

Assignatura: **ESTADÍSTICA.**

Estudiant/a:

Data: 31/05/2021

TEORIA 4

- a) Es tracta d'un disseny de dues mostres aparellades, ja que les observacions $Y_{11}, \dots, Y_{15}$ i $X_{11}, \dots, X_{15}$ corresponen als mateixos individus abans del tractament.

Tenim el següent test d'hipòtesis:

$$\begin{cases} H_0: \mu_D = 5 (\mu_D \leq 5) \\ H_1: \mu_D > 5 \end{cases} \quad \text{Per a resoldre el test s'utilitza}$$

$$\text{l'estadístic} \quad t = \frac{\bar{D} - \mu_0}{S_D / \sqrt{n}}, \quad \text{en aquest cas } n=15,$$

que sota la hipòtesi nula té una distribució $t_{n-1} = t_{14}$.

$$\bar{D} = \bar{X} - \bar{Y}, \quad \mu_D = \mu_X - \mu_Y$$

$$S_D = \sqrt{\sum_{i=1}^n (D_i - \bar{D})^2}, \quad \text{on } D_i = X_i - Y_i.$$

- b) Per resoldre el test fem servir l'estadístic de l'apartat anterior, i tindrem com a regió crítica o de rebuig:

$$R_C = \{X, Y \mid |t| \geq t_{n-1, 1-\alpha}\} \quad \text{Prenem com a nivell de significació } \alpha = 0.05.$$

$$\text{El nostre estadístic } |t| = \frac{\bar{D} - \mu_0}{S_D / \sqrt{n}} = \frac{5.23 - 5}{0.88 / \sqrt{15}} = 2.1883$$

de la taula.

$$t_{14, 0.95} = 1.761.$$

Tenim que $t < 1.761 \Rightarrow$ no rebuigem H_0 .

$\Rightarrow$ No podem determinar que la reducció de pes sigui superior

a 5 kg.

TEORIA 3

$X_1, \dots, X_n$ mas $X \sim \text{Poisson}(\lambda)$.

Estimador de λ $\hat{\lambda} = K \sum_{i=1}^n X_i$

a) Trobar K perquè $\hat{\lambda}$ no tingui biaix.

$$B(\hat{\lambda}) = E[\hat{\lambda}] - \lambda = \lambda K \frac{n(n+1)}{2} - \lambda \neq 0 \Rightarrow \text{evidentment } \hat{\lambda} \text{ té biaix.}$$

$$E[\hat{\lambda}] = K \sum_{i=1}^n i E[X_i] = K \sum_{i=1}^n \lambda i = \lambda K \sum_{i=1}^n i = \lambda K \frac{n(n+1)}{2}$$

línealitat
esperança.

$X_i \sim \text{Poisson}(\lambda)$

$$\downarrow \\ E[X_i] = \lambda$$

$$\boxed{K = \frac{2}{n(n+1)}}$$

Però si $K = \frac{2}{n(n+1)}$, aleshores $B(\hat{\lambda}) = \lambda K \frac{n(n+1)}{2} - \lambda = \lambda - \lambda = 0 \Rightarrow$ sense biaix.

b) Eficent?

$\hat{\lambda}$ serà eficient si aconsegueix la cota de Cramer-Rao, per tant minimitza la varància

$$\text{Còl·lelem } V[\hat{\lambda}] = E[\hat{\lambda}^2] - E[\hat{\lambda}]^2$$

$$V[\hat{\lambda}] = K^2 \sum_{i=1}^n i^2 E[X_i] = K^2 \lambda \sum_{i=1}^n i^2 = K^2 \lambda \frac{n(n+1)(2n+1)}{6}$$

$X_i \sim \text{Poisson}(\lambda)$

$$\downarrow \\ E[X_i] = \lambda$$

per $K = \frac{2}{n(n+1)}$

$$V[\hat{\lambda}] = \frac{4}{n^2(n+1)^2} \lambda \frac{n(n+1)(2n+1)}{6} = \frac{2}{3} \frac{(2n+1)}{n(n+1)} \lambda$$

Còl·lelem la cota de Cramer-Rao

$\Rightarrow$ NO és eficient perquè per a altres valors de K ,

trobariem estadístics amb varància més petita,

per ex. $K = \frac{1}{n^3}$ si per exemple el que trobaríem

no s'atendria a la varància

Assignatura: ESTADÍSTICA

Estudiant/a: _____

Data: 31/05/2024

TEORIA 1

a) mostra de n obs X_k^2 .

Sabem, per definició que $X_k^2 = \sum_{i=1}^k Z^2$ on $Z \sim N(0,1)$.

Per tant, per generar una mostra de n obs X_k^2 farem:

$$\left[\begin{array}{l} m = \text{rnorm}(n) \\ m1 = \text{sum}(i=1, i=k) \} m^2 \end{array} \right.$$

b) mostra de n obs de t_k

$$\frac{Z}{\sqrt{\frac{U}{k}}} \sim t_k \text{ amb } Z \sim N(0,1) \text{ i } U \sim \chi_k^2$$

Ex: Continuam amb el codi d'abans, i.e tenim que $m1$ és una ~~obs~~ mostra de n obs de X_k^2 :

$$\left[\begin{array}{l} m1 = \text{sqr t}(m1/k) \\ m2 = \frac{\text{rnorm}(n)}{m1} \end{array} \right. \rightarrow \text{el codi complet seria: } \left[\begin{array}{l} m = \text{rnorm}(n) \\ m1 = \text{sum}(i=1, i=k) \} m^2 \\ m11 = \text{sqr t}(m1/k) \\ m2 = \frac{m}{m11} \end{array} \right.$$

c) ~~per de~~ mostra de n obs de $F_{k1, k2}$

Sabem, per def que $F_{k1, k2} \sim \frac{\chi_{k1}^2/k1}{\chi_{k2}^2/k2}$

Farem:

$$\left[\begin{array}{l} m = \text{rnorm}(n) \\ mk1 = \text{sum}(i=1, i=k1) \} m^2 \\ mk2 = \text{sum}(i=1, i=k2) \} m^2 \\ m3 = \frac{mk1}{mk2} \end{array} \right.$$

TEORIA 2

$$X \sim N(\mu, \sigma^2)$$

a.) Inversa Gamma prior conjugada per $X \sim N(\mu, \sigma^2)$ pel paràmetre σ^2 .

Cal veure que si tenim distribució a priori $\pi(\sigma^2) \sim \text{Inverse Gamma}(\alpha, \beta)$,

la distribució a posteriori $\pi(\sigma^2 | x) \sim \text{Inverse Gamma}(\tilde{\alpha}, \tilde{\beta})$.

$$\text{on } \pi(\sigma^2 | x) = \frac{f(x | \sigma^2) \pi(\sigma^2)}{\int f(x | \sigma^2) \pi(\sigma^2) dx} \propto f(x | \sigma^2) \pi(\sigma^2)$$

$$f(x | \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

$$\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}$$

$$\pi(\sigma^2 | x) \propto f(x | \sigma^2) \pi(\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\tilde{\alpha}-1} e^{-\frac{\tilde{\beta} - x^2}{2\sigma^2}} e^{-\frac{\beta - 2x^2}{2\sigma^2}} = e^{-\frac{\beta - 2x^2}{2\sigma^2}}$$

0's

Assignatura: ESTADÍSTICA

Estudiant/a:

Data: 21/05/2021

PROBLEMA 1

$Y_1, \dots, Y_n$ i.i.d. $P_\theta(Y=y) = \binom{y-1}{r-1} \theta^r (1-\theta)^{y-r}$ $y = r, r+1, r+2, \dots$, $r \in \mathbb{N}$ conegut $\theta \in (0, 1)$.

a) $L(\theta|Y) = \prod_{i=1}^n P_\theta(Y=y_i) = \prod_{i=1}^n \binom{y_i-1}{r-1} \theta^r (1-\theta)^{y_i-r} = \theta^{nr} \left(\prod_{i=1}^n \binom{y_i-1}{r-1} \right) \cdot (1-\theta)^{\sum_{i=1}^n y_i - nr}$

Funció de versemblança.

$L(\theta|Y) = nr \log \theta + \sum_{i=1}^n \log \binom{y_i-1}{r-1} + (\sum_{i=1}^n y_i - nr) \log(1-\theta) \leftarrow$ funció logversemblança.

$r=2, n=1, Y=2 \quad L(\theta|Y) = \theta^2 \binom{1}{1} (1-\theta)^{2-2} = \theta^2$



b) Mètode moments

$\mu_1 = E[Y] = \frac{r}{\theta} \Leftrightarrow \theta = \frac{r}{\mu_1} \Rightarrow \hat{\theta}_1 = \frac{r}{\bar{Y}} = \frac{r}{\bar{Y}}$ principi substitució

$\mu_2 = E[Y^2] = V[Y] + E[Y]^2 = \mu_2 + \mu_1^2 = \frac{r(1-\theta)}{\theta^2} + \frac{r^2}{\theta^2} = \frac{r-r\theta+r^2}{\theta^2}$
 $V[Y^2] = E[Y^2] - E[Y]^2$

$\Rightarrow \theta^2 \mu_2 = r-r\theta+r^2 \Leftrightarrow \theta^2 \mu_2 + r\theta = r+r^2$

$\theta = \frac{-r \pm \sqrt{r^2 + 4\mu_2(r+r^2)}}{2\mu_2}$ $\rightarrow$ Prenem signe positiu perquè $\theta \in (0,1)$.

$\theta = \frac{-r + \sqrt{(1+4\mu_2)r^2 + 4\mu_2 r}}{2\mu_2} \Rightarrow \hat{\theta}_2 = \frac{-r + \sqrt{(1+4\mu_2)r^2 + 4\mu_2 r}}{2\mu_2}, \mu_2 = \frac{1}{n} \sum_{i=1}^n Y_i^2$

$$\mu_2 = W(Y) = \frac{Y(1-e)}{e^2} \Rightarrow e^2 \mu_2 = Y - Y e \Leftrightarrow \mu_2 e^2 + Y e - Y = 0$$

$$\theta = \frac{-Y \pm \sqrt{Y^2 + 4r\mu_2}}{2\mu_2} \rightarrow \text{Prenem } \oplus \text{ perquè } \theta > 0$$

$$\theta = \frac{-Y + \sqrt{Y^2 + 4r\mu_2}}{2\mu_2}$$

$$\left[\hat{\theta}_3 = \frac{-r + \sqrt{r^2 + 4r\hat{\sigma}^2}}{2\hat{\sigma}^2} \right], \text{ on } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

per $n=2, r=2, Y=2,3$ $\bar{Y} = \frac{2+3}{2} = 2.5$

$$\left[\hat{\theta}_1 = \frac{2}{2.5} = \frac{4}{5} = 0.8 \right]$$

$$\left[\hat{\theta}_2 = \frac{-2 + \sqrt{(4+13)4 + 4 \cdot 13}}{13} = \frac{-2 + \sqrt{108 + 52}}{13} = \frac{-2 + \sqrt{160}}{13} = \frac{-2 + 4\sqrt{10}}{13} \approx \frac{3.55}{13}$$

$$m_2^1 = \frac{1}{2} (2^2 + 3^2) = \frac{13}{2} = 6.5$$

$$\left[\hat{\theta}_3 = \frac{-2 + \sqrt{4 + \frac{8}{4}}}{\frac{2}{4}} = \frac{-2 + \sqrt{6}}{\frac{1}{2}} = -4 + 2\sqrt{6} \approx 0.899 \right]$$

$$\hat{\sigma}^2 = \frac{1}{2} (12 - 2 \cdot 2.5^2 + 13 - 2 \cdot 2.5^2)^2 = \frac{1}{4}$$

c) Estimar màxim versemblant

utilitzem la funció logversemblança de l'apartat a).

$$\ell_Y(\theta) = nr \log \theta + \sum_{i=1}^n (y_i - 1) \log \theta + \left(\sum_{i=1}^n y_i - nr \right) \log(1 - \theta)$$

$$\frac{\partial \ell}{\partial \theta} = \frac{nr}{\theta} + \frac{nr - \sum_{i=1}^n y_i}{1 - \theta} = 0 \Leftrightarrow \frac{nr(1 - \theta) + nr - \sum_{i=1}^n y_i}{\theta(1 - \theta)} = 0 \Leftrightarrow nr - \theta \sum_{i=1}^n y_i = 0$$

$$\rightarrow \left[\hat{\theta}_{MV} = \frac{nr}{\sum_{i=1}^n y_i} = \frac{Y}{Y} \right]$$

Sabem que la distribució asimptòtica de $\hat{\theta}_{MV}$ és

$$\hat{\theta}_{MV} \rightarrow N\left(\theta, \frac{1}{I_X(\theta)}\right) \quad \text{àlculen } I_X(\theta) = \mathbb{E}\left[\frac{-\partial^2 \ell(\theta; Y)}{\partial \theta^2}\right]$$

[Handwritten signature]

Assignatura: **ESTADÍSTICA**

Estudiant/a: XXXXXXXXXX

Data: 31/05/2021

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{-\sum_{i=1}^n y_i \theta (1-\theta) - \theta \sum_{i=1}^n y_i (1-2\theta)}{\theta^2 (1-\theta)^2} = \frac{(nr - \theta \sum_{i=1}^n y_i) - \theta \sum_{i=1}^n y_i (1-2\theta)}{\theta^2 (1-\theta)^2} = \frac{-\theta \sum_{i=1}^n y_i (1-\theta) - nr\theta + \theta \sum_{i=1}^n y_i + 2nr\theta - 2\theta^2 \sum_{i=1}^n y_i}{\theta^2 (1-\theta)^2}$$

$$= \frac{nr(2\theta-1) - \theta^2 \sum_{i=1}^n y_i}{\theta^2 (1-\theta)^2}$$

$$I_Y(\theta) = \mathbb{E} \left[\frac{\theta^2 \sum_{i=1}^n y_i - nr(\hat{\theta}-1)}{\theta^2 (1-\hat{\theta})^2} \right] = \frac{1}{\theta^2 (1-\theta)^2} (-nr(2\theta-1) + \theta^2 \sum_{i=1}^n \mathbb{E}[y_i]) = \frac{-nr(2\theta-1) + nr\theta}{\theta^2 (1-\theta)^2} = \frac{nr - nr\theta}{\theta^2 (1-\theta)^2} = \frac{nr}{\theta^2 (1-\theta)^2}$$

$\mathbb{E}[y_i] = \theta$

Per tant $\left[\hat{\theta}_{MV} \xrightarrow{n \rightarrow \infty} N\left(\theta, \frac{\theta(1-\theta)}{nr}\right) \right]$

1) $IC_{1-\alpha}(\theta)$

Per l'aportat anterior podem construir la quantitat pivotal $Z \sim N(0,1)$

$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

$$\sqrt{n I_Y(\theta)} (\hat{\theta}_{MV} - \theta) \rightarrow N(0,1)$$

$$Z = \sqrt{n I_Y(\theta)} (\hat{\theta}_{MV} - \theta) \sim N(0,1)$$

$$1-\alpha = P\left(Z_{\alpha/2} \leq Z \leq Z_{1-\alpha/2}\right) = P\left(Z_{\alpha/2} \leq \sqrt{n I_Y(\theta)} (\hat{\theta}_{MV} - \theta) \leq Z_{1-\alpha/2}\right) =$$

$$= P\left(\frac{Z_{\alpha/2}}{\sqrt{n I_Y(\theta)}} \leq \hat{\theta}_{MV} - \theta \leq \frac{Z_{1-\alpha/2}}{\sqrt{n I_Y(\theta)}}\right) = P\left(\frac{Z_{\alpha/2} - \hat{\theta}_{MV}}{\sqrt{n I_Y(\theta)}} \leq -\theta \leq \frac{Z_{1-\alpha/2} - \hat{\theta}_{MV}}{\sqrt{n I_Y(\theta)}}\right) =$$

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

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

$$\Rightarrow IC_{1-\alpha}(\theta) = \left[\hat{\theta}_{MV} \pm Z_{1-\alpha/2} \frac{1}{\sqrt{n I_Y(\theta)}} \right]$$

$$[IC_{1-\alpha}(\theta) = \frac{\bar{Y}}{Y} \pm Z_{\frac{\alpha}{2}} \frac{1}{\sqrt{\frac{nr}{\theta(1-\theta)}}}]$$

ja era $I_Y(\theta)$
seu quadrat

-0.2

e) $\alpha = 0.05$

$$\begin{cases} H_0: \theta \geq 0.5 \\ H_1: \theta < 0.5 \end{cases}$$

~~Muntanyes~~ ~~Rubins~~ ~~alternatives~~ ~~competitors~~

Karlin-Rubin

$$\frac{L(\theta_0 | Y)}{L(\theta_1 | Y)} = \frac{\binom{n}{\theta_0} \left(\frac{1-\theta_0}{\theta_0} \right)^{\sum_{i=1}^n Y_i - nr}}{\binom{n}{\theta_1} \left(\frac{1-\theta_1}{\theta_1} \right)^{\sum_{i=1}^n Y_i - nr}}$$

Prenem l'estadística $T(Y) = \sum_{i=1}^n Y_i - nr = n\bar{Y} - nr = n(\bar{Y} - r)$

Volem que la raó de versemblança $\frac{L(\theta_0)}{L(\theta_1)}$ és decreixent en $T(Y)$, ja que $\theta_0 > \theta_1$.

Per tant prendrem per regió crítica $RC = \{Y | T \leq t\}$

amb t t.q. $P(Y \in RC) = 0.05$, i.e. $P(T \leq t) = 0.05$

Aquest test serà UMP, pel teorema de Karlin-Rubin

$$P(T \leq t) = P\left(n\bar{Y} \leq t + nr\right) = P\left(\sum_{i=1}^n Y_i \leq t + nr\right)$$

sigui $X = \sum_{i=1}^n Y_i \Rightarrow P(X \leq t + nr) = 0.05 \Leftrightarrow P(F_X(t + nr) = 0.05$

$$\Leftrightarrow t + nr = F_X^{-1}(0.05) \Leftrightarrow t = F_X^{-1}(0.05) - nr$$

on X és la distribució de $\sum_{i=1}^n Y_i$, amb Y_i variable Negativa (r, θ)
 per $\theta = 0.5$

No cal
 $E(Y) = EY = 0.5$

-0.3

Assignatura: **ESTADÍSTICA**

Estudiant/a: XXXXXXXXXX

Data: 31/05/2020

1) $\alpha = 0.05$

$$H_0: \theta = 0.5 = \theta_0$$

$$H_1: \theta \neq 0.5$$

Aquí utilitzarem el test de mètode de versemblança generalitzada.

$$\Lambda(X) = \frac{\mathcal{L}(\theta_0|Y)}{\max_{\theta \in (0,1)} \mathcal{L}(\theta|Y)} = \frac{\mathcal{L}(\theta_0|Y)}{\mathcal{L}(\hat{\theta}_{MV}|Y)} = \left(\frac{\hat{\theta}_0}{\hat{\theta}_{MV}} \right)^{\sum_{i=1}^n y_i - n\theta_0} \left(\frac{1-\hat{\theta}_0}{1-\hat{\theta}_{MV}} \right)^{n - \sum_{i=1}^n y_i + n\theta_0}$$

amb

La regió crítica d'aquest test és $\{X | \Lambda(X) \leq A_\alpha\}$

on A_α és t.q. $\alpha = 0.05 = P(\Lambda(X) \leq A_\alpha)$

$\theta = 0.5$

Per trobar A_α també podem utilitzar qe $-\log \Lambda(X) \xrightarrow{n \rightarrow \infty} \chi^2_{d=1}$ i utilitzem la distribució χ^2_1

9)

$$Y_i \text{ i.i.d. } \epsilon = 1, -1, 127 \quad r=1$$

Fem servir l'estimador màxim versemblant qe hem trobat a l'apartat c) $\hat{\theta}_{MV} = \frac{r}{n}$

i substituïm les dades per les del problema concret: $\left. \begin{array}{l} r=1 \\ \bar{Y} = \frac{360}{127} \approx 2.835 \end{array} \right\}$

$$\Rightarrow \hat{\theta}_{MV} = \frac{127}{360} \approx 0.35$$

Fem servir també l'interval de confiança trobat a l'apartat d) per θ

$$IC_{95\%}(\theta) = \hat{\theta}_{MV} \pm Z_{1-\alpha/2} \frac{1}{\sqrt{\frac{\hat{\theta}_{MV}}{\theta^2(1-\theta)}}} \approx 0.35 \pm Z_{0.975} \frac{1}{\sqrt{\frac{0.35}{0.35^2(1-0.35)}}} = 0.35 \pm 2.045 \frac{1}{\sqrt{0.44748}} = 0.35 \pm 1.90$$

utilitzem taula de la $N(0,1)$

$$\approx 0.35 \pm 0.6044 = [0.3456, 0.3544]$$

-0.2

h) Faria un test de bondat d'ajust.

$$\begin{cases} H_0: X_i \sim \text{BinNeg}(r, p) \\ H_1: Y_i \sim \text{BinNeg}(r, p) \end{cases}$$



~~Exercici~~

Donada la llista de ~~paràmetres~~ $h_1, 2, 3, 4, \dots, 12$ en subconjunts $I_1, \dots, I_k$.

Comentaria moment $I_1 = 1, I_2 = 2, \dots, I_{12} = 12$.

Aleshores definim $N_j = \{\text{freqüència per } I_j\}$

Calcularia p_j , la probabilitat de que Y pertanyi a I_j sota l'hipòtesi nul·la.

Per a fer-ho, caldrà estimar θ usant $\hat{\theta}_{ML}$ trobat a l'aportat anterior.

Aleshores el valor esperat de N_j és $N \cdot p_j$ on $N = 127$.

Si algun valor esperat donés < 5 reagruparia els subconjunts agrupant-ne alguns.

$$\text{Definim l'estadístic de Pearson } \chi^2 = \sum_{j=1}^K \frac{(N_j - N p_j)^2}{N p_j}$$

Sota H_0 , $\chi^2 \sim \chi^2_{k-1-p}$ on k és el ~~seu~~ nombre de subconjunts.

p el nombre de paràmetres estimats, en aquest cas $p = 1$ (enimem θ i r ja el sabem $r = 1$)

La regió crítica del test serà $EC = \{ \chi^2 \mid \chi^2 \geq \chi^2_{k-1-p, 1-\alpha} \}$, on α és el nivell de significació, normalment $\alpha = 0.05$.

(P1) a) ok

b) ok

(P2) a) ok

b) ok per haver d'interpretar el -0.1 que ven...