Universidad del Valle de Guatemala
Departamento de Matemática
Licenciatura en Matemática Aplicada

Estudiante: Rudik Roberto Rompich
E-mail: rom19857@uvg.edu.gt
Carné: 19857

MM2036 - Estadística Matemática - Catedrático: Paulo Mejía
21 de abril de 2021

Parcial 3

Examen Parcial 3 Instrucciones: Resuelva los siguientes problema. Favor hacer la solución en latex y cargar el archivo latex y pdf en la tarea de Canvas.

Para la resolución de los problemas, se utilizará el libro de Wackerly et al. (2014).

1. Problema 1

Sea $Y_1, Y_2, \dots, Y_n$ una muestra aleatoria de tamaño n de una población cuya densidad está dada por

$$f(y) = \begin{cases} 3\beta^3 y^{-4} & , \quad \beta \leq y \\ 0 & , \text{ en cualquier otro caso} \end{cases}$$

donde $\beta > 0$ es un valor fijo desconocido. Considérese el estimador $\hat{\beta} = \text{mín}\,(Y_1, Y_2, \dots, Y_n)$

1. a) Demuestre que $\hat{\beta}$ es un estimador sesgado de β.

Definición 8.2

Sea $\hat{\theta}$ un estimador fijo para un parámetro θ. Entonces $\hat{\theta}$ es un estimador insesgado si $E(\hat{\theta}) = \theta$. Si $E(\hat{\theta}) \neq \theta$, $\hat{\theta}$ es sesgado.

Sección 6.7 - Estadísticos de Orden (Order Statistics)

Considerando las variables ordenadas aleatorias $Y_1, Y_2, ..., Y_n$ donde $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$. La notación propuesta por el libro:

$$Y_{(1)} = \text{mín}(Y_1, Y_2, \cdots, Y_n)$$

$$Y_{(n)} = \text{máx}(Y_1, Y_2, \cdots, Y_n)$$

Resumiendo, $g_{(1)}(y)$ denota la función densidad de $Y_{(1)}$, entonces:

$$g_{(1)}(y) = n\,[1 - F(y)]^{n-1}\, f(y)$$

Demostración. Se considerará el problema **8,15** del libro de texto, en donde se indica que esta es una de las distribuciones de *Pareto*. Se considera su función de distribución de esta distribución de *Pareto*:

$$F(y) = \begin{cases} 0, & y < \beta \\ 1 - \left(\frac{\beta}{y}\right)^{\alpha}, & y \geq \beta \end{cases}$$

Del cual su función densidad (luego de derivar) es:

$$f(y) = \begin{cases} 0, & y < \beta \\ \alpha\beta^{\alpha}y^{-(\alpha+1)}, & y \geq \beta \end{cases}$$

Ahora bien, por la hipótesis del problema identificamos que tenemos un caso específico de la función densidad de la distribución de *Pareto*; en el cual $\alpha = 3$. Por otra parte, sabemos del **Cuadro de Estadísticos de Orden** que la función densidad de $Y_{(1)} = \text{mín}(Y_1, Y_2, \cdots, Y_n)$ es:

$$\begin{aligned} g_{(1)}(y) &= n\left[1 - F(y)\right]^{n-1} f(y) \\ &= n\left[1 - \left(1 - \left(\frac{\beta}{y}\right)^3\right)\right]^{n-1} 3\beta^3 y^{-(3+1)} \\ &= n\left[\beta^{3(n-1)}y^{-3(n-1)}\right] 3\beta^3 y^{-(4)} \\ &= 3n\beta^{3n}y^{-(3n+1)}, \qquad y \geq \beta \end{aligned} \tag{1}$$

Ahora procedemos a encontrar el valor esperado de $Y_{(1)}$ (considerando $g_{(1)}(y) = f_{(1)}(y)$) por el procedimiento usual:

$$\begin{aligned} E(Y_{(1)}) &= \int_{\infty}^{\infty} y f_{(1)}(y)\ dy \\ &= \int_{\beta}^{\infty} y 3n\beta^{3n}y^{-(3n+1)}\ dy \\ &= 3n\beta^{3n} \lim_{h\to\infty} \int_{\beta}^{h} y^{-3n}\ dy \\ &= 3n\beta^{3n} \lim_{h\to\infty} \left[\frac{y^{-3n+1}}{-3n+1}\right]_{\beta}^{h} \\ &= 3n\beta^{3n}\left[-\frac{\beta^{-3n+1}}{-3n+1}\right] \\ &= \left(\frac{3n}{3n-1}\right)\beta \end{aligned}$$

Ahora, volvemos a la expresión original $Y_{(1)} = \hat{\beta}$. Usando el **Cuadro de la Definición 8.2**, concluimos que $E(\hat{\beta}) \neq 0$. Por lo tanto es un estimador sesgado. □

.

2. b) Determine un múltiplo de $\hat{\beta}$ que constituya un estimador insesgado.

   Solución. Es decir, que nos están preguntando encontrar $E(\hat{\beta}) = \beta$. Por el inciso anterior tenemos que:

$$E(\hat{\beta}) = \left(\frac{3n}{3n-1}\right)\beta$$

   Entonces, despejando β:

$$\left(\frac{3n-1}{3n}\right)E(\hat{\beta}) = \beta$$

$$E\left(\left(\frac{3n-1}{3n}\right)\hat{\beta}\right) = \beta$$

   Es decir, que el múltiplo que constituye un estimador insesgado es:

$$\left(\frac{3n-1}{3n}\right)\hat{\beta} = \left(\frac{3n-1}{3n}\right)Y_{(1)}$$

   □

3. c) Encuentre el sesgo de $\hat{\beta}, \mathrm{B}(\hat{\beta})$ (recuerde que es el estimador original).

   Definición 8.3

   El sesgo de un estimador fijo $\hat{\theta}$ está dado por $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

   Solución. Ahora, considerando la **Definición 8.3** y el inciso **a**, tenemos:

$$\begin{aligned} B(\hat{\beta}) &= E(\hat{\beta}) - \beta \\ &= \left(\frac{3n}{3n-1}\right)\beta - \beta \\ &= \left(\frac{1}{3n-1}\right) \end{aligned}$$

   □

4. d) Encuentre $\mathrm{MSE}(\hat{\beta})$ (recuerde que es el estimador original).

   Definición 8.4

   El error cuadrado medio de un estimador fijo $\hat{\theta}$ es:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

Solución. Considerando la **Definición 8.4**:

$$
\begin{aligned}
MSE(\hat{\beta}) &= E[(\hat{\beta}-\beta)^2] \\
&= E[\hat{\beta}^2 - 2\hat{\beta}\beta + \beta^2] \\
&= E(\hat{\beta}^2) - 2\beta E(\hat{\beta}) + E(\beta^2) \\
&= E(\hat{\beta}^2) - 2\beta E(\hat{\beta}) + \beta^2
\end{aligned}
$$

Nos percatamos que $E(\hat{\beta}^2)$ no ha sido calculado, por lo que se calculará con el método del inciso **a**.

$$
\begin{aligned}
E(\hat{\beta}^2) = E(Y_{(1)}^2) &= \int_{\infty}^{\infty} y^2 f_{(1)}(y)\ dy \\
&= \int_{\beta}^{\infty} y^2 3n\beta^{3n} y^{-(3n+1)}\ dy \\
&= 3n\beta^{3n} \lim_{h\to\infty} \int_{\beta}^{h} y^{-3n+1}\ dy \\
&= 3n\beta^{3n} \lim_{h\to\infty} \left[\frac{y^{-3n+2}}{-3n+2}\right]_{\beta}^{h} \\
&= 3n\beta^{3n} \left[-\frac{\beta^{-3n+2}}{-3n+2}\right] \\
&= \left(\frac{3n}{3n-2}\right)\beta^2
\end{aligned}
$$

Entonces, ahora haciendo una substitución:

$$
\begin{aligned}
MSE(\hat{\beta}) &= E(\hat{\beta}^2) - 2\beta E(\hat{\beta}) + \beta^2 \\
&= \left(\frac{3n}{3n-2}\right)\beta^2 - 2\beta^2\left(\frac{3n}{3n-1}\right) + \beta^2 \\
&= \left(\frac{3n}{3n-2}\right)\beta^2 - \left(\frac{3n}{3n-1}\right)\beta^2 \\
&= \left(\frac{3n}{3n-2} - \frac{3n}{3n-1}\right)\beta^2 \\
&= \left(\frac{2}{(3n-1)(3n-2)}\right)\beta^2
\end{aligned}
$$

□

(Valor 25 puntos).

2. Problema 2

Se puede definir un límite de desviación estándar 2 en el error de estimación con cualquier estimador para el cual se pueda hallar una estimación razonable del error estándar.

1. a) Explique por qué. Supóngase que $Y_1, Y_2, \ldots, Y_n$ representa una muestra aleatoria de una distribución de Poisson con media λ. Se conoce que $\text{VAR}(Y_i) = \lambda$.

Solución. Vamos a considerar el **Cuadro 3.4** del libro de texto:

Table 3.4 Means and variances for some common discrete random variables

Distribution	$E(Y)$	$V(Y)$
Binomial	np	$np(1-p) = npq$
Geometric	$\frac{1}{p}$	$\frac{1-p}{p^2} = \frac{q}{p^2}$
Hypergeometric	$n\left(\frac{r}{N}\right)$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$
Poisson	λ	λ
Negative binomial	$\frac{r}{p}$	$\frac{r(1-p)}{p^2} = \frac{rq}{p^2}$

Además, el **Cuadro 1** del apéndice:

Table 1 Discrete Distributions

Distribution	Probability Function	Mean	Variance	Moment-Generating Function
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!}$; $y = 0, 1, 2, \ldots$	λ	λ	$\exp[\lambda(e^t - 1)]$

Por definición se conoce que:

$$\text{VAR}(Y) = E[Y(Y-1)] + E(Y) - [E(Y)]^2 \tag{2}$$

En donde $E(Y)$:

$$E[Y] = \sum_{y=0}^{\infty} \frac{y\lambda^y e^{-\lambda}}{y!}$$

En donde el primer término ($y = 0$) es 0, por lo cual:

$$= \sum_{y=1}^{\infty} \frac{y\lambda^y e^{-\lambda}}{y!} = \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1} e^{-\lambda}}{(y-1)!}$$

Se propone un cambio de variables $z = y - 1$:

$$= \lambda \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!}$$

Nótese que $p(z) = \frac{\lambda^z e^{-\lambda}}{z!}$ es la función de probabilidad para una variable aleatoria de Poisson; entonces $\sum_{z=0}^{\infty} p(z) = 1$, por lo tanto:

$$= \lambda$$

En donde $E[Y(Y-1)]$:

$$E[Y(Y-1)] = \sum_{y=0}^{\infty} \frac{y(y-1)\lambda^y e^{-\lambda}}{y!}$$

En donde el primer término y segundo término ($y = 0, 1$) son 0, por lo cual:

$$= \sum_{y=2}^{\infty} \frac{y(y-1)\lambda^y e^{-\lambda}}{y!} = \lambda^2 \sum_{y=2}^{\infty} \frac{\lambda^{y-2} e^{-\lambda}}{(y-2)!}$$

Se propone un cambio de variables $z = y - 2$:

$$= \lambda^2 \sum_{z=-2}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!}$$

Nótese que $p(z) = \frac{\lambda^z e^{-\lambda}}{z!}$ es la función de probabilidad para una variable aleatoria de Poisson; entonces $\sum_{z=0}^{\infty} p(z) = 1$, por lo tanto:

$$= \lambda^2$$

Regresando a (1), tenemos:

$$\begin{aligned} \mathrm{VAR}(Y) &= E[Y(Y-1)] + E(Y) - [E(Y)]^2 && (1) \\ &= \lambda^2 + \lambda - \lambda^2 && (2) \\ &= \lambda && (3) \end{aligned}$$

Regresando al problema original, se indica que $Y_1, Y_2, ..., Y_n$ son variables aleatorias que representan una muestra aleatoria de una distribución de Poisson y con media λ. Ahora bien, basándonos en toda la argumentación anterior, la explicación de que $\text{VAR}(Y_i) = \lambda$ es que i es un contador y cada una de las variables aleatorias Y_i son iguales a λ. Análogamente, es la misma argumentación para $E(Y_i) = \lambda$.

$\square$

2. b) Calcule $\text{E}(\overline{Y})$ y $\text{VAR}(\overline{Y})$.

Teorema 5.12

Sea $Y_1, Y_2, \ldots, Y_n$ y $X_1, X_2, \ldots, X_m$ variable aleatorias con $E(Y_i) = \mu_i$ y $E(X_j) = \xi_j$. Se define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

para constantes $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Entonces, lo siguiente se mantiene:

a) $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$.

b) $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum\sum_{1 \le i < j \le n} a_i a_j \operatorname{Cov}(Y_i, Y_j)$, donde la doble suma está sobre todos los pares (i, j) con $i < j$.

c) $\operatorname{Cov}(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \operatorname{Cov}(Y_i, X_j)$.

Solución. Sabemos que $\overline{Y}$ hace referencia a la media muestral:

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

El valor esperado (aplicando las propiedades previamente conocidas):

$$\begin{aligned}
E(\overline{Y}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) &= \frac{1}{n} E\left(\sum_{i=1}^{n} Y_i\right) \\
&= \frac{1}{n} E\left(Y_1 + Y_2 + Y_3 + ... + Y_n\right) \\
&= \frac{1}{n}\left[E(Y_1) + E(Y_2) + E(Y_3) + ... + E(Y_n)\right] \\
&= \frac{1}{n}\left[\lambda + \lambda + \lambda + ... + \lambda\right] \\
&= \frac{1}{n}\left[n\lambda\right] \\
&= \lambda
\end{aligned}$$

La varianza (basándose en el **teorema 5,12**) :

$$\begin{aligned}
\mathrm{VAR}(\overline{Y}) &= \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 \mathrm{VAR}\,(Y_i) + 2\sum\sum_{1\le i<j\le n}\left(\frac{1}{n}\right)^2\left(\frac{1}{n}\right)^2 \mathrm{Cov}\,(Y_i, Y_j) \\
&= \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 \mathrm{VAR}(Y_i) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \mathrm{VAR}(Y_i) \\
&= \left(\frac{1}{n}\right)^2 n\lambda \\
&= \frac{\lambda}{n}
\end{aligned}$$

□

3. c) ¿Cómo emplearía $Y_1, Y_2, \ldots, Y_n$ para estimar λ?

   Solución. Se conoce ampliamente que $\overline{Y}$ es un estimador de una muestra de variables aleatorias $Y_1, Y_2, Y_3, \cdots, Y_n$. Por lo cual:

$$\hat{\lambda} = \overline{Y}$$

□

4. d) ¿Cómo estimaría el error estándar de su estimador propuesto en c)? (Valor 25 puntos).

   Solución. Esto es bastante sencillo, ya que sabemos que:

$$\mathrm{VAR}(\overline{Y}) = \frac{\lambda}{n}$$

que es estimado por medio de:

$$\frac{\hat{\lambda}}{n} = \frac{\overline{Y}}{n}$$

Por lo tanto, el error estándar del estimador es estimado por:

$$\sqrt{\frac{\overline{Y}}{n}}$$

□

3. Problema 3

Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una población con distribución uniforme en el intervalo $(0, \theta)$. Sean $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ y $U = \frac{1}{\theta} Y_{(n)}$.

1. a) Encuentre la función de distribución acumulada $F_U(u)$.

Sección 6.7 - Estadísticos de Orden (Order Statistics)

Considerando las variables ordenadas aleatorias $Y_1, Y_2, ..., Y_n$ donde $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$. La notación propuesta por el libro:

$$Y_{(1)} = \min(Y_1, Y_2, \cdots, Y_n)$$

$$Y_{(n)} = \max(Y_1, Y_2, \cdots, Y_n)$$

Resumiendo

a) $g_{(n)}(y)$ denota la función densidad de $Y_{(n)}$, entonces:

$$g_{(n)}(y) = n\,[F(y)]^{n-1}\,f(y)$$

b) La función distribución de $Y_{(n)}$ es dado por:

$$F_{Y_{(n)}} = P(Y_{(n)} \leq y) = P(Y_1 \leq y)P(Y_2 \leq y)\cdots P(Y_n \leq y) = [F(y)]^n$$

Definición 4.6 - Distribución Uniforme-Densidad

$$f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2 \\ 0, & \text{cualquier otro} \end{cases}$$

Solución. La función densidad de la distribución uniforme en el intervalo $(0, \theta)$ es:

$$f(y) = \begin{cases} \frac{1}{\theta}, & 0 \leq y \leq \theta \\ 0, & \text{cualquier otro} \end{cases}$$

Es decir que integrando la función densidad tenemos la función distribución:

$$F(y) = \begin{cases} \frac{y}{\theta}, & 0 \leq y \leq \theta \\ 0, & \text{cualquier otro} \end{cases}$$

Comenzamos proponiendo el enunciado de la **Sección 6.7** que dice que:

$$\begin{aligned} F_{Y_{(n)}}(y) &= [F(y)]^n \\ &= \left[\frac{y}{\theta}\right]^n, \quad 0 \leq y \leq \theta \end{aligned}$$

Ahora, por otra parte, calculando la función distribución de U, tenemos (usando la defición usual):

$$
\begin{aligned}
F_U(u) &= P(U \leq u) \\
&= P\left(\frac{Y_{(n)}}{\theta} \leq u\right) \\
&= P\left(Y_{(n)} \leq \theta u\right) \\
&= F_{Y_{(n)}}(\theta u)
\end{aligned}
$$

en donde es trivial ver (analizando sus casos) que:

$$
F_U(u) = \begin{cases} 0, & \theta u \leq 0 \\ \left(\frac{\theta u}{\theta}\right)^n, & 0 \leq \theta u \leq \theta \\ 1, & \theta u \geq 1 \end{cases}
$$

Aquí, haremos una suposición para un mejor manejo del problema, como en el **Ejemplo 8,5** del libro de texto, en donde se puede asumir que U es uniformemente distribuido sobre $[0, 1]$. Por lo tanto:

$$
F_U(u) = \begin{cases} 0, & u \leq 0 \\ (u)^n, & 0 \leq u \leq 1 \\ 1, & u \geq 1 \end{cases}
$$

$\square$

2. b) Demuestre que $\frac{1}{\theta}Y_{(n)}$ es una cantidad pivote.

Sección 8.5-Intervalos de confianza - Método del Pivote

El método depende de encontrar una cantidad pivote que posee dos características:

a) Es una función de las medidas de la muestra y el parámetro desconocido θ, donde θ es la única cantidad desconocida.

b) Su función distribución no depende sobre el parámetro θ.

Solución. Entonces, nos están pidiendo (asumiendo h como el intervalo de confianza al que debemos llegar):

$$
\begin{aligned}
P(U \leq a) &= P\left(\frac{Y_{(n)}}{\theta} \leq a\right) \\
&= F_U(a) \\
&= h
\end{aligned}
$$

Entonces $a^n = h$, tal que $a = (h)^{1/n}$. Por lo tanto, $U = \frac{Y_{(n)}}{\theta}$ es una cantidad pivote; ya que es independiente.

$\square$

3. c) Use la cantidad de pivote del inciso b) para hallar un límite de confianza inferior de 95 % para θ $\left(\mathrm{P}\left(\hat{\theta}_L \leq \theta\right) = 1-\alpha\right)$

   Solución. Usando el inciso anterior, en donde, $h = 0{,}95$. Tenemos que $a = (0{,}95)^{1/n}$. Por lo tanto, podemos asumir lo siguiente:

$$
\begin{aligned}
0{,}95 &= P\left(U \leq 0{,}95^{\frac{1}{n}}\right)\\
&= P\left(\frac{Y_{(n)}}{\theta} \leq 0{,}95^{\frac{1}{n}}\right)\\
&= P\left(\theta \geq \frac{Y_{(n)}}{0{,}95^{\frac{1}{n}}}\right)
\end{aligned}
$$

   Concluimos que el límite de confianza de 95 % para θ está definido como:

$$
\theta \in \left\lfloor \frac{Y_{(n)}}{0{,}95^{\frac{1}{n}}}, \infty \right)
$$

□

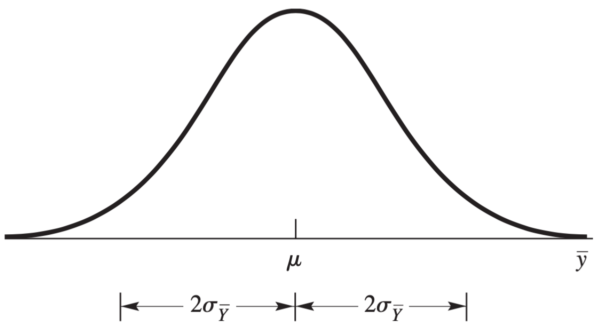
(Valor 25 puntos).

4. Problema 4

Suponga que se quiere estimar el promedio diario de producción μ de un producto farmacéutico con un error de estimación menor que 3 toneladas con probabilidad de 0,95. Además, se sabe que la amplitud de la producción es aproximadamente 120 toneladas. Calcule el tamaño muestral de forma que se tenga cierta seguridad (con un coeficiente de alrededor de 0,95) de que la estimación se encuentra a no más de 3 toneladas del verdadero promedio diario de producción (Para poder realizar dicho procedimiento, incluya las suposiciones pertinentes). (Valor 25 puntos).

Solución. Empezamos analizando los datos proporcionados:

1. Ya que 95 % de las medias muestras no estarán a más de $2\sigma_{\overline{Y}}$ del μ en muestreo repetido, es decir que estamos pidiendo que $2\sigma_{\overline{Y}}$ sea igual a 3 toneladas. Gráficamente:



Es decir, que lo que nos están pidiendo es:

$$\frac{2\sigma}{\sqrt{n}} = 3 \implies n = \frac{4\sigma^2}{9}$$

2. Amplitud de la producción es de 120 toneladas.

Ahora bien, nótese que n no puede ser calculado a menos que sepamos previamente el valor de σ. Tómese en cuenta que la variabilidad asociada con el estimador $\overline{Y}$ depende sobre la variabilidad exhibida de la población muestrada.

Debido a esta incerteza de σ, es necesario encontrar un método, la suposición es la siguiente:

1. Se sabe que el rango es de 4σ. Empíricamente un cuarto de 4σ nos debería dar el valor de σ.

Conociendo previamente que la amplitud de producción es 120, entonces:

$$\sigma \approx \frac{120}{4} = 30$$

Por lo tanto, regresando a la expresión original:

$$n = \frac{4\sigma^2}{9} = \frac{4(30)^2}{9} = 400$$

Lo que quiere decir que usando una muestra de $n = 400$, se puede estar significativamente seguro (con coeficiente de 0.95) de que la estimación se encuentra a no más de 3 toneladas del verdadero promedio diario de producción. □

Referencias

Wackerly, D., Mendenhall, W., and Scheaffer, R. L. (2014). *Mathematical statistics with applications.* Cengage Learning.