Aurora Hermoso Carazo y Mª Dolores Ruiz Medina

Recuperación

1. **(1 punto)** Sean $(X_1,\ldots,X_n),(Y_1,\ldots,Y_n)$ muestras independientes de poblaciones con medias μ_1,μ_2 y varianzas σ_1^2,σ_2^2. Sean $\overline{X},\overline{Y},S_1^2,S_2^2$ las medias y cuasivarianzas muestrales respectivamente. Partiendo de la independencia de los vectores $(\overline{X},\overline{Y}),(S_1^2,S_2^2)$, deducir la distribución de la siguiente variable, especificando y justificando todos los pasos:

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}\sqrt{\frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2}}}\sqrt{n+m-2}$$

2. **(4.25 puntos)**

   a) Enunciar y demostrar el teorema de invarianza de Zehna, detallando previamente las siguientes definiciones:

   - Función de verosimilitud y estimador máximo verosímil de un parámetro.
   - Función de verosimilitud y estimador máximo verosímil de una función paramétrica.

   b) A partir de las condiciones de regularidad de Frechet-Cramer-Rao, deducir de forma justificada la propiedad de aditividad de la función de información.

   c) Sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de una variable X con función de densidad

   $$f_\theta(x)=\frac{e^x}{e^\theta-1},\quad 0<x<\theta.$$

   - Encontrar un estadístico suficiente y completo, probando que lo es.
   - Encontrar, si existe, un UMVUE para e^θ.
   - Encontrar el estimador máximo verosímil de e^θ, justificando su obtención.¿Es insesgado este estimador? Justifícalo.

3. **(2 puntos)**

   a) Sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de una variable $X\to\mathcal{N}(\mu,\sigma_0^2)$ con varianza conocida. Deducir el test de la razón de verosimilitud de tamaño α para contrastar: $\begin{cases} H_0:\mu\geq\mu_0 \\ H_1:\mu<\mu_0\end{cases}$

   b) A partir del test anterior, obtener un intervalo de confianza para μ. Deducir de forma razonada el nivel de confianza del intervalo de confianza.

4. **(1.25 puntos)** Análisis de varianza de una vía. Explicar en qué consiste el problema. Plantearlo en términos de un modelo lineal, especificando todos los componentes del intervalo; comprobar que es de Gauss-Markov y determinar su rango. Especificar el contraste de interés como caso particular de la hipótesis lineal general.

5. **(1.5 puntos)**

   a) Problema de bondad de ajuste. Explicar en qué consiste y describir los test para su resolución explicando la interpretación del estadístico de contraste en cada caso, ventajas e inconvenientes.

   b) A partir de una muestra de 200 empleados, se ha construido la siguiente tabla de frecuencias conjunta de Calidad de Vida (CV) y Rendimiento de Trabajo (RT). Se desea contrastar a nivel $0,05$ si los RT son independiente de la CV de los empleados.

CV/RT	Poca	Media	Mucha
Bajo	62	10	2
Medio	14	38	12
Alto	2	9	51

The End.

AmigoCanarioTeQuiere.pdf

AmigoCanario

Inferencia Estadística

3º Grado en Matemáticas

Facultad de Ciencias
Universidad de Granada

AMIGOCANARIO

1) **1.25 puntos** *Sean* $(X_1,\dots,X_{n_1})$, $(Y_1,\dots,Y_{n_2})$ *muestras aleatorias simples independientes de dos poblaciones,* $\mathcal{N}(\mu_1,\sigma^2)$ *y* $\mathcal{N}(\mu_2,\sigma^2)$, *respectivamente, con medias muestrales* $\overline{X}$ *e* $\overline{Y}$.

$1a)$ **0.75 puntos** *Calcular el percentil* r *de la siguiente variable:*

$$Z=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2}}.$$

$$T=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\to t(n_1+n_2-2)\ \ 0.375$$

$$T=\frac{\sqrt{n_1+n_2-2}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}Z\ \ 0.125$$

$$r/100=P(Z<k)=P\left(T<\frac{\sqrt{n_1+n_2-2}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}k\right)\Rightarrow\frac{\sqrt{n_1+n_2-2}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}k=t_{n_1+n_2-2;1-r/100}\Rightarrow k=\frac{t_{n_1+n_2-2;1-r/100}}{\sqrt{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}.$$

0.25

1.1 $\mu_1-\mu_2=-1;\ n_1=10,\ n_2=8;\ r=99\to t_{16;0.01}=2.5835\Rightarrow k=\frac{2.5835}{4}\sqrt{\frac{1}{10}+\frac{1}{8}}=0.30637.$

1.4 $\mu_2-\mu_1=-1;\ n_1=7,\ n_2=4;\ r=99\to t_{9;0.01}=2.8214\Rightarrow k=\frac{2.8214}{3}\sqrt{\frac{1}{7}+\frac{1}{4}}=0.58947.$

1.5 $\mu_2-\mu_1=-1;\ n_1=8,\ n_2=10;\ r=95\to t_{16;0.05}=1.7459\Rightarrow k=\frac{1.7459}{4}\sqrt{\frac{1}{8}+\frac{1}{10}}=0.20704.$

$1b)$ **0.5 puntos** *Calcular el estimador máximo verosímil de* σ^2 *basado en* $(X_1,\dots,X_{n_1},\ Y_1,\dots,Y_{n_2})$.

$$L_{x_i,y_j}(\sigma^2)=\frac{1}{\sigma^{n_1}(2\pi)^{n_1/2}}e^{-\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2}{2\sigma^2}}\frac{1}{\sigma^{n_2}(2\pi)^{n_2/2}}e^{-\frac{\sum_{j=1}^{n_2}(y_j-\mu_2)^2}{2\sigma^2}}=\frac{1}{\sigma^{n_1+n_2}(2\pi)^{n_1+n_2/2}}e^{-\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2+\sum_{j=1}^{n_2}(y_j-\mu_2)^2}{2\sigma^2}}$$

$$\ln L_{x_i,y_j}(\sigma^2)=-\frac{n_1+n_2}{2}\ln(2\pi)-\frac{n_1+n_2}{2}\ln\sigma^2-\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2+\sum_{j=1}^{n_2}(y_j-\mu_2)^2}{2\sigma^2}$$

$$\frac{d\ln L_{x_i,y_j}(\sigma^2)}{d\sigma^2}=-\frac{n_1+n_2}{2\sigma^2}+\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2+\sum_{j=1}^{n_2}(y_j-\mu_2)^2}{2\sigma^4}=0\Rightarrow\widehat{\sigma}^2=\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2+\sum_{j=1}^{n_2}(y_j-\mu_2)^2}{n_1+n_2}.$$

2.1) **2.5 puntos** *Sea* $(X_1,\ldots,X_n)$ *una muestra aleatoria simple de una variable con función de densidad:*

$$f_\theta(x) = \frac{2x}{(\theta+1)^2},\ 0<x<\theta+1.\ \ (a=2;\ k=-1)$$

a) *Encontrar, si existe, el UMVUE para* θ.

b) *Calcular la función de verosimilitud y encontrar el estimador máximo verosímil de* θ. *¿Es insesgado? (justificar la respuesta).*

$$\Theta = (-1,+\infty),\quad \chi^n = \mathbb{R}^{+n}$$

- *Estadístico suficiente*: $T = \max X_i$. 0.25

$$f_\theta^n(x_1,\ldots,x_n) = \frac{2^n x_1\cdots x_n}{(\theta+1)^{2n}} I_{(0,+\infty)}(\theta+1-\max x_i),\quad (x_1,\ldots,x_n)\in\mathbb{R}^{+n}.$$

- *Distribución de* $T = \max X_i$: 0.25

$$F_\theta(x) = \int_0^x \frac{2t}{(\theta+1)^2}dt = \frac{t^2}{(\theta+1)^2}\Big|_0^x = \frac{x^2}{(\theta+1)^2},\ \ 0<x<\theta+1.\ \Rightarrow\ F_\theta^T(t) = [F_\theta(t)]^n = \frac{t^{2n}}{(\theta+1)^{2n}},\ \ 0<t<\theta+1.$$

$$f_\theta^T(t) = \frac{2n\ t^{2n-1}}{(\theta+1)^{2n}},\ \ 0<t<\theta+1.$$

- *Completitud de* $T = \max X_i$: 0.25

$$E_\theta[g(T)] = \int_0^{\theta+1} g(t)\frac{2nt^{2n-1}}{(\theta+1)^{2n}}dt = 0,\ \ \forall\theta>-1 \Rightarrow \int_0^{\theta+1} g(t)t^{2n-1}dt = 0,\ \ \forall\theta>-1 \Rightarrow g(\theta+1)(\theta+1)^{2n-1} = 0,\ \ \forall\theta>-1.$$

$$\Rightarrow g(t)t^{2n-1} = 0,\ \ \forall t>0 \Rightarrow\ g(t)=0,\ \ \forall t>0\ \Rightarrow\ P_\theta(g(T)=0) \geq P_\theta(T>0) = 1,\ \ \forall\theta>-1.$$

- *Cálculo del UMVUE*:

$$E_\theta[h(T)] = \int_0^{\theta+1} h(t)\frac{2nt^{2n-1}}{(\theta+1)^{2n}}dt = \theta,\ \ \forall\theta>-1\ \Rightarrow\ \int_0^{\theta+1} h(t)2nt^{2n-1}dt = \theta(\theta+1)^{2n},\ \ \forall\theta>-1.$$

Puesto que $\lim\limits_{\theta\to-1}\theta(\theta+1)^{2n} = 0$, $\forall n\in\mathbb{N}$, tiene sentido trabajar con n arbitrario:

- $h(\theta+1)2n(\theta+1)^{2n-1} = \left(\theta(\theta+1)^{2n}\right)' = (\theta+1)^{2n} + 2n\theta(\theta+1)^{2n-1} \Rightarrow h(\theta+1) = \frac{\theta+1}{2n}+\theta,\ \ \forall\theta>-1\ (\theta+1>0)$.

  $\Rightarrow h(T) = \frac{T}{2n} + T - 1,\ \ (T>0)$. 0.75
- $T>0 \Rightarrow h(T) > -1$ es estimador de $\theta > -1$. 0.125
- T está acotado para todo $\theta \Rightarrow h(T)$ está acotado para todo $\theta \Rightarrow$ es de segundo orden. 0.125

$$\text{UMVUE para } \theta\ \rightarrow\ \frac{T}{2n}+T-1.$$

- *Función de verosimilitud y estimador máximo verosímil:*

$$L_{x_1,\ldots,x_n}(\theta) = \begin{cases} 0, & \theta < \max x_i - 1 \\ \dfrac{2^n x_1\cdots x_n}{(\theta+1)^{2n}}, & \theta > \max x_i - 1 \end{cases}\quad 0.25$$

$L_{x_1,\ldots,x_n}(\theta)$ decrece con θ para $\theta > \max x_i - 1 \Rightarrow \widehat{\theta}(X_1,\ldots,X_n) = \max X_i - 1$. 0.25

$\widehat{\theta}(X_1,\ldots,X_n)$ no es insesgado ya que es función del suficiente y completo y, de ser insesgado, coincidiría con el UMVUE. 0.25

3.1) **2.25 puntos** $\boxed{k = 1, \quad \sigma^2 = 2}$

3a) **0.375** *Sea X una variable continua con distribución en una familia regular en el sentido de Fréchet-Cramér-Rao, $X \to \{P_\theta;\ \theta \in \Theta\}$. Probar que si $\int_{\mathcal{X}} \frac{\partial^2 f_\theta(x)}{\partial \theta^2} dx = 0$, la función de información verifica $I_X(\theta) = -E_\theta \left[\frac{\partial^2 \ln f_\theta(X)}{\partial \theta^2}\right]$.*

- $$\frac{\partial \ln f_\theta(X)}{\partial \theta} = \frac{\frac{\partial f_\theta(X)}{\partial \theta}}{f_\theta(X)} \Rightarrow \frac{\partial^2 \ln f_\theta(X)}{\partial \theta^2} = \frac{\frac{\partial^2 f_\theta(X)}{\partial \theta^2} f_\theta(X) - \left(\frac{\partial f_\theta(X)}{\partial \theta}\right)^2}{(f_\theta(X))^2} = \frac{\frac{\partial^2 f_\theta(X)}{\partial \theta^2}}{f_\theta(X)} - \frac{\left(\frac{\partial f_\theta(X)}{\partial \theta}\right)^2}{(f_\theta(X))^2} =$$
$$= \frac{\frac{\partial^2 f_\theta(X)}{\partial \theta^2}}{f_\theta(X)} - \left(\frac{\partial \ln f_\theta(X)}{\partial \theta}\right)^2.$$
- $$E_\theta\left[\frac{\partial^2 \ln f_\theta(X)}{\partial \theta^2}\right] = \int_{\mathcal{X}} \frac{\partial^2 f_\theta(x)}{\partial \theta^2} dx - I_X(\theta) = -I_X(\theta).$$

3b) **1.875** *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable con función de densidad:*

$$f_\theta(x) = \frac{1}{2x\sqrt{\pi}} \exp\left\{-\frac{(\ln x - \theta)^2}{4}\right\}, \quad x > 0.$$

Sabiendo que $E_\theta[\ln X] = \theta$, $Var_\theta[\ln X] = 2$ y que la familia de distribuciones es regular, se pide:

b1) **1.125** *Calcular la función de información asociada a la muestra y encontrar la clase de funciones paramétricas que admiten estimador eficiente y los correspondientes estimadores.*

$$f_\theta(x) = \exp\left\{-\ln(2x\sqrt{\pi}) - \frac{(\ln x - \theta)^2}{4}\right\}, \quad x > 0 \Rightarrow \frac{\partial f_\theta(x)}{\partial \theta} = \frac{1}{2}[\ln x - \theta] f_\theta(x) \ \Rightarrow\ \frac{\partial \ln f_\theta(x)}{\partial \theta} = \frac{1}{2}[\ln x - \theta]$$

- $I_X(\theta) = \frac{1}{4} E_\theta\left[(\ln X - \theta)^2\right] = \frac{Var_\theta[\ln X]}{4} = \frac{1}{2} \Rightarrow I_{(X_1,\ldots,X_n)}(\theta) = n I_X(\theta) = \frac{n}{2}, \quad \forall \theta \in \mathbb{R}.$ *0.375*
- $\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln f_\theta(x_i)}{\partial \theta} = \frac{1}{2}\left[\sum_{i=1}^n \ln x_i - n\theta\right], \quad x_1, \ldots, x_n > 0$ *0.25*
- $a(\theta) = \frac{1}{2}, \quad T(X_1, \ldots, X_n) = \sum_{i=1}^n \ln X_i, \quad g(\theta) = n\theta, \quad g'(\theta) = n \longrightarrow I_{(X_1,\ldots,X_n)}(\theta) = \frac{n}{2} = a(\theta) g'(\theta)$ *0.25*
- $T = \sum_{i=1}^n \ln X_i \in \mathbb{R}$ *estimador de* $g(\theta) = n\theta \in \mathbb{R}$. *0.125*

$$\forall a \neq 0, b \in \mathbb{R}, \ a \sum_{i=1}^n \ln X_i + b \text{ estimador eficiente de } an\theta + b. \ 0.125$$

b2) **0.5** *Calcular $E_\theta[U \ln X_i]$, siendo $U \equiv U(X_1, \ldots, X_n)$ un estimador insesgado de $1/\theta$ y regular.*

U insesgado en $1/\theta$ es regular $\Leftrightarrow -\frac{1}{\theta^2} = E_\theta\left[U \frac{\partial \ln f_\theta^n(X_1, \ldots, X_n)}{\partial \theta}\right] = \sum_{i=1}^n E_\theta\left[U \frac{\partial \ln f_\theta(X_i)}{\partial \theta}\right] = n E_\theta\left[U \frac{\partial \ln f_\theta(X_i)}{\partial \theta}\right]$

$E_\theta\left[U \frac{\partial \ln f_\theta^n(X_i)}{\partial \theta}\right] = \frac{1}{2} E_\theta\left[U(\ln X_i - \theta)\right] = \frac{1}{2} E_\theta\left[U \ln X_i\right] - \frac{1}{2} = -\frac{1}{n\theta^2} \Rightarrow E_\theta\left[U \ln X_i\right] = 1 - \frac{2}{n\theta^2}, \ i = 1, \ldots, n.$

b3) **0.25** *Calcular la cota para la varianza de estimadores regulares, insesgados en $3\theta^2 + 2$, y justificar si se alcanza o no dicha cota.*

$$\frac{(6\theta)^2}{I_{(X_1,\ldots,X_n)}(\theta)} = \frac{72\theta^2}{n} \quad \textit{NO SE ALCANZA}$$

4.1) 2.75 puntos $\boxed{a = 2, \quad k = 2}$

Sea $(X_1, \ldots X_n)$ una muestra aleatoria simple de una variable X con funciń de densidad:

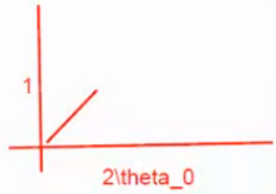
$$f_\theta(x) = \frac{x}{2\theta^2}, \quad 0 < x < 2\theta.$$

a) *Deducir el test de razón de verosimilitudes de tamaño α para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta < \theta_0$, detallando el espacio paramétrico, el espacio muestral, y representando gráficamente el estadístico de contraste. ¿Qué tamaños se alcanzan? Justificar la respuesta.*

$$\begin{matrix} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{matrix} \longrightarrow \Theta = (0, \theta_0], \quad \chi^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^{+n}, \ \text{máx}\, x_i < 2\theta_0\} = (0, \ 2\theta_0)^n. \ 0.25$$

$$(x_1, \ldots, x_n) \in \chi^n \rightarrow L_{x_1,\ldots,x_n}(\theta) = \begin{cases} 0, & 0 < \theta \leq \text{máx}\, x_i/2 \\ \dfrac{x_1 \cdots x_n}{2^n \theta^{2n}}, & \text{máx}\, x_i/2 < \theta \leq \theta_0, \end{cases} \rightarrow \widehat{\theta}(x_1, \ldots, x_n) = \text{máx}\, x_i/2. \ 0.5$$

- $\lambda(x_1, \ldots, x_n) = \dfrac{L_{x_1,\ldots,x_n}(\theta_0)}{\sup\limits_{\theta \leq \theta_0} L_{x_1,\ldots,x_n}(\theta)} = \dfrac{(\text{máx}\, x_i/2)^{2n}}{\theta_0^{2n}} = \left(\dfrac{\text{máx}\, x_i}{2\theta_0}\right)^{2n}, \quad 0 < \text{máx}\, x_i < 2\theta_0.$ 0.375



- $\lambda(x_1, \ldots, x_n)$ crece estrictamente de 0 a 1 con $\text{máx}\, x_i$. 0.125
- $\varphi(X_1, \ldots, X_n) = \begin{cases} 1 & \text{si} \ \ \text{máx}\, X_i < c' \quad (0 < c' \leq 2\theta_0) \\ 0 & \text{si} \ \ \text{máx}\, X_i \geq c' \end{cases}$ 0.125+0.125
- $\alpha = P_{\theta_0}(\text{máx}\, X_i < c') = \left(\displaystyle\int_0^{c'} \frac{x}{2\theta_0^2} dx\right)^n = \left(\dfrac{c'^2}{4\theta_0^2}\right)^n \Rightarrow c' = 2\theta_0 \alpha^{1/2n}.$ 0.25
- $c' = 2\theta_0 \alpha^{1/2n} \leq 2\theta_0, \quad \forall \alpha \in [0, 1]$, por lo que se alcanzan todos los tamaños. 0.25

$$\boxed{\textit{TRV de tamaño } \alpha \in [0,1] \ \rightarrow \ \varphi(X_1, \ldots, X_n) = \begin{cases} 1 & \text{si} \ \ \text{máx}\, X_i < 2\theta_0 \alpha^{1/2n} \\ 0 & \text{si} \ \ \text{máx}\, X_i > 2\theta_0 \alpha^{1/2n}. \end{cases}}$$

b) *Aplicando la definición de intervalo de confianza, deducir el nivel de confianza del intervalo obtenido a partir del test del apartado a).*

- Región de aceptación $\rightarrow 2\theta_0 < \text{máx}\, X_i \alpha^{-1/2n} \ \Rightarrow \ (0, \ \text{máx}\, X_i \alpha^{-1/2n}/2)$ es IC para θ al nivel $1 - \alpha$. 0.25
- Deducción del nivel de confianza: 0.5

$$F_\theta(x) = P_\theta(X \leq x) = \frac{x^2}{(2\theta)^2}, \quad 0 < x < 2\theta, \quad \forall \theta > 0.$$

$$P_\theta\big((0, \ \text{máx}\, X_i \alpha^{-1/2n}/2) \ni \theta\big) = P_\theta\big(\text{máx}\, X_i > 2\theta\alpha^{1/n}\big) = 1 - \big(F_\theta(2\theta\alpha^{1/2n})\big)^n = 1 - (\alpha^{1/n})^n = 1 - \alpha, \quad \forall \theta > 0.$$

5) 1.25 puntos *Para contrastar si la opinión sobre la cobertura de red móvil es la misma en los pueblos de una comarca, se ha efectuado una encuesta en cada uno de ellos. La siguiente tabla muestra las respuestas a la pregunta: ¿Hay buena cobertura de red móvil en su pueblo?*

	Sí	No	No sabe	No contesta
Pueblo 1	n_{11}	n_{12}	n_{13}	n_{14}
Pueblo 2	n_{21}	n_{22}	n_{23}	n_{24}
Pueblo 3	n_{31}	n_{32}	n_{33}	n_{34}

a) *Describir las variables usadas en este problema y especificar formalmente la hipótesis a contrastar en términos de la distribución dichas variables.*

$$X_1, X_2, X_3 : \text{Opinión sobre la cobertura en los pueblos 1, 2, 3.}$$

$$A_j \equiv \text{Sí, NO NS, NC} \rightarrow H_0: \; P(X_1 = A_j) = P(X_2 = A_j) = P(X_3 = A_j), \; j = 1, \ldots, 4. \quad 0.25$$

b) *Especificar qué parámetros deben estimarse para calcular el estadístico de contraste, cuáles son los estimadores usados y justificar por qué se usan estos estimadores.*

Hay que estimar las frecuencias esperadas bajo H_0. Bajo H_0 todas las variables tienen la misma distribución y, por tanto, hay que estimar $p_j^0 = P_{H_0}(X = A_j)$, $j = 1, \ldots, 4$, siendo X una variable con dicha distribución común. Estas probabilidades se estiman por máxima verosimilitud, basándose en el número total de datos en A_j, $N_{.j} = \sum_{i=1}^{3} N_{ij}$. Ya que $N_{.j} \rightarrow B(n, p_j^0)$, siendo n el total de datos observados, el estimador de p_j^0 es $\frac{N_{.j}}{n}$, con . 0.25

c.1) *A la vista de los datos, ¿qué conclusión se obtiene sobre el problema planteado al nivel de significación 0.05? (trabajar con cuatro cifras decimales).* 0.75

	Sí	No	No sabe	No contesta	n_i
Pueblo 1	7	15	1	2	25
Pueblo 2	12	6	3	4	25
Pueblo 3	10	7	7	1	25
$n_{.j}$	29	28	11	7	75

$\frac{n_i \cdot n_{.j}}{n}$	Sí	No	No sabe	No contesta
Pueblo 1	9.6667	9.3333	3.6667	2.3333
Pueblo 2	9.6667	9.3333	3.6667	2.3333
Pueblo 3	9.6667	9.3333	3.6667	2.3333

Se agrupan la dos últimas categorías

	Sí	No	No sabe/No contesta	n_i
Pueblo 1	7	15	3	25
Pueblo 2	12	6	7	25
Pueblo 3	10	7	8	25
$n_{.j}$	29	28	18	75

$e_{ij} = \frac{n_i \cdot n_{.j}}{n}$	Sí	No	No sabe/No contesta
Pueblo 1	9.6667	9.3333	6
Pueblo 2	9.6667	9.3333	6
Pueblo 3	9.6667	9.3333	6

$\frac{(n_{ij}-e_{ij})^2}{e_{ij}}$	Sí	No	No sabe/No contesta
Pueblo 1	0.7356	3.4405	1.5
Pueblo 2	0.5632	1.1905	0.1667
Pueblo 3	0.0115	0.5833	0.6667

$\chi^2_{exp} = 8.8580$ 0.5

$R.C.$ $\chi^2_{exp} > \chi^2_{4,0.05} = 9.4877$. No se rechaza H_0, los datos no muestran evidencias para rechazar que la opinión sobre la cobertura de red móvil es la misma en los tres pueblos estudiados de la comarca. 0.25

ENERO-2018.pdf

ENERO 2018

INFERENCIA ESTADÍSTICA.

1.a) Teoría

b) La distancia en km recorrida por dos tipos de coches A y B con 50 litros de gasolina, sigue, respectivamente una ley $N(\mu_A, 15)$ y $N(\mu_B, 22)$. Se desea encontrar un intervalo de confianza al nivel 0'99 para $\mu_A - \mu_B$, con amplitud máxima de 4 km observando las distancias recorridas con 50 litros de gasolina por un mismo número de coches de cada tipo, elegidos de forma independiente. ¿Cuál es el mínimo número de coches de cada tipo que deben realizar la prueba para obtener dicho intervalo?

X: Distancia en Km del coche A con 50 litros de gasolina $\to N(\mu_A, 15)$

Y: Distancia en Km del coche B con 50 litros de gasolina $\to N(\mu_B, 21)$

$(X_1, \dots, X_{n_1})$ $(Y_1, \dots, Y_{n_2})$ m.a.s. de X e Y, $n_1 = n_2 = n$.

Intervalo de confianza para la diferencia de medias de dos poblaciones normales con varianza conocida:

$$\left(\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

Sea $\alpha = 0{,}01 \Rightarrow \alpha/2 = 0{,}005$ $\quad z_{\alpha/2} = z_{0{,}005} \simeq 1'645$

$\sigma_1^2 + \sigma_2^2 = 36$ $\quad$ luego el intervalo nos queda

$$\left(\bar{X} - \bar{Y} - 1'645\frac{\sqrt{36}}{\sqrt{n}},\ \bar{X} - \bar{Y} + 1'645\frac{\sqrt{36}}{\sqrt{n}} \right)$$

Como queremos una amplitud máxima de 4

$$\left(\cancel{\bar{X}} - \cancel{\bar{Y}} + 1'645\frac{\sqrt{36}}{\sqrt{n}}\right) - \left(\cancel{\bar{X}} - \cancel{\bar{Y}} - 1'645\frac{\sqrt{36}}{\sqrt{n}}\right) \le 4 \Rightarrow 1'645 \cdot 6 \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}}\right) \le 4$$

$$\Rightarrow \frac{2}{\sqrt{n}} \le 0'405268 \Rightarrow \sqrt{n} \ge \frac{2}{0'405268} = 4'935 \Rightarrow n \ge 24'35$$

Como n es natural, necesitamos mínimo 25 coches de cada tipo para realizar la prueba.

2. Teoría

3. Sea $(X_1, \dots, X_n)$ una m.a.s. de una variable aleatoria X con función de densidad $f_\theta(x) = \frac{\sqrt{\theta}}{2} x^{-3/2} \quad x > \theta$

Determinar cuales son los valores de n para los que existe el UMVUE para θ y encontrarlo en tales casos. Calcular el EMV de θ.

En primer lugar buscamos un estadístico suficiente y completo para esta distribución.

$\chi = \bigcup_\theta \chi_\theta = \bigcup_\theta (\theta, +\infty) = \mathbb{R}^+$

$\Theta = \mathbb{R}^+$

$$f_\theta^n(x_1, \dots, x_n) = \frac{(\sqrt{\theta})^n}{2^n} \left(\prod_{i=1}^n x_i\right)^{-3/2} \cdot I_{[\theta, +\infty)}(X_{(1)})$$

Por el ~~tma~~ de factorización, $X_{(1)}$ es un estadístico suficiente, basta tomar $h(x_1, \dots, x_n) = \left(\prod_{i=1}^n x_i\right)^{-3/2}$ $g_\theta(T(x_1, \dots, x_n)) = \frac{(\sqrt{\theta})^n}{2^n} \cdot I_{[\theta, +\infty)}(X_{(1)})$

Veamos si es completo.

Calculamos la distribución de $X_{(1)}$:

$$F(x) = \int_\theta^x \frac{\sqrt{\theta}}{2} t^{-3/2} \, dt = 1 - \frac{\sqrt{\theta}}{\sqrt{x}} \qquad F_{X_{(1)}} = 1 - (1 - F(x))^n = 1 - \left(\frac{\sqrt{\theta}}{\sqrt{x}}\right)^n$$

$$\Rightarrow f_{X_{(1)}}(x) = \frac{n}{2x}\left(\frac{\sqrt{\theta}}{\sqrt{x}}\right)^n = \frac{n}{2x}\left(\frac{\theta}{x}\right)^{n/2} \qquad x > \theta$$

Para que el estadístico sea completo necesito que si tomo cualquier función medible $g(T)$ del estadístico suficiente con $E_\theta(g(T) = 0)$ $\forall \theta \in \Theta$, se cumple que $P_\theta(g(T) = 0) = 1$.

$$E_\theta(g(T)) = \int_\theta^{+\infty} g(t) \frac{n}{2t}\left(\frac{\theta}{t}\right)^{n/2} dt = \frac{n \cdot \theta^{n/2}}{2} \int_\theta^{+\infty} g(t) \frac{1}{t^{1+n/2}} \, dt = 0 \Leftrightarrow$$

$\Leftrightarrow \int_\theta^{+\infty} g(t)/t^{1+n/2} \, dt = 0$ Usando el TFC, sabemos que existe una primitiva, G, de $g(t)/t^{1+n/2}$ que cumple que $G(+\infty) - G(\theta) = 0$ y derivando respecto a θ

Escaneado con CamScanner

Tenemos que $\frac{g(\theta)}{\theta^{1+n/2}} = 0 \iff g(\theta) = 0 \quad \forall \theta \in \Theta$

$\Rightarrow \{t / g(t) = 0\} \supseteq \{t > 0\}$ y tomando probabilidades

$1 \geq P(g(T) = 0) \geq P(T > 0) = 1$ Y por tanto hemos probado que $T = X_{(1)}$ es estadístico suficiente y completo, luego el UMVUE será una función suya insesgada que sea estimador y que tenga momento de segundo orden finito.

Para ver si es UMVUE imponemos insesgadez:

$$E_\theta(h(T)) = \int_\theta^{+\infty} h(t) \frac{n}{2t}\left(\frac{\theta}{t}\right)^{n/2} dt = \theta \iff \frac{n\theta^{n/2}}{2}\int_\theta^{+\infty} h(t)/t^{1+n/2}\, dt = \theta \iff$$

$$\iff \frac{2\,\theta^{1-n/2}}{n} = \int_\theta^{+\infty} h(t)/t^{1+n/2}$$ De nuevo por el TFC tenemos que

$$\frac{h(\theta)}{\theta^{1+n/2}} = \frac{(n-2)\theta^{-n/2}}{n} \Rightarrow h(\theta) = \frac{(n-2)\theta}{n} \Rightarrow h(T) = \frac{(n-2)T}{n}$$ es candidato a UMVUE.

- Claramente $h(T) = \frac{(n-2)T}{n}$ para $n > 2$, ya que como $T \in \mathbb{R}^+$, $h(T) \in \mathbb{R}^+$ cuando $n > 2$
- Veamos que tiene momento de segundo orden finito

$$E_\theta(h(T)^2) = \int_\theta^{+\infty} \left(\frac{n-2}{n}\right)^2 t^2 \frac{n}{2t}\left(\frac{\theta}{t}\right)^{n/2} dt = \frac{(n-2)^2\theta^{n/2}}{2n}\int_\theta^{+\infty} t^{1-n/2}\, dt =$$

$$= \frac{(n-2)^2\theta^{n/2}}{2n}\left[\frac{t^{2-n/2}}{2-n/2}\right]_\theta^{+\infty} < +\infty \iff 2 - n/2 \leq 0 \iff \boxed{n \geq 4}$$

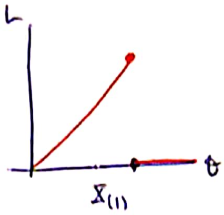
Por tanto existe UMVUE para $n \geq 4$ y es $h(T) = \frac{(n-2)T}{n}$.

Para calcular el EMV, tenemos que la función de verosimilitud es

$$L_{x_1,\dots,x_n}(\theta) = f_\theta^n(x_1,\dots,x_n) = \frac{(\sqrt{\theta})^n}{2^n}\left(\prod_{i=1}^n x_i\right)^{-3/2} \cdot I_{(0,+\infty)}(X_{(1)})$$

Como no es derivable, ya que ~~...~~ el intervalo de definición depende del parámetro, estudiamos la monotonía de la función (donde no se anula.) que definiremos como

$$g_{x_1,\dots,x_n}(\theta) = \frac{(\sqrt{\theta})^n}{2^n}\left(\prod_{i=1}^n x_i\right)^{-3/2} = \underbrace{(\sqrt{\theta})^n}_{\text{Creciente en } \theta} \underbrace{\frac{\left(\prod_{i=1}^n x_i\right)^{-3/2}}{2^n}}_{\text{No depende de } \theta}$$



$\Rightarrow L_{x_1,\dots,x_n}(\theta)$ es creciente cuando

$0<\theta<X_{(1)}$ y se anula en $\theta \geq$, por tanto alcanza su máximo en $X_{(1)}$, así el EMV de θ será $\hat{\theta} = X_{(1)}$.

4. Obtener el test de razón de verosimilitudes de tamaño α para contrastar $H_0: \theta \leq \theta_0$ frente a $H_1: \theta > \theta_0$ basado en una observación de una variable aleatoria con función de distribución exponencial

$$f_\theta(x) = \theta e^{-\theta x} \quad x>0$$

¿Que tamaño alcanza dicho test?

Sabemos que el TRV es de la forma

$$\varphi(X_1,\dots,X_n) = \begin{cases} 1 & \text{si } \lambda(X_1,\dots,X_n) < c \\ 0 & \text{si } \lambda(X_1,\dots,X_n) \geq c \end{cases} \qquad c \in [0,1]$$

donde $$\lambda(X_1,\dots,X_n) = \frac{\sup_{\theta\in\Theta_0} L_{x_1,\dots,x_n}(\theta)}{\sup_{\theta\in\Theta} L_{x_1,\dots,x_n}(\theta)} \qquad \forall (x_1,\dots,x_n) \in \mathcal{X}^n$$

En nuestro caso: $\left.\begin{array}{l}\Theta_0 = (-\infty, \theta_0] \\ \Theta_1 = (\theta_0, +\infty)\end{array}\right\} \Theta = \Theta_0 \cup \Theta_1 = \mathbb{R}$ $\qquad \mathcal{X} = \mathbb{R}^+$, $\mathcal{X}^n = (\mathbb{R}^+)^n$

Función de verosimilitud asociada a $x\in\mathbb{R}^+$: $L_x(\theta)=\theta e^{-\theta x}$ $\quad \theta\in\mathbb{R}$

$\ln L_x(\theta)=\ln(\theta)-\theta x \longrightarrow \dfrac{\partial \ln L_x(\theta)}{\partial\theta}=\dfrac{1}{\theta}-x=0 \Rightarrow \hat{\theta}=\dfrac{1}{x}$

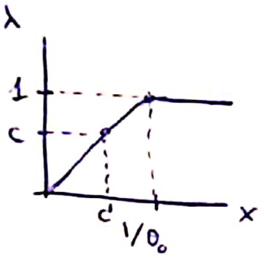
(como tiene un único extremo, es máximo)

$$\lambda(x)=\frac{\sup\limits_{\theta\leq\theta_0} L_x(\theta)}{\sup\limits_{\theta\in\mathbb{R}} L_x(\theta)}=\begin{cases}\dfrac{L_x(1/x)}{L_x(1/x)}=1 & \text{si } 0<\frac{1}{x}\leq\theta_0\\[2ex] \dfrac{L_x(\theta_0)}{L_x(1/x)}=x\theta_0\cdot e^{1-\theta_0 x} & \text{si } 1/x>\theta_0\end{cases}$$

Estudiemos cómo se comporta $\lambda(x)$:

$1/x>\theta_0 \longrightarrow \ln\lambda(x)=\ln(x\theta_0)+1-\theta_0 x \longrightarrow \dfrac{\partial\ln\lambda(x)}{\partial x}=\dfrac{\theta_0}{x\theta_0}-\theta_0=\dfrac{1}{x}-\theta_0>0$

$\Rightarrow \lambda(x)$ crece en $1/x>\theta_0$ y para $1/x\leq\theta_0$ permanece constante



$$\lambda(x)=\begin{cases}1 & \text{si } 0<\frac{1}{\theta_0}\leq x\\ x\theta_0\cdot e^{1-\theta_0 x} & \text{si } 1/\theta_0>x\end{cases}$$

$$\Rightarrow \varphi(x)=\begin{cases}1 & \text{si } x<c' \\ 0 & \text{si } x\geq c'\end{cases}\qquad (0<c'\leq 1/\theta_0)$$

Para calcular c' usamos el tamaño.

- Tamaño: $\sup\limits_{\theta\leq\theta_0} E_\theta(\varphi(X))$

$$\alpha=\sup_{\theta\leq\theta_0}E_\theta(\varphi(X))=\sup_{\theta\leq\theta_0}1\cdot P_\theta(X<c')=\sup_{\theta\leq\theta_0}\int_0^{c'}\theta e^{-\theta x}\,dx=$$

$$=\sup_{\theta\leq\theta_0}\left[-e^{-\theta x}\right]_0^{c'}=\sup_{\theta\leq\theta_0}\left(-e^{-\theta c'}+1\right)=1-e^{-\theta_0 c'}=\alpha\Rightarrow$$

$$\Rightarrow \ln(\alpha-1)=-\theta_0 c'\Rightarrow c'=-\frac{\ln(\alpha-1)}{\theta_0}\ //$$

pero $c'=\underbrace{\dfrac{-\ln(\alpha-1)}{\theta_0}}_{\geq 0\ (*)}\in(0,1/\theta_0] \Leftrightarrow -\dfrac{\ln(\alpha-1)}{\theta_0}\leq 1/\theta_0 \Leftrightarrow -\ln(\alpha-1)\leq 1 \Leftrightarrow$

$\Leftrightarrow \underline{\alpha\geq 1+e^{-1}}$

(*) Como $\alpha\in(0,1)\Rightarrow(1-\alpha)\in(0,1)\Rightarrow \ln(\alpha-1)<0$ y $\theta_0>0$

Por tanto para tamaño $\alpha \geq e^{-1}+1$, el TRV será:

$$\varphi(X) = \begin{cases} 1 & \text{si } x < -\dfrac{\ln(\alpha-1)}{\theta_0} \\ 0 & \text{si } x \geq -\dfrac{\ln(\alpha-1)}{\theta_0} \end{cases}$$

6.b) A partir de los siguientes datos sobre el número de visitas a un durante 100 días, contrasta si el nº medio de visitas diarias sigue una distribución de Poisson de parámetro 2.

Nº visitas	0	1	2	3	4	5	6	7
Nº días	29	51	60	30	19	9	1	1

Se trata de un problema de bondad de ajuste, en el que la variable aleatoria observada y las hipótesis a contrastar son:

X: Nº de visitas diarias $\qquad n = 200$

$$\begin{cases} H_0: X \longrightarrow P(2) \\ H_1: X \not\longrightarrow P(2) \end{cases}$$

Para poder aplicar el test de Pearson tenemos que particionar el conjunto de valores.

N_i: Número de individuos en cada categoría según los datos (en la categoría A_i: $i = 1, \dots, k$)

np_i^0: Número de individuos esperados bajo H_0 (en cada categoría A_i)

Particionamos el conjunto de valores de la distribución hipotética ($\mathbb{N} \cup \{0\}$) en $A_1, \dots, A_8$:

A_i	0	1	2	3	4	5	6	7
prob. esperada p_i^0	$P(X=0)$ 0,1353	$P(X=1)$ 0,2707	$P(X=2)$ 0,2707	0,1804	0,0902	0,0361	0,0120	0,0045
Frec. esperada $200p_i^0$	27,0671	54,1341	54,1341	36,0894	18,0447	7,2179	2,406	0,9068

Escaneado con CamScanner

Como para las categorías $A_i \geq 6$, $200p_i < 5$, y no cumple los requisitos para que el test sea aceptable. Solucionamos esto agrupando las tres últimas categorías para poder aplicar el test.

A_i	0	1	2	3	4	≥ 5
Frec. esperada $200p_i$	27,0671	54,1341	54,1341	36,0894	18,0447	10,5306
Frec. obtenida n_i	29	51	58	39	20	3

$$\chi^2_{exp} = \sum_{i=1}^{6} \frac{(n_i - p_i n)^2}{n p_i} = \frac{(29-27,0671)^2}{27,0671} + \dots + \frac{(3-10,5306)^2}{10,5306} = 2,06947$$

Y com $\chi^2(N_1, \dots, N_6) \underset{H_0}{\simeq} \chi^2(5)$

$$p\text{-nivel} = P_{H_0}(\chi^2(N_1, \dots, N_6) > 2,06947) \in (0'75, 0'8)$$

Se concluye que los datos no proporcionan evidencia para rechazar la hipótesis de que el número de accidentes diarios sigue una distribución $P(2)$.

SEPTIEMBRE-2015.pdf

AzaharaFS

Inferencia Estadística

3º Grado en Matemáticas

Facultad de Ciencias
Universidad de Granada

SEPTIEMBRE 2015

INFERENCIA ESTADÍSTICA

1.- Sean $(X_1,\dots,X_n)$, $(Y_1,\dots,Y_m)$ muestras independientes de poblaciones normales con medias μ_1, μ_2 y varianzas σ_1^2, σ_2^2. Sean $\bar{X}, \bar{Y}, S_1^2, S_2^2$ las medias y cuasivarianzas muestrales respectivamente. Partiendo de la independencia de los vectores $(\bar{X},\bar{Y}), (S_1^2, S_2^2)$, deducir la distribución de la siguiente variable, especificando y justificando todos los pasos:

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\sqrt{\frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2}}}\sqrt{n+m-2}$$

$X \to N(\mu_1, \sigma_1^2)$

$Y \to N(\mu_2, \sigma_2^2)$

Sabemos que $\bar{X} \to N(\mu_1, \sigma_1^2/n)$ y $\bar{Y} \to N(\mu_2, \sigma_2^2/m)$. Por el lema de Fisher tenemos que los estadísticos $\bar{X}$ e $\bar{Y}$ son independientes, además. Además $\bar{X}-\bar{Y} \to N\left((\mu_1-\mu_2), \frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}\right)$ tipificando

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \to N(0,1)$$

Por otro lado $\frac{(n-1)S_1^2}{\sigma_1^2} \to \chi^2(n-1)$ y $\frac{(m-1)S_2^2}{\sigma_2^2} \to \chi^2(m-1)$

Como por el lema de Fisher S_1^2 y S_2^2 son independientes, podemos usar la reproductividad de la χ^2, luego

$$\frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2} \longrightarrow \chi(m+n-2)$$

Sabemos que si A y B son independientes, $A \to N(0,1)$, $B \to \chi^2(n) \Rightarrow T = \frac{A}{\sqrt{B/n}} \to t(n)$

Tomando $A = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$ y $B = \frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2}$

nos queda

$$\frac{A}{\sqrt{B/_{n+m-2}}} = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\sqrt{\frac{(n-1)S_1^2}{\sigma_1^2}+\frac{(m-1)S_2^2}{\sigma_2^2}}}\sqrt{n+m-2} \longrightarrow t(m+n-2)$$

2.- a) Teoría

b) Teoría

c) Sea $(X_1,\dots,X_n)$ una m.a.s. de una variable X con función de densidad

$$f_\theta(x) = \frac{e^x}{e^\theta-1} \quad 0<x<\theta$$

i) Encontrar un estadístico suficiente y completo, probando que lo es.

ii) Encontrar, si existe, un UMVUE para e^θ.

iii) Encontrar el EMV de e^θ, justificando su obtención ¿Es insesgado este estimador? Justifícalo.

i) Ya que $\mathcal{X}_\theta = (0,\theta)$ depende de θ, la familia no es exponencial uniparamétrica, luego para buscar un estadístico suficiente vamos a utilizar el Teorema de Factorización de Neyman-Fisher.

$$f_\theta^n(x_1,\dots,x_n) = \frac{e^{\sum_{i=1}^n x_i}}{(e^\theta-1)^n} \qquad (x_1,\dots,x_n)\in\mathcal{X}^n = (\mathbb{R}^+)^n$$

Como $\mathcal{X} = \bigcup_{\theta\in\Theta} \mathcal{X}_\theta = \bigcup_{\theta\in\mathbb{R}^+}(0,\theta) = \mathbb{R}^+ \Rightarrow \mathcal{X}^n = (\mathbb{R}^+)^n$, siempre se cumple que $x_i > 0$, por tanto $f_\theta^n(x_1,\dots,x_n)$ la podemos expresar como

$$f_\theta^n(x_1,\dots,x_n) = \frac{e^{\sum_{i=1}^n x_i}}{(e^\theta-1)^n}\cdot I_{(-\infty,\theta)}(X_{(n)})$$

y tomando $h(x_1,\dots,x_n) = e^{\sum_{i=1}^n x_i}$ y $g_\theta(T(x_1,\dots,x_n)) = \frac{I_{(-\infty,\theta)}(X_{(n)})}{(e^\theta-1)^n}$

tendríamos que $T = X_{(n)}$ es suficiente.


WELL DONE
NATURE

Para ver que es completo veamos que si tengo cualquier función medible g tal que

$$E_\theta(g(T))=0 \quad \forall \theta\in\Theta=\mathbb{R}^+ \Rightarrow P_\theta(g(T)=0)=1 \quad \forall\theta\in\Theta$$

Así, vamos a calcular primero la función de densidad del estadístico:

$$F_\theta(x)=\int_0^x \frac{e^t}{e^\theta-1}\,dt = \frac{1}{e^\theta-1}\int_0^x e^t\,dt = \frac{e^x-1}{e^\theta-1} \quad \text{con } 0<x<\theta$$

$$\Rightarrow F_\theta^T(x)=(F_\theta(x))^n=\left(\frac{e^x-1}{e^\theta-1}\right)^n \rightarrow f_\theta^T(x)=n\left(\frac{e^x-1}{e^\theta-1}\right)^{n-1}\cdot\frac{e^x}{e^\theta-1}=\frac{ne^x(e^x-1)^{n-1}}{(e^\theta-1)^n}$$

$$E_\theta(g(T))=\int_0^\theta g(t)f_\theta^T(t)\,dt=\int_0^\theta g(t)\cdot\frac{ne^t(e^t-1)^{n-1}}{(e^\theta-1)^n}\,dt=\frac{n}{(e^\theta-1)^n}\int_0^\theta g(t)\underset{\substack{\shortparallel \\ 0}}{e^t(e^t-1)^{n-1}}\,dt$$

$$\Leftrightarrow \int_0^\theta g(t)e^t(e^t-1)^{n-1}\,dt=0$$ Por el TFC sabemos que $g(t)e^t(e^t-1)^{n-1}$ tiene primitiva G cumpliendo que $G(\theta)-G(0)=0$. Derivando respecto de θ, tenemos que $g(\theta)\underbrace{e^\theta(e^\theta-1)^{n-1}}_{\substack{\vee \\ 0}}=0 \Leftrightarrow g(\theta)=0 \quad \forall\theta\in\Theta$

Hemos demostrado que $\{t<\theta\}\subseteq\{t/g(t)=0\}$, y tomando probabilidades tenemos que $1\geq P(g(T)=0)\geq P(T<\theta)=1$

$\Rightarrow \underline{P(g(T)=0)=1}$ Por tanto $T=X_{(n)}$ es completo.

ii) UMVUE?

Sabemos que el UMVUE se puede calcular mediante una función de un estadístico suficiente y completo, luego sea $h(T)$ dicha función, imponemos la insesgadez.

$$E_\theta(h(T)) = \int_0^\theta h(t) f_{X_{(n)}}(t)\,dt = \frac{n}{(e^\theta-1)^n}\int_0^\theta h(t)e^t(e^t-1)^{n-1}dt = e^\theta \Leftrightarrow$$

$$\Leftrightarrow \int_0^\theta h(t)e^t(e^t-1)^{n-1}dt = \frac{e^\theta(e^\theta-1)^n}{n}$$ Aplicando de nuevo el TFC

existe H primitiva de $h(t)e^t(e^t-1)^{n-1}$ tal que $H(\theta)-H(0)=0$,

derivando respecto a θ. $h(\theta)e^\theta(e^\theta-1)^{n-1} = e^{2\theta}(e^\theta-1)^{n-1} + \frac{e^\theta(e^\theta-1)^n}{n}$

$\Rightarrow h(\theta) = e^\theta + \frac{1}{n}$ luego

$h(T) = e^T + \frac{1}{n}$ es candidato a UMVUE

- Veamos que es estimador de e^θ, es decir que toma valores en $\mathbb{R}^+$. Como $T>0 \Rightarrow h(T)>0 \Rightarrow h(T)\in\mathbb{R}^+$, luego sí es estimador

- Veamos por último si tiene momento de segundo orden finito.

$E_\theta(h(T)^2) = \int_0^\theta h(t)^2 f_{X_{(n)}}(t)\,dt < +\infty$ ya que tenemos la integral en un recinto finito ~~de una combinación de exponenciales~~. y $n>0$.

$\Rightarrow h(T) = e^T + \frac{1}{n}$ es UMVUE para e^θ.

iii) Función de verosimilitud:

$$L_{x_1,\dots,x_n}(\theta) \begin{cases} 0 & \text{si } \theta < \max x_i \\ \dfrac{e^{\sum_{i=1}^n x_i}}{(e^\theta-1)^n} & \text{si } \theta > \max x_i \end{cases}$$

Para encontrar el EMV, vamos a buscar el máximo de la función de verosimilitud.

Si $0<\max x_i<\theta$, $L_{x_1,\dots,x_n}(\theta) = \dfrac{e^{\sum_{i=1}^{n} x_i}}{(e^\theta-1)^n}$

con $0<\max x_i$ ^ θ

$\ln L_{x_1,\dots,x_n}(\theta) = \sum_{i=1}^{n} x_i - n\ln(e^\theta-1) \longrightarrow \dfrac{\partial \ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta} = -\dfrac{ne^\theta}{e^\theta-1} < 0$

$\Rightarrow$ cuando $0<X_{(n)}<\theta$, $L_{x_1,\dots,x_n}(\theta)$ decrece, por tanto alcanza su máximo en el mínimo valor de θ, es decir $\cancel{X_{(n)}}$ $\max x_i$, por tanto $\hat\theta = X_{(n)}$. Como nos piden el EMV para e^θ, por el Teorema de invarianza de Zhena tenemos que

$\widehat{e^\theta} = e^{\hat\theta} = e^{X_{(n)}}$ //

¿Es insesgado?

Como ya demostramos que $h(T) = e^T(e^T-1)^{n-2} + \dfrac{(e^T-1)^{n-1}}{n}$ es UMVUE para e^θ, con $T=X_{(n)}$. Como el UMVUE para e^θ es único, entonces se tiene que $\widehat{e^\theta} = e^{X_{(n)}}$ no es UMVUE para e^θ, pues en general $h(T)\neq e^{X_{(n)}}$. Por tanto, tampoco es insesgado, ya que si lo fuera, sería un estimador función del UMVUE para e^θ, contradiciendo así su unicidad.

~~3.- Sea $(X_1,\ldots,X_n)$~~ Teoría

~~b) A partir del test anterior~~

Ej de Febrero 2015 (No sé si está bien)

b) El nº de clientes que visitan una oficina sigue una distribución de Poisson con parámetro λ, $P(\lambda)$, y el nº de visitas al día es independiente un día de otro. Para contrastar que el número medio de visitas por día es 0'5 frente a que es 0'6, se hace un estudio durante 10 días y el número total de visitas es 12. Plantear y resolver el contraste de hipótesis adecuada adecuada a la hipótesis con el test más potente de tamaño 0,005 y ver si se acepta o no la hipótesis nula, H_0.

Utilizaremos el test de máxima potencia o test de Neyman-Pearson, con tamaño 0,005.

$X \equiv$ Nº de clientes que visitan una oficina $\quad X \to P(\lambda)$

$(X_1,\ldots,X_{10}) \quad \sum_{i=1}^{10} X_i = 12 \qquad f_\lambda(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

$$\begin{cases} H_0: \lambda = 0'5 \\ H_1: \lambda = 0'6 \end{cases} \qquad \mathcal{X}_0 = \mathcal{X}_1 = \mathbb{N}\cup\{0\} \Rightarrow \mathcal{X}^n = (\mathbb{N}\cup\{0\})^n \qquad \Theta = \Theta_0 \cup \Theta_1 = \{0,5;\ 0,6\}$$

Para el test de Neyman-Pearson necesitaremos calcular

$$f_j^n(x_1,\ldots,x_n) = \frac{e^{-\lambda_j n}\lambda_j^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \quad (x_1,\ldots,x_n)\in\mathcal{X}^n \quad j=1,0 \quad \text{y } f_0^n(x_1,\ldots,x_n)\neq 0 \ \forall (x_1,\ldots,x_n)\in\mathcal{X}^n$$

Por tanto podemos construir el test de Neyman-Pearson de la siguiente forma:

$$\varphi(x_1,\ldots,x_n) = \begin{cases} 1 & \text{si } \lambda(x_1,\ldots,x_n) > K \\ \gamma & \text{si } \lambda(x_1,\ldots,x_n) = K \\ 0 & \text{si } \lambda(x_1,\ldots,x_n) < K \end{cases}$$

donde $\lambda(x_1,\ldots,x_n) = \dfrac{f_1^n(x_1,\ldots,x_n)}{f_0^n(x_1,\ldots,x_n)}$

$K\in\mathbb{R}^+\cup\{0\}$

y $\gamma(x_1,\ldots,x_n) = \gamma \in [0,1]$

Escaneado con CamScanner

$$\lambda(x_1,\dots,x_n) = \frac{f_1^n(x_1,\dots,x_n)}{f_0^n(x_1,\dots,x_n)} = e^{n(\lambda_0-\lambda_1)} \lambda_1^{\sum_{i=1}^{n} x_i} \lambda_0^{\sum_{i=1}^{n} x_i} \quad \text{con } (x_1,\dots,x_n) \in \mathcal{X}^n$$

Como $n=10$ y $\sum_{i=1}^{10} x_i = 12$ nos queda $\lambda(x_1,\dots,x_n) = e^{10(\lambda_0-\lambda_1)} \lambda_1^{12} \lambda_0^{12}$
$= e^{10(\lambda_0-\lambda_1)} (\lambda_1\lambda_0)^{12}$

Le damos a K el valor que toma $\lambda(x_1,\dots,x_n)$.

Como $\lambda(x_1,\dots,x_n) = K = e^{10(\lambda_0-\lambda_1)} (\lambda_1\lambda_0)^{12}$ $\forall (x_1,\dots,x_n) \in \mathcal{X}^n$ el test de Neyman-Pearson nos queda

$\varphi(x_1,\dots,x_n) = \gamma$ Determinamos γ con el tamaño

~~$E_{\theta_0}(\varphi(X_1,\dots,X_n)) = \gamma$~~ Como $\forall (x_1,\dots,x_n) \in \mathcal{X}^n$ $\varphi(x_1,\dots,x_n) = \gamma$ el tamaño de nuestro test será γ, es decir $E_{\theta_0}(\varphi(X_1,\dots,X_n)) = 0{,}005 = \gamma$ y por tanto el TMP de tamaño $0{,}005$ será $\varphi(x_1,\dots,x_n) = 0{,}005$ $\forall (x_1,\dots,x_n) \in \mathcal{X}^n$

Es decir, ~~rechazaremos~~ aceptaremos H_0 con probabilidad $0{,}995$ (o ~~aceptamos~~ rechazamos con probabilidad $0{,}005$). Los datos no dan evidencia para aceptar H_0.

4.- Teoría

5.- a) Teoría

b) A partir de una muestra de 200 empleados, se ha construido la siguiente tabla de frecuencias conjunta de calidad de vida (CV) y Rendimiento de Trabajo (RT). Se desea contrastar a nivel 0,05 si los RT son independientes de la CV de los empleados.

CV \ RT	Poca	Media	Mucha
Bajo	62	10	2
Medio	14	38	12
Alto	2	9	51

Se trata de contrastar la independencia de las variables "Calidad de vida" y "Rendimiento de trabajo". Ambas variables categóricas:

X = Calidad de vida

Y = Rendimiento de trabajo

$Y_1 \equiv$ "Bajo" $Y_2 \equiv$ "Medio" $Y_3 \equiv$ "Alto"

$X_1 \equiv$ "Poca" $X_2 \equiv$ "Media" $X_3 \equiv$ "Mucha"

$H_0: X, Y$ independientes o equivalentemente

$P(X = X_i, Y = Y_j) = P(X = X_i)P(Y = Y_j) \quad i = 1,2,3 \quad j = 1,2,3$

Debemos estimar las frecuencias esperadas bajo H_0 por máxima verosimilitud $e_{ij} = \frac{n_{i.} n_{.j}}{n}$

RT \ CV	Poca	Media	Mucha	$n_{i.}$
Bajo	62 28,86	10 21,09	2 24,05	74
Medio	14 24,96	38 18,24	12 20,8	64
Alto	2 0,62	9 17,67	51 20,15	62
$n_{.j}$	78	57	65	200 = n

Como todas las frecuencias esperadas son mayores que 2 y que 5, el test χ^2 de independencia es fiable.

$$\chi^2_{exp} = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{(n_{ij} - e_{ij})}{e_{ij}} = \frac{(62-28,86)^2}{28,86} + \frac{(10-21,09)^2}{21,09} + \dots + \frac{(51-20,15)^2}{20,15} =$$

$$= 38,05473 + 5,83158 + 20,21632 + 4,8126 + 21,40667 + 3,72308 +$$
$$+ 3,07161 + 4,25405 + 47,23189 = 148,60253$$

$$\varphi(N_{ij}) = \begin{cases} 1 & \chi^2_{exp} \geq \chi^2_{4;0,05} \\ 0 & \chi^2_{exp} < \chi^2_{4;0,05} \end{cases}$$

Como $\chi^2_{4;0,05} = 9,4877$

Como $\chi^2_{exp} > \chi^2_{4;0,05}$ No se rechaza la hipótesis H_0 a nivel de significación 0,05. Los datos no dan evidencia de que exista relación entre la calidad de vida y el rendimiento de trabajo.