Solución Control 2 MA34B-3 Estadística

Profesores A. Peña, N. Lacourly, R. Assar
pauta: por Diego Díaz Espinoza

1.- Problema 1 (3 puntos) Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra aleatoria i.i.d. proveniente de la distribución Beta de parámetros $(\theta, 1)$, con función de densidad de probabilidad dada por $f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1$, donde $\theta > 0$.

(a) Muestre que el test ms potente de tamaño α para contrastar $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, con $\theta_1 > \theta_0$, rechaza H_0 para valores grandes del estadstico $T(\mathbf{X}) = \sum_{i=1}^n \log X_i$, i.e., tiene región crtica de la forma $Re^* = \{\mathbf{x} \in \mathcal{X} : T(\mathbf{x}) \geq k_\alpha\}$.

(b) Demuestre que $-2\theta \log X_i \sim \chi^2_2$. Deduzca la distribucin de $-2\theta \sum_{i=1}^n \log X_i$.

(c) ¿ Cuál es su decisión en relación a H_0? considerando $\theta_0 = 2$, $\alpha = 0.05$, $n = 10$ y $\sum \log x_i = -5.846$ y un error de tipo I de 5%.

SOLUCION

(a)

$$f(\bar{x}/\theta) = \prod_{i=1}^n \theta x_i^{\theta-1} = \theta^n \prod_{i=1}^n x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta-1}$$

$$\Rightarrow f(\bar{x}/\theta = \theta_0) = \theta_0^n \left(\prod_{i=1}^n x_i \right)^{\theta_0-1}$$

$$\text{y } f(\bar{x}/\theta = \theta_1) = \theta_1^n \left(\prod_{i=1}^n x_i \right)^{\theta_1-1}$$

$$\text{haciendo el cuociente entre ambas } \frac{f(\bar{x}/\theta = \theta_0)}{f(\bar{x}/\theta = \theta_1)} = \left(\frac{\theta_1}{\theta_0} \right)^n \left(\prod_{i=1}^n x_i \right)^{(\theta_1-\theta_0)} \geq k_\alpha$$

$$\text{trabajando un poco } \left(\prod_{i=1}^n x_i \right)^{(\theta_1-\theta_0)} \geq k_\alpha^{'} \text{ pues } \theta_0 \text{ y } \theta_1 > 0$$

$$\text{sacando la raíz } \left(\prod_{i=1}^n x_i \right) \geq k_\alpha^{''} \text{ notar que } \theta_1 - \theta_0 > 0$$

$$\text{aplicando logaritmo natural } ln\left(\prod_{i=1}^n x_i \right) \geq k_\alpha^{'''}$$

$$\text{explotando la productoria } \left(\sum_{i=1}^{n} ln(x_i)\right) \geq k_{\alpha}^{'''}$$

Luego se tiene la región crítica o de rechazo pedida:

$$R^* = \left\{\bar{x} : T(\bar{x}) = \sum_{i=1}^{n} ln(x_i) \geq k_{\alpha}^{'''}\right\}$$

(b) Se puede usar el cambio de variable: $y = -2\theta log(x_i)$ y usar el t.c.v. de probabilidades o usar la f.g.m. que resolveré acá:

$$E\big(e^{-2\theta log(x_i)t}\big) = E\big(e^{log(x_i^{-2\theta t})}\big) = E\big(x_i^{-2\theta t})\big) = \theta \int_0^1 x_i^{\theta-1-2\theta} dx_i$$

$$= \theta \frac{x_i^{\theta-2\theta}}{\theta - 2\theta t}\bigg|_0^1 = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}} \text{ con } n = 2$$

es decir $-2\theta log(x_i) \sim \chi_2^2$ Para la suma se puede decir que se vió en clases que la suma de $\chi_n^2 + \chi_m^2 = \chi_{n+m}^2$

(c) Se puede usar un intervalo de confianza para θ y ver si el valor para H_0 está en el intervalo para dicha confianza. Entonces:

$$P(z > b) = \frac{\alpha}{2} \Rightarrow b = \chi^2_{(0.0025,20)} = 34.2 \text{ y } P(z > a) = 1 - \frac{\alpha}{2} \Rightarrow a = \chi^2_{(0.975,20)} = 9.59$$

Luego

$$p\left(\frac{9.59}{-2\theta log(x_i)} \leq \theta \leq \frac{34.2}{-2\theta log(x_i)}\right) = 0.05$$

Con lo que el IC parar θ con $\alpha = 0.05$ queda $[0.82, 2.93]$ Luego NO se rechaza la hipótesis nula dado que $\theta_0 = 2$ está dentro del intervalo. Ahora el error tipo I:

$$P\Big(x \in R^*/\theta = \theta_0\Big) = \alpha$$

$$= P\Big(\sum log(x_i) > k_\alpha/\theta = \theta_0\Big)$$

$$= P\Big(-2\theta \sum log(x_i) < -2\theta k_\alpha/\theta = \theta_0\Big)$$

$$= P\Big(\chi_{2n}^2 < -2\theta_0 k_\alpha\Big)$$

Recuerden que la tabla muestra los valores $P(z > k)$ luego hay que dar vuelta esta probabilidad para leer:

$$P\Big(\chi_{2n}^2 > -2\theta_0 k_\alpha\Big) = 1 - \alpha = 0.95$$

Entonces $-k2\theta = 10.85 \Rightarrow k = -2.7125$. Ahora el error de tipo II:

$$P\bigg(x \notin R^*/\theta = \theta_1\bigg) = \beta$$

$$= P\bigg(\sum log(x_i) < k_\alpha/\theta = \theta_1\bigg)$$

$$= P\bigg(-2\theta \sum log(x_i) > -2\theta k_\alpha/\theta = \theta_1\bigg)$$

$$= P\bigg(\chi^2_{2n} > -2\theta_1 k_\alpha\bigg) = \beta$$

Entonces $-k_\alpha 2\theta_1 = 15.19$ (usando el k del error tipo I). Luego buscamos el valor en la tabla pero no aparece entonces se estima linealmente con lo que $\beta \in (0.75, 0.73) \Rightarrow \beta \sim 0.73$ o simplemente 0.75 ó 0.5.

2.- Problema 2 3 puntos) Sean $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{Y} = (Y_1, \ldots, Y_m)$ muestras aleatorias mutuamente independientes provenientes de las distribuciones $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, respectivamente.

(a) Proponga un test para contrastar $H_0 : \mu_1 \geq \mu_2$ versus $H_1 : \mu_1 \leq \mu_2$.

(b) ¿Cuál es su decisión con un error de tipo I de 5% considerando los datos $n = 50$, $m = 60$, $\bar{x} = 3.2$, $\bar{y} = 3.7$, $\frac{1}{n}\sum(x_i - \bar{x})^2 = 0.62$ y $\frac{1}{m}\sum(y_i - \bar{y})^2 = 0.79$.

(c) ¿El p-valor del test es menor o mayor que 5%?

SOLUCION

a) Debido a que hay una desigualdad se usa:

$$H_0 : \mu_1 - \mu_2 = c_1, \quad c1 \geq 0$$
$$H_1 : \mu_1 - \mu_2 = c_2, \quad c2 \leq 0$$

Debido a que sabemos que $x_i \sim N(\mu_1, \sigma^2)$ y $x_i \sim N(\mu_2, \sigma^2)$ entonces se forma una $Normal(0, 1)$:

$$\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim N(0, 1)$$

Ahora se forma una χ^2_{n+m-2}

$$\frac{nS_n^2}{\sigma^2} \sim \chi^2_n$$

OJO: El factor que acompaña a S_n^2 TIENE que ser n y no $n-1$ pues el enunciado define claramente $s = \frac{1}{n}\sum(x_i - \bar{x})^2$ con n y no $n-1$. Luego se tiene:

$$\frac{nS_n^2}{\sigma^2} + \frac{mS_m^2}{\sigma^2} \sim \chi^2_{n+m-2}$$

Luego formamos una $t - Student$ con ambas:

$$T(\bar{x},\bar{y}) = \frac{\frac{\bar{x}-\bar{y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}}{\sqrt{\frac{\frac{nS_n^2}{\sigma^2}+\frac{mS_m^2}{\sigma^2}}{(n+m-2)}}} \sim t_{n+m-2}$$

Luego usando este estadístico como pivote se puede obtener una región crítica de la forma:

$$R^* = \{T(\bar{x},\bar{y}) \leq t\} = \alpha$$

b) Hay al menos dos formas. Una de ellas es buscando un IC para $\bar{x}-\bar{y}$ al 95% y viendo si contiene o no a números mayores que cero o menores que cero. Otra forma es usando la región crítica encontrada en a).

Usando la parte a):

$$P(T > t/\bar{x}-\bar{y} = c_1 \geq 0) = \alpha$$

Se ve el valor de la tabla para t y se obtiene que $t \in (1.658, 1.671)$ supongamos $t = 1.658$ (esto es tomando $n = 120$ que es más cercano que $n = 60$ a $n = 108$ que es el valor del experimento). Luego

$$P(\bar{x}-\bar{y} < t\left(\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{nS_n^2+mS_m^2}{n+m-2}}\right) + (\mu_1-\mu_2)) = \alpha$$

Reemplazando los valores se tiene que:

$$P(\bar{x}-\bar{y} < -0.24) = 0.05$$

Usando un IC):

$$P\left(a < \frac{\frac{\bar{x}-\bar{y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}}{\sqrt{\frac{\frac{nS_n^2}{\sigma^2}+\frac{mS_m^2}{\sigma^2}}{(n+m-2)}}} < b\right)$$

$$P(ak < \bar{x}-\bar{y}-(\mu_1-\mu_2) < bk) = \alpha$$

Luego

$$P(\bar{x}-\bar{y}+bk < \mu_1-\mu_2 < \bar{x}-\bar{y}+ak) = \alpha$$

Esto quiere decir que:

$$P(\mu_1-\mu_2 > \bar{x}-\bar{y}+ak) = \frac{\alpha}{2} = 0.025 \text{ y } P(\mu_1-\mu_2 > \bar{x}-\bar{y}+bk) = 1-\frac{\alpha}{2} = 0.975$$

Como la $t-student$ es simétrica:

$$P(\mu_1-\mu_2 > \bar{x}-\bar{y}+ak) = \frac{\alpha}{2} = 0.025 \text{ y } P(\mu_1-\mu_2 > \bar{x}-\bar{y}-ak) = 1-\frac{\alpha}{2} = 0.975$$

Viendo la tabla $a = T(0.975, 120) = -1.980 \Rightarrow b = T(0.025, 120) = 1.980$
Así el IC queda

$$IC(0.0975, \mu_1 - \mu_2) = [-0.82, -0.18]$$

Ahora, viendo el IC se puede ver que $\mu_1 - \mu_2 = \bar{x} - \bar{y} = -0.5$ el intervalo sólo asegura valores negativos para $\mu_1 - \mu_2$ esto quiere decir que se rechaza la hipótesis nula. Por otro lado, usando la parte a), se puede decir que $\bar{x} - \bar{y} \leq 0$ con 95% de confianza. Luego se rechaza $H_0 \Rightarrow \bar{x} < \bar{y}$.

c.- Supongamos n crece mucho y m se hace pequeño pero no tanto como para que $\frac{1}{m}$ no tienda a cero. Entonces: $\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{nS_n^2+mS_m^2}{n+m-2}} \rightarrow \sqrt{\frac{1}{m}}\sqrt{S_n^2} \sim S_n\sqrt{\frac{1}{m}}$. Así la región de rechazo se acerca a -0.5 con lo que cada vez es menos probable que se rechaze H_0.

3.- Problema 3(6 puntos)Un laboratorio requiere determina la concentracin de ADN de una bacteria en un extracto dado. Considerando que se comente errores de medicin, se replica 20 veces la medicin. En la tabla adjunta se presenta los datos obtenidos de manera aleatoria y independiente. Se sabe que las mediciones se ajustan a una distribucin $N(\mu, \sigma^2)$ donde μ y σ^2 son desconocidos y μ es el valor real de la concentracin de ADN.

Replica	Concentracin	Replica	Concentracin	Replica	Concentracin	Replica	Concentracin
1	0.503	2	0.249	3	1.047	4	0.670
5	0.882	6	0.858	7	0.805	8	0.549
9	0.563	10	0.706	11	0.503	12	0.536
13	1.168	14	0.813	15	0.496	16	0.790
18	0.518	19	0.463	20	0.735	21	1.038

La media y la desviacin estndar son respectivamente: $\bar{x} = 0.695$ y $s = 0.232$.

(a) Encuentre el intervalo ms pequeo tal que el valor real μ de la concentracin pertenezca a dicho intervalo con una probabilidad de 95%. ¿Por qu $\bar{x}$ pertenece a este intervalo?

(b) Si queremos un intervalo de largo igual a XXX con un nivel de confianza de 95% ¿cual debe ser el tamao de la muestra? Si queremos un intervalo de largo igual a XXX con una muestra de tamao igual a 20 ¿cual debe ser el nivel de confianza?

(c) Para el laboratorio es ms arriesgado sobre estimar el valor real μ de la concentracin. Encuentre un intervalo [a,b] de modo que la concentracin de ADN pertenezca a dicho intervalo con un nivel de confianza de 95% y tal que $P(\mu < a) = \frac{1}{4}P(\mu > b)$.

(d) Para demostrar al laboratorio la baja calidad de las mediciones, se solicita a Usted que encuentre un intervalo al cual pertenezca σ^2 el valor real de la varianza de la medicin con un nivel de confianza de 90%.

(e) Sea una variable aleatoria con distribucin de Poisson de parmetro $\lambda > 0$ desconocido. Se toma una muestra aleatoria simple $\mathbf{X} = (X_1, \ldots, X_n)$, $n > 0$. Se obtiene $x_i = 0, \quad \forall i$. Dado $\lambda = \lambda_o > 0$, calcule $P(x_1 = 0, ..., x_n = 0|\lambda = \lambda_o)$. Prueba que si λ_o es tal que $e^{-n\lambda_o} \geq \alpha_o$, entonces $\lambda_o = -\frac{1}{n}ln(\alpha_o)$. Concluya entonces que $\lambda_o \in [0, -\frac{1}{n}ln(0.05)]$ si $P(x_1 = 0, ..., x_n = 0|\lambda = \lambda_o) \geq 0.05$. Comente los alcances de este resultado.

SOLUCION

a) Puesto que se tiene $\bar{x}$ y S_n^2 se usa el pivote de una t-student. Pero como S no se ha definido explícitamente en el enunciado puede usarse un S sesgado o insesgado esto es:

Sesgado)

$$S = S_n = \frac{\sum (x - \bar{x})^2}{n}$$

$$\Rightarrow \frac{nS_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

Y el pivote queda:

$$\frac{\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{nS_n^2}{(n-1)\sigma^2}}} = \frac{\bar{x} - \mu}{S_n}\sqrt{(n-1)} \sim t_{n-1}$$

Insesgado)

$$S = S_{n-1} = \frac{\sum (x - \bar{x})^2}{n - 1}$$

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

Y el pivote queda:

$$\frac{\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{(n-1)S_n^2}{(n-1)\sigma^2}}} = \frac{\bar{x}-\mu}{S_n}\sqrt{n} \sim t_{n-1}$$

De todas formas la diferencia en el intervalo de confianza está en el tercer decimal así que no es significativo en el problema, por ello se resolverá acá sólo usano S_n. Esto es:

$$z = \frac{\bar{x}-\mu}{\frac{S_n}{\sqrt{n}}} \sim t_{n-1} = t_{(19,95\%)}$$

Luego:

$$P(a < z < b) = 0.95 = P(\bar{x} - b\frac{S_n}{\sqrt{n}} < \mu < \bar{x} - a\frac{S_n}{\sqrt{n}})$$

Por simetría de la t-student:$a = -b$. Y viendo la tabla: $a = t_{0.975,19} = -t_{0.025,19} = -2.093$ y $b = 2.093$. Así el IC queda:

$$IC(0.95, \mu) = [0.584, 0.806]$$

Luego $\bar{x} \in IC$ pues el intervalo fue construido centrado en $\bar{x}$.

b) Se quiere que

$$\left|\bar{x} - b\frac{S_n}{\sqrt{n}} - \bar{x} + a\frac{S_n}{\sqrt{n}}\right| = 0.12$$

Luego

$$\Rightarrow b\frac{S_n}{\sqrt{n}} - a\frac{S_n}{\sqrt{n}} = 0.12$$

$$a - b = 0.12\frac{\sqrt{n}}{S_n}$$

Por simetría $a = -b$ así

$$\Rightarrow a \sim 1.127 \Rightarrow P(z < 1.127) = \frac{\alpha}{2} \sim 0.125 \Rightarrow \alpha = 0.25$$

c) Se usa el pivote para una Chi

$$z = \frac{nS_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\Rightarrow P(a < z < b) = 0.9$$

$$P(\frac{1}{b} < \frac{\sigma^2}{nS_n^2} < \frac{1}{a}) = 0.9$$
$$\Rightarrow P(\frac{nS_n^2}{b} < \sigma^2 < \frac{nS_n^2}{a}) = 0.9$$
$$\Rightarrow P(\sigma^2 > \frac{nS_n^2}{a}) = 0.05 \Rightarrow \frac{nS_n^2}{a} = 30.1$$
$$\Rightarrow P(\sigma^2 < \frac{nS_n^2}{b}) = 0.05 \Rightarrow \frac{nS_n^2}{b} = 10.12$$

Así el IC queda:

$$IC \sim [0.036, 0.106]$$

d)

$$P(a' < \mu < b') = 0.95 \Rightarrow P(\mu < a') = \frac{1}{4}P(\mu > b')$$

Luego

$$P(\mu < a') = 0.01 \Rightarrow P(\frac{\bar{x} - \mu}{\frac{S_n}{\sqrt{n}}}) = 0.01$$

$$\Rightarrow b' = 2.539 \text{ , } a' = 1.8$$

Así

$$a \sim 0.563 \text{ y } b \sim 0.79$$

Entonces el IC queda:

$$[0, 563, 0.79]$$

e) $z \sim Poisson(\lambda)$, $\lambda > 0$

$$P(x_1 = 0, x_2 = 0 \ldots x_n = 0/\lambda) = P(x_i = 0/\lambda)^n = (e^{-\lambda})^n = e^{-n\lambda}$$

Luego por enunciado:

$$e^{-n\lambda} \geq \alpha_0 \Rightarrow -n\lambda \geq ln(\alpha_0) \Rightarrow \lambda \leq \frac{-1}{n}ln(\alpha_0)$$

Así

$$P(x_i = 0/\lambda) \geq 0.05 \text{ pero } \lambda \leq \frac{-1}{n}ln(\alpha_0 = 0.05)$$
$$\text{pero } e^{-n\lambda} \text{ estrictamente decreciente}$$
$$\Rightarrow \lambda \leq 0 \Rightarrow \lambda \in (0, \frac{-1}{n}ln(0.05)]$$

Que era lo pedido. Esto es se puede usar $P(x_i = 0, \forall i/\lambda)$ como forma de calcular un IC.

Diego Díaz Espinoza