{aligned} E(\bar{X}) &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] \\ &= \frac{n}{n}\mu = \mu \end{aligned}$$

8.2 Inciso b

$$
\begin{aligned}
V(\bar{X}) &= V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left(X_i, X_j\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \text{ dado que si } X_i \sim i.i.d. \Rightarrow Cov(X_i, X_j) = 0 \ \forall \ i \neq j \\
&= \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}
$$

8.3 Inciso c

$$
\begin{aligned}
E(S^2) &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right] \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} E[X_i^2] - 2E\left[\bar{X}\sum_{i=1}^{n} X_i\right] + \sum_{i=1}^{n} E[\bar{X}^2]\right) \\
&= \frac{1}{n-1}\left(nE[X_i^2] - 2nE\left[\bar{X}^2\right] + nE[\bar{X}^2]\right) \\
&= \frac{1}{n-1}\left(nE[X_i^2] - nE\left[\bar{X}^2\right]\right)
\end{aligned}
$$

Usando la definición de varianza: $\mathrm{Var}(X) = E(X^2) - E(X)^2$:

$$
\begin{aligned}
E(S^2) &= \frac{n}{n-1}\left(E[X_i^2] - E\left[\bar{X}^2\right]\right) \\
&= \frac{n}{n-1}\left[(\mathrm{Var}(X) + E(X)^2) - (\mathrm{Var}(\bar{X}) + E(\bar{X})^2)\right] \\
&= \frac{n}{n-1}\left[(\sigma^2 + \mu^2) - (\mathrm{Var}(\bar{X}) + \mu^2)\right]
\end{aligned}
$$

Usando que $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$:

$$
\begin{aligned}
E(S^2) &= \frac{n}{n-1}\left[\sigma^2 - \frac{\sigma^2}{n}\right] \\
E(S^2) &= \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2
\end{aligned}
$$

9 Ejercicio 9

9.1 Inciso a

Empecemos calculando $E(T)$:

$$E(T) = E\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} E[Y_i] = nE[Y_i]$$

Para obtener $E(Y_i)$ utilizamos:

$$\begin{aligned}
E[Y_i] &= 1 * P(Y_i = 1) + 0 * P(Y_i = 0) \\
E[Y_i] &= 1 * P(X \geq \mu) + 0 * P(X_i < \mu) \\
E[Y_i] &= P(X \geq \mu) = 0.5 \Rightarrow \boxed{E(T) = 0.5n}
\end{aligned}$$

Ahora, para calcular $\text{Var}(T)$, utilizando que las Y_i son independientes:

$$\text{Var}(T) = \text{Var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \text{Var}(Y_i) = n\text{Var}(Y_i)$$

Para obtener $\text{Var}(Y_i)$ usamos:

$$\text{Var}(Y_i) = E(Y_i^2) - E(Y_i)^2$$

Ya tenemos $E(Y_i) = 0.5$. Para $E(Y_i^2)$ utilizamos la misma definición de antes:

$$\begin{aligned}
E[Y_i^2] &= 1^2 * P(Y_i = 1) + 0^2 * P(Y_i = 0) \\
E[Y_i^2] &= P(X \geq \mu) = 0.5
\end{aligned}$$

Entonces,

$$\text{Var}(Y_i) = 0.5 - 0.5^2 = 0.25 \Rightarrow \boxed{\text{Var}(T) = 0.25n}$$

9.2 Inciso b

Queremos obtener la función de densidad[4] asociada a $T = k$ para algún número $k \in [0, n]$.

Notar que para obtener $T = k$ en una secuencia de variables aleatorias $\{Y_i\}_{i=1}^{n}$, requiero observar $(Y_i = 1)$ k veces e $(Y_i = 0)$ $n - k$ veces. Por ejemplo, si las primeras k observaciones son iguales a 1 y las $n - k$ restantes son iguales a cero, tenemos:

$$f(\{Y_i\}_{i=1}^{k} = 1; \{Y_i\}_{i=k+1}^{n} = 0) = f(Y_i = 1)^k f(Y_i = 0)^{n-k}$$

Pero existen $\binom{n}{k}$ maneras de combinar dichas variables aleatorias de manera que el total de éxitos sea igual a k. Por lo tanto, la función de densidad asociada a $T = k$ es:

$$f(T = k) = \binom{n}{k} f(Y_i = 1)^k f(Y_i = 0)^{n-k}$$

Por la definición de Y_i sabemos que $f(Y_i = 1) = P(X \geq \mu) = 0.5$, por lo tanto:

$$f(T = k) = \binom{n}{k} 0.5^k 0.5^{n-k} = \binom{n}{k} 0.5^n$$

Alternativamente, podríamos haber notado que Y_i es una variable Bernoulli y T es una suma de éxitos de variables Bernoulli, con lo cual se distribuye en forma Binomial.

[4]Dado que T es una variable discreta, esto es lo mismo que decir la función de probabilidad.

10 Ejercicio 10

10.1 Inciso a

Sabemos que para una variable $X \sim U(0, 10)$:

$$\mu_X = \frac{a+b}{2} = 5$$
$$\sigma_X = \sqrt{\frac{(b-a)^2}{12}} = \frac{5}{\sqrt{3}}$$

Por el TLC, sabemos que $\sqrt{n}(\bar{X} - \mu_X)/\sigma_X \xrightarrow{D} N(0, 1)$. Sea Z una variable normal estándar, aproximamos usando:

$$P(4.5 \leq \bar{X} \leq 5.5) \approx P\left(\frac{4.5 - \mu_X}{\sigma_X/\sqrt{n}} \leq Z \leq \frac{5.5 - \mu_X}{\sigma_X/\sqrt{n}}\right)$$

Reemplazando μ_X, σ_X:

$$P(4.5 \leq \bar{X} \leq 5.5) \approx P\left(-0.1\sqrt{3n} \leq Z \leq 0.1\sqrt{3n}\right) = 2\Phi(0.1\sqrt{3n}) - 1$$

Donde $\Phi(\cdot)$ es la función de distribución normal estándar. La última igualdad vale por simetría de la distribución normal.

Para la segunda probabilidad:

$$P(|\bar{X} - 5| > 1) = P(\bar{X} > 6) + P(\bar{X} < 4)$$

Por simetría de la distribución normal, basta con calcular sólo uno de los dos términos. Dado que es una distribución continua. Tomemos el segundo.

$$P(\bar{X} < 4) = P(X > 6) \approx \Phi\left(\frac{4-5}{5/\sqrt{3n}}\right) = \Phi\left(\frac{-\sqrt{3n}}{5}\right)$$

Por lo tanto:

$$P(|\bar{X} - 5| > 1) \approx 2\Phi\left(\frac{-\sqrt{3n}}{5}\right)$$

10.2 Inciso b y c

Para $P(4.5 \leq \bar{X} \leq 5.5)$ vemos que:

$$P(4.5 \leq \bar{X} \leq 5.5) \approx (2\Phi(0.1\sqrt{3n}) - 1)$$

La expresión del lado derecho es creciente en n y tiende a 1 a medida que $n \to \infty$. Es decir, la probabilidad de que la media muestral se encuentre dentro del intervalo que contiene a la media poblacional es creciente en el tamaño de la muestra y tiende a 1 a medida que esta se va a infinito.

Para $P(|\bar{X} - 5| > 1)$ vemos que:

$$P(|\bar{X} - 5| > 1) \approx 2\Phi\left(\frac{-\sqrt{3n}}{5}\right)$$

La expresión del lado derecho es decreciente en n y tiende a 0 a medida que $n \to \infty$. Es decir, la probabilidad de que la distancia entre la media muestral y la media poblacional sea mayor a uno decrece a medida que crece el tamaño de muestra, y se vuelve cero a medida que esta se va a infinito.

11 Ejercicio 11

11.1 Inciso a

Para calcular los la esperanza y varianza de $X_{(1)}$, primero tenemos que calcular su función de densidad. Para ello, usamos que $f_{X_{(1)}} = F'_{X_{(1)}}$.

$$\begin{aligned}
F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) \\
&= 1 - P(X_{(1)} \geq x) \\
&= 1 - \prod_{i=1}^{n} P(X_i \geq x) \\
&= 1 - P(X \geq x)^n \\
&= 1 - [1 - P(X \leq x)]^n \\
F_{X_{(1)}}(x) &= 1 - (1-x)^n
\end{aligned}$$

Entonces:

$$f_{X_{(1)}}(x) = \frac{\partial F_{X(1)}(x)}{\partial x} = n(1-x)^{n-1}$$

Para calcular $E(X_{(1)})$:

$$E(X_{(1)}) = \int_0^1 x f_{X_{(1)}}(x)dx = \int_0^1 xn(1-x)^{n-1}dx$$

Integrando por partes, obtenemos:

$$\boxed{E(X_{(1)}) = \frac{1}{n+1}}$$

Para obtener la varianza del mínimo, usamos $V(X_{(1)}) = E(X_{(1)}^2) - E(X_{(1)})^2$. Para calcular $E(X_{(1)}^2)$:

$$E(X_{(1)}^2) = \int_0^1 x^2 f_{X_{(1)}}(x)dx = \int_0^1 x^2 n(1-x)^{n-1}dx$$

Una vez más, integrando por partes:

$$E(X_{(1)}^2) = \frac{2}{(n+1)(n+2)}$$

Entonces, obtenemos:

$$\boxed{V(X_{(1)}) = \frac{n}{(n+1)^2(n+2)}}$$

11.2 Inciso b

Repetimos los pasos anteriores para el estadístico de orden $X_{(n)}$.

$$\begin{aligned}
F_{X_{(n)}}(x) &= P(X_{(n)} \leq x) \\
&= \prod_{i=1}^{n} P(X_i \leq x) \\
&= P(X \leq x)^n \\
F_{X_{(n)}}(x) &= x^n \\
\Rightarrow f_{X_{(n)}}(x) &= nx^{n-1}
\end{aligned}$$

Entonces para $E(X_{(n)})$:

$$\boxed{E(X_{(n)}) = \int_0^1 nx^n dx = \frac{n}{n+1}}$$

Para $V(X_{(n)}) = E(X_{(n)}^2) - E(X_{(n)})^2$:

$$E(X_{(n)}^2) = \int_0^1 nx^{n+1} dx = \frac{n}{n+2}$$
$$V(X_{(n)}^2) = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 \Rightarrow \boxed{V(X_{(n)}^2) = \frac{n}{(n+1)^2(n+2)}}$$

11.3 Inciso c

En caso de que $X \sim \text{Exp}(\theta)$, tenemos que $F_X(x) = 1 - e^{-\theta x}$, con lo cual las funciones de distribución y densidad de $X_{(1)}$ y $X_{(n)}$ se vuelven:

$$F_{X_{(1)}} = 1 - e^{-n\theta x} \qquad f_{X_{(1)}} = n\theta e^{-n\theta x}$$
$$F_{X_{(n)}} = (1 - e^{-\theta x})^n \qquad f_{X_{(n)}} = n\theta(1 - e^{-\theta x})^{n-1} e^{-\theta x}$$

Notar que $X_{(1)}$ se distribuye exponencial con parámetro $n\theta$.

12 Ejercicio 12

12.1 Inciso a

Podemos realizar aproximaciones de la esperanza y varianza de $g(\bar{X}) = \bar{X}^2$ usando el método delta:

$$E(g(T)) \approx g(\theta)$$
$$V(g(T)) \approx \left(g^{(1)}(\theta)\right)^2 V(T)$$

Para poder aplicar esta aproximación de primer orden, requerimos que $g^{(1)}(\theta) = 2\mu \neq 0$. De ser este el caso, tenemos:

$$E(\bar{X}^2) \approx \mu^2$$
$$V(\bar{X}^2) \approx \frac{(2\mu\sigma)^2}{n}$$

12.2 Inciso b

Sabemos por TLC que $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$. Entonces por método delta y lo visto en el anterior inciso:

$$\sqrt{n}(\bar{X}^2 - \mu^2) \xrightarrow{D} N\left(0, (2\mu\sigma)^2\right)$$

12.3 Inciso c

Si $\mu = 0$, entonces $g^{(1)}(\theta) = 2\mu = 0$, con lo cual no podemos aplicar la aproximación de primer orden utilizada en el inciso anterior. No obstante, sí tenemos que $g^{(2)}(\theta) = 2 \neq 0$, por lo cual podemos utilizar una aproximación de segundo orden[5] tal que:

$$\sqrt{n}\bar{X}^2 \xrightarrow{D} \sigma^2 \chi^2$$

[5] Ver página 41 de los slides de la primera clase.

12.4 Inciso d

Para $h(\bar{X})$ tenemos:

$$h(\theta) = e^{\mu}$$
$$g(\theta)^{(1)} = e^{\mu}$$

Con lo cual:

$$\sqrt{n}(e^{\bar{X}} - e^{\mu}) \xrightarrow{D} N(0, e^{2\mu}\sigma^2)$$

12.5 Inciso e

Sabemos que $T = \bar{X}$ es un estadístico insesgado para $\theta = E(X) = \mu$. Sea $V(X) = \sigma^2$, sabemos que por el ejercicio 6.d que:

$$V(\bar{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n} < \infty$$

13 Ejercicio 13

13.1 Inciso a

El Teorema de Factorización de Fisher-Neyman nos dice que $T(X)$ será un estadístico suficiente para θ si podemos factorizar la función de densidad $f_X(x;\theta)$ de la siguiente manera:

$$f_{\mathbf{X}}(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)h(\mathbf{x})$$

Entonces, tenemos:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\theta) &= \prod_{i=1}^{n} \theta^{-1} \exp\left(-\frac{x_i}{\theta}\right) \\ &= \theta^{-n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right) \end{aligned}$$

Entonces, podemos ver que para $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$, tenemos:

$$g(T(\mathbf{x});\theta) = \theta^{-n} \exp\left(-\frac{T(\mathbf{x})}{\theta}\right) \qquad h(\mathbf{x}) = 1$$

Con lo cual $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$ es suficiente para θ.

13.2 Inciso b

$f(x,\theta)$ es la función de densidad de una variable que se distribuye Exponencial con parámetro θ. En el ejercicio 3.b mostramos que la distribución Exponencial pertenece a la familia exponencial, con $t(x) = x$. Por lo visto en clase (página 11 de Slide 2), el estadístico:

$$T(\mathbf{X}) = \sum_{i=1}^{n} t(X_i)$$

Es suficiente para θ en una familia exponencial de 1 parámetro.

14 Ejercicio 14

En este caso:

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x};\theta) &= \prod_{i=1}^{n} \theta x_i^{\theta-1} \\
&= \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta-1}
\end{aligned}
$$

Entonces $T(\mathbf{x}) = \prod_{i=1}^{n} x_i$ es suficiente para θ con:

$$
g(T(\mathbf{x});\theta) = \theta^n T(\mathbf{x})^{\theta} \qquad h(\mathbf{x}) = T(\mathbf{x})^{-1}
$$

15 Ejercicio 15

15.1 Inciso a

Primero que nada, notemos que dado que $\theta \notin \{0,1\}$, el soporte de $f(x;\theta)$ no depende de θ. Luego, intentamos expresar $f(x;\theta)$ en la forma de familia exponencial vista en el ejercicio 2:

$$
\begin{aligned}
f(x;\theta) &= \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|} \\
\ln f(x;\theta) &= |x| \ln\left(\frac{\theta}{2}\right) + (1-|x|)\ln(1-\theta) \\
\exp(\ln f(x;\theta)) &= \exp\left[|x| \ln\left(\frac{\theta}{2}\right) + (1-|x|)\ln(1-\theta) \right] \\
f(x;\theta) &= \exp\left[|x| \left(\ln\left(\frac{\theta}{2}\right) - \ln(1-\theta) \right) + \ln(1-\theta) \right]
\end{aligned}
$$

Esta expresión se ajusta a la forma exponencial vista en el ejercicio 2 con:

$$
\begin{aligned}
a(x) &= t(x) = |x| \\
b(\theta) &= \left(\ln\left(\frac{\theta}{2}\right) - \ln(1-\theta) \right) \\
m(x) &= 0 \\
-d(\theta) &= \ln(1-\theta)
\end{aligned}
$$

15.2 Inciso b

Al verificar que $f(x;\theta)$ pertence a una familia exponencial, sabemos que:

$$
T(\mathbf{x}) = \sum_{i=1}^{n} t(x_i) = \sum_{i=1}^{n} |x_i|
$$

Es suficiente para θ.

15.3 Inciso c

Sabemos que al ser un estadístico suficiente proveniente de una familia exponencial, podemos afirmar que $T(X) = \sum_{i=1}^{n} |x_i|$ es completo (ver página 21 de Slide 2).

16 Ejercicio 16

16.1 Inciso a

Para una variable $X_i \sim U(0, \theta)$, tenemos:

$$f(x, \theta) = \frac{1}{\theta} \mathbb{1}_{(x_i \in [0,\theta])}$$

Entonces,

$$\begin{aligned}
f(\mathbf{x}, \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{(x_i \in [0,\theta])} \\
&= \theta^{-n} \prod_{i=1}^{n} \mathbb{1}_{(x_i \in [0,\theta])} \\
&= \theta^{-n} \mathbb{1}_{(x_{(1)} \geq 0)} \mathbb{1}_{(x_{(n)} \leq \theta)}
\end{aligned}$$

Donde $x_{(1)}$ y $x_{(n)}$ son el mínimo y máximo de la secuencia $\{x_i\}_i^n$ respectivamente. El último paso vale dado que sabemos que el resto de las observaciones estarán en el intervalo $[0, \theta]$ y por lo tanto las funciones indicadoras relacionadas con las mismas serán iguales a 1. Entonces, por Neyman-Fisher, $T(\mathbf{X}) = X_n = \max\{X_i\}_{i=1}^n$ es suficiente para θ con:

$$\begin{aligned}
h(\mathbf{x}) &= \mathbb{1}_{(x_{(1)} \geq 0)} \\
g(T(\mathbf{x}); \theta) &= \theta^{-n} \mathbb{1}_{(T(\mathbf{x}) \leq \theta)}
\end{aligned}$$

16.2 Inciso b

Del ejercicio 11.b sabemos que la función de distribución del máximo de una secuencia de variables aleatorias independientes está dada por:

$$F_{X_{(n)}}(x) = F_X(x)^n$$

Para una variable $X \sim U(0, \theta)$: $F_X(x) = \frac{x}{\theta}$. Entonces:

$$F_{X_{(n)}}(x) = \left(\frac{x}{\theta}\right)^n \Rightarrow f_{X_{(n)}}(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}$$

Entonces, queremos ver que $E(\frac{n+1}{n} X_{(n)}) = \theta$:

$$\begin{aligned}
E\left(\frac{n+1}{n} X_{(n)}\right) &= \frac{n+1}{n} \int_0^\theta x \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} dx \\
&= \frac{n+1}{\theta^n} \int_0^\theta x^n dx \\
&= \theta
\end{aligned}$$

16.3 Inciso c

Sí, cualquier función biyectiva (en este caso lineal) de un estadístico suficiente da como resultado otro estadístico suficiente (ver página 14 de Slide 2).

17 Ejercicio 17

El Teorema de Basu afirma que un estadístico *minimal suficiente* para θ es **independiente** de todos los estadísticos *ancillares* de θ cuando el modelo estadístico es **completo**.

La distribución normal pertenece a la familia exponencial con $t_\mu = x$ y $t_{\sigma^2} = -x^2/2$ (puede verificarlo). Por propiedad de familia exponencial, $\bar{X} = n^{-1}\sum_{i=1}^n x_i$ es un estadístico completo y minimal suficiente para μ.

Entonces, sólo falta verificar que S^2 sea *ancillar* para μ. Por definición, S^2 será ancillar para μ si su distribución no depende de μ, lo cual se verifica[6] dado que $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$.

Por lo tanto, el Teorema de Basu nos permite afirmar que para una población normal, $\bar{X}$ y S^2 serán independientes.

18 Ejercicio 18

18.1 Inciso a

La función de verosimilitud para una población Poisson con $n = 10$ está dada por:

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{10} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$L(\lambda|\mathbf{x}) = \frac{e^{-10\lambda}\lambda^{\sum_{i=1}^{10} x_i}}{\prod_{i=1}^{10} x_i!}$$

18.2 Inciso b

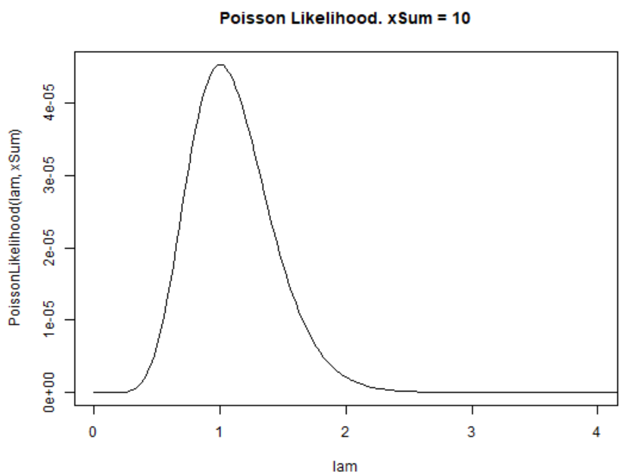
Analizando la expresión obtenida en el inciso anterior, parecería imposible graficar $L(\lambda|\mathbf{X})$ sólamente a partir del conocimiento de $\sum_{i=1}^{10} x_i$. No obstante, sólo nos interesa el valor *relativo* de la función de verosimilitud para una muestra dada de $\mathbf{X} = \mathbf{x}$. Es fácil[7] ver que el valor máximo de esta función a medida que varía λ sólo depende de $\sum_{i=1}^{10} x_i$, y no del denominador. Por lo tanto, podemos reescribir:

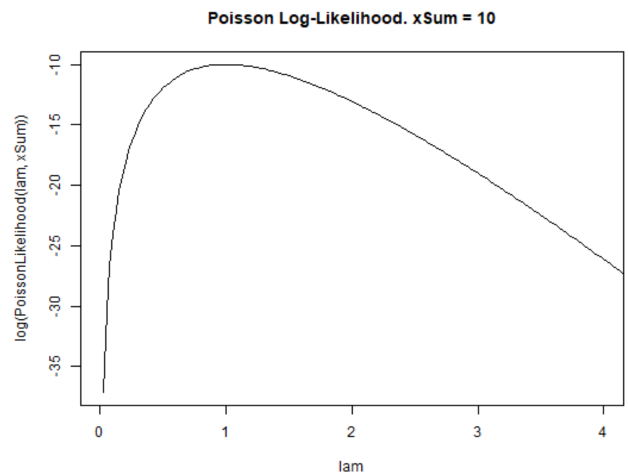
$$L(\lambda|\mathbf{x}) = e^{-10\lambda}\lambda^{\sum_{i=1}^{10} x_i}$$

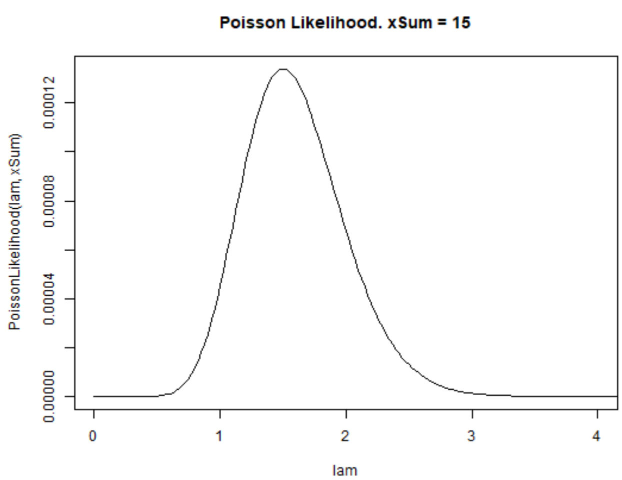
Entonces, grafiquemos $L(\lambda|\mathbf{x})$ y $l(\lambda|\mathbf{x})$ como función de λ para los diferentes valores de $\sum_{i=1}^{10} x_i$. Yo utilicé `R`, los códigos están en el campus.

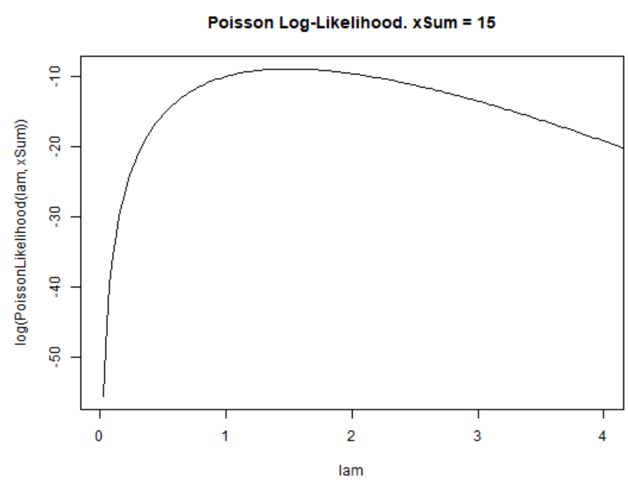
[6] Este resultado proviene del teorema de Cochran.

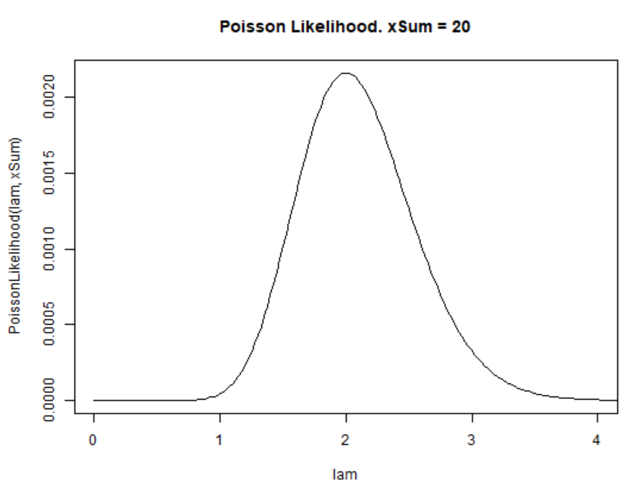
[7] Esto se verá más claramente la estudiar estimación por máxima verosimilitud. De lo contrario, pruebe graficar $L(\lambda|\mathbf{x})$ para dos muestras que tengan el mismo valor de $\sum_{i=1}^{10} x_i$ y podrá verificarlo

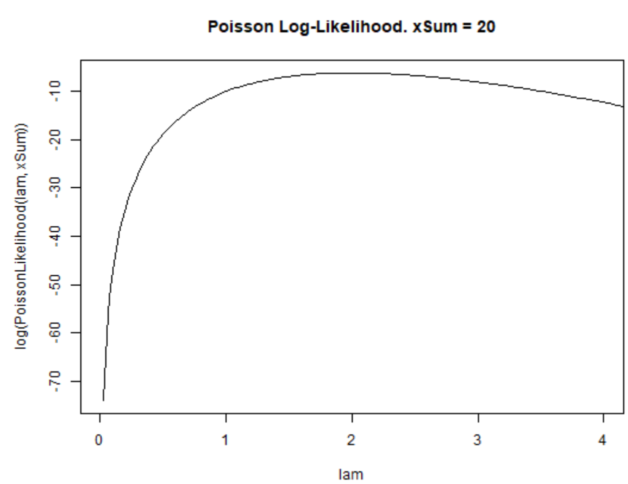

Poisson Likelihood. xSum = 10
PoissonLikelihood(lam, xSum)
0e+00
1e-05
2e-05
3e-05
4e-05
0
1
2
3
4
lam


Poisson Log-Likelihood. xSum = 10
log(PoissonLikelihood(lam, xSum))
-35
-30
-25
-20
-15
-10
0
1
2
3
4
lam


Poisson Likelihood. xSum = 15
PoissonLikelihood(lam, xSum)
0.00000
0.00004
0.00008
0.00012
0
1
2
3
4
lam


Poisson Log-Likelihood. xSum = 15
log(PoissonLikelihood(lam, xSum))
-50
-40
-30
-20
-10
0
1
2
3
4
lam


Poisson Likelihood. xSum = 20
PoissonLikelihood(lam, xSum)
0.0000
0.0005
0.0010
0.0015
0.0020
0
1
2
3
4
lam


Poisson Log-Likelihood. xSum = 20
log(PoissonLikelihood(lam, xSum))
-70
-60
-50
-40
-30
-20
-10
0
1
2
3
4
lam

18.3 Inciso c

Analizando la figura de $L(\lambda|\mathbf{x})$ para $\sum_{i=1}^{10} x_i = 10$, es fácil ver que la verosimilitud es máxima cuando $\lambda = 1$, con lo cual es más factible que esta muestra provenga de una población Poisson con $\lambda = 1$ que de una con cualquier otro valor del parámetro.

Más formalmente, podemos responder a la pregunta utilizando el ratio de verosimilitudes o *likelihood ratio*. Calculamos:

$$\frac{L\left(\lambda = 1 | \sum_{i=1}^{10} x_i = 10\right)}{L\left(\lambda = 2 | \sum_{i=1}^{10} x_i = 10\right)} = \frac{e^{-10}1^{10}}{e^{-20}2^{10}} \approx 21.51 > 1$$

Con lo cual es más factible que la muestra provenga de una población Poisson con $\lambda = 1$.

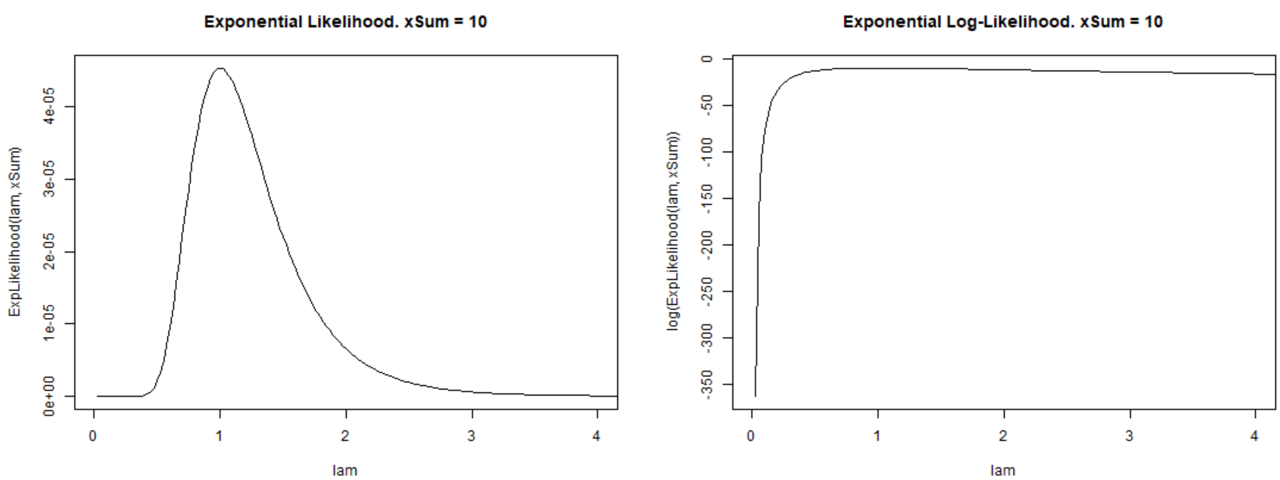
18.4 Inciso d

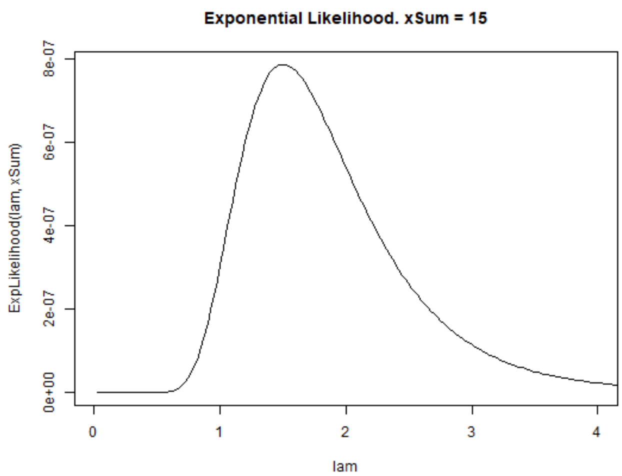
Inciso a: Función de verosimilitud para población Exponencial:

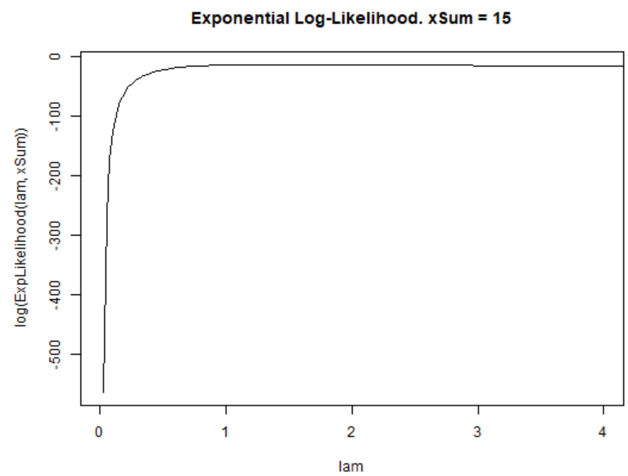
$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{10} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x_i}$$

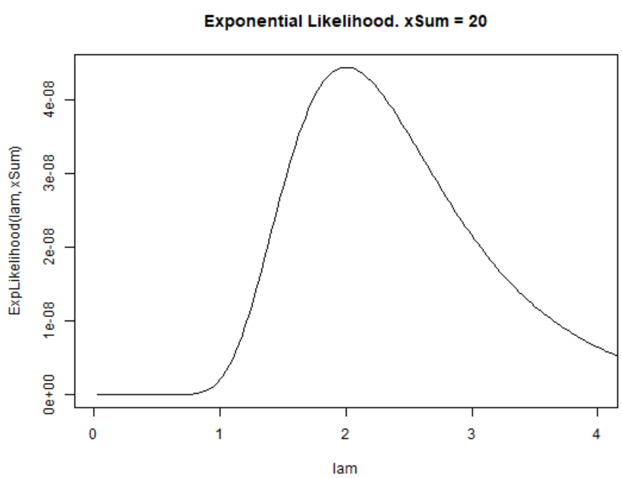
$$L(\lambda|\mathbf{x}) = \lambda^{-10} \exp\left(-\frac{1}{\lambda}\sum_{i=1}^{10} x_i\right)$$

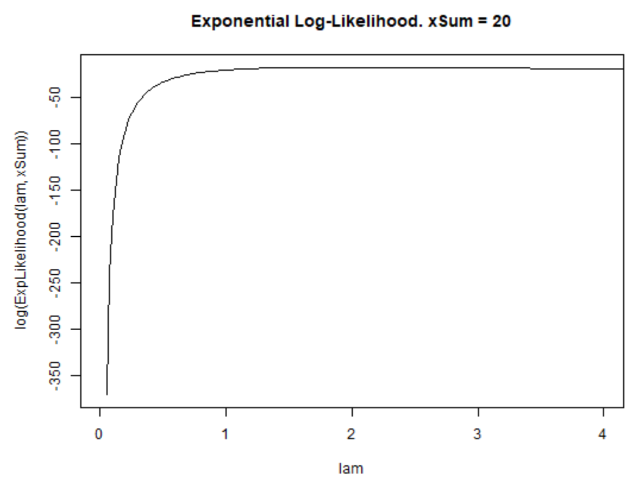
Inciso b: Gráficos de verosimilitud y log-verosimilitud.











Inciso c: Ratio de verosimilitud.

$$\frac{L\left(\lambda = 1 | \sum_{i=1}^{10} x_i = 10\right)}{L\left(\lambda = 2 | \sum_{i=1}^{10} x_i = 10\right)} = \frac{1^{-10} e^{-\frac{10}{1}}}{2^{-10} e^{-\frac{10}{2}}} \approx 6.89 > 1$$

Con lo cual una vez más es más factible que la muestra provenga de una población con $\lambda = 1$.