ITAM
Departamento de Estadística

Inferencia Estadística– Laboratorio #6
Estimación por Intervalos de Confianza Parte 2

Otoño 2021

1. Suponga que $Y_1, .., Y_n$ es una muestra aleatoria de tamaño n para una distribución exponencial con media θ. Utilice el método de f.g.m. para demostrar que $2\sum_1^n Y/\theta$ es una cantidad pivote y encuentre su distribución.

Sabemos que la fgm de una $X \sim Exp(\theta)$

$$m_X(t) = E[e^{tx}] = \int_0^\infty e^{tx}\frac{1}{\theta}e^{-x/\theta}\,dx = \frac{1}{\theta}\int_0^\infty e^{-x\left(\frac{1-\theta t}{\theta}\right)}dx$$

$$= \frac{1}{\cancel{\theta}}\left[\frac{-e^{-x\left(\frac{1-\theta t}{\theta}\right)}}{\frac{1-\theta t}{\cancel{\theta}}}\right]\Bigg|_0^\infty = \frac{1}{1-\theta t}\left[-\cancel{e^{-\infty}}^{\to 0} + \cancel{e^{0}}^{\to 1}\right] = \frac{1}{1-\theta t} = (1-\theta t)^{-1}$$

Sea $T = 2\sum Y_i/\theta$

$$\Rightarrow m_T(t) = E\left[e^{t2\sum Y_i/\theta}\right] = E\left[\exp\left\{\frac{2t}{\theta}(\overbrace{Y_1 + \dots + Y_n}^{\text{independientes}})\right\}\right]$$

$$= E\left[e^{\frac{2tY_1}{\theta}}\right]\cdots E\left[e^{\frac{2tY_n}{\theta}}\right] = \left\{m_Y\left(\frac{2t}{\theta}\right)\right\}^n = \left[1-\left(\frac{2t}{\theta}\right)\theta\right]^{-n}$$

$$= [1-2t]^{-n} \sim \chi^2_{2n}$$

Como θ es libre $\Rightarrow$ T es una cantidad PIVOTE.

$$\therefore \text{IC de } (1-\alpha)\times 100\% \Rightarrow \left[\chi^2_{(2n,1-\alpha/2)} < \frac{2\sum Y_i}{\theta} < \chi^2_{(2n,\alpha/2)}\right]$$

$$\Rightarrow \left[\frac{2n\bar{Y}}{\chi^2_{(2n,\alpha/2)}}, \frac{2n\bar{Y}}{\chi^2_{(2n,1-\alpha/2)}}\right]$$

2. Suponga que $Y_1, .., Y_n$ es una muestra aleatoria de tamaño n para una distribución Gamma con $\alpha = 2$ y β desconocida.

   (a) Utilice el método de f.g.m. para demostrar que $2\sum_1^n Y/\beta$ es una cantidad pivote y encuentre su distribución.

Sabemos que $m_X(t) = (1-\beta t)^{-\alpha n}$ si $X \sim \Gamma(\alpha, \beta)$

Sea $T = \dfrac{2\sum Y_i}{\beta}$ para $\alpha = 2$

Porqué?

$\Rightarrow m_T(t) = E[\exp\{2t\sum Y_i/\beta\}] = E\left[e^{\frac{2t}{\beta}Y_1}\right] \cdots E\left[e^{\frac{2t}{\beta}Y_n}\right]$

$= \left[m_Y\left(\frac{2t}{\beta}\right)\right]^{2n} = (1-2t)^{-2n} = (1-2t)^{-4n/2}$

Recordar $W \sim \chi^2_k$ tiene $m_W(t) = (1-2t)^{-k/2}$

Usando este hecho podemos decir que $T = \dfrac{2\sum Y_i}{\beta} \sim \chi^2_{4n}$

Dado que β es libre $\Rightarrow$ T es Q Pivote

   (b) Construya un Intervalo de Confianza de 95 % para β, si $n = 10$, $\bar{Y} = 5{,}39$

IC al 95% $\longrightarrow \left[\chi^2_{(4n, 1-\frac{\alpha}{2})} < \dfrac{2\sum Y_i}{\beta} < \chi^2_{(4n, \alpha/2)}\right]$

$\Rightarrow \left[\dfrac{2n\bar{Y}}{\chi^2_{(4n,\alpha/2)}}, \dfrac{2n\bar{Y}}{\chi^2_{(4n,1-\alpha/2)}}\right]$

```r
n<- 20; ybarra<- 5.39
alpha<- .05

ic.inf<-2*n*ybarra/qchisq(alpha/2,4*n,lower.tail = FALSE)
ic.sup<-2*n*ybarra/qchisq(1-alpha/2,4*n,lower.tail = FALSE)
 paste0("El intervalo es [", ic.inf, ",", ic.sup, "]")

# Ahora vam rgamma(n, shape, rate = 1, scale = 1/rate) as alpha=2
m<- 10000
datos<- rgamma(m,2, rnorm(1) )
mean(datos)

hist(datos, breaks= 100,probability = T)
curve(dgamma(x,2,1), col= 'red', lwd=2, add= T)
abline(v=mean(datos), col= 'blue', lwd =3, add=T)

xbarra<- sum(datos)/m

ic.inf<-2*m*xbarra/qchisq(alpha/2,4*n,lower.tail = FALSE)
ic.sup<-2*m*xbarra/qchisq(1-alpha/2,4*n,lower.tail = FALSE)
paste0("El intervalo es [", ic.inf, ",", ic.sup, "]")

#Crees que beta esté en el intervalo?
#Se que alpha es 2 y ya tengo mean datos entonces beta sería??????
```

3. Dos marcas de refrigeradores, denotadas por A y B, están garantizadas por 1 año. En una muestra aleatoria de 50 refrigeradores de la marca A, se observó que 12 de ellos fallaron antes de terminar el periodo de garantía. Una muestra aleatoria independiente de 60 refrigeradores de la marca B también reveló 12 fallas durante el período de garantía. Calcule la diferencia real $(p_1 - p_2)$ entre las proporciones de fallas durante el período de garantía, con un coeficiente de confianza de aproximadamente ,98 si $\hat{p_1} = ,24$, $\hat{p_2} = ,20$.

Si nos acordamos del lab pasado, ahora mi IC está dado por

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Datos $\hat{p}_1 = 24$, $\hat{p}_2 = 20$ $\Rightarrow$ $\hat{q}_1 = 1 - \hat{p}_1 = .76$, $\hat{q}_2 = 1 - \hat{p}_2 = .80$

$n_1 = 50$, $n_2 = 60$, $\alpha = .02 \Rightarrow \alpha/2 = .01$, $z_{.01} = 2.33$

Sustituimos valores y listo!

$$(.24 - .20) \pm z_{.01} \sqrt{\frac{(.24)(.76)}{50} + \frac{(.20)(.80)}{60}}$$

$$\Rightarrow [-.1451, .2251]$$

4. (a) Una pequeña cantidad selenio, de 50 a 200 microgramos (μg) por día, es considerada esencial para una buena salud. Suponga que se seleccionaron muestras aleatorias independientes, de $n_1 = n_2 = 30$ adultos provenientes de dos regiones de Estados Unidos y que se registró para cada persona una ingesta diaria de selenio tanto de líquidos como de sólidos. La media y la desviación estándar de la ingesta diaria de selenio para los 30 adultos de la región 1 fueron $\bar{Y_1} = 167{,}1\mu g$ y $s_1 = 24{,}3\mu g$, respectivamente. Los 1 estadísticos correspondientes para los 30 adultos de la región 2 fueron $\bar{Y_2} = 140{,}9\mu g$ y $s_2 = 17{,}6\mu g$. Encuentre un intervalo de confianza de 95 % para la diferencia en la ingesta media de selenio para las dos regiones.

Sabemos que el IC es $(\bar{Y}_1 - \bar{Y}_2) \pm z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

Sustituimos valores y obtenemos $[15.46, 36.94]$

(b) Consulte la comparación de la ingesta diaria de selenio para un adulto en dos diferentes regiones de Estados Unidos del inciso anterior. Suponga que se desea calcular la diferencia en la ingesta diaria promedio entre las dos regiones, con un error máximo de no más de

$5\mu g$, con probabilidad ,90. Si se planea seleccionar un número igual de adultos de las dos regiones (esto es, si $\mu_1 = \mu_2$), ¿qué tan grandes deben ser n_1 y n_2?

Dado a que $1-\alpha = .90 \Rightarrow \alpha = .10$

Sabemos que $s_1 = 24.3$, $s_2 = 17.6$

$$\Rightarrow z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 5\mu g. \quad \text{y } n_1 = n_2$$

$$\Rightarrow z_{\frac{.10}{2}}\sqrt{\frac{s_1^2 + s_2^2}{n}} = 5 \qquad z_{\frac{.10}{2}} \approx 1.645$$

$$\Rightarrow \sqrt{\frac{s_1^2 + s_2^2}{n}} = 5\left(1/z_{\frac{.10}{2}}\right) = 3.03951$$

$$\Rightarrow \frac{\sqrt{s_1^2 + s_2^2}}{3.0391} = n^{1/2} \Rightarrow \frac{s_1^2 + s_2^2}{3.0391^2} = n \approx \underline{\underline{98}}$$

5. Una muestra aleatoria de n observaciones en la que $Y_i \sim Unif(0, \theta)$ donde θ es desconocida. Use el método de momoentos para estimar el parámetro.

Sabemos que el primer momento μ_1 de una uniforme es su media i.e. $\mu_1 = \mu = \frac{\theta}{2}$

(Recordar $X \sim U(a,b)$ $E[X] = \frac{a+b}{2}$, $V(X) = \frac{(b-a)^2}{12}$)

El primer momento MUESTRAL es $m_1 = \frac{1}{n} \sum_1^n Y_i = \bar{Y}$

$$\Rightarrow \mu_1 = \frac{\theta}{2} = \bar{Y} = m_1 \quad \therefore \quad \boxed{\hat{\theta} = 2\bar{Y}}$$

6. Una muestra aleatoria de n observaciones en la que $Y_i \sim Poissonf(\lambda)$. Use el método de momoentos para estimar λ.

Primer momento

$$\mu_1 = E[Y] = \sum_{y=1}^{\infty} y \frac{e^{-\lambda}\lambda^y}{y!} = e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^y}{(y-1)!} = e^{-\lambda}\lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!}$$

$$= e^{-\lambda}\lambda \underbrace{\sum_{y=0}^{\infty} \frac{\lambda^y}{y!}}_{e^{\lambda}} = \frac{e^{\lambda}}{e^{\lambda}}\lambda = \lambda$$

$$m_1 = \bar{Y} \quad \Rightarrow \quad \lambda = \bar{Y} \quad \therefore \quad \boxed{\hat{\lambda} = Y}$$

7. Suponga que $Y_1, .., Y_n$ es una muestra aleatoria de tamaño n para una distribución normal con media μ y varianza σ^2, encuentre los estimadores de ambos parámetros por el método de momentos.

Sabemos perfectamente que $\mu_1 = E[Y] = \mu$

$$\mu_2 = E[(Y-\mu)^2] = \quad ... = \sigma^2$$

Qué es esto

Ahora

$$m_1 = \bar{Y} \Rightarrow \hat{\mu} = \bar{Y}$$

$$m_2 = \frac{1}{n-1} \sum_1^n (Y_i - \bar{Y})^2 \Rightarrow \mu_2 = \sigma^2 = m_2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n-1} \sum_1^n (Y_i - \bar{Y})$$

segundo momento muestral

8. Introducción al Estimador de Máxima Verosimilitud (EMV).

(a) Idea

(b) Algoritmo

(c) Sea $X_1, ..X_n$ una m.a. tal que $X_i \sim N(0, \theta)$. Encuentra el EMV

Estimador de máxima verosimilitud **MLE** maximum likelihood estimation

Introduction: $X \sim f(x;\theta)$ with $\theta = \{1, 2\}$.

$x = 2$ # observations.

$f(2;\theta=1) = .8$
$f(2;\theta=2) = .2$

nuestro caso: $\hat{\theta} = 1$ porque tiene una proba mayor.

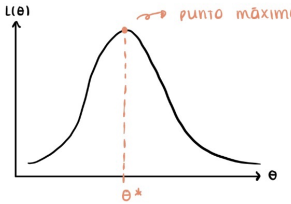
$X_1, \dots, X_n \overset{iid}{\sim} f(x;\theta)$

$\boxed{\text{Likelihood Function}}$ **1** $L(\theta \mid x) = L(\theta; x) = L(\theta) = f(x;\theta) \overset{iid}{=} \prod^{n} f(x_i;\theta)$

variable aleatoria / observaciones

θ: rand. var
x_i: are fixed.

Page 34



punto máximo (derivamos para encontrar el máximo)

$\hat{\theta}^* = \hat{\theta}(x_1, \dots, x_n)$ is the MLE of θ.

2 $\ell(\theta) = \log L(\theta)$

Then: $\dfrac{\partial \ell(\theta)}{\partial \theta} = 0 \implies \hat{\theta}$

3 $\left.\dfrac{\partial^2 L(\theta)}{\partial \theta^2}\right|_{\theta = \hat{\theta}} < 0$

El parámetro debe ser continuo

Remarks:

- If the MLE is unique then its a function of suff stat.
- MLE's may not even exist.
- let $\hat{\theta}$ be the MLE(θ): $E(\hat{\theta})$ is not nec. equal to θ.
- MLE are invariants if MLE(θ) = $\hat{\theta}$ then MLE($\tau(\theta)$) = $\tau(\hat{\theta})$.

Q.6 Let $X_1, \dots, X_n$ be a random sample from $N(0, \theta)$. ($\to \sigma^2$)

(a) Find the ML estimator.

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} e^{-x_i^2/2\theta^2} = \left(\frac{1}{2\pi\theta}\right)^{-n/2} e^{-\sum_i^2 x_i^2/2\theta}$$

$$\ell(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta) - \frac{1}{2\theta}\sum x_i^2$$

$$\ell'(\theta) = \frac{-n}{2\theta} + \frac{\sum x_i^2}{2\theta^2} = 0 \iff \sum x_i^2 - n(\theta) = 0$$

Multiply by $2\theta^2$

$$\iff \boxed{\hat{\theta}_{MLE} = \frac{\sum x_i^2}{n}}$$

$$\frac{d^2}{d\theta^2} = \left.\frac{n}{2\theta^2} - \frac{2\sum x_i^2}{2\theta^3}\right|_{\theta = \hat{\theta}_{MLE}} = \frac{n^3}{2[\sum x_i^2]} - \frac{2n^3}{2[\sum x_i^2]} < 0$$