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:

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

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 $\boldsymbol{\theta} = \{\lambda, k\}$.

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

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 $\mu = \beta_0 + h(X)$.

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

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

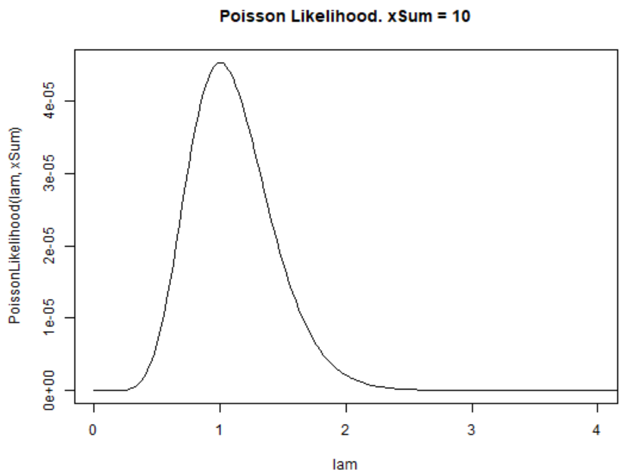
Analizando la expresión obtenida en el inciso anterior, parecería imposible graficar $L(\lambda|\mathbf{X})$ sólamente a partir del conocimiento de $\sum_{i=1}^{10} x_i$. No obstante, sólo nos interesa el valor *relativo* de la función de verosimilitud para una muestra dada de $\mathbf{X} = \mathbf{x}$. 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:

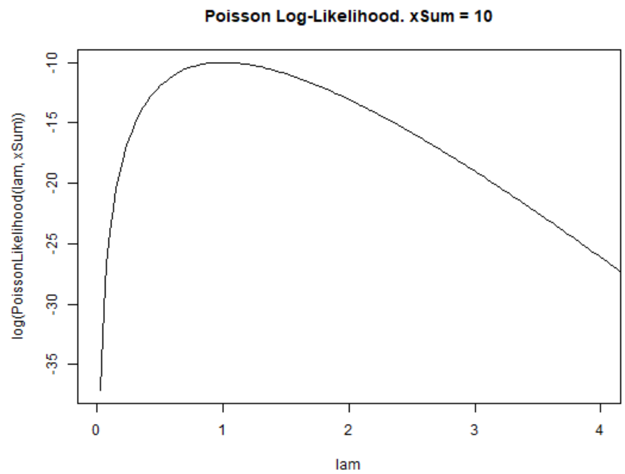
$$L(\lambda|\mathbf{x}) = e^{-10\lambda}\lambda^{\sum_{i=1}^{10} x_i}$$

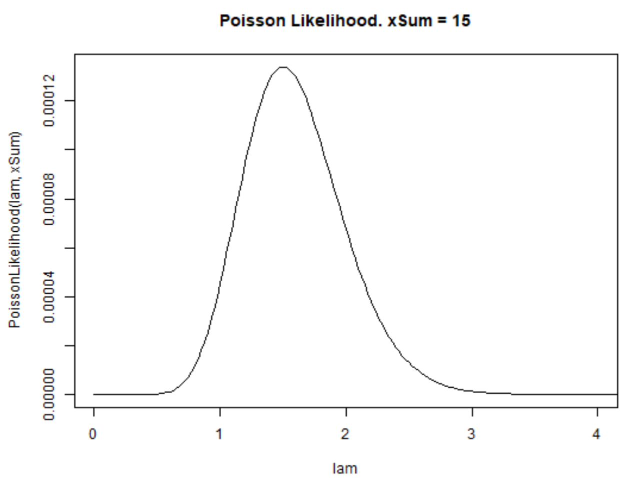
Entonces, grafiquemos $L(\lambda|\mathbf{x})$ y $l(\lambda|\mathbf{x})$ como función de λ para los diferentes valores de $\sum_{i=1}^{10} x_i$. Yo utilicé `R`, los códigos están en el campus.

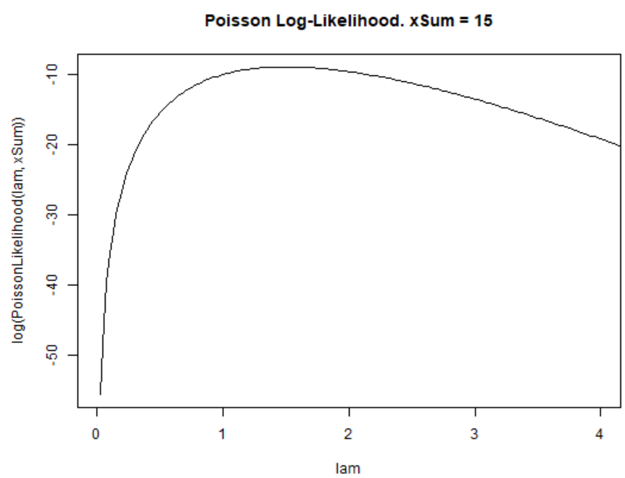
[6] Este resultado proviene del teorema de Cochran.

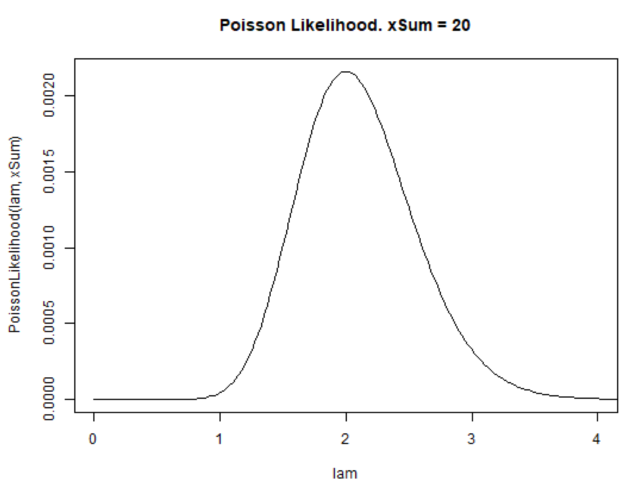
[7] Esto se verá más claramente la estudiar estimación por máxima verosimilitud. De lo contrario, pruebe graficar $L(\lambda|\mathbf{x})$ para dos muestras que tengan el mismo valor de $\sum_{i=1}^{10} x_i$ y podrá verificarlo

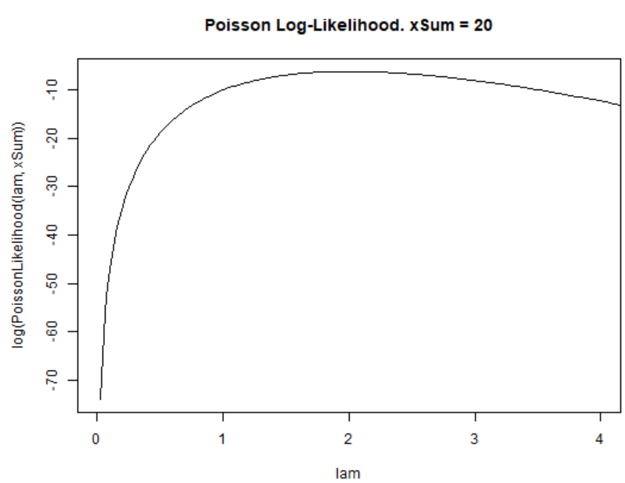

Poisson Likelihood. xSum = 10
PoissonLikelihood(lam, xSum)
0e+00
1e-05
2e-05
3e-05
4e-05
0
1
2
3
4
lam


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 $\lambda = 1$.

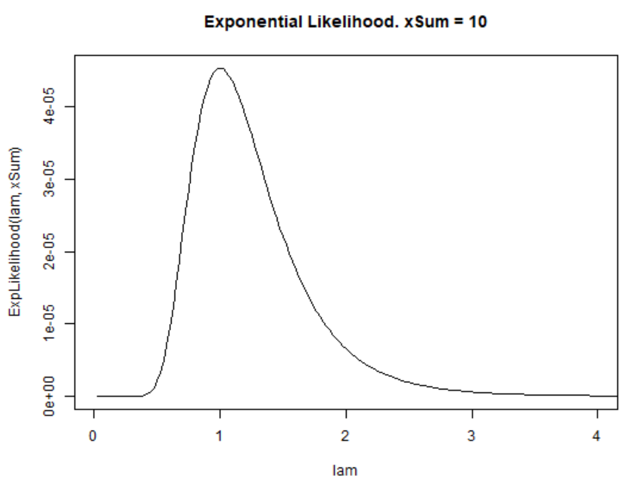
18.4 Inciso d

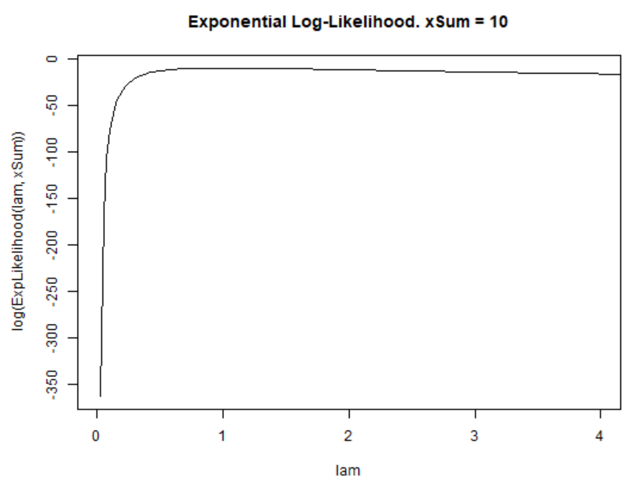
Inciso a: Función de verosimilitud para población Exponencial:

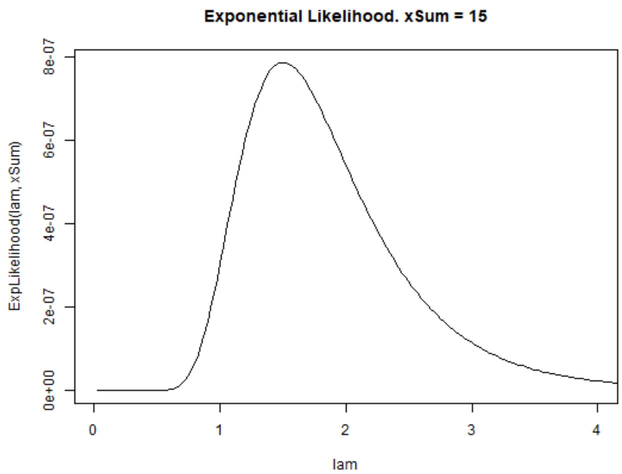
$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{10} \frac{1}{\lambda} e^{-\frac{1}{\lambda} x_i}$$

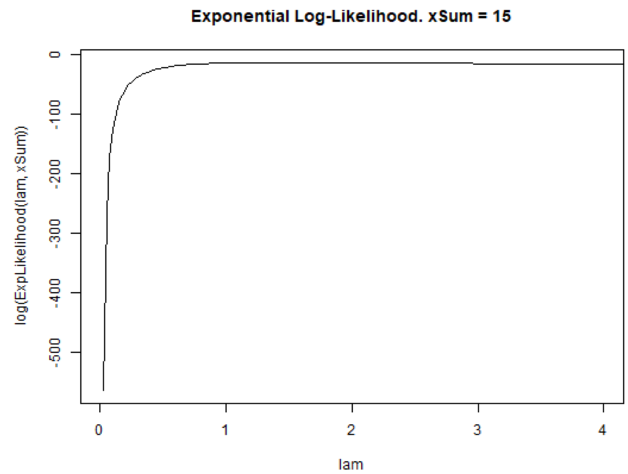
$$L(\lambda|\mathbf{x}) = \lambda^{-10} \exp\left(-\frac{1}{\lambda} \sum_{i=1}^{10} x_i\right)$$

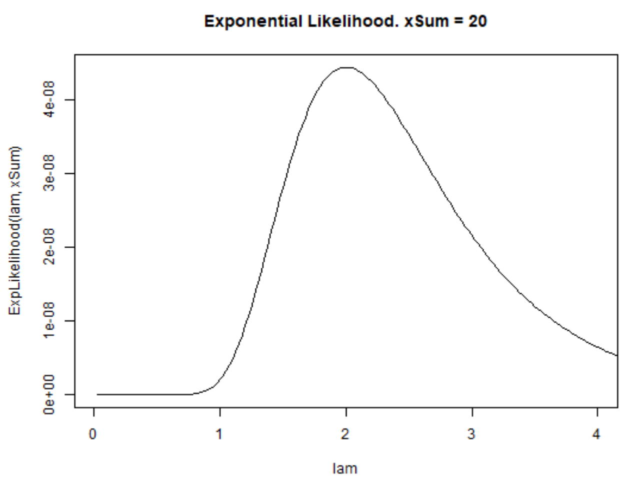
Inciso b: Gráficos de verosimilitud y log-verosimilitud.

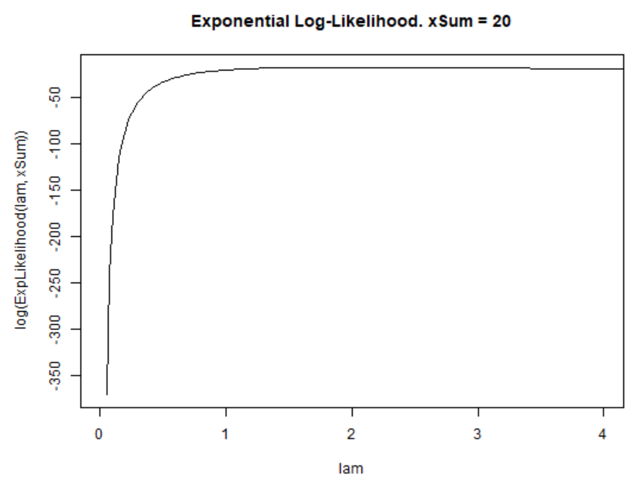












Inciso c: Ratio de verosimilitud.

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

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

Inferencia Estadística - Guia 2

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

Junio 2020

1 Ejercicio 1

1.1 Inciso a

A partir de ahora, utilizaremos $M_k(\mathbf{X})$ para referirnos al *momento muestral* de orden k de la variable aleatoria (o secuencia de variables aleatoria) $\mathbf{X}$. Por lo tanto:

$$M_k(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} x_i^k \tag{1}$$

Para un vector de parámetros $\boldsymbol{\theta}$ de longitud k, el estimador del método de momentos $\tilde{\boldsymbol{\theta}}$ debe ser la solución al sistema:

$$\begin{aligned} M_1(X) &= E(X) \\ M_2(X) &= E(X^2) \\ &\vdots \\ M_k(X) &= E(X^k) \end{aligned}$$

En el caso de una variable uniforme discreta X con soporte en $\{0, 1, ..., \theta - 1, \theta\}$:

$$M_1(X) = \frac{\tilde{\theta} + 0}{2} \Rightarrow \tilde{\theta} = 2\bar{X}$$

Este estimador no resulta muy coherente, dado que $2\bar{X}$ no necesariamente será un número entero.

1.2 Inciso b

Para los datos provistos:

$$\bar{X} = 0.7 \Rightarrow \tilde{\theta} = 1.4$$

1.3 Inciso c

Si X se distribuye uniforme discreta con límites $(-\theta, \theta)$:

$$E_\theta(X) = \frac{\theta + (-\theta)}{2} = 0$$

Y por lo tanto el estimador de momentos $\tilde{\theta}$ no está definido.

2 Ejercicio 2

2.1 Inciso a

Sea $X \sim Bi(\theta, k)$, tenemos:

$$E(X) = k\theta$$
$$E(X^2) = V(X) + E(X)^2 = k\theta(1-\theta) + (k\theta)^2$$

Para obtener los estimadores de momentos $\tilde{\theta}$ y $\tilde{k}$ resolvemos el sistema:

$$M_1(x) = \tilde{\theta}\tilde{k}$$
$$M_2(x) = \tilde{\theta}\tilde{k}(1-\tilde{\theta}) + (\tilde{\theta}\tilde{k})^2$$

Entonces:

$$\tilde{\theta} = 1 - \frac{(n^{-1}\sum_{i=1}^{n} x_i^2) - \bar{x}^2}{\bar{x}}$$
$$\tilde{k} = \frac{\bar{x}}{\tilde{\theta}}$$

2.2 Inciso b

Notar que por descomposición de varianza podemos escribir:

$$(n^{-1}\sum_{i=1}^{n} x_i^2) - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Entonces, $\tilde{\theta}$ se transforma en:

$$\tilde{\theta} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}{\bar{x}}$$

Usando los datos provistos, obtenemos:

$$\tilde{\theta} = \tilde{k} = -1$$

Lo cual no resulta coherente dado que esperaríamos que ambos sean valores positivos.

3 Ejercicio 3

Para una variable Gamma(α, λ) tenemos:

$$E(X) = \frac{\alpha}{\lambda}$$
$$E(X^2) = \frac{\alpha}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2$$

Por lo tanto, igualando con sus contrapartes muestrales y resolviendo el sistema obtenemos:

$$\tilde{\lambda} = \frac{\bar{x}}{\bar{x}^2 + n^{-1}\sum_{i=1}^{n} x_i^2}$$
$$\tilde{\alpha} = \bar{x}\tilde{\lambda}$$

Para los datos provistos:

$$\tilde{\lambda} = \frac{5}{25 + 144/10} = \frac{50}{394} \approx 0.1269$$
$$\tilde{\alpha} = 5 * \frac{50}{394} \approx 0.6345$$

4 Ejercicio 4

A partir de ahora, utilizaremos $\hat{\theta}_{ML}$ para referirnos al estimador de máxima verosimilitud de θ, mientras que usaremos $\tilde{\theta}$ para el estimador de momentos.

4.1 Inciso a

Dado que una variable $X \sim Pois(\lambda)$ tiene $E(X) = \lambda$, el estimador de momentos de λ está dado por:

$$\tilde{\lambda} = \bar{X}$$

A partir de la función de verosimilitud para un modelo Poisson calculada en el Ejercicio 18 de la Guía 1 podemos calcular la log-verosimilitud como:

$$l(\lambda|\mathbf{x}) = -n\lambda + \sum_{i=1}^{n} x_i \ln(\lambda) - \sum_{i=1}^{n} \ln(x_i!)$$

Para obtener el estimador de máxima verosimilitud, utilizamos la condición de primer orden:

$$\frac{\partial l(\lambda|\mathbf{x})}{\partial \lambda}(\hat{\lambda}_{ML}) = 0$$

De la cual obtenemos:

$$\hat{\lambda}_{ML} = \bar{x}$$

Por lo tanto, para los datos de la muestra provista:

$$\tilde{\lambda} = \hat{\lambda}_{ML} = 2.5$$

4.2 Inciso b

Dado que una variable aleatoria $X \sim Exp(\frac{1}{\lambda})$ tiene $E(X) = \lambda$, el estimador de momentos $\tilde{\lambda}$ está dado por:

$$\tilde{\lambda} = \bar{X}$$

Una vez más, de lo visto en el ejercicio 18 de la Guía 1 podemos calcular la log-verosimilitud para este modelo:

$$l(\lambda|\mathbf{x}) = -n\ln(\lambda) - \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

De la condición de primer orden para la maximización de la log-verosimilitud obtenemos:

$$\hat{\lambda}_{ML} = \bar{X}$$

Para los datos de la muestra provista:

$$\tilde{\lambda} = \hat{\lambda}_{ML} = 2.5$$

4.3 Inciso c

Para poder obtener el estimador de momentos λ en esta distribución, primero calculamos la esperanza de X utilizando:

$$E(X) = \int_{\lambda}^{\infty} \frac{1}{\lambda} e^{1-\frac{x}{\lambda}} x \, dx$$

Al integrar esta expresión por partes[1], se obtiene $E(X) = 2\lambda$. Por lo tanto, el estimador de momentos de λ será

$$\tilde{\lambda} = \frac{\bar{X}}{2}$$

Por otro lado, la función de verosimilitud y log-verosimilitud para este modelo están dadas por:

$$L(\lambda|\mathbf{x}) = \lambda^{-n} e^{n - \frac{1}{\lambda}\sum_{i=1}^{n} x_i}$$
$$l(\lambda|\mathbf{x}) = -n\ln(\lambda) + n - \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

De la condición de primer orden para la maximización de la log-verosimilitud obtenemos:

$$\hat{\lambda}_{ML} = \bar{X}$$

Para los valores provistos para esta muestra:

$$\tilde{\lambda} = 1.25$$
$$\hat{\lambda}_{ML} = 2.5$$

4.4 Inciso d

Para una variable $X \sim \Gamma(\lambda, k)$, $E(X) = \frac{k}{\lambda}$. Por lo tanto, el estimador de momentos $\tilde{\lambda}$ con $k = 2$ será:

$$\tilde{\lambda} = \frac{2}{\bar{X}}$$

Para el modelo Gamma, la función de verosimilitud y log-verosimilitud están dadas por (con $k = 2$):

$$L(\lambda|\mathbf{x}) = \lambda^{2n} e^{-\lambda\sum_{i=1}^{n} x_i} \frac{x^n}{\Gamma(2)^n}$$
$$l(\lambda|\mathbf{x}) = 2n\ln(\lambda) - \lambda\sum_{i=1}^{n} x_i + n\left(\ln(x) - \ln(\Gamma(2))\right)$$

De la condición de primer orden para la maximización de la verosimilitud obtenemos:

$$\hat{\lambda}_{ML} = \frac{2}{\bar{X}}$$

5 Ejercicio 5

Tanto el modelo Poisson como Exponencial pertenecen a familias exponenciales, y sus estimadores máximo verosímiles son insesgados. Por lo tanto, podemos utilizar este hecho para calcular fácilmente la Cota de CR para la varianza de los EMV y analizar si esta es alcanzada. Recordar que para familias exponenciales:

$$I(\theta) = -nE\left(\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right)$$

Para el caso del modelo **Poisson**:

$$I(\lambda) = -nE\left(\frac{x}{\lambda^2}\right) = \frac{n}{\lambda}$$

Dado que:

$$\mathrm{Var}(\hat{\lambda}) = \frac{\mathrm{Var}(x)}{n} = \frac{\lambda}{n} = I(\theta)^{-1}$$

Con lo cual se alcanza la Cota de CR y $\hat{\lambda}$ es eficiente para el modelo Poisson.

[1] También deberá usar la Regla de L'Hopital, la cual afirma que:

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}$$

6 Ejercicio 6

6.1 Inciso a

Para obtener el estimador de momentos para θ, primero calculamos $E(X)$:

$$E(X) = \int_{-1}^{1} x \left(\frac{1+\theta x}{2} \right) \, dx = \frac{\theta}{3}$$

Entonces, el estimador de momentos $\tilde{\theta}$ debe satisfacer $E_{\tilde{\theta}}(X) = \bar{X}$. Por lo tanto:

$$\tilde{\theta} = 3\bar{X}$$

6.2 Inciso b

Dado que $\tilde{\theta}$ es un estimador insesgado para θ (verifíquelo), tenemos $ECM(\tilde{\theta}, \theta) = \mathrm{Var}(\tilde{\theta})$. Por lo tanto, querríamos una estimación para la varianza asintótica del estimador.

Dado que las X_i se distribuyen independientemente y poseen varianza finita (verifíquelo[2]), se cumplen los supuestos del TCL, por lo tanto:

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{D} N(0, 9\mathrm{Var}(X))$$

6.3 Inciso c

Dado que $\tilde{\theta}$ es insesgado y su varianza converge a cero a medida que $n \to \infty$, tenemos que:

$$\lim_{n\to\infty} ECM(\tilde{\theta}, \theta) = 0$$

Lo que equivale a afirmar que $\tilde{\theta}$ es un estimador consistente de θ.

6.4 Inciso d

Para la muestra provista, el estimador de momentos $\tilde{\theta}$ toma el valor:

$$\tilde{\theta} = 3\left(\frac{-0.2}{4} \right) = -0.15$$

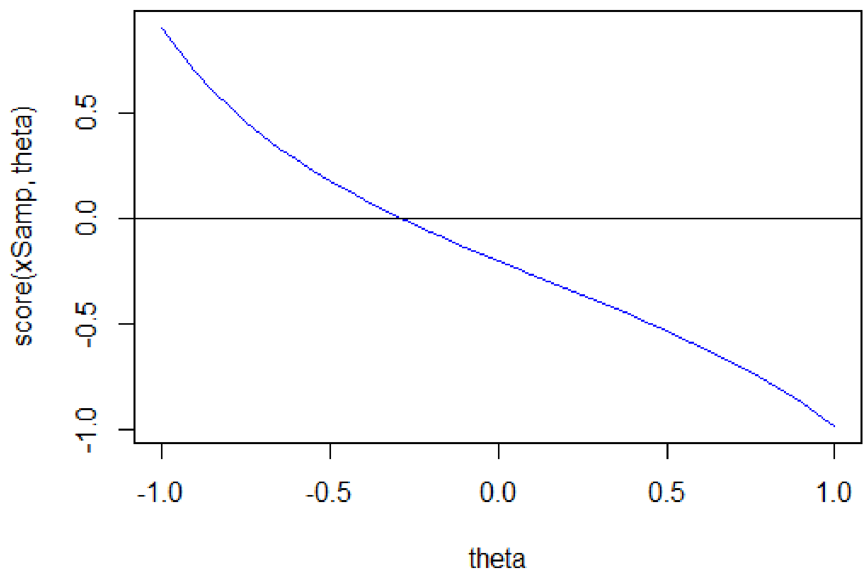
Para obtener el estimador máximo verosímil $\hat{\theta}_{ML}$, debemos igualar a cero la función score (lo que equivale a la condición de primer orden del problema de maximización de la log-verosimilitud):

$$\begin{aligned} S(\hat{\theta}_{ML}) &= \frac{\partial}{\partial \theta} \log f(\mathbf{x}; \hat{\theta}_{ML}) = 0 \\ &= \sum_{i=1}^{n} \frac{x_i}{1 + \hat{\theta}_{ML} x_i} = 0 \end{aligned}$$

A continuación se adjunta un gráfico del valor de esta función para la muestra provista en la consigna.

[2] $\mathrm{Var}(X) = \frac{1}{3} - \frac{\theta^2}{9}$

Figure 1: Función Score(θ) para muestra provista



Puede evidenciarse que la función alcanza valores cercanos a 0 para $\theta \approx 0.3$. Podemos obtener una estimación más precisa utilizando métodos numéricos. Por ejemplo, podemos implementar el método Newton-Raphson para encontrar una única raíz para la función Score utilizando la librería `uniroot` en `R`. El código utilizado estará subido al campus. Utilizando dicho método obtenemos $\hat{\theta}_{ML} \approx 0.285$.

7 Ejercicio 7

Notar que podemos escribir:

$$\psi(\lambda) = P(X = 2|\lambda) = \frac{\lambda^2 e^{-2\lambda}}{2!}$$

Por invarianza del estimador de máxima verosimilitud:

$$\hat{\psi}_{ML}(\lambda) = \psi(\hat{\lambda}_{ML})$$

Sabemos que por el ejercicio 4 que para una variable Poisson, $\hat{\lambda}_{ML} = \bar{X}$. Con los datos provistos, tenemos:

$$\bar{X} = \frac{116}{55} \Rightarrow \hat{\psi}_{ML} \approx 0.2698$$

8 Ejercicio 8

8.1 Inciso a

Sea $l(\mathbf{x}; \mu, \sigma^2)$ la función de log-verosimilitud del modelo normal, los estimadores máximo verosímiles $\hat{\mu}_{ML}, \hat{\sigma^2}_{ML}$ deben satisfacer las condiciones de primer orden:

$$\frac{\partial l(\cdot)}{\partial \mu}(\hat{\mu}_{ML}) = 0$$
$$\frac{\partial l(\cdot)}{\partial \sigma^2}(\hat{\sigma^2}_{ML}) = 0$$

A su vez, es condición necesaria que la matriz hessiana:

$$H(\hat{\theta}_{ML}) = \begin{bmatrix} \frac{\partial^2 l(\cdot)}{\partial \mu^2} & \frac{\partial^2 l(\cdot)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 l(\cdot)}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 l(\cdot)}{\partial (\sigma^2)^2} \end{bmatrix}$$

Sea semidefinida negativa. Para el modelo normal, la log-verosimilitud está dada por:

$$l(\mathbf{x}; \mu, \sigma^2) = -\frac{n}{2}\left(\ln(\sigma^2 2\pi)\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Entonces, calculamos las derivadas a utilizar:

$$\begin{aligned}
\frac{\partial l(\cdot)}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \\
\frac{\partial l(\cdot)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 \\
\frac{\partial^2 l(\cdot)}{\partial \mu^2} &= -\frac{n}{\sigma^2} \\
\frac{\partial^2 l(\cdot)}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma^6} \\
\frac{\partial l(\cdot)}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \\
\frac{\partial l(\cdot)}{\partial \mu \partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \mu)
\end{aligned}$$

Entonces, de las condiciones de primer orden obtenemos:

$$\begin{aligned}
\hat{\mu}_{ML} &= \bar{X} \\
\hat{\sigma^2}_{ML} &= \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n}
\end{aligned}$$

Notar que para obtener la segunda expresión utilizamos el principio de invarianza de estimadores maximo verosímiles para reemplazar $\hat{\mu}_{ML} = \bar{X}$. Para corroborar que $H(\hat{\theta}_{ML})$ es semidefinida negativa, analizamos el signo de sus primeros menores principales, D_1 y D_2, cuando los evaluamos en $\hat{\mu}_{ML}$ y $\hat{\sigma^2}_{ML}$:

$$\begin{aligned}
D_1 &= \frac{\partial^2 l(\cdot)}{\partial \mu^2} < 0 \\
D_2 &= \frac{\partial^2 l(\cdot)}{\partial \mu^2}\frac{\partial^2 l(\cdot)}{\partial (\sigma^2)^2} - 2 * \frac{\partial l(\cdot)}{\partial \mu \partial \sigma^2} > 0
\end{aligned}$$

Dado que su primer menor principal es negativo y el segundo positivo, la matriz es definida negativa y se cumplen las condiciones necesarias y suficientes para la maximización de la verosimilitud.

8.2 Inciso b

Con los valores provistos para la muestra obtenemos la estimación (recuerde utilizar la descomposición de la varianza muestral):

$$\begin{aligned}
\hat{\mu}_{ML} &= 1.7 \\
\hat{\sigma^2}_{ML} &= 5.21
\end{aligned}$$

8.3 Inciso c

Empecemos trabajando con el estimador insesgado S^2. Dado que trabajamos con una distribución normal, podemos usar de lo visto en clase que la varianza de S^2 (que al ser insesgado es igual a su ECM) es:

$$ECM(S^2, \sigma^2) = \mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

El EMV de σ^2 está dado por:

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2 = \frac{n-1}{n}S^2$$

Utilizando su relación con S^2, obtenemos:

$$E(\hat{\sigma}^2_{ML}) = \frac{n-1}{n}\sigma^2$$
$$\mathrm{Var}(\hat{\sigma}^2_{ML}) = \frac{2(n-1)}{n^2}\sigma^4$$

Por lo tanto, utilizando la descomposición del error medio cuadrático:

$$ECM(\hat{\sigma}^2_{ML}, \sigma^2) = \frac{\sigma^4}{n^2} + \frac{2(n-1)\sigma^4}{n^2} = \frac{2n-1}{n^2}\sigma^4$$

Se puede ver que $ECM(S^2, \sigma^2) > ECM(\hat{\sigma}^2_{ML}, \sigma^2)$ se cumple si y sólo si:

$$\frac{2}{n-1} > \frac{2}{n} - \frac{1}{n^2}$$

Lo cual vale dado que:

$$\frac{2}{n-1} > \frac{2}{n} > \frac{2}{n} - \frac{1}{n^2}$$

Para cualquier $n > 1$. Entonces, el estimador insesgado posee un riesgo más alto (es decir, es menos eficiente en términos de ECM).

8.4 Inciso d

Para el caso multiparámetro en familias exponenciales, el elemento ij de la matriz de información $\mathbf{I}(\boldsymbol{\theta})$ está dado por:

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = -E\left(\frac{\partial^2}{\partial\theta_i\theta_j}\log f(\mathbf{X}, \theta)\right)$$

A partir de los cálculos del inciso 8.a podemos ver que para este modelo:

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix}$$

Dado que la matriz de información es diagonal, su inversa se obtiene de invertir los elementos en su diagonal principal. Por lo tanto, la cota de Crámer Rao será simplemente:

$$V(\boldsymbol{\theta}) \geq \mathbf{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}$$

8.5 Inciso e

Sabemos que el estimador máximo verosímil de la media está dado por $\bar{X}$, y que su varianza es $\text{Var}(\bar{X}) = \sigma^2/n$. Comparando este valor con el primer elemento de la Cota de Crámer Rao, podemos evidenciar que el estimador máximo verosímil de la media es eficiente.

En contraste, para el estimador máximo verosímil de σ^2 sólo será eficiente si la media es conocida, dado que de lo contrario su varianza está dada por:

$$\text{Var}(\hat{\sigma}^2) = \frac{2(n-1)}{n^2}\sigma^4$$

Lo cual se puede ver del inciso 8.3

8.6 Inciso f

Utilizando lo visto en clase para estimadores máximo verosímiles[3], sabemos que para un modelo parametrizado por un vector $\boldsymbol{\theta}$ de dimensión $p = 2$:

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\mathbf{I}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \underset{F}{\rightarrow} \chi^2_2$$

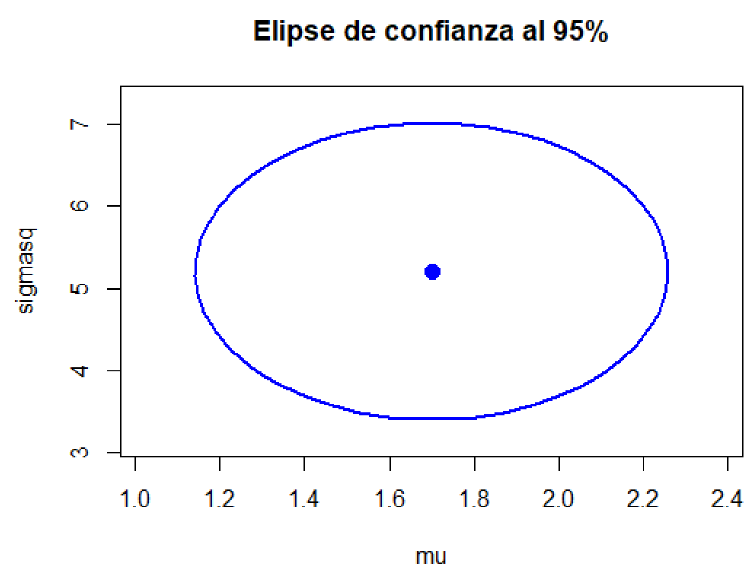
En este caso tenemos $\boldsymbol{\theta} = [\mu, \sigma^2]$. Para la matriz de información encontrada en el inciso 8.d:

$$\frac{n(\hat{\mu} - \mu)^2}{\hat{\sigma}^2} + \frac{n(\hat{\sigma}^2 - \sigma^2)^2}{2\hat{\sigma}^4} \underset{F}{\rightarrow} \chi^2_2$$

Por lo tanto, para los valores obtenidos en el inciso 8.b para $\hat{\mu}$ y $\hat{\sigma}^2$, definimos la elipse de confianza al 95% como:

$$R_{0.95} = \left\{(x, y) : \frac{100(1.7 - x)^2}{5.21} + \frac{100(5.21 - y)^2}{2(5.21)^2} \leq c\right\}$$

Donde c es el cuantil $1 - \alpha$ de una variable chi-cuadrado con dos grados de libertad ($c = 5.99$). Si graficamos en `R`:



[3] Ver por ejemplo página 22 de Slide 5.

9 Ejercicio 9

Sea $\{X_i\}_{i=1}^n$ una muestra aleatoria de una población Uniforme (continua) con soporte en $[0, \theta]$, nos interesa encontrar $\hat{\theta}_{ML}$. Sabemos que para cualquier X_i:

$$f(x;\theta) = \begin{cases} 1/\theta & \text{si } 0 \leq x \leq \theta \\ 0 & \text{en otro caso.} \end{cases}$$

Por lo cual la función de verosimilitud será:

$$f(\mathbf{x}, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \text{ para } x_i \in [0, \theta]$$

Al derivar esta expresión[4] en búsqueda de una condición de primer orden obtenemos:

$$\frac{\partial f(\mathbf{x}, \theta)}{\partial \theta} = -\frac{n}{\theta^{n-1}} < 0 \ \forall \ n, \theta > 0$$

Por lo tanto, la verosimilitud es siempre decreciente como función de θ para un dado n. Por lo tanto, querríamos elegir el valor de θ más bajo posible como su estimador máximo verosímil. No obstante, la función de verosimilitud sólo está definida para $x_i : x_i < \theta$. Por lo tanto, el estimador máximo verosímil será $\hat{\theta}_{ML} = X_{(n)} = \max \{X_i\}_{i=1}^n$.

10 Ejercicio 10

Dado un estimador W para θ, definimos el error cuadrático medio asociado a W como:

$$\begin{aligned}
ECM(W, \theta) &= E[(W-\theta)^2] \\
&= E[(W - E(W) + E(W) - \theta)^2] \\
&= E[(W - E(W))^2 + 2(W - E(W))(E(W) - \theta) + (E(W) - \theta)^2] \\
&= E[(W - E(W))^2] + 2E[(W - \underbrace{E(W)}_{\text{constante}})\underbrace{(E(W) - \theta)}_{\text{constante}}] + E[(E(W) - \theta)^2] \\
&= E[(W - E(W))^2 + 2(E(W) - E(W))(E(W) - \theta) + (E(W) - \theta)^2] \\
&= \underbrace{E[(W - E(W))^2]}_{=\text{Var}(W)} + \underbrace{E[(E(W) - \theta)^2]}_{Sesgo^2_\theta(W)}
\end{aligned}$$

Que era lo que queríamos demostrar.

11 Ejercicio 11

Para la ver que $a + bW$ es insesgado para $a + b\theta$:

$$E[a + bW] = a + bE[W] = a + b\theta$$

En el caso de W^2:

$$E[W^2] = \text{Var}(W) + E[W]^2 = \text{Var}(W) + \theta^2$$

Entonces,

$$\text{Sesgo}(E[W^2], \theta^2) = E[W^2] - \theta^2 = \text{Var}(W)$$

[4] Generalmente trabajamos con la *log-verosimilitud* porque facilita las cuentas, en este caso, no es necesario.

12 Ejercicio 12

Para cualquier estimador W de θ podemos descomponer su error cuadrático medio $R(W,\theta)$ como:

$$R(W,\theta) = \text{Sesgo}_\theta^2(W) + \text{Var}(W)$$

En este ejercicio trabajaremos con una secuencia de variables $\{X_i\} \overset{iid}{\sim} Bi(p)$ con un tamaño muestral dado. La suma de variables aleatorias binomiales es en sí una variable aleatoria binomial con un tamaño de experimento igual a la suma del tamaño de las muestras asociadas a cada variable. Por lo tanto, en este ejercicio, n se refiere a la suma del tamaño de cada uno de los experimentos binomiales X_i, y $\bar{X}$ a la proporción de éxitos en la suma de experimentos.

Para el estimador máximo verosímil $\hat{p}_1 = \bar{X}$:

$$\begin{aligned}
\text{Sesgo}_p(\hat{p}_1) &= E(\bar{X}) - p = 0 \\
\text{Var}(\hat{p}_1) &= \frac{p(1-p)}{n} \\
R(\hat{p}_1, p) &= \frac{p(1-p)}{n}
\end{aligned}$$

Para el estimador bayesiano $\hat{p}_1 = (\sqrt{n/4} + n\bar{X})/(n+\sqrt{n})$:

$$\begin{aligned}
\text{Sesgo}_p(\hat{p}_2) &= \frac{\sqrt{n/4} + nE(\bar{X})}{n+\sqrt{n}} - p \\
&= \frac{\sqrt{n}(0.5-p)}{n+\sqrt{n}} \\
\text{Var}(\hat{p}_2) &= \left(\frac{n}{n+\sqrt{n}}\right)^2 \text{Var}(\bar{X}) \\
&= \frac{np(1-p)}{(n+\sqrt{n})^2} \\
R(\hat{p}_2, p) &= \frac{n(0.5-p)^2}{(n+\sqrt{n})^2} + \frac{np(1-p)}{(n+\sqrt{n})^2} \\
&= \frac{n}{(n+\sqrt{n})^2}\left((0.5-p)^2 + p(1-p)\right)
\end{aligned}$$

12.1 Inciso b

Notar que para $p = 0.5$:

$$\begin{aligned}
R(\hat{p}_1, 0.5) - R(\hat{p}_2, 0.5) &= 0.25\left(\frac{1}{n} - \frac{n}{(n+\sqrt{n})^2}\right) \\
&= 0.25\frac{2\sqrt{n}+1}{(n+\sqrt{n})^2} > 0 \text{ para todo } n > 0
\end{aligned}$$

Es decir, para un n fijo, siempre podemos encontrar un valor de p arbitrariamente cerca de 0.5 tal que $R(\hat{p}_1, p) > R(\hat{p}_2, p)$. Dicho límite queda caracterizado por la expresión $R(\hat{p}_1, p) = R(\hat{p}_2, p)$

12.2 Inciso c

Dado que el sesgo del estimador $\hat{p}_2$ tiende a cero a medida que n tiende a infinito y su varianza es siempre menor, lo preferiríamos en caso de que n sea lo suficientemente grande. Si n es pequeño, el sesgo puede ser significativo (particularmente si p se encuentra muy lejos de 0.5), por lo cual preferiríamos el estimador $\hat{p}_1$.

13 Ejercicio 13

13.1 Inciso a

Sabemos que S^2 es un estimador insesgado de la varianza[5]. A su vez, sabemos que para muestras provenientes de una población normal se cumple[6]:

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Por lo tanto:

$$\text{Var}(S^2) = \left(\frac{\sigma^2}{n-1}\right)^2 \underbrace{\text{Var}(\chi^2_{n-1})}_{=2(n-1)} = \frac{2\sigma^4}{n-1}$$

Entonces:

$$\begin{aligned}
\text{Sesgo}_{\sigma^2}(bS^2) &= (b-1)\sigma^2 \\
\text{Var}(bS^2) &= b^2\frac{2\sigma^4}{n-1} \\
R(bS^2, \sigma^2) &= (b-1)^2\sigma^4 + b^2\frac{2\sigma^4}{n-1} \\
&= \sigma^4\left((b-1)^2 + \frac{2b^2}{n-1}\right)
\end{aligned}$$

13.2 Inciso b

Quiero elegir el b que minimice $R(bS^2, \sigma^2)$. Usando la condición de primer orden:

$$\begin{aligned}
\frac{\partial R(bS^2, \sigma^2)}{\partial b} &= 0 \\
\sigma^4\left(2(b-1) + \frac{4b}{n-1}\right) &= 0 \Rightarrow b^* = \frac{n-1}{n+1}
\end{aligned}$$

14 Ejercicio 14

14.1 Inciso a

Al buscar la condición de primer orden del problema de maximización de verosimilitud, encontramos que no podemos obtener una solución analítica para el estimador máximo verosímil de θ:

$$\begin{aligned}
S(\hat{\theta}_{ML}) &= \frac{\partial}{\partial\theta}\log f(\mathbf{x}; \hat{\theta}_{ML}) = 0 \\
&= 2\sum_{i=1}^{n} \frac{(x_i - \hat{\theta}_{ML})}{1 + (x_i - \hat{\theta}_{ML})^2} = 0
\end{aligned}$$

[5] Ejercicio 8.c de Guía 1.

[6] Ver página 32 de Slide 1.

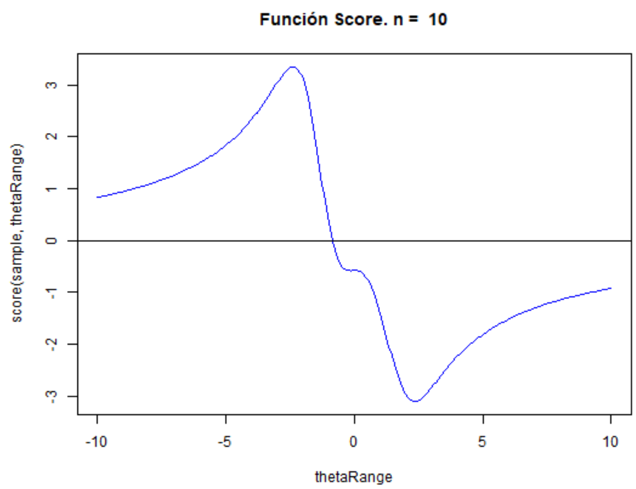
14.2 Inciso b

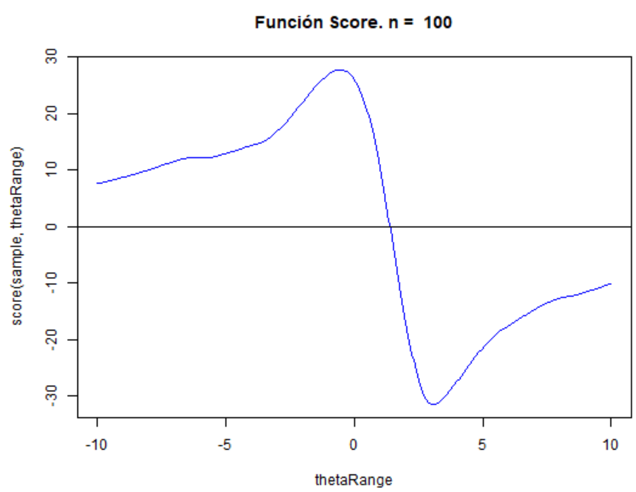
Ya encontramos la función $S(\theta)$ en el inciso anterior. Para obtener la función hessiano $H(\theta)$ simplemente derivamos una vez más.

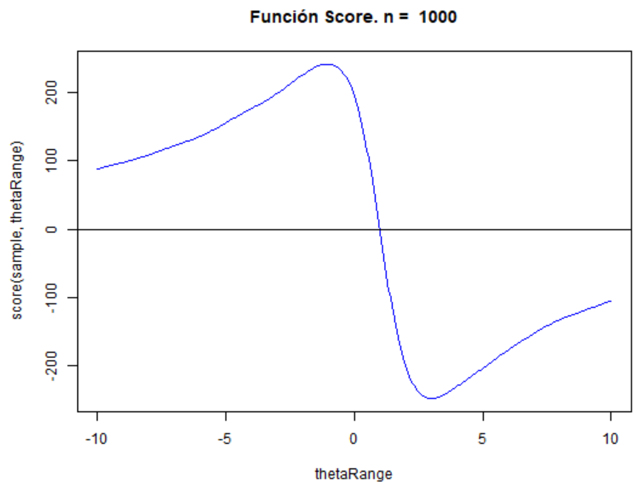
$$
\begin{aligned}
H(\theta) &= \frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}; \hat{\theta}_{ML}) = 0 \\
&= 2 \sum_{i=1}^{n} \frac{(x_i - \theta)^2 - 1}{\left(1 + (x_i - \theta)^2\right)^2}
\end{aligned}
$$

14.3 Inciso c

A continuación se presentan gráficos de la función Score para cada una de las poblaciones Cauchy creadas utilizando R.







Puede evidenciarse que la función muestra un comportamiento mas "suave" a medida que la población crece. Al igual que en el ejercicio anterior, utilizamos la función `uniroot` para obtener los estimadores MVE por métodos numéricos. Los códigos usados serán subidos al campus. Los valores obtenidos fueron:

$$\hat{\theta}_{n=10} = -0.847$$
$$\hat{\theta}_{n=100} = 1.381$$
$$\hat{\theta}_{n=1000} = 0.98$$

Notar que el valor estimado se acerca al valor verdadero a medida que n crece.

15 Ejercicio 15

15.1 Inciso a

Estimador de momentos

Dado $\{X\}_{i=1}^{n} \sim U(0, \theta)$, sabemos por el ejercicio 1.a de esta guía que el estimador de momentos de θ está dado por $\tilde{\theta} = 2\bar{X}$

Para este estimador:

$$\text{Sesgo}^2_\theta(\tilde{\theta}) = \left[E(2\bar{X}) - \theta\right]^2 = 0$$
$$\text{Var}(\tilde{\theta}) = \frac{4Var(X)}{n} = \frac{\theta^2}{3n}$$

Por lo tanto:

$$ECM(\tilde{\theta}, \theta) = \frac{\theta^2}{3n}$$

Estimador alternativo

Sea $X_{(n)} = \max\{X_i\}_{i=1}^{n}$, podemos aplicar los pasos del inciso 11.b de la Guía 1, para obtener su función de densidad, la cual necesitaremos para calcular la esperanza y varianzas de este estimador. Entonces tenemos:

$$\begin{aligned}
f_{X_{(n)}} &= n\frac{x^{n-1}}{\theta^n} \\
E(X_{(n)}) &= \frac{n}{n+1}\theta \\
\mathrm{Var}(X_{(n)}) &= \frac{2n\theta^2}{(n+2)(n+1)^2}
\end{aligned}$$

Llamemos $\hat{\theta} = \frac{n+1}{n}X_{(n)}$, entonces:

$$\begin{aligned}
\mathrm{Sesgo}^2_\theta(\hat{\theta}) &= \left[E(\hat{\theta}) - \theta\right]^2 = 0 \\
\mathrm{Var}(\hat{\theta}) &= \left(\frac{n+1}{n}\right)^2 \mathrm{Var}(X_{(n)}) = \frac{2\theta^2}{n(n+2)}
\end{aligned}$$

Por lo tanto:

$$ECM(\hat{\theta}, \theta) = \frac{2\theta^2}{n(n+2)}$$

15.2 Inciso b

Ambos estimadores son insesgados y su varianza tiende a cero a medida que n tiende a infinito, con lo cual son consistentes.

15.3 Inciso c

Dado que ambos estimadores son insesgados, su eficiencia relativa está dada por:

$$e(\tilde{\theta}, \hat{\theta}) = \frac{\mathrm{Var}(\tilde{\theta})}{\mathrm{Var}(\hat{\theta})} = \frac{\theta^2}{3n}\frac{n(n+2)}{2\theta^2} = \frac{n+2}{6}$$

15.4 Inciso d

Dado que ambos estimadores son insesgados, de acuerdo al criterio de eficiencia relativa nos quedaríamos con el estimador $\hat{\theta}$ siempre que $n > 4$ dado que posee menor varianza.

16 Ejercicio 16

16.1 Inciso a

Dado que $E(X) = \theta$, estamos trabajando con una variable X tal que:

$$\begin{aligned}
f_X(x;\theta) &= \frac{e^{-\frac{x}{\theta}}}{\theta} \\
F_X(x) &= 1 - e^{-\frac{x}{\theta}}
\end{aligned}$$

Para el EMV $\hat{\theta}$ tenemos:

$$\begin{aligned}
\mathrm{Sesgo}^2_\theta(\hat{\theta}) &= \left[E(\bar{X}) - \theta\right]^2 = 0 \\
\mathrm{Var}(\tilde{\theta}) &= \frac{\mathrm{Var}(\bar{X})}{n} = \frac{\theta^2}{n}
\end{aligned}$$

Por lo tanto:

$$ECM(\hat{\theta},\theta) = \frac{\theta^2}{n}$$

Siguiendo los pasos del ejercicio 11 de la Guía 1 para el mínimo de variables aleatorias, podemos llegar a la siguiente función de densidad para $X_{(1)}$:

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

Comparando esta función de densidad con la correspondiente a X, podemos ver que $X_{(1)} \sim \text{Exp}(\theta/n)$ con $E(X_{(1)}) = \frac{\theta}{n}$ y $\text{Var}(X_{(1)}) = \left(\frac{\theta}{n}\right)^2$. Por lo tanto, para el estimador $W = nX_{(1)}$:

$$Sesgo_\theta^2(W) = \left[nE(X_{(1)}) - \theta\right]^2 = 0$$
$$\text{Var}(W) = n^2 \text{Var}(X_{(1)}) = \theta^2$$

Por lo tanto:

$$ECM(W,\theta) = \theta^2$$

Ambos estimadores son insesgados, pero la varianza de $\hat{\theta}$ es menor para cualquier $n > 1$, por lo cual es preferible utilizar el primer estimador.

16.2 Inciso b

En este caso, la varianza de W no converge a cero a medida que crece la población, con lo cual sólo $\hat{\theta}$ es consistente.

17 Ejercicio 17

17.1 Inciso a

Utilizando que el modelo Beta pertenece a la familia exponencial (ver ejercicio 4.g de Guía 1), podemos utilizar la siguiente definición para encontrar la Información de Fisher:

$$\begin{aligned} I(\theta) &= ni(\theta) \\ &= -nE\left(\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right) \\ &= -nE\left(-\frac{1}{\theta^2}\right) \\ &= \frac{n}{\theta^2} \end{aligned}$$

17.2 Inciso b

Para este modelo, la verosimilitud y log-verosimilitud están dadas por:

$$L(\theta,\mathbf{x}) = \theta^n \left(\prod_{i=1}^n x_i\right)^{(\theta-1)}$$

$$l(\theta,\mathbf{x}) = n\log(\theta) + (\theta-1)\sum_{i=1}^n \log x_i$$

Por lo tanto, la función score será:

$$S(\theta|\mathbf{x}) = \frac{\partial}{\partial\theta} l(\theta|\mathbf{x}) = \frac{n}{\theta} + \sum_{i=1}^n \log x_i$$

Igualando $S(\theta|\mathbf{x}) = 0$, encontramos que el estimador máximo verosimil $\hat{\theta}$ es:

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \log x_i}$$

Si este estimador fuese insesgado, podríamos utilizar el hecho de que es función de un estadístico suficiente para afirmar que es UMVUE (por Lehman-Scheffé). No obstante, se puede demostrar que dicha propiedad no se cumple[7].

17.3 Inciso c

Por normalidad asintótica del estimador máximo verosímil:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, i(\theta)^{-1}\right)$$

Entonces, para cualquier estimador máximo verosímil podemos construir un intervalo de confianza aproximado para θ de significancia α como:

$$P\left(\hat{\theta} - \frac{Z_{1-\alpha/2}}{\sqrt{ni(\hat{\theta})}} \leq \theta \leq \hat{\theta} + \frac{Z_{1-\alpha/2}}{\sqrt{ni(\hat{\theta})}}\right) \approx 1 - \alpha$$

Notar que podemos reemplazar $ni(\theta) = I(\theta)$. Por lo tanto, los límites del intervalo de confianza aproximado para θ de significancia α son:

$$IC_\alpha(\theta) \approx \hat{\theta} \pm Z_{1-\alpha/2}\frac{\hat{\theta}}{\sqrt{n}} = \hat{\theta}\left(1 \pm \frac{Z_{1-\alpha/2}}{\sqrt{n}}\right)$$

Donde $Z_{1-\alpha/2}$ es tal que:

$$\Phi(Z_{1-\alpha/2}) = 1 - \alpha/2$$

Siendo Φ la función de distribución de una variable normal estándar.

17.4 Inciso d

Para un nivel de significancia $\alpha = 5\%$, tenemos $Z_{1-\alpha/2} = 1.96$. Por lo tanto, para $n = 100$ y $\hat{\theta} = 7$ tenemos:

$$IC_{\alpha=0.05}(\theta) \approx 7\left(1 \pm \frac{1.96}{\sqrt{100}}\right) = [5.628, 8.372]$$

18 Ejercicio 18

18.1 Inciso a

Dado que las x están fijas, es fácil ver que Y es un modelo de locación respecto a ε con $\mu = \beta x$. Por lo tanto,

$$Y \sim N(\beta x, \sigma^2)$$

Dado que las x están dadas, el modelo está parametrizado por β y σ^2.

[7]Para ver una demostración de su sesgo, puede ver el ejemplo 6.2.4 de Hogg & Craig. *Introduction to Mathematical Statistics*

18.2 Inciso b

Dada la distribución de Y, la verosimilitud y log-verosimilitud para $\theta = (\beta, \sigma^2)$ están dada por:

$$L(\theta|\mathbf{y},\mathbf{x}) = \left(\frac{1}{2\sqrt{\sigma^2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2\right)$$
$$l(\theta|\mathbf{y},\mathbf{x}) = -\frac{n}{2}\left[\log 2\pi + \log\sigma^2\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2$$

Entonces, de las condiciones de primer orden respecto a β, σ^2:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}x_i)^2$$

Utilizando que $\sum_{i=1}^{n} x_i^2/n = 1$, podemos simplificar estas expresiones a:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{n}$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i^2) - \hat{\beta}^2$$

18.3 Inciso c

Utilizando que el modelo normal es una familia exponencial, calculamos $I(\theta)$ a partir de la segunda derivada de la log-verosimilitud:

$$\begin{aligned} I(\beta) &= -E\left[\frac{\partial^2}{\partial\beta^2}\log f(\mathbf{x};\beta,\sigma^2)|\mathbf{x}\right] \\ &= -E\left[-\frac{\sum_{i=1}^{n} x_i^2}{\sigma^2}|\mathbf{x}\right] \\ &= E\left[\frac{n}{\sigma^2}|\mathbf{x}\right] = \frac{n}{\sigma^2} \end{aligned}$$

Donde el último paso proviene de $\sum_{i=1}^{n} x_i^2/n = 1$. Notar que $\mathbf{x}$ está fijo, con lo cual condicionamos sobre su valor al tomar esperanza y calcular varianza.

Para saber si $\hat{\beta}$ es eficiente, calculamos su varianza y la comparamos con la Cota de Crámer Rao.

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i | \mathbf{x}\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} x_i^2 \mathrm{Var}(y_i|\mathbf{x}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} x_i^2 \mathrm{Var}(\beta x_i + \varepsilon|\mathbf{x}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} x_i^2 \left(\mathrm{Var}(\beta x_i|\mathbf{x}) + \mathrm{Var}(\varepsilon_i|\mathbf{x})\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} x_i^2 \sigma^2 \\
&= \frac{\sigma^2}{n} = I(\beta)^{-1}
\end{aligned}$$

Con lo cual $\mathrm{Var}(\hat{\beta})$ alcanza la Cota de CR, por lo cual es eficiente.

18.4 Inciso c

Para encontrar la distribución de $\hat{\beta}$ notar que:

$$\begin{aligned}
\hat{\beta} &= \frac{1}{n}\sum_{i=1}^{n} y_i x_i \\
&= \frac{1}{n}\sum_{i=1}^{n} \beta x_i^2 + x_i \varepsilon_i \\
&= \beta + \sum_{i=1}^{n} \frac{x_i}{n}\varepsilon_i
\end{aligned}$$

Entonces, condicional en $\mathbf{x}$, $\hat{\beta}$ es una combinación lineal de ε_i provenientes de una población $N(0,1)$. Por lo tanto, $\hat{\beta}$ se distribuye normal con media y varianza:

$$\begin{aligned}
E(\hat{\beta}|\mathbf{x}) &= \beta + \sum_{i=1}^{n} \frac{x_i}{n} E(\varepsilon_i|\mathbf{x}) = \beta \\
\mathrm{Var}(\hat{\beta}|\mathbf{x}) &= \frac{\sigma^2}{n} \text{ (por el inciso anterior)}
\end{aligned}$$

18.5 Inciso e

Dado que $\hat{\beta}$ es insesgado, su ECM es igual a:

$$ECM(\hat{\beta}, \beta) = \mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{n}$$

Dado que su varianza tiende a cero cuando $n \to \infty$, $\hat{\beta}$ es consistente.

19 Ejercicio 19

19.1 Inciso a

Para este modelo, la log-verosimilitud de θ está dada por:

$$l(\theta|\mathbf{x}) = -n\log 2 + 3n\log\theta + 2\sum_{i=1}^{n}\log x_i - \theta\sum_{i=1}^{n}x_i$$

Entonces, la condición de primer orden de la maximización de la log-verosimilitud está dada por:

$$\begin{aligned}\frac{\partial}{\partial\theta}l(\theta|\mathbf{x}) &= 0\\ \frac{3n}{\theta} - \sum_{i=1}^{n}x_i &= 0\\ \Rightarrow \hat{\theta} &= \frac{3}{\bar{x}}\end{aligned}$$

19.2 Inciso b

Sabemos por normalidad asintótica del estimador máximo verosímil que a medida que $n \to \infty$ su varianza converge a la inversa de la matriz de información. Es fácil mostrar que este modelo corresponde a una familia exponencial. Por tanto:

$$\begin{aligned}I(\theta) &= -E\left[\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{x};\theta)|\mathbf{x}\right]\\ &= -E\left[-\frac{3n}{\theta^2}\right]\\ &= \frac{3n}{\theta^2}\end{aligned}$$

Entonces, la varianza asintótica de $\hat{\theta}$ es igual a $I(\theta)^{-1} = \theta^2/3n$.

19.3 Inciso c

Utilizando la normalidad asintótica del estimador máximo verosímil:

$$\sqrt{I(\hat{\theta})}(\hat{\theta}-\theta) \xrightarrow{D} N(0,1)$$

Por lo tanto, podemos construir un intervalo de confianza para θ con un nivel de significancia de $\alpha = 5\%$ como:

$$\hat{\theta} \pm Z_{1-\alpha/2}\frac{\hat{\theta}}{\sqrt{3n}}$$

Reemplazando $Z_{0.975} = 1.96$ y operando, los límites del intervalo pueden escribirse como:

$$\hat{\theta}\left(1 \pm \frac{1.96}{\sqrt{3n}}\right)$$

19.4 Inciso d

Para los datos provistos, $\bar{x} = 1.728 \Rightarrow \hat{\theta} \approx 1.736$. Por lo tanto, el intervalo de confianza estimado se vuelve:

$$1.736\left(1 \pm \frac{1.96}{\sqrt{3*250}}\right) \approx [1.61, 1.86]$$

20 Ejercicio 20

Para este ejercicio, podemos proponer el modelo $\{X_i\}_{i=1}^n \overset{iid}{\sim} Ber(p)$, donde cada variable aleatoria representa la intención de voto de un habitante de la CABA y el éxito se define como intención de votar por A.

20.1 Inciso a

Dado que para el modelo propuesto $n_{si} = \sum_{i=1}^n x_i$, podemos escribir al estimador $\hat{p}_{JP}$ como:

$$\hat{p}_{JP} = \frac{1}{n+20}\left(10 + \sum_{i=1}^n x_i\right)$$

Por lo tanto:

$$\begin{aligned} E(\hat{p}_{JP}) &= \frac{1}{n+20}\left(10 + \sum_{i=1}^n E(x_i)\right) \\ &= \frac{1}{n+20}\left(10 + \sum_{i=1}^n p\right) \\ &= \frac{np+10}{n+20} \end{aligned}$$

El sesgo de $\hat{p}_{JP}$ está dado por:

$$\text{Sesgo}_p(\hat{p}_{JP}) = \frac{np + 10 - np - 20p}{n+20} = \frac{10-20p}{n+20} \neq 0 \text{ para todo } p \neq 0.5$$

La varianza de $\hat{p}_{JP}$ está dada por:

$$\begin{aligned} \text{Var}(\hat{p}_{JP}) &= \left(\frac{1}{n+20}\right)^2 \sum_{i=1}^n \text{Var}(x_i) \\ &= \frac{np(1-p)}{(n+20)^2} \end{aligned}$$

Por lo tanto, a partir de la descomposición del error medio cuadrático:

$$\begin{aligned} ECM(\hat{p}_{JP}, p) &= \text{Sesgo}_p^2(\hat{p}_{JP}) + \text{Var}(\hat{p}_{JP}) \\ &= \left(\frac{10-20p}{n+20}\right)^2 + \frac{np(1-p)}{(n+20)^2} \end{aligned}$$

Finalmente, queremos comparar el error medio cuadrático de ambos estimadores. En el caso del estimador insesgado $\hat{p}$, su error medio cuadrático es igual a su varianza:

$$ECM(\hat{p}, p) = \text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

La diferencia de ECM es igual a:

$$\begin{aligned} ECM(\hat{p}_{JP}, p) - ECM(\hat{p}, p) &= \left(\frac{10-20p}{n+20}\right)^2 + \frac{np(1-p)}{(n+20)^2} - \frac{p(1-p)}{n} \\ &= \left(\frac{10-20p}{n+20}\right)^2 + p(1-p)\left(\frac{n^2-(n+20)^2}{(n+20)^2 n}\right) \\ &= \left(\frac{10-20p}{n+20}\right)^2 + p(1-p)\left(\frac{n^2-(n^2+20n+400)}{(n+20)^2 n}\right) \\ &= \frac{1}{(n+20)^2}\left((10-20p)^2 - p(1-p)\frac{400+20n}{n}\right) \end{aligned}$$

Reemplazando $n = 1000$ de acuerdo a la información provista en el ejercicio, la diferencia se vuelve:

$$ECM(\hat{p}_{JP}, p) - ECM(\hat{p}, p) = \frac{1}{(1020)^2}\left((10-20p)^2 - p(1-p)\frac{20400}{1000}\right)$$

Si igualamos a cero y buscamos las raíces de la ecuación, encontramos que los estimadores son equivalentes en términos de ECM para valores aproximados de p de 0.38 y 0.61. Valores por fuera de este intervalo favorecen al estimador insesgado, mientras que valores mas cercanos a 0.5 favorecen a $\hat{p}_{JP}$.

21 Ejercicio 21

21.1 Inciso a

Para este modelo, la log-verosimilitud está dada por:

$$l(\alpha|\mathbf{t}) = n\log\alpha + \sum_{i=1}^{n}\log t_i - \frac{\alpha}{2}\sum_{i=1}^{n} t_i^2$$

Entonces, de la condición de primer orden del problema de maximización, se obtiene:

$$\hat{\alpha} = \frac{2n}{\sum_{i=1}^{n}\log t_i^2}$$

21.2 Inciso b

Para los valores provistos en el ejercicio, tenemos:

$$\hat{\alpha} = \frac{500}{510.58} \approx 0.979$$

Inferencia Estadística - Guia 3

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

Junio 2020

1 Ejercicio 1

Para este ejercicio, podemos pensar a la intención de pagar de cada pasajero como una variable Bernoulli parametrizada por p, la proporción de individuos dispuestos a pagar 5 dólares por acceso ilimitado a Internet durante vuelos de cabojate. Dado que el estimador máximo verosímil para p para un modelo Binomial está dado por $\hat{p} = \bar{X}$, podemos utilizar el Teorema del Límite Central para construir un intervalo de confianza del 95% aproximado para p.

Entonces, usando que:

$$\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{D} N(0,1)$$

Podemos afirmar que:

$$P\left(Z_{\alpha/2} \leq \frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \leq Z_{1-\alpha/2}\right) \approx 1-\alpha$$

Donde $Z_{\alpha/2}$ es el cuantil $\alpha/2$ de la distribución normal estándar.

Por lo tanto, (utilizando la simetría de la distribución normal) el intervalo de confianza aproximado del 95% para p es:

$$IC_{95\%}(p) = \hat{p} \pm Z_{0.975}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

Reemplazando $\hat{p} = 0.625$, $n = 200$ y $Z_{0.975} \approx 1.96$:

$$IC_{95\%}(p) = [0.558, 0.692]$$

La calidad de la aproximación depende de que n sea suficientemente grande y p no tome valores cerca de los extremos (0 o 1). Se suele establecer el siguiente criterio para juzgar la calidad de la aproximación:

$$\begin{cases} n\hat{p} \geq C \\ n(1-\hat{p}) \geq C \end{cases}$$

Donde C es alguna constante (generalmente 5 o 10). En nuestro caso, $n(1-\hat{p}) = 75$.

2 Ejercicio 2

Estamos interesados en construir un intervalo de confianza para la varianza de puntajes de los exámenes de una compañía, la cual denominaremos σ^2. Dado que desconocemos las características de la población bajo estudio, asumimos que los puntajes de los exámenes siguen una distribución $N(\mu, \sigma^2)$. De esta manera, podemos utilizar la definición de un intervalo de confianza para σ^2 en una distribución normal (página 20 de Slide 6):

$$IC_{(1-\alpha)\%}(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]$$

En este caso tenemos $s^2 = 108.16$, $n = 18$ y $\alpha = 0.1$. Los valores críticos de la distribución chi-cuadrado con 17 grados de libertad son:

$$\chi^2_{17,0.05} \approx 8.67$$
$$\chi^2_{17,0.95} \approx 27.58$$

Por lo tanto, el intervalo de confianza será:

$$IC_{90\%}(\sigma^2) = [66.65, 212.03]$$

Para arribar a este resultado, debimos asumir que los puntajes de los exámenes siguen una distribución normal. Algunas de las estrategias disponibles para verificar el cumplimiento de este supuesto son análisis gráfico de la distribución empírica (ej. histograma o *QQ-plot*) o algún test de normalidad como el Jarque-Bera o Chi-Cuadrado de Pearson.

3 Ejercicio 3

3.1 Inciso a

Dada la naturaleza aleatoria del experimento propuesto, resulta razonable asumir que los tiempos de reacción observados para las luces rojas y verdes resultan independientes. Por lo tanto, si asumimos que las poblaciones de tiempos de respuesta a luces de cada color se distribuyen normalmente, podemos escribir:

$$\{X_i\}_{i=1}^{n} \overset{iid}{\sim} N(\mu_X, \sigma_X^2)$$
$$\{Y_i\}_{i=1}^{m} \overset{iid}{\sim} N(\mu_Y, \sigma_Y^2)$$

Donde X e Y representan la población de tiempos de respuesta a luces de color roja y verde respectivamente. El parámetro de interés para el cual queremos obtener un intervalo de confianza es la diferencia de medias $\Delta = \mu_X - \mu_Y$.

3.2 Inciso b

Si estamos dispuestos a asumir que la varianza de los tiempos de respuesta no difiere según se trate de luces rojas o verdes (es decir, $\sigma_X^2 = \sigma_Y^2$), podemos utilizar la versión simplificada de la página 22 del Slide 6 para construir el intervalo de confianza:

$$P\left(\widehat{\Delta} - t_{m+n-2,\alpha/2}\tilde{S}\sqrt{\frac{1}{n} + \frac{1}{m}} \leq \Delta \leq \widehat{\Delta} + t_{m+n-2,\alpha/2}\tilde{S}\sqrt{\frac{1}{n} + \frac{1}{m}}\right) = 1 - \alpha$$

Donde $\hat{\Delta} = \bar{X} - \bar{Y}$, m y n son el tamaño de cada muestra, y $\tilde{S}^2 = \frac{(n-1)S_X^2 + (m-1)s_M^2}{n+m-2}$.
A partir de los valores provistos, calculamos:

$$\bar{X} = 0.37, \quad \bar{Y} = 0.4325$$
$$n = 8, \quad m = 8$$
$$s_X^2 \approx 0.0126, \quad S_Y^2 \approx 0.0137$$

Entonces, para un valor crítico[1] de $t_{14,0.975} \approx 2.144$, el intervalo de confianza del 95% para Δ es:

$$IC_{95\%}(\Delta) = -0.0625 \pm 2.144 * 0.114\sqrt{1/8} = [-0.023, 0.149]$$

[1] Recordar que la distribución t es simétrica.

3.3 Inciso c

Al igual que en el ejercicio 2, deberíamos corroborar la normalidad de la población de tiempos de respuesta. Por otro lado, también asumimos que la varianza de ambas poblaciones era igual. Para justificar este supuesto, podríamos utilizar un test F de igualdad de varianzas. Bajo la hipótesis nula de igualdad de varianzas, el ratio de varianzas muestrales de dos poblaciones normales debería seguir una distribución $F_{n-1,m-1}$.

4 Ejercicio 4

Llamemos X_i e Y_i las mediciones de colesterol para el individuo i antes y después del tratamiento con la droga. Para este experimento, está claro que estas variables no son independientes, dado que el nivel de colesterol después del aplicar el tratamiento estará relacionado con la condición inicial del paciente.

Adicionalmente, dado que se nos informa que por experiencias anteriores la distribución del cambio en la cantidad de colesterol sigue una distribución normal, podemos utilizar la metodología vista en clase para intervalos de confianza para diferencia de medias en muestras apareadas[2].

Sea $D_i = X_i$, podemos construir el pivote:

$$g\left(\bar{D}; S_D^2, \mu_D\right) = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} \to_F N(0,1)$$

Donde $\bar{D}$ es la el promedio de diferencias en la muestra y S_D el estimador insesgado de la varianza de D. Por lo tanto, podemos construir el intervalo de confianza usando:

$$P\left(\bar{D} - z_{\alpha/2} S_D/\sqrt{n} \leq \mu_D \leq \bar{D} + z_{\alpha/2} S_D/\sqrt{n}\right) = 1 - \alpha$$

Notar que este intervalo no es aproximado, dado que se asume que se satisface el supuesto de normalidad. Nos interesa encontrar el valor de n tal que la *precisión* del intervalo sea igual a 5mg. En este caso, la precisión P del intervalo es igual a:

$$P = \left(\bar{D} + z_{\alpha/2} S_D/\sqrt{n}\right) - \left(\bar{D} - z_{\alpha/2} S_D/\sqrt{n}\right) = 2\frac{z_{\alpha/2} S_D}{\sqrt{n}}$$

Entonces, para $S_D = 4$ y $z_{0.975} \approx 1.96$, encontramos que:

$$5 = 2 * \frac{1.96 * 4}{\sqrt{n}} \iff n = 9.8334496$$

Por lo tanto, cualquier valor de n superior a 10 nos asegurará un intervalo de confianza al 95% con una precisión menor a 5mg.

5 Ejercicio 5

En este escenario, estamos tratando de estimar el intervalo de confianza para la diferencia de medias en muestras independientes. Si asumimos normalidad de las variables aleatorias (la cual podemos tratar de verificar utilizando alguno de los test mencionados en el ejercicio 2), podemos aplicar la metodología planteada en la página 31 del Slide 6. Sea $\hat{\Delta} = \bar{x}_H - \bar{x}_M = 14$, podemos construir el pivote para Δ:

$$g(\widehat{\Delta}, \Delta) = \frac{\widehat{\Delta} - \Delta}{\sqrt{\frac{s_H^2}{n} + \frac{s_M^2}{m}}} \xrightarrow{F} t_v$$

$$\text{con } v = \frac{\left(\frac{s_H^2}{n_H} + \frac{s_M^2}{n_M}\right)^2}{\frac{s_H^4}{n_H^2(n_H-1)} + \frac{s_M^4}{n_M^2(n_M-1)}}$$

[2] Ver página 29 del Slide 6.

Por lo tanto, el intervalo de confianza queda definido por:

$$P\left(\widehat{\Delta} - t_{v,\alpha/2}\sqrt{\frac{S_H^2}{n_H} + \frac{S_M^2}{n_M}} \leq \Delta \leq \widehat{\Delta} + t_{v,\alpha/2}\sqrt{\frac{S_H^2}{n_H} + \frac{S_M^2}{n_M}}\right) \approx 1 - \alpha$$

Reemplazando $s_H = 19.13$, $s_M = 18.83$, $n_H = 151$ y $n_M = 108$, tenemos $v \approx 232$. Por lo tanto, $t_{232,0.975} \approx 1.97$ y los límites del intervalo de confianza de interés serán:

$$IC(\Delta)_{95\%} = 14 \pm 1.97\sqrt{\frac{19.13^2}{151} + \frac{18.83^2}{108}} \approx [9.29,\ 18.70]$$

6 Ejercicio 6

Este ejercicio es similar al Ejercicio 2, en tanto queremos calcular un intervalo de confianza para la varianza poblacional. Al igual que en aquel ejercicio, asumimos que las medidas de contaminación atmosférica siguen una distribución normal y construimos el intervalo de confianza para σ^2.

En este caso, $\alpha = 0.05$, $n = 10$, $s^2 \approx 1.409$. Para estos valores de α y n, los valores críticos de la distribución Chi-cuadrado son:

$$\chi^2_{9,0.025} \approx 2.7$$
$$\chi^2_{9,0.975} \approx 19.02$$

Siguiendo lo visto en el ejercicio 2, el intervalo de confianza al 95% para σ^2 es:

$$IC_{95\%}(\sigma^2) = [0.666, 4.695]$$

Dado que la distribución Chi-cuadrado no es simétrica, no podemos invocar el teorema de la página 17 del Slide 6 para afirmar que sea el intervalo de máxima precisión.

Al graficar la distribución chi-cuadrado con 9 grados de libertad[3] podemos ver que un intervalo con mayor precisión podría construirse corriendo los límites hacia la izquierda.

7 Ejercicio 7

Podemos escribir:

$$P(L \leq \theta \leq U) = P(\theta \leq U) - P(\theta \leq L)$$

Por lo tanto, dado que $P(\theta \leq L) = 1 - P(L \leq \theta) = \alpha_L$, se ve fácilmente que:

$$P(L \leq \theta \leq U) = 1 - \alpha_U - \alpha_L$$

8 Ejercicio 8

8.1 Inciso a

Modelo 1

En el caso del modelo 1, es fácil ver que $X \sim U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Por lo tanto, podemos construir el pivote:

$$g(X, \theta) = X - \theta \sim U\left(-\frac{1}{2}, \frac{1}{2}\right).$$

[3] En R, puede utilizar el comando: `curve(dchisq(x, 9), from=0, to=40)`

Modelo 2

Por otro lado, para el modelo 2, podemos utilizar nuestro conocimiento de modelos de locación-escala para proponer el pivote:

$$g(X, \theta) = \frac{X}{\theta} \sim f_g(x) = \theta f_2(\theta x)$$

Por lo cual:

$$f_g(x) = \begin{cases} 2x & \text{si } 0 < \frac{x}{\theta} < 1, \text{ si } \theta > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

8.2 Inciso b

Modelo 1

Dado que la distribución uniforme es simétrica, podemos construir el intervalo de confianza para θ a partir de:

$$P(q_{\alpha/2} \leq X - \theta \leq q_{1-\alpha/2}) = 1 - \alpha$$

Donde $q_{\alpha/2}, q_{1-\alpha/2}$ son los cuantiles relevantes para la distribución $U(-0.5, 0.5)$. Es fácil ver que para esta distribución:

$$q_{1-\alpha/2} = \frac{1-\alpha}{2}$$

Por lo tanto:

$$IC_{1-\alpha}(\theta) = X \pm \frac{1-\alpha}{2}$$

Modelo 2

Para el modelo 2, tenemos:

$$P(q_{\alpha/2} \leq X/\theta \leq q_{1-\alpha/2}) = 1 - \alpha$$

Podemos utilizar nuestro conocimiento de $f_g(x)$ para obtener la función de distribución:

$$F_g(x) = x^2$$

Notar que esta distribución no es simétrica. Por lo tanto, los cuantiles críticos serán:

$$q_{\alpha/2} = \sqrt{\alpha/2}$$
$$q_{1-\alpha/2} = \sqrt{1-\alpha/2}$$

Por lo tanto, el intervalo de confianza de nivel $1 - \alpha$ está dado por:

$$IC_{1-\alpha}(\theta) = \left[\frac{x}{\sqrt{1-\alpha/2}} \ , \ \frac{x}{\sqrt{\alpha/2}} \right]$$

9 Ejercicio 9

9.1 Inciso a

Notar que este modelo se trata de la distribución Geométrica, que modela la cantidad de fracasos ocurridos hasta obtener un éxito en una secuencia de experimentos Bernoulli.

Utilizamos la normalidad asintótica de los estimadores máximo verosímiles para encontrar un pivote aproximado $g(\hat{\theta}, \theta)$ de la forma:

$$g(\hat{\theta}, \theta) = \sqrt{ni(\theta)}(\hat{\theta} - \theta) \xrightarrow{D} N(0, 1)$$

La log-verosimilitud para este modelo está dada por:

$$L(p|\mathbf{x}) = \left(\sum_{i=1}^{n} x_i - n \right) \log(1-p) + n \log p$$

Maximizando esta función encontramos que $\hat{p} = \bar{x}^{-1}$.

Es fácil ver que este modelo pertenece a una familia exponencial, por lo cual podemos utilizar la siguiente expresión para la $\hat{p}$ Información de Fisher para una observación:

$$i(p) = -E\left(\frac{\partial^2}{\partial p^2}\log f(x,p)\right) = \frac{1}{p(1-p)} + \frac{1}{p^2} = \frac{1}{p^2(1-p)}$$

Por lo tanto, encontramos el pivote y su distribución asintótica:

$$g(\hat{p},p) = \frac{\sqrt{n}}{p\sqrt{1-p}}(\hat{p}-p) \xrightarrow{D} N(0,1)$$

9.2 Inciso b

Dado que la $I(p) = ni(p)$ es una función continua para $p \in (0,1]$, podemos utilizar lo visto en clase[4] para definir el intervalo de confianza de nivel $1-\alpha$:

$$P\left(\hat{p} - z_{1-\alpha/2}\frac{\hat{p}\sqrt{1-\hat{p}}}{\sqrt{n}} \leq p \leq \hat{p} + z_{\alpha/2}\frac{\hat{p}\sqrt{1+\hat{p}}}{\sqrt{n}}\right) = 1-\alpha$$

Reemplazando $\hat{p} = \bar{x}^{-1}$ encontramos que los límites del intervalo son:

$$IC(p)_{1-\alpha} = \bar{x}^{-1}\left(1 \pm z_{1-\alpha/2}\sqrt{\frac{1-\bar{x}^{-1}}{n}}\right)$$

9.3 Inciso c

Para los valores provistos, $\hat{p} = \bar{x}^{-1} = 0.02$ y $z_{0.975} \approx 1.96$, por lo tanto:

$$IC(p)_{95\%} = 0.02\left(1 \pm 1.96\sqrt{\frac{0.98}{100}}\right) \approx [0.0161, 0.0238]$$

10 Ejercicio 10

10.1 Inciso a

Supongamos que observamos $Y = X + \mu$ tal que:

$$Y \sim f_Y(y) = f_X(y-\mu)$$

Por lo tanto, por propiedad de familia exponencial:

$$Y - \mu \sim f_X(y)$$

Por lo tanto, $g(Y,\mu) = Y - \mu$ es una función continua cuya distribución no depende de μ, por lo cual califica como pivote para μ. Si definimos el cuantil q_α tal que:

$$F_X(q_\alpha) = \alpha$$

Podemos escribir el intervalo de confianza para μ de nivel $1-\alpha$ como:

$$IC_{1-\alpha}(\mu) = [y - q_{1-\alpha/2},\ y - q_{\alpha/2}]$$

[4] Ver página 26 de Slide 6.

10.2 Inciso b

En este caso, tenemos $Y = \sigma X$ tal que:

$$Y \sim f_Y(y) = f_X(y/sigma)/sigma$$

Utilizando propiedades de familia exponencial:

$$Y/\sigma \sim f_X(y)$$

Con lo cual Y/σ es un pivote válido. Entonces, para la anterior definición del cuantil q_α, tenemos:

$$IC_{1-\alpha}(\sigma) = \left[\frac{y}{q_{1-\alpha/2}},\ \frac{y}{q_{\alpha/2}}\right]$$

11 Ejercicio 11

11.1 Inciso a

Para este modelo, la log-verosimilitud es igual a:

$$l(\theta|\mathbf{x}) = n\log(\theta+1) + \theta\sum_{i=1}^{n}\log x_i$$

De la condición de primer orden la maximización de la log- verosimilitud obtenemos:

$$\hat{\theta} = -\frac{\sum_{i=1}^{n}\log x_i}{n} - 1$$

11.2 Inciso b

Para construir el intervalo de confianza aproximado para θ, notar que la Información de Fisher $I(\theta)$ es igual a:

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} l(\theta|\mathbf{x})\right] = \frac{n}{\theta+1}$$

Por lo tanto, utilizando lo visto en clase[5], podemos plantear el intervalo de confianza aproximado de nivel $1-\alpha$ a partir de:

$$P\left(\hat{\theta} - z_{1-\alpha/2}\frac{\hat{\theta}+1}{\sqrt{n}} \leq \theta \leq \hat{\theta} + z_{1-\alpha/2}\frac{\hat{\theta}+1}{\sqrt{n}}\right) = 1-\alpha$$

Por lo tanto, el intervalo de confianza queda dado por:

$$IC_{1-\alpha}(\theta) = \hat{\theta} \pm z_{1-\alpha/2}\frac{\hat{\theta}+1}{\sqrt{n}}$$

11.3 Inciso c

Para los valores provistos, $\hat{\theta} = -0.1$ y $z_{0.975} \approx 1.96$. Por lo tanto, el intervalo de confianza de nivel 95% será:

$$IC_{95\%}(\theta) = -0.1 \pm 1.96\frac{0.9}{\sqrt{500}} \approx [-0.178,\ -0.021]$$

[5]Ver página 26 de Slide 6

12 Ejercicio 12

12.1 Inciso a

Utilizando lo provisto en la consigna, sabemos que:

$$P\left(\frac{2\sum_{i=1}^{n} x_i}{q_{1-\alpha/2}} \leq \theta \leq \frac{2\sum_{i=1}^{n} x_i}{q_{\alpha/2}}\right) = 1-\alpha$$

Donde definimos q_α como el cuantil α de una distribución χ^2_{2n}. Por lo tanto, los límites del intervalo serán:

$$IC_{1-\alpha}(\theta) = \left[\frac{2\sum_{i=1}^{n} x_i}{q_{1-\alpha/2}},\ \frac{2\sum_{i=1}^{n} x_i}{q_{\alpha/2}}\right]$$

12.2 Inciso b

Para los valores provistos:

$$\begin{aligned} n &= 20 \\ \sum_{i=1}^{20} x_i &= 2023 \\ q_{0.975} &\approx 59.34 \\ q_{0.025} &\approx 24.43 \end{aligned}$$

Por lo cual el intervalo de confianza será:

$$IC_{95\%}(\theta) = [68.18,\ 165.61]$$

13 Ejercicio 13

13.1 Inciso a

Estamos interesados en conocer la distribución del error de predicción $\bar{X} - X_{n+1}$. Sabemos de lo visto al inicio de la materia[6] que si $X \sim N(\mu, \sigma^2)$, entonces $\bar{X} \sim N(\mu, \sigma^2/n)$. Por otro lado, $X_{n+1} \sim N(\mu, \sigma^2)$.

Se puede demostrar que para dos variables aleatorias independientes $X_1 \sim N(\mu_1, \sigma_1^2)$ y $X_2 \sim N(\mu_2, \sigma_2^2)$:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$$

En nuestro caso[7]:

$$\bar{X} - X_{n+1} \sim N\left(0, \frac{\sigma^2}{n} + \sigma^2\right)$$

13.2 Inciso b

Sabemos por definición de la distribución t que para una secuencia de variables aleatorias normales se cumple:

$$\frac{\bar{X} - \mu}{\sqrt{\mathrm{Var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \Rightarrow \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

[6] Ver página 32 de Slide 1.

[7] Notar que $(-X_{n+1}) \sim N(-\mu, \sigma^2)$.

Dado lo aprendido para la error de predicción en el inciso a, tenemos:

$$\frac{(\bar{X} - X_{n+1}) - E(\bar{X} - X_{n+1})}{\sqrt{\mathrm{Var}(\bar{X} - X_{n+1})}} = \frac{(\bar{X} - X_{n+1}) - 0}{\sqrt{\mathrm{Var}(\bar{X} - X_{n+1})}} = \frac{\bar{X} - X_{n+1}}{\sigma\sqrt{n^{-1}+1}} \sim N(0,1)$$

Por lo tanto:

$$\frac{\bar{X} - X_{n+1}}{S\sqrt{n^{-1}+1}} \sim t_{n-1}$$

Entonces, $c = (\sqrt{n^{-1}+1})^{-1}$.

13.3 Inciso c

Dado lo aprendido en el inciso anterior, sabemos que podemos construir el siguiente pivote para X_9:

$$g(X_9, \bar{X}) = \frac{\bar{X} - X_9}{S\sqrt{8^{-1}+1}} \sim t_7$$

Por lo tanto, para $\alpha = 0.2$

$$P\left(t_{7,0.1} \leq \frac{\bar{X} - X_9}{S\sqrt{8^{-1}+1}} \leq t_{7,0.9}\right) = 0.8$$

Donde $t_{7,\alpha}$ es el cuantil α de una distribución t con 7 grados de libertad. Operando y utilizando la simetría de la distribución t, obtenemos:

$$P(\bar{X} - t_{7,0.9}S\sqrt{8^{-1}+1} \leq X_9 \leq \bar{X} + t_{7,0.9}S\sqrt{8^{-1}+1}) = 0.8$$

Con lo cual $k = t_{7,0.9}\sqrt{8^{-1}+1}$.

14 Ejercicio 14

De las propiedades de suma de variables normales vistas en el ejercicio anterior, es fácil ver que[8].

$$g(\widehat{\Delta}, \Delta) = \frac{\widehat{\Delta} - \Delta}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1) \tag{1}$$

Pero dado que $\sigma_X = 3\sigma_Y$:

$$g(\widehat{\Delta}, \Delta) = \frac{\widehat{\Delta} - \Delta}{\sigma_X\sqrt{n^{-1} + (3m)^{-1}}} \sim N(0,1) \tag{2}$$

Por lo tanto, podemos construir el pivote alternativo:

$$g(\widehat{\Delta}, \Delta) = \frac{\widehat{\Delta} - \Delta}{\tilde{S}_X\sqrt{n^{-1} + (3m)^{-1}}} \sim t_{n+m-2} \tag{3}$$

Donde $\tilde{S}_X^2 = \frac{(n-1)S_X^2 + (m-1)3S_Y^2}{m+n-2}$ es un promedio ponderado de estimadores insesgados de σ_X^2. De esta manera, volvemos al caso general para medias equivalentes. Por lo tanto, la forma general del intervalo de confianza del 95% quedará determinada por:

$$P\left(\widehat{\Delta} - t_{m+n-2,0.975}\tilde{S}\sqrt{\frac{1}{n} + \frac{1}{3m}} \leq \Delta \leq \widehat{\Delta} + t_{m+n-2,0.975}\tilde{S}\sqrt{\frac{1}{n} + \frac{1}{3m}}\right) = 95\% \tag{4}$$

[8] Alternativamente, ver página 22 del Slide 6. Un análisis detallado del caso puede encontrarse en el artículo: Niwitpong & Niwitpong (2010) "Confidence Interval for the Difference of Two NormalPopulation Means with a Known Ratio of Variances".

15 Ejercicio 15

15.1 Inciso a

De acuerdo a la distribución provista para el pivote, tenemos:

$$P\left(\frac{S_B^2}{S_A^2}q_{\alpha/2} \leq \tau \leq \frac{S_B^2}{S_A^2}q_{1-\alpha/2}\right) = 1-\alpha$$

Por lo cual el intervalo de confianza estimado de nivel α será:

$$IC_{1-\alpha}(\tau) = \left[\frac{S_B^2}{S_A^2}q_{\alpha/2},\ \frac{S_B^2}{S_A^2}q_{1-\alpha/2}\right]$$

15.2 Inciso b

El intervalo de confianza de nivel $1-\alpha$ construido en el punto anterior no es único, dado que podemos elegir diferentes cuantiles tales que el intervalo contenga $(1-\alpha)\%$ de los valores dentro. Dado que la distribución F de Snedecor suele no ser simétrica, sabemos por lo visto en clase que el intervalo de confianza construido probablemente no sea el de mayor precisión.

15.3 Inciso c

Para los valores provistos:

$$\begin{aligned}\frac{s_B^2}{s_A^2} &\approx 2.35\\ q_{0.025} &\approx 0.659\\ q_{0.975} &\approx 1.531\end{aligned}$$

Por lo tanto el intervalo construido será:

$$IC_{95\%}(\tau) = [1.55,\ 3.59]$$

15.4 Inciso d

Sabemos que el parámetro τ es igual al ratio entre varianzas de las notas de cada población. Si ambas poblaciones fueran equivalentes en términos de varianza de calificationes, tendríamos $\tau = 1$. Por definición de intervalo de confianza, sabemos que el 95% de los intervalos construidos en forma equivalente al obtenido en el inciso anterior contendrán el verdadero valor de τ. Por lo tanto, el resultado obtenido seria indicativo de que la varianza de notas de la población B es mayor que la de la población A. No obstante, debemos ser conscientes de que dichos valores pueden corresponder al 5% de intervalos que no contiene al verdadero valor de τ.

16 Ejercicio 17

Recordemos que un intervalo de confianza de **Wald** de nivel $1-\alpha$ para θ está dado por:

$$IC_{1-\alpha}(\theta) = \hat{\theta} \pm \frac{z_{1-\alpha/2}}{\sqrt{I(\hat{\theta})}}$$

Por otro lado, el intervalo de confianza construido en base al **método de verosimilitud** es igual a:

$$IC_{1-\alpha}(\theta) = \left\{\theta : l(\theta|\mathbf{x}) \geq l(\hat{\theta}|\mathbf{x}) - \frac{1}{2}c_1(\alpha)\right\}$$

Donde $c_1(\alpha)$ es el cuantil α de una distribución Chi cuadrado con 1 grado de libertad. Notar que tras ciertas operaciones algebráicas, podemos escribir este intervalo como:

$$IC_{1-\alpha}(\theta) = \left\{\theta : 2\log\left(\frac{L(\hat{\theta}|\mathbf{x})}{L(\theta|\mathbf{x})}\right) \leq c_1(\alpha)\right\}$$

Esta expresión introduce el concepto del logaritmo de ratio de verosimilitudes (*log likelihood-ratio*), que se hará presente más adelante cuando estudiemos tests de hipótesis.

16.1 Modelo Exponencial

Para el modelo exponencial caracterizado por la función de densidad $f(x) = \lambda e^{-\lambda x}$ tenemos:

$$\begin{aligned} l(\lambda|\mathbf{x}) &= \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right) \\ \hat{\lambda} &= \bar{x}^{-1} \\ I(\lambda) &= \frac{n}{\lambda^2} \end{aligned}$$

Reemplazando en la expresión para el intervalo de **Wald**, tenemos:

$$IC_{1-\alpha}(\lambda) = \bar{x}^{-1}\left(1 \pm \frac{z_{1-\alpha/2}}{\sqrt{n}}\right)$$

Por otro lado, el intervalo construido en base al método de verosimilitud queda determinado por los valores de λ que satisfacen:

$$IC_{1-\alpha}(\lambda) = \left\{\lambda : \lambda^n \exp\left(-\lambda\sum_{i=1}^{n} x_i\right) \geq \bar{x}^{-n}\exp\left(-\bar{x}^{-1}\sum_{i=1}^{n} x_i\right) - \frac{1}{2}c_1(\alpha)\right\}$$

Estos dos ejemplos son representativos de las ventajas y desventajas de cada método:

- Los **intervalos de Wald** resultan fáciles de calcular y otorgan una buena aproximación si el tamaño de la muestra es suficientemente grande. No obstante, si la muestra no es lo suficientemente grande o la significatividad del intervalo (α) es muy pequeña, podría ser que el intervalo contenga valores inverosímiles del parámetro bajo análisis. En nuestro ejemplo, ello podría ocurrir si $z_{1-\alpha/2}/\sqrt{n}$ resultase mayor a 1.
- A diferencia de los intervalos de Wald, los **intervalos del método de verosimilitud** sólo se encuentran definidos para valores razonables del parámetro bajo análisis, al depender directamente de la función de verosimilitud. No obstante, en la mayoría de los casos se requiere utilizar métodos numéricos para encontrar los valores del parámetro que representan los límites del intervalo.

16.2 Modelo Poisson

Para el modelo Poisson caracterizado por $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, tenemos[9]:

$$\begin{aligned} l(\lambda|\mathbf{x}) &\propto -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln(\lambda) \\ \hat{\lambda} &= \bar{x} \\ I(\lambda) &= \frac{n}{\lambda} \end{aligned}$$

[9]Omitimos el término asociado a $x!$ dado que no depende de λ.

Por lo tanto, el intervalo de Wald estará dado por:

$$IC_{1-\alpha}(\lambda) = \bar{x} \pm z_{1-\alpha/2}\sqrt{\frac{\bar{x}}{n}}$$

El intervalo del método de verosimilitud estará dado por (luego de un poco de álgebra):

$$IC_{1-\alpha}(\lambda) = \left\{\lambda : -n\left(\lambda - \bar{x}\right) + \left(\sum_{i=1}^{n} x_i\right)\ln\left(\frac{\lambda}{\bar{x}}\right) \geq -\frac{1}{2}c_1(\alpha)\right\}$$

16.3 Modelo Binomial

Para el modelo binomial caracterizado por $f(x) = \binom{n}{x}p^x(1-p)^{n-x}$, tenemos[10]:

$$\begin{aligned} l(p|\mathbf{x}) &\propto \left(\sum_{i=1}^{n} x_i\right)\ln(p) + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p) \\ \hat{p} &= \bar{x} \\ I(\lambda) &= \frac{n}{p(1-p)} \end{aligned}$$

El intervalo de Wald esta dado por:

$$IC_{1-\alpha}(\lambda) = \bar{x} \pm z_{1-\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

El intervalo por método de verosimilitud estará dado por:

$$IC_{1-\alpha}(p) = \left\{p : \left(\sum_{i=1}^{n} x_i\right)\ln\left(\frac{p}{\bar{x}}\right) + \left(n - \sum_{i=1}^{n} x_i\right)\ln\left(\frac{1-p}{1-\bar{x}}\right) \geq -\frac{1}{2}c_1(\alpha)\right\}$$

[10]Omitimos el término que incluye la combinatoria dado que no depende de p. A su vez n representa el número de experimentos en el *total* de experimentos Bernouilli asociado a la secuencia de variables binomiales.

Inferencia Estadística - Guia 4

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

Agosto 2020

1 Ejercicio 1

1.1 Inciso a

Falso. El nivel de significación (α) es igual a la probabilidad de rechazar la hipótesis nula dado que esta sea cierta, es decir, la probabilidad de cometer un error de tipo I.

1.2 Inciso b

Falso. No podemos cuantificar la probabilidad de que H_0 sea cierta, sólo la probabilidad de alcanzar conclusiones erróneas condicional en que esta sea cierta o no.

1.3 Inciso c

Falso. Ya definimos anteriormente a α como la probabilidad de cometer un error de tipo I. La potencia de un test es la probabilidad de rechazar la hipótesis nula cuando la alternativa es cierta, es decir, el complemento de la probabilidad de cometer un error de tipo II.

1.4 Inciso d

Verdadero (a medias). Un error de tipo I ocurre cuando se rechaza *erróneamente* la hipótesis nula, es decir, condicional en que ésta sea cierta.

1.5 Inciso e

Verdadero. Si elijo un nivel de significancia mas pequeño, lo cual hará más probable aceptar la hipótesis nula, debo enfrentar una probabilidad mas alta de estár comentiendo un error de tipo II.

2 Ejercicio 2

2.1 Inciso a

Para $Q = 0$, la probabilidad de cometer un error de tipo I es igual a la probabilidad de rechazar la hipótesis nula si esta es cierta, es decir:

$$\alpha(\theta) = P_{\theta \leq 0}(\delta = 1) = P_{\theta \leq 0}(\bar{X} > 0)$$

Sabemos que $\frac{\bar{X}-\theta}{\sigma/\sqrt{n}} \sim N(0, 1)$, por lo tanto:

$$\alpha(\theta) = P_{\theta \leq 0}(\bar{X} > 0) = P_{\theta \leq 0}\left(\frac{\bar{X}-\theta}{\sigma/\sqrt{n}} > \frac{0-\theta}{\sigma/\sqrt{n}}\right) = 1 - P_{\theta \leq 0}\left(Z \leq -\theta\sqrt{n}\right) = 1 - \Phi_{\theta \leq 0}(-\theta\sqrt{n})$$

Donde Z es una variable normal estándar y Φ su CDF. Dado que $\alpha(\theta)$ es creciente en θ, podemos afirmar que la cota superior de $\alpha(\theta)$ es:

$$\sup_{\theta\leq 0}\alpha(\theta) = 1-\Phi(0) = 0.5$$

2.2 Inciso b

Para hallar la potencia del test, primero estimamos la probabilidad de error de tipo II:

$$\beta(\theta) = P_{\theta>0}(\delta = 0) = P_{\theta>0}(\bar{X} < Q)$$

Utilizando una vez más la normalidad de la media muestral de variables normales:

$$\beta(\theta) = P_{\theta>0}\left(\frac{\bar{X}-\theta}{\sigma/\sqrt{n}} < \frac{Q-\theta}{\sigma/\sqrt{n}}\right) = P_{\theta>0}\left(Z \leq \frac{Q-\theta}{\sigma/\sqrt{n}}\right) = \Phi_{\theta>0}\left(\frac{Q-\theta}{\sigma/\sqrt{n}}\right)$$

La potencia estará dada por $1-\beta(\theta)$.

2.3 Inciso c

De los incisos anteriores, tenemos:

$$\alpha(\theta) = 1-\Phi_{\theta\leq 0}\left(\frac{Q-\theta}{\sigma/\sqrt{n}}\right)$$
$$1-\beta(\theta) = 1-\Phi_{\theta>0}\left(\frac{Q-\theta}{\sigma/\sqrt{n}}\right)$$

Dado que el término que resta en ambas expresiones es creciente en Q, podemos afirmar que incrementar Q eleva tanto la probabilidad de cometer un ET I como la potencia del test. Esto

2.4 Inciso d

Queremos encontrar el valor de Q tal que:

$$\alpha(\theta) = 1-\Phi_{\theta\leq 0}\left((Q-\theta)\sqrt{n}\right) = 0.05$$

Sabemos que el máximo de $\alpha(\theta)$ se alcanza en $\theta = 0$, por lo tanto, el valor de Q que nos otorgue una significancia de 5% será aquel tal que:

$$Q_{\alpha=5\%} = \{Q : \Phi(Q\sqrt{n}) = 0.95\}$$

Dado que sabemos que el cuantil 0.95 de una distribución normal estándar es aproximadamente 1.645:

$$Q_{\alpha=5\%} \approx \frac{1.645}{\sqrt{n}} \tag{1}$$

Ahora supon que observamos $\bar{x} = 0.5$ para una muestra de tamaño $n = 10$. Queremos calcular el p-valor, es decir, la probabilidad de observar un valor de $\bar{x}$ al menos tan grande como el observado bajo el supuesto de que se cumple H_0. En términos formales:

$$\sup_{\theta\leq 0} P(\bar{X} > 0.5)$$

Entonces, utilizando la normalidad de la media muestral:

$$P(\bar{X} > 0.5) = P\left(\frac{\bar{X}-\theta}{\sigma/\sqrt{n}} > \frac{0.5-\theta}{\sigma/\sqrt{n}}\right) = P\left(Z > \frac{0.5-\theta}{\sigma/\sqrt{n}}\right)$$

Reemplazando $\sigma = 1$ y $n = 10$:

$$P(\bar{X} > 0.5) = P(Z > (0.5-\theta)\sqrt{10}) = 1-\Phi((0.5-\theta)\sqrt{10})$$

Dado que esta expresión es creciente en θ, tenemos:

$$\sup_{\theta \leq 0} P(\bar{X} > 0.5) = 1 - \Phi(0.5\sqrt{10}) \approx 0.0569$$

Dado que el p-valor es mayor a 5%, no recharazíamos $H_0 : \theta \leq 0$. Notar que un mayor tamaño de muestra u observar un valor más grande de $\bar{x}$ resultaría en un p-valor más bajo y por lo tanto, nos llevaría a rechazar la hipótesis nula.

3 Ejercicio 3

3.1 Inciso a

Siguiendo lo visto en clase para contrastes de localización aproximados[1], podemos utilizar el siguiente estadístico para determinar la región de rechazo en este test bilateral:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \overset{D}{\sim} N(0,1)$$

Dada la distribución de este estadístico, sabemos que:

$$P\left(\frac{|\hat{p} - p_0|}{\sqrt{p_0(1-p_0)/n}} \geq z_{\alpha/2}\right) \approx \alpha$$

Por lo tanto, para $\alpha = 0.05$ y $p_0 = 1/2$, la región de rechazo estará dada por:

$$R = \left\{\hat{p} : |\hat{p} - 1/2| \geq z_{0.025}\sqrt{\frac{0.25}{n}}\right\}$$

3.2 Inciso b

Para calcular la potencia, primero calculamos la probabilidad de cometer error de tipo II como:

$$\beta(p) = P_{p \neq 1/2}(\hat{p} \notin R)$$

Entonces, de lo visto en el inciso anterior:

$$\beta(p) = P_{p \neq 1/2}\left(|\hat{p} - 1/2| \leq z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

Y la potencia estará dada por $1 - \beta(p)$. Por lo tanto, si n crece, la región de no rechazo se hace más pequeña, $\beta(p)$ decrece y aumenta la potencia del test. A medida de que p se acerca a $1/2$, la varianza de $\hat{p}$ crece, con lo cual aumenta $\beta(p)$ y disminuye la potencia.

3.3 Inciso c

Para los datos provistos, calculamos el p-valor como:

$$\begin{aligned}
P_{p=1/2}(|\hat{p} - p| \geq 0.04) &= 2 * P_{p=1/2}\left(Z \geq \frac{0.04}{\sqrt{p(1-p)/n}}\right) \\
&= 2 * P\left(Z \geq \frac{0.04 * 10}{\sqrt{0.25}}\right) \\
&= 2 * [1 - \Phi(0.8)] \approx 0.423
\end{aligned}$$

Con lo cual no rechazaríamos la hipótesis nula de $p = 1/2$ para $\alpha = 0.5$.

[1] Página 33 de Slide 7.

4 Ejercicio 4

Para obtener una expresión de la probabilidad de error de tipo I, tenemos que encontrar la distribución del estadístico $T = X_1X_2$ bajo $H_0 : \theta = 1$. Para ello:

$$\begin{aligned} F_T(t) = P(X_1X_2 \leq t) &= P\left(X_2 \leq \frac{t}{X_1}\right) \\ &= \int_0^1 P\left(X_2 \leq \frac{t}{X_1}\right) \underbrace{f_{X_1}(x)}_{=1} dx \\ &= \int_0^1 F(t/x)dx \end{aligned}$$

Donde $f_{X_1}(x) = 1$ bajo $H_0 : X_1, X_2 \sim U(0, 1)$. Notar que:

$$F(t/x) = \begin{cases} 1 & \text{si } t/x \geq 1 \text{ (dado que } F_{X_2}(x) = 1 \text{ si } x \geq 1) \\ t/x & \text{si } t/x \leq 1 \end{cases}$$

Entonces:

$$\begin{aligned} F_T(t) &= \int_0^1 F(t/x)dx \\ &= \int_0^t 1\ dx + \int_t^1 t/x\ dx \\ &= t + [t\ln(x)]_{x=t}^1 \\ &= t - t\ln(t) \end{aligned}$$

Entonces, podemos calcular la probabilidad de error de tipo I como:

$$\begin{aligned} P(T \geq 3/4|\theta = 1) &= 1 - P(T \leq 3/4|\theta = 1) \\ &= 1 - F_T\left(\frac{3}{4}\right) \\ &= 1 - \frac{3}{4} + \frac{3}{4}\ln\left(\frac{3}{4}\right) \approx 0.034 \end{aligned}$$

Para calcular la probabilidad de error de tipo II, tenemos que calcular la función de distribución bajo $H_1 : \theta = 2$. En este caso $f(x) = 2x$, con lo cual, aplicando el mismo razonamiento que antes:

$$\begin{aligned} F_T(t) &= \int_0^1 \underbrace{P\left(X_2 \leq \frac{t}{X_1}\right)}_{F_{X_2}(x)} \underbrace{f_{X_1}(x)}_{=2x} dx \\ &= \int_0^1 F(t/x)2x\ dx \end{aligned}$$

En este caso:

$$F(t/x) = \begin{cases} 1 & \text{si } t/x \geq 1 \text{ (dado que } F_{X_2}(x) = 1 \text{ si } x \geq 1) \\ (t/x)^2 & \text{si } t/x \leq 1 \end{cases}$$

Por lo tanto:

$$\begin{aligned} F_T(t) &= \int_0^1 F(t/x)2xdx \\ &= \int_0^t 2x\ dx + \int_t^1 \left(\frac{t}{x}\right)^2 2x\ dx \\ &= t^2 + \left[2t^2\ln(x)\right]_{x=t}^1 \\ &= t^2 - 2t^2\ln(t) \end{aligned}$$

Por lo tanto, la probabilidad de error de tipo II será:

$$P(T \leq 3/4|\theta = 2) = \left(\frac{3}{4}\right)^2 - 2\left(\frac{3}{4}\right)^2 \ln\left(\frac{3}{4}\right) \approx 0.886$$

5 Ejercicio 5

5.1 Inciso a

Dado que sabemos que $T \sim Pois(8\theta)$, queremos calcular la probabilidad asociada a un error de tipo I, es decir:

$$\alpha(\theta) = P_{\theta \leq 0.5}(T \geq 8) = 1 - P_{\theta \leq 0.5}(T \leq 7)$$

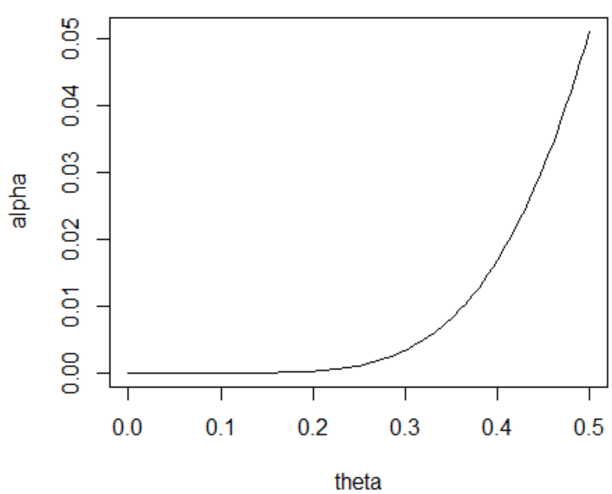
Para $T \sim Pois(8\theta)$, la CDF está dada por:

$$F_T(t) = e^{-8\theta} \sum_{i=1}^{t} \frac{(8\theta)^i}{i!}$$

Por lo tanto, la expresión que buscamos para $\alpha(\theta)$ es:

$$\alpha(\theta) = \left[1 - e^{-8\theta} \sum_{i=1}^{7} \frac{(8\theta)^i}{i!} |\ \theta \leq 0.5\right]$$

Si graficamos esta expresión en R[2], vemos que el máximo valor de α se alcanza para $\theta = 0.5$, con $\alpha \approx 0.05$.



[2]Puede utilizar el siguiente comando: `curve(1-ppois(7, 8*x), from = 0, to = 0.5, xlab = 'theta', ylab = 'alpha')`

Entonces, evaluando la expresión anterior en $\theta = 0.5 \Rightarrow \alpha(\theta) \approx 0.0511$.

5.2 Inciso b

Para la potencia del test, tenemos:

$$1 - \beta(\theta) = 1 - P_{\theta>0.5}(T \leq 7)$$

Con lo cual encontramos la misma expresión que la correspondiente a la probabilidad de error tipo I pero con $\alpha > 0.5$:

$$1 - \beta(\theta) = \left[1 - e^{-8\theta} \sum_{i=1}^{7} \frac{(8\theta)^i}{i!} \mid \theta > 0.5\right]$$

6 Ejercicio 6

Sea la función de riesgo para este test:

$$\delta = \begin{cases} 1 & \text{si } T(\mathbf{X}) \geq Q, \\ 0 & \text{en otro caso.} \end{cases}$$

Para el error de tipo I, queremos calcular la probabilidad de que $\delta = 1$ condicional en $\theta \leq 0.5$, es decir:

$$\alpha(\theta) = P_{\theta \leq 0.5}(T(\mathbf{X}) \geq Q) = 1 - P_{\theta \leq 0.5}(T(\mathbf{X}) \leq Q - 1)$$

Dado que sabemos que $T(\mathbf{X}) = \sum_{i=1}^{5} X_i \sim Bi(5, \theta)$:

$$\alpha(\theta) = 1 - \left[\sum_{i=1}^{Q-1} \binom{5}{i} \theta^i (1-\theta)^{5-i} \;\Big|\; \theta \leq 0.5\right]$$

Para la función de potencia, tendremos:

$$1 - \beta(\theta) = 1 - P_{\theta>0.5}(T(\mathbf{X}) \leq Q - 1)$$

Que será la misma expresión que la de $\alpha(\theta)$ pero evaluada en valores de θ mayores a 0.5.

7 Ejercicio 7

7.1 Inciso a

Para poder construir los tests de Wald y ratio de verosimiltudes, debemos encontrar expresiones para la verosimilitud, estimador máximo verosímil e Información de Fisher asociadas al modelo Exponencial. Dado que se especifica que $E(X) = \lambda^{-1}$, sabemos que estamos trabajando con el modelo especificado por la función de densidad:

$$f(x) = \lambda e^{-\lambda x}$$

Por lo tanto, para este modelo tenemos:

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

$$\ell(\lambda|\mathbf{x}) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

$$i(\lambda) = \frac{1}{\lambda^2}$$

Test de Wald

Para construir un test de Wald, utilizamos nuestro conocimiento de la distribución asintótica del estimador máximo verosímil, el cual nos indica que:

$$Z = \sqrt{ni(\hat{\lambda})}(\hat{\lambda} - \lambda) \xrightarrow{F} N(0,1)$$

Por lo tanto, la región de rechazo para un test de tamaño α de dos colas estará dada por[3]:

$$R_\alpha = \left\{ \hat{\lambda} : Z = \sqrt{ni(\hat{\lambda})}|\hat{\lambda} - \lambda_0| \geq z_{\alpha/2} \right\}$$

Reemplazando $\hat{\lambda}$ e $i(\hat{\lambda})$:

$$R_\alpha = \left\{ \hat{\lambda} : Z = \frac{|(\bar{x}^{-1} - \lambda_0)|}{\bar{x}/\sqrt{n}} \geq z_{\alpha/2} \right\}$$

Likelihood-ratio test

De lo visto en clase[4], sabemos que para familias exponenciales se cumple:

$$-2\log\Lambda(\mathbf{X}) = -2\left(\ell(\lambda_0|\mathbf{X}) - \ell(\hat{\lambda}|\mathbf{X})\right) \xrightarrow{F} \chi^2_1$$

Por lo cual la región de rechazo será:

$$R_\alpha = \left\{ \mathbf{X} : -2\left(\ell(\lambda_0|\mathbf{X}) - \ell(\hat{\lambda}|\mathbf{X})\right) \geq q_\alpha \right\}$$

Donde q_α es el cuantil α de una distribución χ^2_1. Reemplazando la función de log-verosimilitud y operando, obtenemos:

$$R_\alpha = \{\mathbf{X} : -2\left[n(\ln(\lambda_0\bar{x}) - \lambda_0\bar{x} + 1)\right] \geq q_\alpha\}$$

7.2 Inciso b

Supon que observamos:

$$\begin{aligned} n &= 50 \\ \bar{x} &= 3 \\ \lambda_0 &= 0.4 \end{aligned}$$

Para estos valores, el estadístico del **test de Wald** es igual a:

$$Z = \frac{|1/3 - 0.4|}{3/\sqrt{50}} \approx 0.157$$

Si $\alpha = 0.05$ tal que $z_{0.975} \approx 1.96$, no rechazamos la hipótesis nula.

Por otro lado, el estadístico del **LRT** será:

$$-2[50(\ln(0.4 * 3) - 0.4 * 3 + 1)] \approx 1.767$$

Si $\alpha = 0.05$ tal que $q_{0.95} \approx 3.84$, no rechazamos la hipótesis nula.

[3]Notar que en este caso, $z_{\alpha/2}$ indica el cuantil tal que $\alpha/2\%$ de los valores de una distribución normal yacen a la derecha.

[4]Página 50 de Slide 7.

8 Ejercicio 8

Sea μ_A el puntaje de la encuesta en la población de empresarios argentinos. Planteamos como hipótesis nula que $\mu_A \leq 4$. De esta manera, rechazar la hipótesis nula es un argumento a favor de que los empresarios argentinos se preocupan más que los chilenos por la justicia social. Entonces, planteamos el test:

$$\begin{cases} H_0: & \mu_A \leq 4 \\ H_1: & \mu_A > 4 \end{cases}$$

Sea $\bar{X}$ la media muestral del resultado de las encuestas realizadas. Entonces, utilizando la normalidad asintótica de la media muestral $\bar{X}$, planteamos el estadístico:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \xrightarrow{F} N(0,1)$$

Dado que desconocemos el verdadero valor de σ, utilizamos el siguiente estadístico T para realizar el test:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \xrightarrow{F} t_{n-1}$$

Para los valores provistos, tenemos:

$$T = \frac{4.27 - 4}{1.32/\sqrt{200}} \approx 2.89$$

Sea $q_{0.95}$ el cuantil 0.95 de una distribución t_{199}, tenemos $q_{0.95} \approx 1.65$. Dado que $T > q_{0.95}$, rechazamos H_0.

9 Ejercicio 9

Antes de empezar, repasemos los conceptos de *sensibilidad* y *especificidad.*
Sensibilidad: Probabilidad de que sujeto enfermo tenga un resultado positivo en la prueba. Es decir,

$$Sensibilidad = \frac{VP}{VP + FN}$$

Especificidad: Probabilidad de que sujeto sano tenga un resultado negativo en la prueba. Es decir,

$$Especificidad = \frac{VN}{VN + FP}$$

9.1 Inciso a

Utilizando la ayuda provista, definimos las variables aleatorias:

- $\{X_i\}_{i=1}^{n} \overset{iid}{\sim} Bin(\theta_X)$: secuencia de variables aleatorias que representa el resultado del test de COVID-19 de la población de contagiados. Entonces, θ_X es la proporción de resultados positivos en la población de contagiados, es decir, la *sensibilidad.*

- $\{Y_i\}_{i=1}^{m} \overset{iid}{\sim} Bin(\theta_Y)$: secuencia de variables aleatorias que representa el resultado del test de COVID-19 de la población de sanos. Entonces, θ_Y es la proporción de resultados positivos en la población de sanos, es decir, el complemento de la *especificidad.*

9.2 Inciso b

Si tuviésemos un medio alternativo de confirmación que nos permitiese identificar una muestra "certera" de contagiados y sanos, simplemente podríamos utilizar el estimador máximo verosímil de θ, es decir, la proporción de resultados positivos en cada muestra.

Por otro lado, para realizar estimación por intervalos podemos utilizar pivotes aproximados de Wald o estimación por método máximo verosímil. No obstante, necesitaríamos que la muestra sea considerablemente grande, dado que idealmente esperaríamos que los valores de θ se acerquen a los extremos 0 o 1 si el test tiene un comportamiento deseable.

9.3 Inciso c

Si queremos encontrar un test que nos asegure una sensibilidad mayor a 990 por cada mil testeados, propondríamos el test:

$$\begin{cases} H_0: & \theta_X \leq 0.99 \\ H_1: & \theta_X > 0.99 \end{cases}$$

De esta manera, rechazar la hipótesis nula es un argumento a favor de que la *sensibilidad* es mayor a 990 por mil.

Si queremos encontrar un test que nos asegure una especificidad mayor a 950 por cada mil testeados, propondríamos el test:

$$\begin{cases} H_0: & (1-\theta_Y) \leq 0.95 \\ H_1: & (1-\theta_Y) > 0.95 \end{cases}$$

De esta manera, rechazar la hipótesis nula es un argumento a favor de que la *especificidad* es mayor a 950 por mil.

9.4 Inciso d

Para resolver este ejercicio utilizamos un test de Wald. Recordemos que para un test de Wald de 1 cola de tamaño α, la región de rechazo será:

$$R_\alpha = \left\{ \hat{\theta} : Z = \sqrt{ni(\hat{\theta})}(\hat{\theta} - \theta_0) \geq z_\alpha \right\}$$

Recordemos que para el modelo binomial, el EMV y la Información de Fisher son iguales a:

$$\hat{\theta} = \bar{X}$$
$$i(\theta) = \frac{1}{\theta(1-\theta)}$$

<u>**Test de Sensibilidad**</u>
Para $n = 500$ y $\bar{X} = 0.995$, el estadístico de Wald será igual a:

$$Z = \sqrt{\frac{500}{0.995(0.005)}}(0.995 - 0.99) \approx 1.58$$

De acuerdo al valor de este estadístico, no deberíamos rechazar la hipótesis nula para un test de tamaño $\alpha = 0.05$, dado que $z_{0.95} \approx 1.64$.

<u>**Test de Especificidad**</u>
Para $n = 300$ y $(1 - \bar{Y}) = 0.998$, el estadístico de Wald será igual a:

$$Z = \sqrt{\frac{300}{0.998(0.002)}}(0.998 - 0.95) \approx 18.6$$

De acuerdo al valor de este estadístico, deberíamos rechazara la hipótesis nula incluso para un test de tamaño $\alpha = 0.01$, dado que $z_{0.99} \approx 2.32$.

De acuerdo a nuestro análisis, el test satisface los requerimientos de ANMAT en términos de *especificidad*, pero no así en términos de *sensibilidad*. No obstante, nuestras conclusiones dependen del tamaño de los tests utilizados y el supuesto de que la población es lo suficientemente grande como para justificar que las contrapartes muestrales de los parámetros se distribuyen normalmente. Este último supuesto es cuestionable, particularmente en el caso del análisis de especificidad.

10 Ejercicio 10

Sean μ_C y μ_{SC} la probabilidad de fraude estimada por las poblaciones de auditores que tienen y no tienen acceso a información sobre flujos de caja de empresas fraudulentas respectivamente. Dado que nos interesa analizar si tener acceso a información de flujos de caja de mejora la capacidad de reconocimiento de fraude, podemos plantear el test de hipótesis como:

$$\begin{cases} H_0: & \mu_C - \mu_{SC} \leq 0 \\ H_1: & \mu_C - \mu_{SC} > 0 \end{cases}$$

Es decir, nuestra hipótesis nula es que observar flujos de caja no incrementa el reconocimiento de fraude. Dado que los 36 auditores que observaron flujos de caja son distintos de los 36 auditores que no lo hicieron, podemos asumir que las muestras son independientes y utilizar el estadístico provisto en la página 37 del Slide 7:

$$Z = \frac{\bar{X}_C - \bar{X}_{SC}}{\sqrt{\frac{S_C^2}{n} + \frac{S_{SC}^2}{m}}}$$

En nuestro caso:

$$\bar{X}_C = 66.21,\ S_C = 22.93$$
$$\bar{X}_{SC} = 47.56,\ S_{SC} = 27.56$$
$$n = 36,\ m = 36$$

Reemplazando:

$$Z \approx 3.12$$

Por lo tanto, para un test de tamaño $\alpha = 0.05$ ($z_{0.95} \approx 1.645$), rechazaríamos la hipótesis nula de que observar flujos de caja no aporta a la capacidad de reconocer fraude.

11 Ejercicio 11

11.1 Inciso a

Sea $\{X_i\} \sim N(\mu, \sigma^2)$ la secuencia de variables aleatorias que representan la generación diaria de energía del nuevo molino de viento. Dado que asumimos que se considera una mayor generación de poder como algo positivo, tiene sentido realizar un test unilateral de la forma:

$$\begin{cases} H_0: & \mu < 800 \\ H_1: & \mu \geq 800 \end{cases}$$

De esta manera, rechazar la hipótesis nula es un argumento a favor de la capacidad de generación del nuevo molino. Dado que se trata de una población normal con varianza desconocida, podemos utilizar el estadístico:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Por lo tanto, definimos la región de rechazo de tamaño α como:

$$R_\alpha = \left\{ \bar{X} : \frac{\bar{X} - \mu_0}{S/\sqrt{n}} > q_\alpha \right\}$$

Donde $q_\alpha : P(t_{n-1} > q_\alpha) = \alpha$. Reemplazando $\mu_0 = 800$, $S = 120$, $n = 100$ y $t_{99}(0.05) \approx 1.66$:

$$R_{\alpha=5\%} = \left\{ \bar{X} : \frac{\bar{X} - 800}{120/10} > 1.66 \right\}$$

Por lo tanto, el valor mínimo observado de $\bar{X}$ tal que se rechaza H_0 para este test es:

$$\bar{X}^* = 1.66 * 12 + 800 \approx 820$$

11.2 Inciso b

Para este test, la probabilidad de cometer error de tipo II si la media poblacional fuese $\mu = 820$ es igual a:

$$\begin{aligned}
\beta(\mu) &= P_{\mu=820}\left(\bar{X} \leq 820\right) \\
&= P\left(\frac{\bar{X} - 820}{120/10} \leq \frac{820 - 820}{120/10}\right) \\
&= P(T \leq 0)
\end{aligned}$$

Dado que $T \sim t_{99}$, $P(T \leq 0) = F_T(0) = 0.5$.

11.3 Inciso c

A diferencia del accionista de la empresa, podríamos imaginar que el veedor externo quiere decidir si dejar entrar a la empresa o no a la licitación. En este sentido, está interesado en verificar si la capacidad de generación promedio del molino se encuentra por debajo de los 800 kilovatios diarios, en cuyo caso declinaría la participación de la empresa. Por lo tanto, plantearía el test unilateral:

$$\begin{cases} H_0 : & \mu \geq 800 \\ H_1 : & \mu < 800 \end{cases}$$

La región de rechazo para un test de tamaño $\alpha = 0.05$ será:

$$R_{\alpha=5\%} = \left\{ \bar{X} : \frac{\bar{X} - 800}{120/10} < 1.66 \right\}$$

El límite superior por debajo del cual se afirma que la empresa no es apta para participar de la licitación será:

$$\bar{X}^* \approx 820$$

11.4 Inciso d

Para este escenario, la probabilidad de error de tipo II asumiendo que $\mu = 770$ será[5]:

$$\begin{aligned}
\beta(\mu) &= P_{\mu=770}\left(\bar{X} \geq 820\right) \\
&= 1 - P\left(\frac{\bar{X} - 770}{120/10} \leq \frac{820 - 770}{120/10}\right) \\
&= 1 - P(T \leq 4.16) = 1 - F_T(4.16) \approx 0.00003
\end{aligned}$$

[5] Dado que trabajamos con una distribución continua, aproximamos la desigualdad estricta por desigualdad débil.

12 Ejercicio 12

Sea $\{X_i\} \sim N(\mu = 1.6, \sigma^2 = 0.05^2)$ la secuencia de variables aleatorias que representa el diámetro de los cables producidos. Tenemos una muestra de tamaño $n = 16$ para la cual $\bar{X} = 1.615$ y $S = 0.086$.

12.1 Inciso a

Nos interesa testear:

$$\begin{cases} H_0: & \mu = 1.6 \\ H_1: & \mu \neq 1.6 \end{cases}$$

Dado que partimos del supuesto de normalidad y varianza conocida, podemos utilizar el estadístico:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Por lo tanto, construimos la región de rechazo de tamaño α para este test de dos colas:

$$R_\alpha = \left\{ \bar{X} : \frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \geq z_{\alpha/2} \right\}$$

Donde $z_\alpha : P(Z > z_\alpha) = \alpha$. Reemplazando $\mu_0 = 1.6$, $\sigma = 0.05$, $n = 16$ y $z_{0.05} \approx 1.645$:

$$R_{\alpha=10\%} = \left\{ \bar{X} : \frac{|\bar{X} - 1.6|}{0.0125} \geq 1.645 \right\}$$

Para $\bar{X} = 1.615$, tenemos:

$$\frac{|1.615 - 1.6|}{0.0125} = 1.2 \leq 1.645$$

Con lo cual no rechazamos H_0 para un test de significancia $\alpha = 10\%$. Para encontrar el menor nivel de significación tal que se rechace la hipótesis nula, sólo debemos buscar el valor de α tal que $z_{\alpha/2} = 1.2$, es decir:

$$1 - \Phi(1.2) = \alpha/2$$
$$1 - 0.885 = \alpha/2 \Rightarrow \alpha^* \approx 0.23$$

Notar que este valor de α es equivalente al p-valor asociado a $\bar{X} = 1.615$.

12.2 Inciso b

Queremos testear:

$$\begin{cases} H_0: & \sigma \leq 0.05 \\ H_1: & \sigma > 0.05 \end{cases}$$

Siguiendo lo visto en clase para contrastes sobre σ^2 en poblaciones normales[6], proponemos el estadístico:

$$\chi = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Con lo cual definimos la región de rechazo de tamaño α como:

$$R_\alpha = \left\{ S^2 : \frac{(n-1)S^2}{\sigma_0^2} > q_\alpha \right\}$$

Donde q_α es tal que $P(\chi^2_{n-1} > q_\alpha) = \alpha$. Reemplazando $S = 0.086$, $\sigma_0 = 0.05$, $n = 16$:

$$\chi = \frac{15 * (0.086^2)}{0.05^2} \approx 44.3$$

Con lo cual, para $q_{0.1} = 22.3$, rechazamos la hipótesis nula.

[6] Ver página 30 de Slide 7.

13 Ejercicio 13

Sea $\{X_i\} \sim Ber(p)$ la secuencia de variables aleatorias que representa la capacidad de los clientes de distinguir vino de jugo de uva congelado. Nos interesa testear:

$$\begin{cases} H_0: & p \leq 0.09 \\ H_1: & p > 0.09 \end{cases}$$

La regla de decisión propuesta se puede expresar como la región de rechazo: $R = \{\hat{p} : \hat{p} > 0.14\}$. Para resolver este ejercicio, proponemos utilizar el siguiente estadístico de contraste:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \xrightarrow{D} N(0,1)$$

13.1 Inciso a

Para una muestra de tamaño $n = 100$ queremos calcular:

$$\begin{aligned}
\alpha(p) &= P_{p \leq 0.09}\left(\hat{p} > 0.14\right) \\
&= P_{p \leq 0.09}\left(Z > \frac{0.14 - p}{\sqrt{p(1-p)/n}}\right) \\
&= 1 - P_{p \leq 0.09}\left(Z \leq \frac{0.14 - p}{\sqrt{p(1-p)/n}}\right) \\
&= 1 - P\left(Z \leq \frac{0.14 - 0.09}{\sqrt{0.09(0.91)/100}}\right) \\
&\approx 1 - \Phi\left(1.747\right) \approx 0.04
\end{aligned}$$

Entonces, la probabilidad de rechazar la hipótesis nula cuando esta es verdadera para esta regla de decisión es de aproximadamente el 4%.

13.2 Inciso b

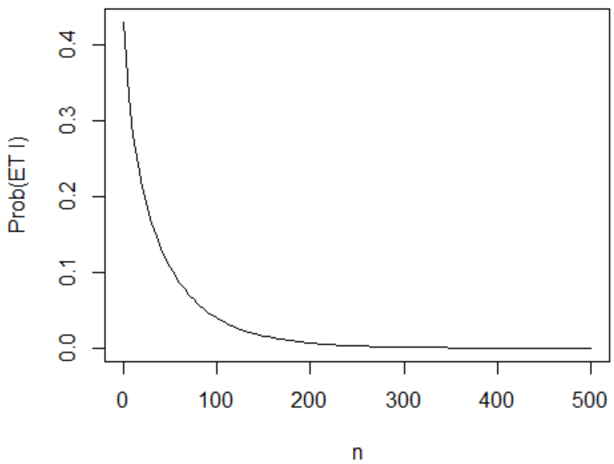
Para una muestra de $n = 400$, tenemos:

$$\begin{aligned}
\alpha(p) &= 1 - P\left(Z \leq \frac{0.14 - 0.09}{\sqrt{0.09(0.91)/400}}\right) \\
&\approx 1 - \Phi(3.494) \approx 0.0002
\end{aligned}$$

A medida que aumenta n, la varianza de la proporción muestral utilizada para realizar el contraste decrece, haciendo más concluyente cualquier inferencia realizada en base al mismo. El hecho de que $n >> 0$ nos permite intuir que el haber observado un valor de $\hat{p}$ que nos lleve a rechazar la hipótesis nula se relaciona con el hecho de que la verdadera proporción en la población es mayor a 0.09, y no meramente a la varianza del estadístico.

Para verlo en términos gráficos, graficamos $\alpha(p)$ como función de n[7]:

[7]En R: `curve(1-pnorm(0.05/sqrt(0.09*0.91/x)), from = 1, to = 500, xlab = 'n', ylab = 'Prob(ET I)')`



13.3 Inciso c

Queremos calcular:

$$\begin{aligned}\beta(p) &= P_{p=0.2}(\hat{p} \leq 0.14)\\ &= P_{p=0.2}\left(Z \leq \frac{0.14-p}{\sqrt{p(1-p)/n}}\right)\\ &\approx \Phi\left(\frac{0.14-0.2}{\sqrt{0.2(0.8)/100}}\right) \approx 0.066\end{aligned}$$

13.4 Inciso d

Ahora se propone la siguiente región de rechazo: $R = \{\hat{p} : \hat{p} > 0.16\}$.

13.4.1 Parte i

Al elevar el umbral a partir del cual rechazamos la hipótesis nula, la probabilidad de rechazar H_0 condicional en que H_0 sea cierta será menor que en el inciso a).

13.4.2 Parte ii

Al elevar el umbral a partir del cual rechazamos la hipótesis nula, la probabilidad de no rechazar H_0 cuando H_1 es cierta será mayor que en el inciso c).

14 Ejercicio 14

Sea $\{X_i\} \sim N(\mu, \sigma)$ la secuencia de variables aleatorias que representa las mediciones de resistencia a la compresión de las mezclas de ceniza pulverizadas producidas por esta empresa.

14.1 Inciso a

El ente regulador plantea el siguiente test de hipótesis:

$$\begin{cases} H_0: & \mu \leq 1300 \\ H_1: & \mu > 1300 \end{cases}$$

Nuestra valoración sobre el test propuesto puede variar según consideremos la decisión del ente regulador como una de *aprobación* o una de *rechazo*. En el primer caso, el test propuesto sería correcto, dado que el ente supondría que una muestra no es apta a menos que se pruebe lo contrario. En cambio, si la decisión del ente fuese una de *rechazo*, el test opuesto sería más apto, ya que buscaríamos descartar aquellas muestras que no cumplen con los estándares propuestos.

Para el test propuesto, las consecuencias de los errores de tipo I y II serían las siguientes:

- **Error Tipo I**: Significaría afirmar que la mezcla cumple con la normativa especificada cuando en realidad no lo hace. Este debería ser el principal tipo de error que queremos evitar, dado que es el que podría resultar en la utilización de un producto no apto para uso.
- **Error Tipo II**: Significaría no aprobar una mezcla que resulta apta para uso. En este caso, el costo sería de carácter económico, pero no estaríamos generando una situación de inseguridad por el uso de un producto de baja calidad.

14.2 Inciso b

Se propone la región de rechazo: $R = \{\bar{X} : \bar{X} > 1331.26\}$. Dado que sabemos que la población de mediciones de resistencia a la compresión se distribuye Normal con desvío $\sigma = 60$, proponemos el estadístico:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Queremos encontrar el nivel de este test, es decir, la probabilidad de cometer error de tipo I:

$$\begin{aligned} \alpha(\mu) &= P_{\mu \leq 1300}(\bar{X} > 1331.26) \\ &= P_{\mu \leq 1300}\left(Z > \frac{1331.26 - \mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - P_{\mu \leq 1300}\left(Z \leq \frac{1331.26 - \mu}{\sigma/\sqrt{n}}\right) \end{aligned}$$

Reemplazando $\sigma = 60$, $n = 20$, y utilizando que $\alpha(\mu)$ es creciente en μ:

$$\sup_{\mu \leq 1300} \alpha(\mu) = 1 - \Phi\left(\frac{1331.26 - 1300}{60/\sqrt{20}}\right) \approx 0.01$$

14.3 Inciso c

Para este inciso, nos interesa calcular la *potencia* del test propuesto, es decir, la probabilidad de rechazar la hipótesis nula condicional en que H_1 cierta ($\mu = 1350$ cae dentro del espacio de parámetros correspondiente a H_1). Ello es lo mismo que calcular el complemento de la probabilidad de error de tipo II, es decir:

$$\begin{aligned} 1 - \beta(\mu) &= 1 - P_{\mu = 1350}(\bar{X} \leq 1331.26) \\ &= 1 - \Phi\left(\frac{1331.26 - 1350}{60/\sqrt{20}}\right) \approx 0.92 \end{aligned}$$

Intuitivamente, explicaríamos a los empresarios que a pesar de que el proceso productivo utilizado es tal que el requisito del ente regulador se supera con cierto margen (dado $\mu = 1350$), existe aproximadamente un 8% de chances que, dada la varianza presente en el proceso productivo, la muestra de 20 paquetes represente una tirada mala para la cual no se supere la prueba impuesta por el regulador.

14.4 Inciso d

Como ya anticipásemos en el inciso *c*), calculamos la *potencia* del test para $\mu = 1350$.

14.5 Inciso e

Dado que tomamos $\mu_0 = 1300$, $\mu = 1350$ y $\sigma = 60$ como valores fijos, sólo podemos hacer variar Q (el valor que determina la región de rechazo) y n (el tamaño de la muestra utilizada para el test). Entonces, queremos encontrar un par $\{Q, n\}$ tal que:

$$\sup_{\mu \leq 1300} \alpha(\mu) = 0.01$$
$$1 - \beta(\mu) = 0.95$$

Para las cuentas obtenidas en los incisos *b*) y *c*), ello implica encontrar valores de $\{Q, n\}$ tal que:

$$1 - \Phi\left(\frac{Q - 1300}{60/\sqrt{n}}\right) = 0.01$$
$$1 - \Phi\left(\frac{Q - 1350}{60/\sqrt{n}}\right) = 0.95$$

Entonces, queremos resolver el sistema de ecuaciones:

$$\frac{Q - 1300}{60/\sqrt{n}} = z_{0.99}$$
$$\frac{Q - 1350}{60/\sqrt{n}} = z_{0.05}$$

Donde $z_\alpha : \Phi(z_\alpha) = \alpha$. Utilizando $z_{0.99} = 2.326$ y $z_{0.05} \approx 1.645$, obtenemos:

$$n = \left(\frac{6}{5}(z_{0.99} - z_{0.05})\right)^2 \approx 23$$
$$Q = 1300 + \frac{z_{0.99}}{\sqrt{n}/60} \approx 1329$$

Por lo tanto, elevando la muestra y reduciendo levemente el límite de rechazo podemos obtener los valores de significatividad y potencia deseados.

15 Ejercicio 15

En este ejercicio, nos interesa testear la hipótesis nula de que la proporción de mujeres a favor de la ILE es mayor que la de hombres:

$$\begin{cases} H_0: & p_M \geq p_H \\ H_1: & p_M < p_H \end{cases}$$

De esta manera, rechazar la hipótesis nula es equivalente a verificar si lo observado en los datos es estadísticamente significativo. Entonces, utilizamos el siguiente estadístico propuesto en clase:

$$Z = \frac{(\widehat{p}_M - \widehat{p}_H) - c}{\sqrt{\frac{\widehat{p}_H \times (1 - \hat{p}_H)}{n} + \frac{\widehat{p}_M \times (1 - \hat{p}_M)}{m}}} \xrightarrow{D} N(0, 1)$$

Dado que desconocemos los valores exactos de n y m, asumimos que $n = 727$ y $m = 758$ tal que las mujeres representan poco más de un 51% de la muestra (aproximadamente su participación en la población de Argentina).

Entonces, la región de rechazo de nivel α estará dada por $R_\alpha = \{(\hat{p}_H, \hat{p}_M) : Z < z_\alpha\}$ donde z_α es el cuantil α de la distribución normal estándar. En este caso, al calcular Z para los valores obtenidos en la muestra y $c = 0$, obtenemos:

$$Z = \frac{(0.516 - 0.556)}{\sqrt{\frac{0.556 \times 0.444}{727} + \frac{0.516 \times 0.484}{758}}} \approx -1.546 > z_{0.05}$$

Por lo tanto, no rechazamos la hipótesis nula de que la proporción de mujeres a favor de la ILE es al menos tan grande como la de hombres.

Inferencia Estadística - Guia 5

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

Agosto 2020

1 Ejercicio 1

Sea $X|\theta \sim \text{Pois}(\theta)$, $P(\theta = 2) = 1/3$ y $P(\theta = 3) = 2/3$. Supón que observamos $x_1 = 2$ y $x_2 = 4$. Queremos calcular la probabilidad a posteriori de θ:

$$P(\theta|X_1 = 2, X_2 = 4) = \frac{P(X_1 = 2, X_2 = 4|\theta)}{P(X_1 = 2, X_2 = 4)}P(\theta)$$

Recordar que para el modelo Poisson:

$$P(\mathbf{X}|\theta) = \frac{\theta^{\sum x_i} e^{-n\theta}}{\prod x_i!} \Rightarrow P(X_1 = 2, X_2 = 4|\theta) = \frac{\theta^6 e^{-2\theta}}{2 * 4!}$$

Por otro lado, usando la ayuda provista:

$$\begin{aligned} P(X_1 = 2, X_2 = 4) &= P(X_1 = 2, X_2 = 4|\theta = 2)P(\theta = 2) + P(X_1 = 2, X_2 = 4|\theta = 3)P(\theta = 3) \\ &= \frac{2^6 e^{-2*2}}{2 * 4!} * \frac{1}{3} + \frac{3^6 e^{-2*3}}{2 * 4!} * \frac{2}{3} \\ &\approx 0.033 \end{aligned}$$

Por lo tanto, podemos escribir la probabilidad a posteriori de θ como:

$$P(\theta|X_1 = 2, X_2 = 4) = \frac{P(\theta)}{P(X_1 = 2, X_2 = 4)} \frac{\theta^6 e^{-2\theta}}{2 * 4!}$$

Notar que $P(X_1 = 2, X_2 = 4)$ es independiente de θ, por lo tanto, podemos escribir:

$$P(\theta|X_1 = 2, X_2 = 4) \propto \theta^6 e^{-2\theta} P(\theta)$$

Para realizar inferencia sobre θ, podemos reportar θ_{MAP}, el valor que maximiza la probabilidad a posteriori. Calculando los valores de $P(\theta|\mathbf{X})$ para cada caso:

$$\begin{aligned} P(\theta = 2|X_1 = 2, X_2 = 4) &\propto 2^6 e^{-2*2} * \frac{1}{3} \approx 0.39 \\ P(\theta = 3|X_1 = 2, X_2 = 4) &\propto 3^6 e^{-2*3} * \frac{2}{3} \approx 1.2 \end{aligned}$$

Por lo tanto, $\theta_{MAP} = 3$.

2 Ejercicio 2

Definimos la probabilidad a posteriori de θ condicional en haber observado $\{\mathbf{x}_1, \mathbf{x}_2\}$ como:

$$\pi(\theta|\mathbf{x}_1, \mathbf{x}_2) \propto L(\theta|\mathbf{x}_1, \mathbf{x}_2)\pi(\theta)$$

Si las observaciones provienen de una secuencia de variables aleatorias independientemente distribuidas, podemos escribir $L(\theta|\mathbf{x}_1, \mathbf{x}_2) = L(\theta|\mathbf{x}_1)L(\theta|\mathbf{x}_2)$. Por lo tanto:

$$\begin{aligned}\pi(\theta|\mathbf{x}_1, \mathbf{x}_2) &\propto L(\theta|\mathbf{x}_2)\underbrace{L(\theta|\mathbf{x}_1)\pi(\theta)}_{\pi(\theta|\mathbf{x}_1)} \\ &\propto L(\theta|\mathbf{x}_2)\pi(\theta|\mathbf{x}_1)\end{aligned}$$

Que era lo que queríamos mostrar. Intuitivamente, la independencia entre variables aleatorias nos permite calcular la probabilidad a posteriori de toda la muestra "actualizando" secuencialmente probabilidades a posteriori.

3 Ejercicio 3

3.1 Inciso a

Queremos calcular:

$$\pi(\theta|\mathbf{X}) \propto L(\theta|\mathbf{x})\pi(\theta)$$

Para el modelo Poisson(θ):

$$L(\theta|\mathbf{X}) \propto \theta^{\sum x_i} e^{-n\theta}$$

Por lo tanto, utilizando que $\theta \sim \Gamma(\alpha, \beta)$:

$$\begin{aligned}\pi(\theta|\mathbf{X}) &\propto \theta^{\sum x_i} e^{-n\theta} \frac{\theta^{\alpha-1} e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha} \\ &\propto \theta^{(\alpha+\sum x_i)-1} e^{-\theta(n+1/\beta)} \\ &\propto \theta^{\alpha_n-1} e^{-\theta/\beta_n}\end{aligned}$$

Esta expresión corresponde (por proporcionalidad) a la verosimilitud de una variable $\Gamma(\alpha_n, \beta_n)$. En el segundo paso, podemos eliminar el denominador porque α, β están dados. Para el último paso, notar que:

$$n + \frac{1}{\beta} = \frac{\beta n + 1}{\beta} = \frac{1}{\beta_n}$$

3.2 Inciso b

Para $\theta \sim \Gamma(\alpha_n, \beta_n)$ definida como en la consigna, tenemos:

$$\begin{aligned}E(\theta) &= \alpha_n\beta_n \\ V(\theta) &= \alpha_n\beta_n^2\end{aligned}$$

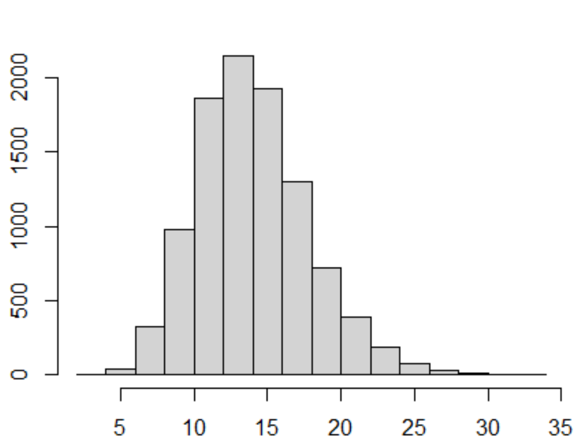
Notar que podemos escribir $\alpha_n = \alpha + n\bar{x}$. Por lo tanto:

$$\begin{aligned}\lim_{n\to\infty} E(\theta) &= (\alpha + n\bar{x})\left(\frac{\beta}{\beta n + 1}\right) = \bar{x} = \hat{\theta}_{MLE} \\ \lim_{n\to\infty} V(\theta) &= (\alpha + n\bar{x})\left(\frac{\beta}{\beta n + 1}\right)^2 = 0\end{aligned}$$

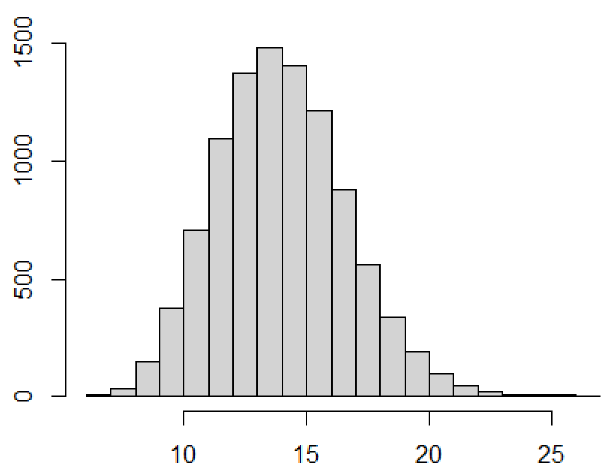
Puede verse entonces que a medida que n crece la distribución a priori "pierde peso" a favor de la información contenida en la muestra.

3.3 Inciso c

A la hora de elegir valores de α, β, querríamos comunicar nuestra creencia previa respecto a la distribución de los posibles valores de la media de crímenes diarios. Dado que observamos $\bar{x} = 14$, resulta razonable alguna combinación de $\{\alpha, \beta\}$ tal que $\alpha * \beta = 14$. No obstante, existen infinitas combinaciones que satisfacen dicha condición, por lo tanto, debemos tomar cierta postura respecto a la varianza del parámetro. Si consideramos que existe un amplio rango de valores posibles para θ, favoreceríamos una mayor valor de β. Presentamos por ejemplo los histogramas asociados a dos muestras generadas con diferentes parametrizaciones:



(a) $\alpha = 14$, $\beta = 1$



(b) $\alpha = 28$, $\beta = 1/2$

La primera parametrización resulta en una media igual pero mayor varianza, por lo tanto, la elegiríamos si considerásemos que existe mayor "incertidumbre" sobre θ. Si quisiesemos una prior cercana a ser "no informativa", elegiríamos un valor de β muy grande en relación a α. Para resolver el resto del ejercicio elegimos arbitrariamente $\alpha = 14$ y $\beta = 1$.

Dado lo visto en el inciso a, sabemos que la distribución a posteriori de θ será $\Gamma(\alpha_n, \beta_n)$. Reemplazando nuestros valores a priori para $(\alpha,\, \beta)$ y la información de la muestra:

$$\pi(\theta|\mathbf{x}) \propto \Gamma(\alpha_n = 154, \beta_n = 1/11)$$

3.4 Inciso d

Por lo visto en el inciso b):

$$E(\theta) = \alpha_n \beta_n = \frac{154}{11} = 14$$
$$V(\theta) = \alpha_n \beta_n^2 = \frac{154}{11^2} \approx 1.272$$

3.5 Inciso e

En este inciso, queremos encontrar el intervalo de credibilidad de mayor densidad a posteriori, es decir el *highest posterior density* (HPD) credible set. Este conjunto puede definirse como:

$$HPD = \{\theta \in \Theta : \pi(\theta|\mathbf{x}) \geq \nu_\alpha\}$$

Donde ν_α es el valor más alto tal que:

$$P(\theta \in HPD(\nu_\alpha)|\mathbf{x}) = 1 - \alpha$$

Existen diferentes maneras en las cuales podríamos aproximar el intervalo de credibilidad HPD. Si la distribución es unimodal pero no simétrica, podríamos por ejemplo computar todos los intervalos que acumulan $(1-\alpha)\%$ de las observaciones y elegir el más estrecho[1].

No obstante, si la distribución es unimodal y aproximadamente simétrica (como en este caso) alrededor de la moda M, será suficiente buscar los cuantiles $\alpha/2$ y $(1-\alpha/2)$ de la distribución, ya que estos concentrarán $(1-\alpha)\%$ de los valores posibles con la mayor densidad. En este caso, para una variable $\Gamma(154, 11^{-1})$[2]

$$HPD_{95\%} = [q_{0.025}, q_{0.975}] = [11.876, 16.295]$$
$$HPD_{99\%} = [q_{0.005}, q_{0.995}] = [11.264, 17.076]$$

Bajo el enfoque bayesiano, podemos afirmar que la probabilidad a posteriori condicional en haber observado $\mathbf{x}$ de que θ caiga en el intervalo HPD de cobertura $(1-\alpha)$ es igual a $(1-\alpha)\%$.

4 Ejercicio 4

Una vez más, queremos obtener:

$$\pi(\theta|\mathbf{X}) \propto L(\theta|\mathbf{x})\pi(\theta)$$

El modelo $Ber(\theta)$ se puede representar con la verosimilitud:

$$L(\theta|\mathbf{X}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

Para la distribución uniforme continua, $\pi(\theta) = 1$ para todo θ, por lo tanto:

$$\pi(\theta|\mathbf{X}) \propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} = \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}$$

4.1 Inciso a

En este caso, utilizar una prior $U(0,1)$ resultaría *no informativo*, en tanto asignamos probabilidad equivalente a todos los valores posibles de θ. Notar, entonces, que toda la información referente a la distribución de θ contenida en la probabilidad a posteriori es la proveniente de la muestra.

4.2 Inciso b

Recordar que la función de densidad de una variable aleatoria $X \sim \text{Beta}(\alpha, \beta)$ es igual a:

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

Donde $B(\alpha, \beta)$ es la función beta incompleta que no depende de x. Por lo tanto, la probabilidad a posteriori calculada anteriormente es equivalente a una distribución $\text{Beta}\,(n\bar{x}+1, n(1-\bar{x})+1)$.

Para una variable aleatoria $X \sim \text{Beta}(\alpha, \beta)$:

$$E(X) = \frac{\alpha}{\alpha+\beta}$$
$$V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Por lo tanto, reemplazando α y β por los valores obtenidos anteriormente:

$$E(\theta|\mathbf{X}) = \frac{n\bar{x}+1}{n+2}$$
$$V(\theta|\mathbf{X}) = \frac{n^2\bar{x}(1-\bar{x})+n+1}{(n+2)^2(n+3)}$$

[1] Si quieren ver un ejemplo de esto, adjunto un código de R en el campus.

[2] En `R` puede calcular los cuantiles con la funcion `qgamma`. Recuerde que `R` parametriza el modelo Gamma como $\Gamma(\alpha, 1/\beta)$.

5 Ejercicio 5

$$
\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2 - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(\theta-x_i)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta^2-2\theta\mu_0+\mu_0^2) - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(\theta^2-2\theta x_i+x_i^2)\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta^2-2\theta\mu_0+\mu_0^2) - \frac{1}{2\sigma_0^2}\left(n\theta^2-2\theta\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}x_i^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta^2-2\theta\mu_0+\mu_0^2) - \frac{1}{2\sigma_0^2}\left(n\theta^2-2\theta n\bar{x}+\sum_{i=1}^{n}x_i^2\right)\right\}
\end{aligned}
$$

Cualquier término dentro del exponencial que no incluya θ puede ser considerado un factor de proporcionalidad, por lo cual:

$$
\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto \exp\left\{-\frac{1}{2\tau_0^2}(\theta^2-2\theta\mu_0) - \frac{n}{2\sigma_0^2}\left(\theta^2-2\theta\bar{x}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\tau_0^2}(\theta^2-2\theta\mu_0) + \frac{n}{\sigma_0^2}\left(\theta^2-2\theta\bar{x}\right)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\theta^2\frac{\sigma_0^2+n\tau_0^2}{\tau_0^2\sigma_0^2} - 2\theta\frac{\mu_0\sigma_0^2+n\bar{x}\tau_0^2}{\tau_0^2\sigma_0^2}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma_0^2}\right) - 2\theta\left(\frac{\mu_0}{\tau_0^2}+\frac{n\bar{x}}{\sigma_0^2}\right)\right]\right\}
\end{aligned}
$$

Reemplazando las definicionnes de μ_n y σ_n^2:

$$
\pi(\theta|\mathbf{x}) \propto \exp\left\{-\frac{1}{2\sigma_n^2}\left(\theta^2-2\theta\mu_n\right)\right\}
$$

Dado que μ_n y σ_n^2 no dependen de θ, podemos completar el cuadrado con un factor de proporcionalidad:

$$
\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto \exp\left\{-\frac{1}{2\sigma_n^2}\left(\theta^2-2\theta\mu_n+\mu_n^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_n^2}\left(\theta-\mu_n\right)^2\right\}
\end{aligned}
$$

Que era el resultado que queríamos obtener.

5.1 Inciso a

Notar que:

$$
\begin{aligned}
\lim_{n\to\infty}\mu_n &= \bar{x} \\
\lim_{n\to\infty}\sigma_n^2 &= 0
\end{aligned}
$$

5.2 Inciso b

Este resultado nos indica que a medida que crece n, la distribución a posteriori del parámetro θ se degenera a la información provista en la muestra, en este caso, a través de $\bar{x}$.

5.3 Inciso c

Recordar que la distribución normal resulta ser unimodal, simétrica y que su moda coincide con su esperanza. Entonces, dado que θ sigue una distribución a posteriori $N(\mu_n, \sigma_n)$, podemos afirmar que el intervalo de credibilidad HPD estará dado por:

$$HPD_{95\%} = \mu_n \pm \sigma_n z_{0.975}$$

Donde z_α es el cuantil α de la distribución $N(0, 1)$.

6 Ejercicio 6

6.1 Inciso a

Tal como en escenarios anteriores en los cuales analizamos el desempleo, podemos plantear el parámetro de interés θ como la probabilidad de éxito asociada a una secuencia de variables aleatorias $\{X_i\} \sim \text{Ber}(\theta)$.

Tal como en el ejercicio 3, una opción *no informativa* sería utilizar una distribución $U(0, 1)$ como prior para θ. No obstante, ello resulta poco razonable, dado que sabemos que la tasa de desempleo suele encontrarse entre 5% y el 15%. Por lo tanto, otra alternativa sería utilizar una distribución $U(a, b)$ con límites que consideremos razonables, por ej., en base a datos históricos.

Otra opción sería utilizar una prior $\text{Beta}(\alpha, \beta)$. Esta elección tiene la ventaja de que forma un modelo *conjugado*, lo cual significa que la distribución a posteriori también será un modelo Beta. Además, nos permite ajustar los parámetros α y β para definir cuan informativa queremos que sea nuestra prior. Para el resto de este ejercicio, elegimos esta distribución y trabajamos sin fijar valores de α y β.

6.2 Inciso b

Para $\pi(\theta) = \text{Beta}(\alpha, \beta)$:

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Por otro lado, la verosimilitud para el modelo Bernouilli es:

$$L(\theta|\mathbf{X}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

Por lo tanto, la distribución a posteriori será:

$$\begin{aligned}\pi(\theta|\mathbf{X}) &\propto \theta^{\alpha+\sum x_i-1}(1-\theta)^{\beta+n-\sum x_i-1} \\ &\propto \theta^{\alpha_n-1}(1-\theta)^{\beta_n-1}\end{aligned}$$

Entonces, $\pi(\theta|\mathbf{X}) = \text{Beta}(\alpha_n, \beta_n)$, con:

$$\begin{aligned}\alpha_n &= \alpha + n\bar{x} \\ \beta_n &= \beta + n(1-\bar{x})\end{aligned}$$

6.3 Inciso c

Dado que sabemos que la distribución a posteriori es Beta(α_n, β_n), podemos expresar sus momentos como:

$$E(\theta|\mathbf{X}) = \frac{\alpha_n}{\alpha_n + \beta_n} = \frac{\alpha + n\bar{x}}{\alpha + \beta + n}$$
$$M(\theta|\mathbf{X}) = \frac{\alpha_n - 1}{\alpha_n + \beta_n - 2} = \frac{\alpha + n\bar{x} - 1}{\alpha + \beta + n - 2}$$
$$V(\theta|\mathbf{X}) = \frac{\alpha_n\beta_n}{(\alpha_n + \beta_n)^2(\alpha_n + \beta_n + 1)} = \frac{(\alpha + n\bar{x})(\beta + n(1-\bar{x}))}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}$$

La expresión utilizada para la moda es válida para valores de α, β mayores a uno, de lo contrario, la distribución sería bimodal.

6.4 Inciso d

Si consideramos que la distribución de la tasa de desempleo es unimodal, elegiríamos valores de α, β mayores a 1 para nuestra prior . No obstante, la distribución β sólo preserva la simetría en tanto $\alpha = \beta$. Por lo tanto, preservar la simetría implica que la distribución se centre alrededor de 1/2, lo cual no resulta razonable para la tasa de desempleo.

No obstante, si graficamos la distribución por ejemplo para $\alpha = 20, \beta = 180$, vemos que la misma es aproximadamente simétrica alrededor del 10%:

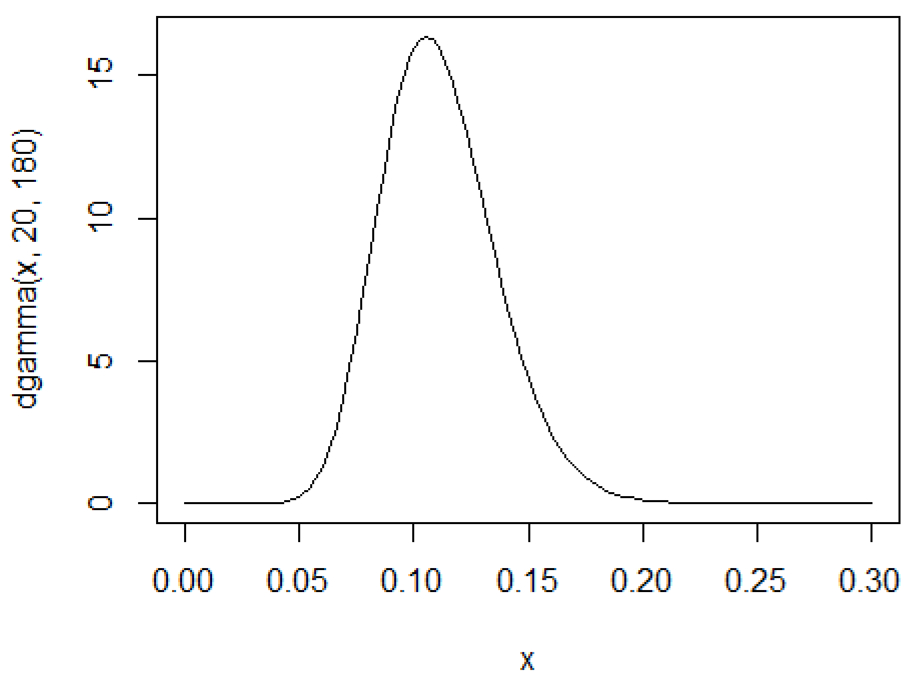


Figure 2: Función de densidad para Beta(20, 180)

Entonces, una posible metodología es elegir valores de α y β tales que la distribución se centre alrededor de algún valor de θ que consideremos razonable. Luego, si escalamos ambos factores en forma equivalente, podemos modificar la dispersión y simetría de la distribución[3].

Una vez elegidos los valores de nuestra prior, podemos calcular los valores a posteriori y analizar si los mismos resultan en una distribución razonablemente simétrica y unimodal. De ser así, podríamos simplemente reportar los cuantiles de la distribución:

$$HPD_{95\%} = [q_{0.025}, q_{0.975}]$$

Alternativamente, si la distribución resulta ser muy asimétrica, podemos aplicar la metodología comentada en el inciso 3.e.

6.5 Inciso e

Si nuestro amigo eligió una prior considerablemente diferente a la nuestra, ya sea por ser más o menos informativa o encontrarse centrada en algún valor diferente de θ, los resultados reportados podrían diferir.

6.6 Inciso f

Al incrementar el tamaño de la muestra, el valor de los parámetros de la distribución a posteriori tenderán a verse más influenciados por lo observado en los datos que por la elección de prior. Por lo tanto, esperaríamos que la "distancia" entra las conclusiones obtenidas sea menor.

[3] Valores más grandes resultarán en una distribución con menor varianza pero más simétrica.