Ingeniería Matemática
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
Probabilidades y Estadística 09-2

Profesor: Felipe Célèry C.
Profesor Auxiliar: Víctor Riquelme F.

Control 3

P1. Las marca de ampolletas *Rayito de Sol* ha recibido críticas acerca de la vida útil de su producto. La compañia asegura que la vida útil sus ampolletas sigue una v.a. exponencial de parámetro $1/\theta$ con $\theta = 5/4[\frac{\text{año}}{\text{amp}}]$ (i.e. cada ampolleta dura, teóricamente, 5/4 de año). Por otra parte un estudio muestra que el parámetro $\theta = 1$. Se hace un encuesta en 100 hogares obteniéndose un promedio muestral de 1,13 años para la duración de las ampolletas.

(a) Encuentre el estimador de máxima verosimilitud de θ.

(b) Calcule la esperanza y varianza del estimador anterior y averigüe si es consistente.

(c) Si usted fuese abogado de los clientes en contra de *Rayito de Sol* tomaría como hipótesis nula "El estudio tiene razón" en vez de la hipótesis "La compañía tiene razón". ¿Por qué? Explique en base a los errores Tipo I y Tipo II.

(d) Muestre que para el contraste (test)

$$H_0: \text{ "El estudio tiene razón"}$$
$$H_1: \text{ "La compañía tiene razón"}$$

la región más potente es de la forma $R^* = \{\sum_{i=1}^n x_i \geq k\}$.

(e) Usando un tamaño de test $\alpha = 0,05$ calcule el valor de k.
Indicación: Justifique sus cálculos usando T.C.L.

(e) Decida si rechaza o no la hipótesis nula.

P2. Supongamos que en una fábrica hay dos máquinas: tipo 1 y tipo 1. Se quiere estudiar el costo semanal de reparación de ambas máquinas. Para esto suponemos que el costo de mantención de la máquina 1 sigue sigue una distribución normal de media μ_1 y varianza σ^2 y que el costo de mantención de la máquina 2 sigue una normal de media μ_2 y varianza $3\sigma^2$. Denotando por X e Y los costos de las máquinas 1 y 2 respectivamente podemos decir que

$$X \sim \text{Normal}(\mu_1, \sigma^2)$$
$$Y \sim \text{Normal}(\mu_2, 3\sigma^2).$$

Además se calcula (no lo demuestre) que el costo total esperado es $\mu = 2\mu_1 + \mu_2$.

Dadas dos m.a.s. independientes $X_1, \ldots, X_n$ de X y $Y_1, \ldots, Y_m$ de Y se quiere, **aprovechando toda la información**, encontrar un intervalo de confianza para μ.

(a) Suponga σ^2 conocido y calcule el intervalo de confianza para μ con nivel de confianza $(1-\alpha)\times 100\,\%$.

(b) Encuentre un estimador insesgado para σ^2 usando ambas m.a.s. Denote a este estimador por S^2.
Indicación: Primero estime la varianza de X y la varianza de Y denotando por S_X^2 y S_Y^2 a los respectivos estimadores.

(c) Averigüe la distribución de S^2 encontrado anteriormente. Justifique brevemente.

(d) Calcule un intervalo de confianza para μ considerando a σ^2 como desconocido y estimándolo por S^2.

09 de noviembre de 2009
Consultas sólo de enunciado
Tiempo: 3:00 hrs.

Pauta Problema 1 Control 3 Probabilidades y Estadística Pauta '08 FC; VR

Problema 1

a) La fn. de verosimilitud es

$$L(\theta) = f_{\vec{X}}(x_1 \dots x_n / \theta) = \prod_{i=1}^{n} f_{X_i}(x_i/\theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}} \mathbb{1}_{[0,\infty)}(x_i)$$

$$= \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} \mathbb{1}_{[0,\infty)^n}(\vec{x}).$$

Como el dominio de la densidad conjunta no depende de θ, podemos tomar log y derivar

$$l(\theta) = \ln(L(\theta)) = \ln\left(\theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}\right)$$

$$= -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

$$\partial_\theta l(\theta) = 0 \iff \frac{-n}{\theta} - \frac{(-1)}{\theta^2}\sum_{i=1}^{n} x_i = 0 \qquad \text{con } \theta \neq 0.$$

$$\boxed{\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}_n}$$ el promedio muestral.

b) Como $(X_i)_{i=1}^{n}$ es iid, $X_i \sim Exp(1/\theta)$, entonces sabemos que $\sum_{i=1}^{n} X_i \sim Gamma(n, 1/\theta)$

$$\Rightarrow \mathbb{E}(\hat{\theta}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\, n\, \frac{1}{1/\theta}$$

$$\mathbb{E}(\hat{\theta}) = \theta \qquad \text{(Es insesgado)}$$

y $$Var(\hat{\theta}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\frac{n}{(1/\theta)^2} = \frac{\theta^2}{n} \xrightarrow{n\to\infty} 0$$

Entonces, el estimador $\hat{\theta}$ de máxima verosimilitud de θ es insesgado, y varianza tiende a 0, por lo que es consistente.

c) El lema de Neyman-Pearson encuentra la mejor región crítica (con menor error tipo 2) tal que acota el error tipo I por α. (Este es el error importante)

Si defiendo al cliente, es mas grave {que se rechace que "el cliente tiene la razon" dado que en realidad la tiene"} que {se acepte que el cliente tiene la razón dado que no la tiene}.

Por ello, la hipótesis nula es "El cliente tiene la razon". (Nos interesa acotar por α la prob de rechazar que el cliente tiene la razon dado que en realidad la tiene)

d) Por el lema de Neyman Pearson, la región de rechazo es de la forma

$$R^* = \{\vec{x} \in \overline{X} \mid r(\vec{x}) \geq k\} \qquad \overline{X} = \text{espacio muestral}$$

con $r(\vec{x}) = \dfrac{f_{\vec{x}}(\vec{x}/\theta_1)}{f_{\vec{x}}(\vec{x}/\theta_0)}$. (H_0 es equivalente a $\theta = \theta_0 = 1$; H_1 es equivalente a $\theta = \theta_1 = 1{,}25$)

$$= \frac{\frac{1}{\theta_1^n} e^{-\frac{1}{\theta_1}\sum_{i=1}^n x_i}}{\frac{1}{\theta_0^n} e^{-\frac{1}{\theta_0}\sum x_i}} = \left(\frac{\theta_0}{\theta_1}\right)^n e^{-\sum_{i=1}^n x_i \left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)}$$

Imponiendo que $r(\vec{x}) \geq k$, $\left(\dfrac{\theta_0}{\theta_1}\right)^n e^{-\sum_{i=1}^n x_i \left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)} \geq k$ (las constantes serán llamadas k)

$$\Leftrightarrow e^{-\sum_{i=1}^n x_i \left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)} \geq k \qquad / \ln(\cdot)$$

$$\Leftrightarrow -\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right) \sum_{i=1}^n x_i \geq k \qquad \text{con } \frac{1}{\theta_1} - \frac{1}{\theta_0} = \frac{4}{5} - 1 < 0$$

$$\Rightarrow -\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right) > 0$$

y la desigualdad no cambia.

$$\Leftrightarrow \sum_{i=1}^n x_i \geq k$$

Luego, $R^* = \{\vec{x} \in \overline{X} \mid \sum_{i=1}^n x_i \geq k\}$

c) Imponemos que el error tipo I sea α

$$\mathbb{P}(R^* / H_0 \text{ cierta}) = \mathbb{P}\left(\sum_{i=1}^{n} X_i \geq k \mid \theta = 1\right) = \alpha$$

$$= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \frac{k}{n} \mid \theta = 1\right) = \alpha$$

$$\mathbb{P}\left(\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \geq \frac{\frac{k}{n} - \mu}{\sigma/\sqrt{n}} \mid \theta = 1\right) = \alpha$$

donde, dado que $\theta = 1$, $X_1 \sim Exp(1)$, por lo que para aplicar TCL usamos

$$\mu = \mathbb{E}(X_1 / \theta = 1) = \frac{1}{1/\theta}\Big|_{\theta=1} = 1$$

$$\sigma^2 = Var(X_1 / \theta = 1) = \frac{1}{(1/\theta)^2}\Big|_{\theta=1} = 1 \quad , n = 100 \; ; \; \alpha = 0{,}05$$

Usamos TCL, por lo que la distribución de $Z = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}}$ es aproximadamente una $N(0,1)$

$$\Rightarrow \mathbb{P}\left(Z \geq \frac{\frac{k}{n} - \mu}{\sigma/\sqrt{n}}\right) = \alpha \Rightarrow \frac{\frac{k}{n} - \mu}{\sigma/\sqrt{n}} = z_{1-\alpha}$$

Reemplazando los datos, $\dfrac{k/100 - 1}{1/\sqrt{100}} = 1{,}645$

$$10\left(\frac{k}{100} - 1\right) = 1{,}645 \Rightarrow \frac{k}{100} - 1 = 0{,}1645$$

$$\frac{k}{100} = 1{,}1645$$

$$\boxed{k = 116{,}45}$$

La región de rechazo (de nivel de significancia 0,05) es

$$R^* = \left\{\vec{x} \in \overline{X} \mid \sum_{i=1}^{n} x_i \geq 116{,}45\right\}$$

(o sea, rechazamos H_0 si $\sum_{i=1}^{n} x_i$ es mayor o igual a 116,45).

f) Como $n=100$, $\bar{x}_n = 1{,}13$, entonces $\sum_{i=1}^{n} x_i = n\,\bar{x}_n = 113$.

Dado que $113 < 116{,}45$, no rechazamos H_0, osea,

no rechazamos que los clientes tengan la razón.