Universidad Nacional Agraria La Molina
Departamento de Estadística e Informática
Curso: Inferencia Estadística
Profesor: Clodomiro Fernando Miranda Villagómez

Estimación de parámetros con intervalos de confianza

Nota: Empezar esta Guía investigando el Script R (desde la fila 1 hasta la fila 387)

Intervalo aleatorio

$$P\left(T_1 \le \tau\left(\theta\right) \le T_2\right) = \gamma = 1-\alpha$$

Observación

La longitud del intervalo aleatorio $I = \left[T_1, T_2\right]$ es $L = T_2 - T_1$ y la longitud esperada es $E\left(L\right)$ si es que existe.

Ejemplo 1 Sea $X_1,\ldots,X_n$ una m.a de una población $N\left(\mu,\sigma^2\right)$, con σ^2 conocida. Sea $I = \left[\bar{X} - 1.4\frac{\sigma}{\sqrt{n}}; \bar{X} + 1.4\frac{\sigma}{\sqrt{n}}\right]$ un intervalo aleatorio para $\tau\left(\mu\right) = \mu$.

c. Asuma que $\sigma = 2$ y $n = 120$, entonces se obtiene $I = \left[\bar{X} - 0.2556; \bar{X} + 0.2556\right]$. Este intervalo aleatorio tiene una probabilidad de 0.8384867 de incluir a la media μ.

a. Halle el coeficiente de confianza para este intervalo.

$$1-\alpha = \gamma = P\left[\bar{X} - 1.4\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.4\frac{\sigma}{\sqrt{n}}\right] = P\left[-1.4 \le \frac{\left(\bar{X}-\mu\right)}{\sigma}\sqrt{n} \le 1.4\right]$$

$$1-\alpha = \gamma = P\left[-1.4 \le Z \le 1.4\right] = 0.8384867$$

b. Determine la longitud del intervalo y la longitud esperada.

$$L = \left(\bar{X} + 1.4\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X} - 1.4\frac{\sigma}{\sqrt{n}}\right) = 2.8\frac{\sigma}{\sqrt{n}}$$

$$E\left(L\right) = 2.8\frac{\sigma}{\sqrt{n}}$$

c. Asuma que $\sigma = 2$ y $n = 120$, entonces se obtiene $I = \left[\bar{X} - 0.2556; \bar{X} + 0.2556\right]$. Este intervalo aleatorio tiene una probabilidad de 0.8384867 de incluir a la media μ.

d. Si en una muestra particular la estimación puntual de la media es $\bar{x} = 30$, entonces el intervalo confidencial [29.7444, 30.2556] ya no es un intervalo aleatorio. Interpretamos este intervalo confidencial afirmando que tenemos un 95% de confianza de que la media poblacional está entre 29.7444 y 30.2556. También se dice que, si se toman N muestras de tamaño n, en cada muestra se obtendrán intervalos confidenciales diferentes y el 83.84867% de estos intervalos aleatorios incluirán a la media poblacional.

NOTA: Ver el SCRIPT R.

Ejemplo 2: Si X_1, X_2 una m.a de una población Uniforme$\left(4\theta - \frac{1}{2}; 4\theta + \frac{1}{2}\right)$.

a. Halle el coeficiente de confianza si se cumple que $P\left[\frac{Y_1}{4} - \frac{1}{32} \le \theta \le \frac{Y_2}{4} + \frac{1}{32}\right] = \gamma$.

$$\gamma = P\left[\frac{X_1}{4} - \frac{1}{32} \le \theta \le \frac{X_2}{4} + \frac{1}{32}\right] + P\left[\frac{X_2}{4} - \frac{1}{32} \le \theta \le \frac{X_1}{4} + \frac{1}{32}\right] =$$

$$\gamma = P\left[\frac{X_1}{4} - \frac{1}{32} \le \theta\right] P\left[\frac{X_2}{4} + \frac{1}{32} \ge \theta\right] + P\left[\frac{X_2}{4} - \frac{1}{32} \le \theta\right] P\left[\frac{X_1}{4} + \frac{1}{32} \ge \theta\right] =$$

$$\gamma = P\left[X_1 \le 4\theta + \frac{1}{8}\right] P\left[X_2 \ge 4\theta - \frac{1}{8}\right] + P\left[X_2 \le 4\theta + \frac{1}{8}\right] P\left[X_1 \ge 4\theta - \frac{1}{8}\right] =$$

$$\gamma = 2\left[\int_{4\theta - \frac{1}{2}}^{4\theta + \frac{1}{8}} 1dt\right]\left[\int_{4\theta - \frac{1}{8}}^{4\theta + \frac{1}{2}} 1dt\right] = 2\left[\frac{5}{8}\right]\left[\frac{5}{8}\right] = \frac{25}{32}$$

b. Obtenga la longitud del intervalo y la longitud esperada.

Longitud del intervalo

$$L = \left[\frac{Y_2}{4} + \frac{1}{32}\right] - \left[\frac{Y_1}{4} - \frac{1}{32}\right] = \frac{1}{4}\left(Y_2 - Y_1\right) + \frac{1}{16}$$

Pero:

$L = \frac{1}{4}\left(Y_2 - Y_1\right) + \frac{1}{16} > 0 \rightarrow Y_2 - Y_1 > -\frac{1}{4}$ se cumple con todos los reales.

Longitud esperada

$$F(x) = \int_{4\theta - \frac{1}{2}}^{x} 1dt = x - 4\theta + \frac{1}{2}$$

$$g(y_1) = 2\left[1 - y_1 + 4\theta - \frac{1}{2}\right] \times 1 = 2\left[\frac{1}{2} - y_1 + 4\theta\right] \rightarrow E(Y_1) = 4\theta - \frac{1}{6}$$

$$g(y_2) = 2\left[y_2 - 4\theta + \frac{1}{2}\right] \times 1 = 2\left[y_2 - 4\theta + \frac{1}{2}\right] \rightarrow E(Y_2) = 4\theta + \frac{1}{6}$$

$$E(L)=E\left[\frac{1}{4}Y_2-\frac{1}{4}Y_1+\frac{1}{16}\right]=\frac{1}{4}E(Y_2)-\frac{1}{4}E(Y_1)+\frac{1}{16}=$$

$$E(L)=\frac{1}{4}E(Y_2)-\frac{1}{4}E(Y_1)+\frac{1}{16}=\frac{1}{4}\left(4\theta+\frac{1}{6}\right)-\frac{1}{4}\left(4\theta-\frac{1}{6}\right)+\frac{1}{16}=\frac{7}{48}$$

c. Halle el coeficiente de confianza si se cumple que: $P\left[Y_1\le\theta\le Y_2\right]=\gamma$

$$\gamma=P\left[X_1\le\theta\le X_2\right]+P\left[X_2\le\theta\le X_1\right]=$$
$$\gamma=P\left[X_1\le\theta\right]P\left[X_2\ge\theta\right]+P\left[X_2\le\theta\right]P\left[X_1\ge\theta\right]=$$
$$\gamma=2\left[\int_{4\theta-\frac{1}{2}}^{\theta}1dt\right]\left[\int_{\theta}^{4\theta+\frac{1}{2}}1dt\right]=\frac{(1-6\theta)(1+6\theta)}{2}=\frac{1-36\theta^2}{2}$$

Pero:

$$0<\frac{1-36\theta^2}{2}<1\rightarrow 0<\theta<\frac{1}{6}$$

d. Obtenga la longitud esperada de la subpregunta c.

Aplicando la definición y los resultados de b:

$$E(L)=E(Y_2-Y_1)=E(Y_2)-E(Y_1)=\left(4\theta+\frac{1}{6}\right)-\left(4\theta-\frac{1}{6}\right)=\frac{1}{3}$$

Cantidad Pivotal (Pivote)

Sea $X_1,\ldots,X_n$ una m.a extraída de una población con densidad $f(x;\theta)$, y sea $Q=q(X_1,\ldots,X_n;\theta)$ una función de $X_1,\ldots,X_n$ y θ, $\theta\in\Theta\subset\mathbb{R}$. Si la distribución de Q no depende de θ entonces a Q se le llama cantidad pivotal o pivote. En general Q es función de una estadística suficiente.

Definición: Sea X una variable aleatoria con densidad de probabilidad $f(x;\theta)$, se dice que θ es un parámetro de escala si la distribución de la variable $\frac{X}{\theta}$ o θX no depende de θ.

Definición: Sea X una variable aleatoria con densidad de probabilidad $f(x;\theta)$, se dice que θ es un parámetro de localización si la distribución de la variable $X+\theta$ o $X-\theta$ no depende de θ.

Definición: Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución con densidad $f(x;\theta)$, y sea T un estimador máximo verosímil de θ, entonces:

- Si θ es un parámetro de escala, por lo tanto, $\frac{T}{\theta}$ o θT es una variable pivote para θ.
- Si θ es un parámetro de localización, por lo tanto, $T+\theta$ o $T-\theta$ es una variable pivote para θ.

Ejemplo 3: Sea $X_1,\ldots,X_n$ una m.a extraída de una población $N\left(\mu,\sigma^2=20\right)$entonces:

a. $Q=\bar{X}-\mu$ es una cantidad pivotal porque $Q=\bar{X}-\mu \sim N\left(0,\frac{20}{n}\right)$. Se aprecia que la distribución no depende de $\theta=\mu$.

b. $Q=\frac{\left(\bar{X}-\mu\right)\sqrt{n}}{\sqrt{20}}$ es una cantidad pivotal porque $Q=\frac{\left(\bar{X}-\mu\right)\sqrt{n}}{\sqrt{20}} \sim N\left(0,1\right)$. Se aprecia que la distribución no depende de $\theta=\mu$.

c. $Q=\frac{\bar{X}}{\mu}$ no es una cantidad pivotal porque $Q=\frac{\bar{X}}{\mu} \sim N\left(1,\frac{20}{n\mu^2}\right)$. Se aprecia que la distribución depende de $\theta=\mu$.

Ejemplo 4: Sea $X_1,\ldots,X_n$ una m.a extraída de una población $\text{Uniforme}\left(0,\theta\right)$**.**

a. ¿Es $Q=\frac{Y_n}{\theta}$ una cantidad pivotal?

$$f\left(x\right)=\frac{1}{\theta}I_{(0,\theta)}\left(x\right)\to F\left(x\right)=\frac{x}{\theta}I_{(0,\theta)}\left(x\right)$$

Y_n es una estadística suficiente.

$$f_{Y_n}\left(y\right)=n\left[F\left(y\right)\right]^{n-1}f\left(y\right)=n\left[\frac{y}{\theta}\right]^{n-1}\left(\frac{1}{\theta}\right)=\frac{ny^{n-1}}{\theta^n}I_{(0,\theta)}\left(y\right)$$

$$Q=\frac{Y_n}{\theta}\to q=\frac{y}{\theta}\to y=\theta q\to J=\frac{dy}{dq}=\theta\to$$

$$f_Q\left(q\right)=f_{Y_n}\left(\theta q\right)\left|J\right|=\frac{n\left(\theta q\right)^{n-1}}{\theta^n}\left|\theta\right|=nq^{n-1}I_{(0,1)}\left(q\right)$$

La distribución de Q no depende de $\theta\to Q=\frac{Y_n}{\theta}$ es cantidad pivotal.

b. ¿Es $T=\left(\frac{Y_n}{\theta}\right)^n$ una cantidad pivotal?

$$F_{Y_n}(y)=\int_0^y \frac{nt^{n-1}}{\theta^n}dy=\left[\frac{y}{\theta}\right]^n \sim \text{Uniforme}(0,1)$$

Se demuestra que la distribución de T es $f_T(t)=1I_{(0,1)}(t)$ y es independiente de θ por lo tanto T es una cantidad pivotal.

Ejemplo 5: Sea $X_1,\ldots,X_n$ una m.a extraída de la distribución

$f(x)=\frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)I_{(0,\infty)}(x)$.

a. Si n=1 halle el pivote apropiado.

$$f(x)=\underbrace{\frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)}_{g_\theta(x)}\times \underset{h(x)}{1}$$

$\rightarrow x$ es una estadística suficiente $\therefore$ se puede suponer que:

$Q=\frac{x}{\theta}$ es un pivote. Con el método del Jacobiano se obtiene:

$f_Q(q)=e^{-q}I_{(0,\infty)}(t)$ que no depende de $\theta \rightarrow Q=\frac{x}{\theta}$ es pivote.

b. Para cualquier n halle el pivote apropiado.

Se sabe que con una distribución exponencial $\sum_{i=1}^n X_i$ es una estadística suficiente de θ. Como la muestra de tamaño n se extrae de una distribución exponencial con media θ entonces $\sum_{i=1}^n X_i \sim \text{Gamma}\left(r=n;\lambda=\frac{1}{\theta}\right)$ por lo tanto $T=2\left(\sum_{i=1}^n X_i\right)\frac{1}{\theta}\sim\chi^2_{(2n)}$. Como la distribución de T no depende de θ se concluye que $T=2\left(\sum_{i=1}^n X_i\right)\frac{1}{\theta}$ es un pivote.

<u>Método del pivote para construir intervalos de confianza</u>

$$P(q_1\le Q\le q_2)=\gamma=1-\alpha\ .$$

Si para cada valor muestral $x_1,\ldots,x_n$ se puede invertir el pivote, esto es, si $q_1\le Q\le q_2$ es invertido de la forma:

$$t_1=t_1(x_1,\ldots,x_n)\le\tau(\theta)\le t_2=t_2(x_1,\ldots,x_n)$$

Observaciones

1. Los números reales $q_1 \quad y \quad q_2$ son independientes de θ, pues Q es un pivote cuya distribución no depende de θ.

2. Para diferentes valores de $q_1 \quad y \quad q_2$ se obtienen diferentes valores de $t_1 \; y \; t_2$.

3. Se puede construir una familia de intervalos de confianza con el mismo $\gamma = 1-\alpha$. La elección de $q_1 \quad y \quad q_2$ es óptima cuando la longitud del intervalo de confianza es mínima.

Ejemplo 6: Sea $X_1,\ldots,X_n$ una m.a de una población $N\left(\mu,\sigma^2\right)$, con σ^2 conocida. Encuentre un intervalo aleatorio de longitud mínima para μ.
Solución

El pivote $Q = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N\left(0,1\right)$

Se eligen $q_1 \quad y \quad q_2$ tales que $P\left(q_1 \le \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le q_2 \right) = \gamma = 1-\alpha$

Invirtiendo se tiene $P\left(\bar{X} - q_2 \dfrac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} - q_1 \dfrac{\sigma}{\sqrt{n}} \right) = \gamma = 1-\alpha$

El intervalo aleatorio para μ será: $I = \left[\underbrace{\bar{X} - q_2 \dfrac{\sigma}{\sqrt{n}}}_{T_1}, \underbrace{\bar{X} - q_1 \dfrac{\sigma}{\sqrt{n}}}_{T_2} \right]$

La longitud del intervalo aleatorio es: $L = T_2 - T_1 = \dfrac{\sigma}{\sqrt{n}}\left(q_2 - q_1\right)$

Suponiendo que $q_2 = h\left(q_1\right)$ minimicemos L:

$$\frac{\partial L}{\partial q_1} = \frac{\sigma}{\sqrt{n}}\left(\frac{\partial q_2}{\partial q_1} - 1 \right) = 0 \quad \rightarrow \quad \frac{\partial q_2}{\partial q_1} = 1$$

De otro lado: $P\left(q_1 \le Q \le q_2\right) = P\left(Q \le q_2\right) - P\left(Q \le q_1\right) = \gamma = 1-\alpha$

Entonces:

$$P\left(Q \le q_2\right) = \gamma + P\left(Q \le q_1\right)$$

$$F_Q\left(q_2\right) = \gamma + F_Q\left(q_1\right)$$

Derivando con respecto a q_1 : $f_Q\left(q_2\right)\underbrace{\frac{\partial q_2}{\partial q_1}}_{Uno} = f_Q\left(q_1\right) \rightarrow f_Q\left(q_2\right) = f_Q\left(q_1\right)$

De donde: $\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}q_2^2\right\} = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}q_1^2\right\} \rightarrow \exp\left\{\frac{1}{2}\left(q_1^2 - q_2^2\right)\right\} = 1 \rightarrow$

$\rightarrow q_1^2 - q_2^2 = 0 \rightarrow \left(q_1 - q_2\right)\left(q_1 + q_2\right) = 0 \rightarrow q_1 = q_2 \quad y \quad \underbrace{q_1 = -q_2}_{\text{Resultado útil}}$

Por lo tanto, el intervalo aleatorio de longitud mínima para μ será:

$$I = \left[\underbrace{\bar{X} - q_2\frac{\sigma}{\sqrt{n}}}_{T_1}, \underbrace{\bar{X} + q_2\frac{\sigma}{\sqrt{n}}}_{T_2}\right] \quad , \quad q_2 > 0$$

$$I = \left[\underbrace{\bar{X} - z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}}_{T_1}, \underbrace{\bar{X} + z_{\left(1-\frac{\alpha}{2}\right)}\frac{\sigma}{\sqrt{n}}}_{T_2}\right] \rightarrow L = \frac{2z_{\left(1-\frac{\alpha}{2}\right)}\sigma}{\sqrt{n}}$$

La longitud L se hace más pequeña si:

- Disminuye el nivel de confianza $1-\alpha$, ósea cuando aumenta α. Cuando disminuye $1-\alpha$ el intervalo será más pequeño y preciso.
- Aumenta el tamaño de muestra n. Graficando la función $\frac{1}{\sqrt{n}}$ se observa que para valores de n cercanos a 20, la disminución en la longitud, por cada incremento en una unidad de n, empieza a ser pequeña. Una recomendación heurística (buscar la solución de un problema con métodos no rigurosos, reglas empíricas) es considerar que un tamaño de muestra es grande si es mayor de 30.

NOTA: Ver el Script

Ejemplo 7: Sea $X_1,\ldots,X_n$ una m.a de una población $\text{Uniforme}\left(0,\theta\right)$.

a. Con el método de la cantidad pivotal construya un intervalo de confianza de 90% para θ.

Y_n es una estadística suficiente para θ.

Se sabe que $Q = \frac{Y_n}{\theta}$ es un pivote y que $f_Q\left(q\right) = nq^{n-1}I_{(0,1)}\left(q\right) \rightarrow F_Q\left(q\right) = q^n I_{(0,1)}\left(q\right)$.

Usando el método del pivote: $P\left(a \le \frac{Y_n}{\theta} \le b\right) = 0.90$ y un posible intervalo sería el siguiente:

$F_Q(a) = a^n = 0.05 \to a = \sqrt[n]{0.05}$

$F_Q(b) = b^n = 0.95 \to b = \sqrt[n]{0.95}$

$\therefore P\left(\sqrt[n]{0.05} \le \frac{Y_n}{\theta} \le \sqrt[n]{0.95}\right) = 0.90 \to P\left(\frac{Y_n}{\sqrt[n]{0.95}} \le \theta \le \frac{Y_n}{\sqrt[n]{0.05}}\right)$

Un posible intervalo para θ es:

$\left[\frac{Y_n}{\sqrt[n]{0.95}}; \frac{Y_n}{\sqrt[n]{0.05}}\right]$

b. Encuentre un posible intervalo de confianza del 90% para $\left(1 - \frac{1}{\theta}\right)$.

Utilizando el resultado de la pregunta a.

$P\left(\frac{Y_n}{\sqrt[n]{0.95}} \le \theta \le \frac{Y_n}{\sqrt[n]{0.05}}\right) = 0.90 \to P\left(1 - \frac{\sqrt[n]{0.95}}{Y_n} \le 1 - \frac{1}{\theta} \le 1 - \frac{\sqrt[n]{0.05}}{Y_n}\right) = 0.90$

Un posible intervalo de confianza de $\left(1 - \frac{1}{\theta}\right)$ es:

$\left[1 - \frac{\sqrt[n]{0.95}}{Y_n}; 1 - \frac{\sqrt[n]{0.05}}{Y_n}\right]$

c. Halle un intervalo de confianza de longitud mínima para θ con 90% de confianza.

Se sabe que $f_{Y_n}(y) = n\left[F(y)\right]^{n-1} f(y) = n\left[\frac{y}{\theta}\right]^{n-1}\left(\frac{1}{\theta}\right) = \frac{ny^{n-1}}{\theta^n} I_{(0,\theta)}(y)$.

El intervalo de longitud mínima estará limitado superiormente por θ.

$P\left[k_1(\theta) \le Y_n \le \theta\right] = 0.90$

$P\left[k_1(\theta) \le Y_n \le \theta\right] = \int_{k_1(\theta)}^{\theta} \frac{ny^{n-1}}{\theta^n} dy = \frac{y^n}{\theta^n}\Big|_{k_1(\theta)}^{\theta} = 1 - \left[\frac{k_1(\theta)}{\theta}\right]^n = 0.90$

$\to k_1(\theta) = \theta\sqrt[n]{0.10}$

$P\left[k_1(\theta) \le Y_n \le \theta\right] = P\left[\theta\sqrt[n]{0.10} \le Y_n \le \theta\right] = P\left[Y_n \le \theta \le \frac{Y_n}{\sqrt[n]{0.10}}\right]$

Ejemplo 8: Sea $X_1, \ldots, X_n$ una muestra aleatoria de $f(x) = \frac{3x^2}{\theta^3} I_{(0,\theta)}(x)$.

a. Con el método del pivote halle un posible intervalo de 90% de confianza para θ.

Y_n es una estadística suficiente para θ.

Se demuestra que $f_{Y_n}(y)=\frac{3n}{\theta^{3n}}y^{3n-1}I_{(0,\theta)}(y)$ y que $F_{Y_n}(y)=T=\left(\frac{Y_n}{\theta}\right)^{3n}\sim \text{Uniforme}(0,1)$ es un pivote pues su distribución no depende de θ.

Usando el método del pivote: $P\left[a\le\left(\frac{Y_n}{\theta}\right)^{3n}\le b\right]=0.90$ y un posible intervalo sería el siguiente:

$$P(T\le a)=\int_0^a 1dt=a=0.05$$

$$P(T\le b)=\int_0^b 1dt=b=0.95$$

Entonces:

$$P\left[0.05\le\left(\frac{Y_n}{\theta}\right)^{3n}\le 0.95\right]=0.90\rightarrow P\left[\frac{Y_n}{0.95^{\frac{1}{3n}}}\le\theta\le\frac{Y_n}{0.05^{\frac{1}{3n}}}\right]$$

Un intervalo de confianza para θ será:

$$\left[\frac{Y_n}{0.95^{\frac{1}{3n}}};\frac{Y_n}{0.05^{\frac{1}{3n}}}\right]$$

b. Si se considera $T=\frac{Y_n}{\theta}\rightarrow f_T(t)=3nt^{3n-1}I_{(0,1)}(t)\rightarrow T$ es un pivote porque su distribución no depende de θ.

Usando el método del pivote: $P\left[a\le\frac{Y_n}{\theta}\le b\right]=0.90$ y un posible intervalo sería el siguiente:

$$P(T\le a)=\int_0^a 3nt^{3n-1}dt=a^{3n}=0.05\rightarrow a=0.05^{\frac{1}{3n}}$$

$$P(T\le b)=\int_0^b 3nt^{3n-1}dt=b^{3n}=0.95\rightarrow b=0.95^{\frac{1}{3n}}$$

Entonces:

$$P\left[0.05^{\frac{1}{3n}}\le\frac{Y_n}{\theta}\le 0.95^{\frac{1}{3n}}\right]=0.90\rightarrow P\left[\frac{Y_n}{0.95^{\frac{1}{3n}}}\le\theta\le\frac{Y_n}{0.05^{\frac{1}{3n}}}\right]$$

Un intervalo de confianza para θ será:

$$\left[\frac{Y_n}{0.95^{\frac{1}{3n}}};\frac{Y_n}{0.05^{\frac{1}{3n}}}\right]$$

c. Halle un intervalo de longitud mínima para θ con 90% de confianza.

$F(x) = \left(\frac{x}{\theta}\right)^3 I_{[0,\theta)}(x)$ entonces:

Se sabe que $f_{Y_n}(y) = n\left[F(y)\right]^{n-1} f(y) = n\left[\left(\frac{y}{\theta}\right)^3\right]^{n-1}\left(\frac{3y^2}{\theta^3}\right) = \frac{3ny^{3n-1}}{\theta^{3n}} I_{(0,\theta)}(y)$.

El intervalo de longitud mínima estará limitado superiormente por θ.

$P\left[k_1(\theta) \le Y_n \le \theta\right] = 0.90$

$P\left[k_1(\theta) \le Y_n \le \theta\right] = \int_{k_1(\theta)}^{\theta} \frac{3ny^{3n-1}}{\theta^{3n}} dy = \frac{y^{3n}}{\theta^{3n}}\Big|_{k_1(\theta)}^{\theta} = 1 - \left[\frac{k_1(\theta)}{\theta}\right]^{3n} = 0.90$

$\rightarrow k_1(\theta) = \theta\sqrt[3n]{0.10}$

$P\left[k_1(\theta) \le Y_n \le \theta\right] = P\left[\theta\sqrt[3n]{0.10} \le Y_n \le \theta\right] = P\left[Y_n \le \theta \le \frac{Y_n}{\sqrt[3n]{0.10}}\right]$

Ejemplo 9: Sea $X_1, \ldots, X_n$ una muestra aleatoria de $f(x) = e^{\theta - x} I_{(\theta,\infty)}(x)$.

a. Con el método del pivote halle un posible intervalo de 90% de confianza para θ.

Y_1 es una estadística suficiente para θ.

Se demuestra que $f_{Y_1}(y) = ne^{-n(y-\theta)} I_{(\theta,\infty)}(y)$ y que $T = Y_1 - \theta \sim$ Exponencial $\left(\text{con media } \frac{1}{n}\right)$ es un pivote pues su distribución no depende de θ.

$\rightarrow F_T(t) = \left(1 - e^{-nt}\right) I_{[0,\infty)}(t)$

Usando el método del pivote: $P\left[a \le Y_1 - \theta \le b\right] = 0.90$ y un posible intervalo sería el siguiente:

$P(T \le a) = F_T(a) = 1 - e^{-na} = 0.05 \rightarrow a = -\frac{\ln 0.95}{n}$

$P(T \le b) = F_T(b) = 1 - e^{-nb} = 0.95 \rightarrow b = -\frac{\ln 0.05}{n}$

Entonces:

$P\left[-\frac{\ln 0.95}{n} \le Y_1 - \theta \le -\frac{\ln 0.05}{n}\right] = 0.90 \rightarrow P\left[Y_1 + \frac{\ln 0.05}{n} \le \theta \le Y_1 + \frac{\ln 0.95}{n}\right]$

Un intervalo de confianza para θ será:

$\left[Y_1 + \frac{\ln 0.05}{n} ; Y_1 + \frac{\ln 0.95}{n}\right]$

b. Determine un intervalo de confianza del 90% para $e^{-\theta}$.

Respuesta: $\left[e^{-\left(Y_1+\frac{\ln 0.95}{n}\right)} ; e^{-\left(Y_1+\frac{\ln 0.05}{n}\right)} \right]$

c. Halle un intervalo de longitud mínima para θ con 90% de confianza.

$f_{Y_1}(y)=ne^{-n(y-\theta)}I_{(\theta,\infty)}(y)$

El intervalo de longitud mínima estará limitado inferiormente por θ.

$P\left[\theta \le Y_1 \le k_2(\theta)\right] = 0.90$

$$P\left[\theta \le Y_1 \le k_2(\theta)\right] = \int_{\theta}^{k_2(\theta)} ne^{n(\theta-y)}dy = -e^{n(\theta-y)}\Big|_{\theta}^{k_2(\theta)} = 1-e^{n(\theta-k_2(\theta))} = 0.90$$

$$\rightarrow k_2(\theta) = \theta - \frac{\ln 0.1}{n}$$

$$P\left[\theta \le Y_1 \le k_2(\theta)\right] = P\left[\theta \le Y_1 \le \theta - \frac{\ln 0.1}{n}\right] = P\left[Y_1 + \frac{\ln 0.1}{n} \le \theta \le Y_1\right]$$

Ejemplo 10: Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución de mínimos de Gumbel que tiene la siguiente densidad de probabilidades: $f(x) = \alpha \exp\left[\alpha(x-\mu)\right]\exp\left\{-\exp\left[\alpha(x-\mu)\right]\right\} I_{(-\infty,\infty)}(x),\ \mu \in R,\ \alpha > 0.\ \alpha$ es conocida.

a. Con el método del pivote halle un posible intervalo de 90% de confianza para μ.

Se puede demostrar $S = \exp(\alpha x_i) \sim \text{Exponencial}\left[\text{con media } \exp(\alpha\mu)\right]$ y que $T = \sum_{i=1}^{n} d(x_i) = \sum_{i=1}^{n}\exp(\alpha x_i)$ es una estadística suficiente y completa y que $T \sim \text{Gamma}\left[n, \exp(-\alpha\mu)\right]$, por lo tanto $Q = 2T\exp(-\alpha\mu) \sim \chi^2_{(2n)}$ es un pivote porque su distribución no depende de μ.

Usando el método del pivote: $P\left[a \le 2T\exp(-\alpha\mu) \le b\right] = 0.90$ y un posible intervalo sería el siguiente:

$P(Q \le a) = F_Q(a) \rightarrow a = \chi^2(0.05, 2n)$

$P(Q \le b) = F_Q(b) \rightarrow b = \chi^2(0.95, 2n)$

De $P\left[a \le 2T\exp(-\alpha\mu) \le b\right] = 0.90$ se obtiene:

$$P\left[-\frac{\ln\left(\frac{b}{2T}\right)}{\alpha} \le \mu \le -\frac{\ln\left(\frac{a}{2T}\right)}{\alpha} \right] = 0.90$$

b. Encuentre un intervalo de 90% de confianza para $\exp(\mu)$.

$$\left[\exp\left[-\frac{\ln\left(\frac{b}{2T}\right)}{\alpha}\right];\exp\left[-\frac{\ln\left(\frac{a}{2T}\right)}{\alpha}\right]\right]$$

c. Usando un pivote que se base en el mínimo $Y_1^* = \min\left[\exp(\alpha x_1),\ldots,\exp(\alpha x_n)\right]$ halle un intervalo de confianza de 90% para μ.

$$S = \exp(\alpha x_i) \sim \text{Exponencial}\left[\text{con media } \exp(\alpha\mu)\right] \rightarrow$$
$$F_S(s) = 1 - \exp\left\{-\left[\exp(-\alpha\mu)\right]s\right\}$$
$$f_{Y_1^*}(y) = nf_S(y)\left[1-F_S(y)\right]^{n-1} = n\exp(-\alpha\mu)\left\{\exp\left[\left(-n\exp(-\alpha\mu)\right)y\right]\right\}I_{(0,\infty)}(y)$$
$$Y_1^* \sim \text{Exponencial}\left[\text{con media } \frac{\exp(\alpha\mu)}{n}\right] \rightarrow$$
$$F_{Y_1^*}(y) = 1 - \exp\left[\left(-n\exp(-\alpha\mu)\right)y\right] \sim \text{Uniforme}(0,1) \rightarrow F_{Y_1^*}(y) \text{ es pivote.}$$

De: $P\left\{a \le 1 - \exp\left[\left(-n\exp(-\alpha\mu)\right)y\right] \le b\right\}$ se obtiene:

$$P\left\{-\frac{\ln\left[-\frac{\ln(1-b)}{ny}\right]}{\alpha} \le \mu \le -\frac{\ln\left[-\frac{\ln(1-a)}{ny}\right]}{\alpha}\right\} = 0.90$$

donde: $a = 0.05$ y $b = 0.95$

d. Encuentre un intervalo de 90% de confianza para la mediana de la distribución.

Sea M la mediana entonces:

$$P(X < M) = \int_{-\infty}^{M} \alpha\exp\left[\alpha(x-\mu)\right]\exp\left\{-\exp\left[\alpha(x-\mu)\right]\right\} =$$
$$= -\exp\left\{-\exp\left[\alpha(x-\mu)\right]\right\}\Big|_{-\infty}^{M} = 0.5$$

Por lo tanto:

$$M = \frac{\ln\left[-\ln(0.5)\right]}{\alpha} + \mu \rightarrow$$

de acuerdo con la pregunta a:

$$P\left[-\frac{\ln\left(\frac{b}{2T}\right)}{\alpha}+\frac{\ln\left[-\ln(0.5)\right]}{\alpha}\leq M\leq-\frac{\ln\left(\frac{a}{2T}\right)}{\alpha}+\frac{\ln\left[-\ln(0.5)\right]}{\alpha}\right]=0.90$$

<u>Observaciones</u>

Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución continua (esto implica que su distribución acumulada $F(x)=F(x;\theta)$ es continua en x) entonces:

1. Para una muestra de tamaño n=1, $Y=F(x;\theta)\sim \text{Uniforme}(0,1)$, entonces Y es un pivote.

Prueba

$$Y=F(x)\rightarrow\frac{dY}{dx}=f(x)\rightarrow J=\frac{dx}{dY}=\frac{1}{f(x)}\text{ pero}$$

$x=F^{-1}(Y)\rightarrow$ usando el jacobiano:

$$f_Y(y)=f\left(x=F^{-1}(Y)\right)\left|\frac{1}{f\left(x=F^{-1}(Y)\right)}\right|=1I_{(0,1)}(y)\rightarrow$$

$$f_Y(y)\sim\text{Uniforme}(0,1)$$

2. Para una muestra de tamaño n=1, $U=-\ln F(x)=-\ln Y\sim\text{Exponencial}(\text{con media }1)$, entonces U es un pivote.

Prueba

$$F_U(u)=P(U\leq u)=P(-\ln Y\leq u)=P(\ln Y\geq -u)=$$

$$F_U(u)=P\left(Y\geq e^{-u}\right)=\int_{e^{-u}}^{1}1dy=1-e^{-u}\rightarrow$$

$$f_U(u)=e^{-u}I_{(0,\infty)}(u)$$

3. Para una muestra de tamaño n: $Q=-\sum_{i=1}^{n}\ln F(x_i)=-\sum_{i=1}^{n}\ln Y_i\sim\text{Gamma}(\text{n},1)$, entonces:

Q es un pivote porque no depende de θ. También:

$$Q_1=2\left(-\sum_{i=1}^{n}\ln F(x_i)\right)(1)\sim\chi^2(2n)\text{ es un pivote.}$$

Ejemplo 11: Sea $X_1,\ldots,X_n$ una muestra aleatoria de la siguiente distribución $f(x;\theta)=\theta x^{\theta-1}I_{(0,1)}(x)$.

a. Para una muestra de tamaño n=1 se cumple que $Y = F(x;\theta) = x^{\theta} I_{(0,1)}(x) \sim \text{Uniforme}(0,1)$, entonces Y es un pivote.

b. Para una muestra de tamaño n=1 se cumple que $U = -\ln F(x;\theta) = -\ln Y = -\theta \ln x \sim \text{Exponencial}(1)$, entonces U es un pivote.

c. Para una muestra de tamaño n: $Q = -\sum_{i=1}^{n} \ln F(x_i;\theta) = -\sum_{i=1}^{n} \ln Y_i$

$Q = -\sum_{i=1}^{n} \theta \ln x_i \sim \text{Gamma}(\text{n},1)$, entonces:

Q es un pivote porque no depende de θ. También:

$Q_1 = -2\left(\sum_{i=1}^{n} \theta \ln x_i\right)(1) \sim \chi^2(2n)$ es un pivote.

Ejemplo 12: Sea $X_1,\ldots,X_n$ una muestra aleatoria de la siguiente distribución $f(x;\theta) = \dfrac{2x}{\theta^2} I_{(0,\theta)}(x)$.

a. Con la muestra unitaria x=1.5 encuentre un intervalo del 98% de confianza para $P\left(X > \sqrt{\theta}\right)$. No use la estadística suficiente.

$$F(x) = P(X \le x) = \int_0^x \frac{2t}{\theta^2}\,dt = \frac{x^2}{\theta^2} I_{[0,\theta)}(x) \sim \text{Uniforme}(0,1)$$

$$P\left(X > \sqrt{\theta}\right) = 1 - P\left(X \le \sqrt{\theta}\right) = 1 - \frac{\left(\sqrt{\theta}\right)^2}{\theta^2} = 1 - \frac{1}{\theta}$$

$$P\left(a \le \frac{x^2}{\theta^2} \le b\right) = P\left(0.01 \le \frac{1.5^2}{\theta^2} \le 0.99\right) = 0.98 \rightarrow$$

$$P\left(0.33668 \le 1 - \frac{1}{\theta} \le 0.93333\right) = 0.98$$

$$\text{I. de Conf.}\left(1 - \frac{1}{\theta}\right) = [0.33668; 0.93333]$$

b. Con la muestra de tamaño cinco: 1.2, 1.5, 1.1, 1.3 y 1.4; construya un intervalo de 90% de confianza para θ. No use la estadística suficiente.

Se sabe que $F(x) = P(X \le x) = \dfrac{x^2}{\theta^2} I_{[0,\theta)}(x) \sim \text{Uniforme}(0,1) \rightarrow$

$$-\ln\frac{x^2}{\theta^2}\sim \text{Exponencial}(1)\rightarrow -\sum_{i=1}^{5}\ln\frac{x_i^2}{\theta^2}\sim \text{Gamma}(5,1)\rightarrow$$

$$2\left[-\sum_{i=1}^{5}\ln\frac{x_i^2}{\theta^2}\right](1)\sim\chi^2(10)$$

$$P\left\{a\le 2\left[-\sum_{i=1}^{5}\ln\frac{x_i^2}{\theta^2}\right](1)\le b\right\}=P\left[a\le\chi^2(10)\le b\right]=0.90$$

$$a=\chi^2(0.05;10)=3.94$$

$$b=\chi^2(0.95;10)=18.307$$

$$P\left\{a\le 2\left[-\sum_{i=1}^{5}\ln\frac{x_i^2}{\theta^2}\right](1)\le b\right\}=P\left[a\le -2\sum_{i=1}^{5}\left(\ln x_i^2-2\ln\theta\right)\le b\right]=$$

$$=P\left[\exp\left(\frac{a+4\sum_{i=1}^{5}\ln x_i}{20}\right)\le\theta\le\exp\left(\frac{b+4\sum_{i=1}^{5}\ln x_i}{20}\right)\right]$$

Reemplazando los valores de $a,\ b$ y $\sum_{i=1}^{5}\ln x_i=1.28193$

en la probabilidad anterior se obtiene:

$$P(1.57363\le\theta\le 3.22759)$$

c. Con la muestra de tamaño 5 de la subpregunta b y usando un pivote asociado a la estadística suficiente halle un intervalo de 90% de confianza para θ.

Y_n es la estadìstica suficiente para θ.

Se demuestra que $f_{Y_n}(y)=\dfrac{10y^9}{\theta^{10}}I_{(0,\theta)}(y)\rightarrow$

$$F_{Y_n}(y)=\frac{y^{10}}{\theta^{10}}I_{[0,\theta)}(y)\sim \text{Uniforme}(0,1)$$

$$P\left(a\le\frac{y^{10}}{\theta^{10}}\le b\right)=P\left(0.05\le\frac{1.5^{10}}{\theta^{10}}\le 0.95\right)=0.90\rightarrow$$

$$P(1.5077\le\theta\le 2.0239)=0.90$$

$$\text{I. de Conf.}(\theta)=[1.50771;2.02392]$$

Este intervalo es más preciso, tiene menor longitud, que
el de la pregunta b ya que se utiliza una estadìstica suficiente.

Nota: Se llega al mismo resultado utilizando el pivote $Q = \frac{Y_n}{\theta}$ pues este pivote tiene como densidad $f_Q(q) = 10q^9 I_{(0,1)}(q)$

Método estadístico para obtener intervalos aleatorios

Sea $X_1, \dots, X_n$ una m.a de la distribución $f(x;\theta) \quad \forall \theta \in \Theta$. Sea T una estadística que puede ser elegida teniendo en cuenta lo siguiente:

1. Si existe una estadística suficiente para θ entonces esta es elegida.
2. Si no existe una estadística suficiente para θ entonces el estimador máximo verosímil puede ser elegido.
3. La elección de T puede depender de la facilidad con la cual pueden ser ejecutadas las operaciones que se necesitan para obtener el intervalo aleatorio. Una de estas operaciones puede ser la obtención de la densidad de T.

Si X es una variable aleatoria continua

Sea $f_T(t;\theta)$ la densidad de T y $k_1(\theta), k_2(\theta)$ dos funciones monótonas crecientes tales que:

$$\int_{-\infty}^{k_1(\theta)} f_T(t;\theta)\,dt = p_1 \quad , \quad \int_{k_2(\theta)}^{\infty} f_T(t;\theta)\,dt = p_2$$

Siendo $p_1 > 0$, $p_2 > 0$, $p_1 + p_2 < 1 \quad y \quad k_1(\theta) < k_2(\theta)$, además $p_1 \ y \ p_2$ son dos números fijados.

Si X es una variable aleatoria discreta la integral se cambia por sumatoria.

Por lo expuesto se tiene que:

$P\left(k_1(\theta) \le T \le k_2(\theta)\right) = 1 - p_1 - p_2$, invirtiendo se obtiene:

$P\left(v_1(X_1, \dots, X_n) \le \theta \le v_2(X_1, \dots, X_n)\right) = 1 - p_1 - p_2$

Luego el intervalo $[v_1, v_2]$ es el intervalo aleatorio para θ con $(1 - p_1 - p_2)$ de probabilidad.

Ejemplo 13: Sea $X_1,\ldots,X_n$ una muestra aleatoria de la siguiente distribución $f(x;\theta)=\frac{1}{\theta}I_{(0,\theta)}(x)$. Utilizando el método estadístico determine un intervalo de $100(1-p_1-p_2)\%$ para θ.

Y_n es una estadística suficiente para θ.

Se sabe que $f_{Y_n}(y)=n\left[F(y)\right]^{n-1}f(y)=n\left[\frac{y}{\theta}\right]^{n-1}\left(\frac{1}{\theta}\right)=\frac{ny^{n-1}}{\theta^n}I_{(0,\theta)}(y)$.

$$p_1=\int_0^{k_1(\theta)}\frac{ny^{n-1}}{\theta^n}dy=\left[\frac{k_1(\theta)}{\theta}\right]^n\rightarrow k_1(\theta)=\theta\sqrt[n]{p_1}$$

$$p_2=\int_{k_2(\theta)}^{\theta}\frac{ny^{n-1}}{\theta^n}dy=1-\left[\frac{k_2(\theta)}{\theta}\right]^n\rightarrow k_2(\theta)=\theta\sqrt[n]{1-p_2}$$

$$P\left[k_1(\theta)\le Y_n\le k_2(\theta)\right]=P\left[\theta\sqrt[n]{p_1}\le Y_n\le\theta\sqrt[n]{1-p_2}\right]=1-p_1-p_2\rightarrow$$

$$P\left[\frac{Y_n}{\sqrt[n]{1-p_2}}\le\theta\le\frac{Y_n}{\sqrt[n]{p_1}}\right]=1-p_1-p_2$$

$$\text{Int. de Conf.}(\theta)=\left[\frac{Y_n}{\sqrt[n]{1-p_2}};\frac{Y_n}{\sqrt[n]{p_1}}\right]$$

Ejemplo 14: En el distrito de Lince se estudia si una familia tiene mascota. Si 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1 es una muestra aleatoria de 20 familias de esa distribución $\text{Binomial}(1,p)$ y considerando $p_1=0.05095$ y $p_2=0.015902$, con el método estadístico determine un intervalo de 93.3148% de confianza para p.

$X_i\sim\text{Binomial}(n=1,p)$

$T=\sum_{i=1}^{n}X_i$ es una estadística suficiente de *p*.

$$t_0=\sum_{i=1}^{20}X_i=4$$

$$T=\sum_{i=1}^{20}X_i\sim\text{Binomial}(n=20,p)\rightarrow f_T(t)=\binom{20}{t}p^t(1-p)^{20-t}I_{\{0,1,\ldots,20\}}(t)$$

$$p_1=0.05095=\sum_{t=0}^{t_0=k_1(p)}f_T(t)=\sum_{t=0}^{4}\binom{20}{t}p^t(1-p)^{20-t}\rightarrow p=0.4$$

$$p_2=0.015902=\sum_{t=t_0=k_2(p)}^{20}f_T(t)=\sum_{t=4}^{20}\binom{20}{t}p^t(1-p)^{20-t}\rightarrow p=0.05$$

$$\text{Int. de Conf.}(p)=[0.05,0.4]$$

Observación 1: Si $X \sim \text{Binomial}(n,\pi)$ entonces $Y \sim \text{Beta}(x+1,n-x)$ por lo tanto se cumple que $P(X \geq x+1) = P(Y < \pi)$. La demostración se hace integrando por partes.

Como ejemplo de esta observación 1 consideremos lo siguiente: Si $X \sim \text{Binomial}(n=8,\pi=0.25)$ calcule de dos maneras $P(X \geq 4)$.

Solución

Por la observación se sabe que $Y \sim \text{Beta}(x+1,8-x) \sim \text{Beta}(3+1,8-3) \sim \text{Beta}(4,5)$.

Como $P(X \geq x+1) = P(Y < \pi) \rightarrow P(X \geq 4) = P(Y < 0.25) = 0.1138153076$.

Observación 2: Si $X \sim F(m,n)$. Se verifica que $Y = \dfrac{\frac{m}{n}X}{1+\frac{m}{n}X} \sim \text{Beta}\left(\dfrac{m}{2},\dfrac{n}{2}\right)$. La demostración se puede hacer con el jacobiano.

Como ejemplo de esta observación 2 consideremos lo siguiente: Si $X \sim F(m=8,n=25)$ calcule de dos maneras $P(X > 1.88)$.

Solución

Primera manera

$P(X > 1.88) = 0.10877$

Segunda manera

$$Y = \frac{\frac{m}{n}X}{1+\frac{m}{n}X} \sim \text{Beta}\left(\frac{m}{2},\frac{n}{2}\right) = \text{Beta}\left(\frac{8}{2},\frac{25}{2}\right) = \text{Beta}(4,12.5)$$

$$P(X > 1.88) = P\left(\frac{\frac{m}{n}X}{1+\frac{m}{n}X} > \frac{\frac{8}{25}\times 1.88}{1+\frac{8}{25}\times 1.88}\right) = P(Y > 0.3756243756) = 0.1087699622$$

Observación 3: Una generalización del ejemplo 14 es la siguiente, dada la observación y de $Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n,p)$, entonces:

$$\frac{\alpha}{2} = \sum_{k=0}^{y} \binom{n}{k} p^k (1-p)^{n-k}$$

$$\frac{\alpha}{2}=\sum_{k=y}^{n}\binom{n}{k}p^k\left(1-p\right)^{n-k}$$

Considerando las observaciones 1 y 2 y teniendo en cuenta las siguientes distribuciones F:

$F \sim F\left[2\left(y+1\right),2\left(n-y\right)\right]$

$F_1 \sim F\left[2y,2\left(n-y+1\right)\right]$

Se obtiene el siguiente intervalo aleatorio de $100\left(1-\alpha\right)\%$ de confianza para p :

$$\left\{\frac{\left[\dfrac{y}{n-y+1}\right]F\left[\dfrac{\alpha}{2},2y,2\left(n-y+1\right)\right]}{1+\left[\dfrac{y}{n-y+1}\right]F\left[\dfrac{\alpha}{2},2y,2\left(n-y+1\right)\right]},\frac{\left[\dfrac{y+1}{n-y}\right]F\left[1-\dfrac{\alpha}{2},2\left(y+1\right),2\left(n-y\right)\right]}{1+\left[\dfrac{y+1}{n-y}\right]F\left[1-\dfrac{\alpha}{2},2\left(y+1\right),2\left(n-y\right)\right]}\right\}=\left[\text{Li,Ls}\right]$$

Con los datos del ejemplo 14 el intervalo de 95% de confianza con $\frac{\alpha}{2}=0.025$ en cada cola es $\left[0.05733398925, 0.4366140051\right]$.

Ejemplo 15: En la fábrica de TV Sony se estudia el número de TV defectuosos fabricados por día. Sea 1, 0, 0, 0, 2, 2, 0, 0, 0, 1 una muestra aleatoria de 10 días de esa distribución que es una $\text{Poisson}\left(\lambda\right)$. Considerando $p_1=0.050066054$ y $p_2=0.050521526$ y con el método estadístico determine un intervalo de 89.941242% de confianza para λ.

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

$T=\sum_{i=1}^{n}X_i$ es una estadística suficiente de λ.

$$t_0=\sum_{i=1}^{10}X_i=6$$

$$T=\sum_{i=1}^{10}X_i \sim \text{Poisson}\left(10\lambda\right)\rightarrow f_T\left(t\right)=\frac{e^{-10\lambda}\left(10\lambda\right)^t}{t!}I_{\{0,1,\ldots\}}\left(t\right)$$

$$p_1=0.050066054=\sum_{t=0}^{t_0=k_1(\lambda)}f_T\left(t\right)=\sum_{t=0}^{6}\frac{e^{-10\lambda}\left(10\lambda\right)^t}{t!}\rightarrow\lambda=1.184$$

$$p_2=0.050521526=\sum_{t=t_0=k_2(\lambda)}^{\infty}f_T\left(t\right)=\sum_{t=6}^{\infty}\frac{e^{-10\lambda}\left(10\lambda\right)^t}{t!}\rightarrow\lambda=0.262$$

$\text{Int. de Conf.}\left(\lambda\right)=\left[0.262,1.184\right]$

Ejemplo 16: Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución $\text{Poisson}(\lambda)$. Usando el método estadístico y la relación que hay entre las distribuciones Poisson y Gamma, halle un intervalo de $100(1-p_1-p_2)\%$ de confianza para λ.

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

$T=\sum_{i=1}^{n} X_i$ es una estadística suficiente de λ.

$$t_0=\sum_{i=1}^{n} X_i$$

$$T=\sum_{i=1}^{n} X_i \sim \text{Poisson}(n\lambda) \rightarrow f_T(t)=\frac{e^{-n\lambda}(n\lambda)^t}{t!} I_{\{0,1,\ldots\}}(t)$$

$$\text{Int. de Conf.}(\lambda)=\left\{\frac{1}{2n}\chi^2(2t_0,p_2);\frac{1}{2n}\chi^2\left[2(t_0+1),1-p_1\right]\right\}$$

Ejemplo 17: En la fábrica de TV Sony se estudia el número de TV defectuosos fabricados por día. Sea 1, 0, 0, 0, 2, 2, 0, 0, 0, 1 una muestra aleatoria de 10 días de esa distribución que es una $\text{Poisson}(\lambda)$. Considerando $p_1=0.050066054$ y $p_2=0.050521526$ y utilizando la relación que hay entre las distribuciones Poisson y Gamma determine un intervalo de 89.941242% de confianza para λ.

$$t_0=\sum_{i=1}^{10} X_i=6 \ , \ n=10$$

$$\text{Int. de Conf.}(\lambda)=\left\{\frac{1}{2n}\chi^2(2t_0,p_2);\frac{1}{2n}\chi^2\left[2(t_0+1),1-p_1\right]\right\}$$

$$\text{Int. de Conf.}(\lambda)=\left\{\frac{1}{20}\chi^2(12,0.050521526);\frac{1}{20}\chi^2\left[2(6+1),1-0.050066054\right]\right\}$$

$$\text{Int. de Conf.}(\lambda)=\left[\frac{1}{20}\chi^2(12,0.050521526);\frac{1}{20}\chi^2(14,0.949933946)\right]$$

$$\text{Int. de Conf.}(\lambda)=\left[\frac{5.24}{20};\frac{23.68}{20}\right]=[0.262;1.184]$$

Se llega a la misma respuesta que se halló en la pregunta 15.

Ejemplo 18: El investigador A estudió entre otras la variable X definida como el número de vicuñas hembras analizadas hasta obtener la tercera con un contenido de hemoglobina mayor de 15.8 g/DL (gramos por decilitro). Si 7, 8, 9, 6 y 5 son las observaciones de una muestra aleatoria de la distribución en estudio, determine un intervalo de confianza de $100(1-p_1-p_2)\%$ para la proporción de vicuñas hembras con un contenido de hemoglobina mayor de 15.8 g/DL si $p_1=0.0730689545$ y $p_2=0.0204540062$.

Solución

$X_i \sim \text{Pascal}(k=3,\pi)$, $x_i = 3,4,5,6,\ldots$

$T=\sum_{i=1}^{n} X_i$ es una estadística suficiente de π.

$t_0=\sum_{i=1}^{5} X_i = 35$

$$T=\sum_{i=1}^{5} X_i \sim \text{Pascal}(nk=15,\pi) \rightarrow f_T(t)=\binom{t-1}{15-1}\pi^{15}(1-\pi)^{t-15} I_{\{15,16,17,18,\ldots\}}(t)$$

$$p_1 = 0.0730689545 = \sum_{t=15}^{t_0=35}\binom{t-1}{15-1}\pi^{15}(1-\pi)^{t-15} \rightarrow \pi = 0.3$$

$$p_2 = 0.0204540062 = \sum_{t=t_0=35}^{\infty}\binom{t-1}{15-1}\pi^{15}(1-\pi)^{t-15} = 1-\sum_{t=15}^{34}\binom{t-1}{15-1}\pi^{15}(1-\pi)^{t-15} =$$

$$p_2 = 0.0204540062 = 1-0.9795459938 \rightarrow \pi = 0.6$$

Respuesta: $[0.3, 0.6]$

Intervalo de confianza para el parámetro θ de una población con muestras pequeñas y $ECM(\theta)$ conocido

Según Chebichev se establece que $P\left(\left|\theta-\theta\right| \le k\sqrt{ECM(\theta)}\right) \ge 1-\frac{1}{k^2}$. Utilizando la definición de intervalos se obtiene. $P\left(\left|\theta-\theta\right| \le k\sqrt{ECM(\theta)}\right) \ge 1-\alpha$.

Donde la cota inferior del nivel de confianza $1-\alpha$ es tal que $1-\alpha = 1-\frac{1}{k^2} \Rightarrow k=\frac{1}{\sqrt{\alpha}}$ derivándose el intervalo $\left[\theta-\frac{1}{\sqrt{\alpha}}\sqrt{ECM(\theta)};\theta+\frac{1}{\sqrt{\alpha}}\sqrt{ECM(\theta)}\right]$.

Ejemplo 18: Se tiene una población de 4 elementos y se ha estudiado la variable X que toma los siguientes valores $\{4;5;5.5;6\}$. Si se extraen todas las **muestras aleatorias simples** de tamaño 2.

a. Obtenga todos los intervalos de por lo menos 95% de confianza para la media poblacional si se utiliza la media aritmética muestral como estimador.

$$f(x) = \frac{1}{4} I_{\{4;5;5.5;6\}}(x) \rightarrow E(X) = \mu = 5.125, \; Var(X) = \sigma^2 = 0.546875$$

$$E(\bar{X}) = \mu = 5.125, \; Var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{0.546875}{2}\left(\frac{4-2}{4-1}\right) = 0.18229$$

$$ECM(\bar{X}) = Var(\bar{X}) + b^2(\bar{X}) = 0.18229 + 0 = 0.18229$$

$$IC(\mu) = \bar{X} \pm \frac{1}{\sqrt{\alpha}}\sqrt{ECM(\bar{X})} = \bar{X} \pm \frac{1}{\sqrt{0.05}}\sqrt{0.18229}$$

Muestra	$\bar{X}$	Probab.	$IC(\mu)$
(4;5)	4.5	1/6	[2.59060,6.40940]
(4;5.5)	4.75	1/6	[2.84060,6.65940]
(4;6)	5	1/6	[3.09060,6.90940]
(5;5.5)	5.25	1/6	[3.34060,7.15940]
(5;6)	5.5	1/6	[3.59060,7.40940]
(5.5;6)	5.75	1/6	[3.84060,7.65940]

b. Calcule todos los intervalos de por lo menos 95% de confianza para la media poblacional si se utiliza la media geométrica muestral como estimador.

$$f(x) = \frac{1}{4} I_{\{4;5;5.5;6\}}(x) \rightarrow E(X) = \mu = 5.125, \; Var(X) = \sigma^2 = 0.546875$$

$$E(\bar{X}_G) = \frac{1}{6}(4.47214 + \ldots + 5.74456) = 5.087895, \; Var(\bar{X}_G) = 0.196667$$

$$b(\bar{X}_G) = E(\bar{X}_G) - \mu = 5.087895 - 5.125 = -0.037105$$

$$ECM(\bar{X}_G) = Var(\bar{X}_G) + \left[b(\bar{X}_G)\right]^2 = 0.19804$$

$$IC(\mu) = \bar{X}_G \pm \frac{1}{\sqrt{\alpha}}\sqrt{ECM(\bar{X}_G)} = \bar{X}_G \pm \frac{1}{\sqrt{0.05}}\sqrt{0.19804}$$

Muestra	$\bar{X}_G$	Probab.	$IC(\mu)$
(4;5)	4.47214	1/6	[2.48196,6.46232]
(4;5.5)	4.69042	1/6	[2.70024,6.68060]
(4;6)	4.89898	1/6	[2.90880,6.88916]
(5;5.5)	5.24404	1/6	[3.25386,7.23422]
(5;6)	5.47723	1/6	[3.48705,7.46741]
(5.5;6)	5.74456	1/6	[3.75438,7.73474]

Si se comparan los intervalos de las subpreguntas anteriores se concluye que $\bar{X}$ es más preciso que $\bar{X}_G$ (con $\bar{X}$ se obtienen intervalos de menor longitud). Esto se debe a las propiedades que tiene $\bar{X}$.

c. Halle todos los intervalos de por lo menos 98% de confianza para la media armónica poblacional utilizando la media geométrica muestral, la media armónica muestral, la media aritmética muestral y el máximo de la muestra como estimadores (Queda como ejercicio).

d. Obtenga los intervalos de confianza de por lo menos 98% de confianza para la variancia poblacional utilizando $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ y después $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$. Diga qué estimador es más preciso (Queda como ejercicio).

Ejercicios

1. Si X_1, X_2 es una muestra aleatoria de una distribución $Unif\left(10\theta - 0.5, 10\theta + 0.5\right)$.
   a. Calcule el coeficiente de confianza en la siguiente igualdad $P\left(0.1Y_1 - 0.01 \leq \theta \leq 0.1Y_2 + 0.01\right) = \gamma = 1 - \alpha$.
   b. Obtenga la longitud esperada en la pregunta a).
2. El Quitosano se encuentra en el exoesqueleto ("cáscara") de crustáceos e insectos. Esta sustancia, se descubrió en el año 1859. Se puede usar en agricultura como fungicida y en la industria vitivinícola para evitar el deterioro del vino. En medicina se usa a veces como añadido en vendajes para reducir el sangrado y disminuir la cantidad de infecciones. Según la investigadora Florence Nightingale el porcentaje de quitosano X extraído con respecto al peso inicial del exoesqueleto en langostinos azules tiene distribución Pareto$\left(\alpha, \beta = 20\right)$ y su densidad es $f\left(x\right) = \frac{\beta\alpha^{\beta}}{x^{\beta+1}} I_{[\alpha,\infty)}\left(x\right)$ donde $\alpha > 0$. Con la muestra con la muestra 49.8, 49.4, 51.4, 51.6 y 51.0.
   a. Obtenga el intervalo óptimo de 98% de confianza para α.
   b. Con el método estadístico halle un intervalo de confianza de 98% para α.
   c. Utilizando la distribución acumulada de X determine un intervalo de 98% para α.
3. El crecimiento relativo anual (%) de la población de Lima tiene distribución Logística$\left(\alpha = 0, \beta\right)$, esto quiere decir que su densidad es $f\left(x\right) = \frac{\exp\left[-\left(\frac{x-\alpha}{\beta}\right)\right]}{\beta\left\{1 + \exp\left[-\left(\frac{x-\alpha}{\beta}\right)\right]\right\}^2}$. Usando el método del pivote determine un intervalo de 98% de confianza para β con la muestra 4.5, 4.9, 5.2, 5.8 y 4.8 %.
4. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la función densidad de Laplace $f\left(x;\theta\right) = \frac{1}{2\theta}\exp\left\{-\frac{|x|}{\theta}\right\} I_{\left(-\infty,\infty\right)}\left(x\right)$, $\theta \rangle 0$. Halle un intervalo de confianza de 90% para θ.

5. Si 10.5, 15.6, 10.4, 18.7 y 17.3 son las observaciones de una m.a de tamaño 5 de $f(x)=\frac{4x^3}{\theta^4} I_{(0,\theta)}(x)$. Con el método de la cantidad pivotal halle un intervalo de confianza de 95% para θ.
6. La Sicóloga Leonor Respeto hizo un estudio de las barras del fútbol. Entre otras variables estudió el tiempo, en minutos, que demoran las barras en proferir el primer insulto racista después de empezado un encuentro. Con la asesoría de un Ingeniero Estadístico encontró que la densidad de ese tiempo es $f(x)=\frac{1}{\theta x^{\frac{1}{\theta}+1}} I_{(1,\infty)}(x)$ con $0<\theta<1$. Si 1.08, 1.14, 1.18 y 1.08 son las observaciones de una muestra aleatoria de 4 partidos. Basado en la estadística suficiente encuentre un intervalo de confianza de 90% para el percentil 44 de la distribución.
7. Sean $(X_1,\cdots,X_n)$ y $(Y_1,\cdots,Y_m)$ dos m.as independientes de distribuciones exponenciales con medias μ_X y μ_Y respectivamente. Si $\rho=\frac{\mu_X}{\mu_Y}$ demuestre que $\frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i+\rho\sum_{i=1}^{m} y_i}$ es un pivote de ρ. Si las m.as son (0.8, 1.1, 0.9) y (1.2, 0.7) respectivamente, halle un intervalo de confianza de 95% para ρ.
8. Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución $N(\mu,\sigma^2)$, con σ^2 conocida. Halle un intervalo de confianza de longitud mínima para μ.
9. A 20 estudiantes universitarios se les preguntó si habían visto la película 8 $\frac{1}{2}$ del inmortal Federico Fellini y 4 respondieron afirmativamente. Halle un intervalo de 95% de confianza para la verdadera proporción de estudiantes que ha visto esta obra del Genio de la cinematografía universal Federico Fellini.
10. El número de clientes atendidos en una ventanilla, cada 5 minutos, del banco "AVerFirma" tiene distribución $\text{Poisson}(\lambda)$. Con la muestra 1, 0, 0, 0, 1, 3, 0, 0, 0, 2 halle un intervalo de confianza de $100(1-p_1-p_2)\%$ para λ considerando $p_1=0.0540282$ y $p_2=0.0045339$:
    a. Con el método estadístico.
    b. Con la relación que existe entre las distribuciones Gamma y Poisson.
11. El investigador Francis Galton estudió entre otras la variable X definida como el número de exoesqueletos de langostino azul analizados hasta obtener el cuarto con un porcentaje de quitosano extraído, con respecto al peso inicial del exoesqueleto, mayor de 51.2%. Si 7, 8, 9, 6 y 5 son las observaciones de una muestra aleatoria de la distribución en estudio, determine un intervalo de confianza de $100(1-p_1-p_2)\%$ para la proporción de langostinos azules con más de 51.2% de quitosano en su exoesqueleto si $p_1=0.03004887$ y $p_2=0.00122598$.

12. El tiempo en años que dura cierta calculadora hasta que necesita reparación tiene distribución exponencial con parámetro λ. Si con una muestra aleatoria de calculadoras se obtuvo los siguientes tiempos: 14.4, 10.2, 9.3, 5.1 y 11.1. Estime con un intervalo de 90% de confianza a la variancia poblacional.

13. Suponga que las v.as $X_1,\ldots,X_n$ son tales que $\sum_{i=1}^{n}\ln X_i \sim N\left(0, n\beta^2\right)$. Estime con un 95% de confianza a β.

14. El tiempo en minutos que demora un ratón en recorrer un laberinto tiene distribución $\text{Gamma}(1,\beta)$. Sea $X_1,..,X_n$ una muestra aleatoria de la distribución en mención. Primero halle un intervalo general de 98% de confianza para el valor del primer cuartil de la distribución y luego obtenga un intervalo particular con la muestra: 1.2, 1.5, 1.8, 1.0, 1.4.

15. El investigador Diego Ollero Carmona considera que la demanda diaria, en toneladas, para Lima Metropolitana y el Callao de camote amarillo procedente de Barranca tiene una distribución $\text{Normal}\left(\mu = 38, \sigma^2\right)$. Si 10.8, 25.2, 38.4, 58.6 y 48.4 es una muestra aleatoria de esa distribución utilice el método de la cantidad pivotal para encontrar un intervalo de confianza de 98% para $4\sigma^3$.

16. El contenido de vitamina B6 en 100 gr de tomate de árbol tiene la siguiente densidad $f(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}(x-\theta)} I_{(\theta;\infty)}(x)$. Si $X_1,\ldots,X_n$ es una muestra aleatoria de esa densidad y utilizando estadísticas suficientes:
   a. Halle un intervalo de 95% de confianza para λ.
   b. Obtenga el intervalo óptimo de 95% de confianza para θ.

17. El Ingeniero Mario Lanza afirma que la temperatura X del suelo, en cierta región, tiene densidad $f(x) = \frac{4x^3}{\theta^4} I_{(0,\theta]}(x)$. Considerando una muestra aleatoria de tamaño n, utilice el método estadístico para hallar un intervalo de 95% de confianza para estimar a θ.

18. La demanda máxima diaria en un periodo de una semana, en miles de kilos, de pollo en pie en Lima Metropolitana y el Callao según los centros de acopio tiene distribución de valor extremo de tipo II o de Fréchet y su densidad es:

$$f(x) = \frac{k}{\mu}\left(\frac{\mu}{x}\right)^{k+1} e^{-\left(\frac{\mu}{x}\right)^k} I_{[0,\infty[}(x)$$, asuma que $k = 4$.

Suponga que los datos de la Tabla 1 corresponden a la demanda máxima diaria de una muestra aleatoria de 20 semanas.

Tabla

Demanda máxima diaria por semana (en miles de kg)

177.8	175.4	176.2	178.1	175.8	176.9	174.8	174.9	177.4	176.2
178.2	177.3	176.6	175.9	177.2	174.8	177.2	179.0	173.4	175.9

a. Utilizando el estimador suficiente halle e interprete un intervalo de 90% de confianza para μ.
b. Utilizando el estimador basado en un mínimo apropiado halle e interprete un intervalo de 90% de confianza para μ.
c. Suponga que otro investigador trabaja con la distribución anterior pero truncada en (μ,∞), y asumiendo que $k=4$. En estas condiciones utilice el método estadístico para estimar con un 90% de confianza a μ.

19. El caudal máximo anual de un rio pequeño $\left(en\ pies^3\right)$ tiene como densidad $f(x)=\left(\dfrac{k}{w-\mu}\right)\left(\dfrac{w-x}{w-\mu}\right)^{k-1}\exp\left[-\left(\dfrac{w-x}{w-\mu}\right)^{k}\right]I_{(-\infty,w)}(x)$, con k y w conocidas. Halle e interprete un intervalo de confianza de 90% para el caudal modal máximo anual. Asuma en esta pregunta que los parámetros $w=108$ y $k=2$ y que una m.a de tamaño 4 de $f(x)$ es 99, 105, 98 y 102 $pies^3$.

20. Bacterias Psicrófilas: Son aquellas capaces de vivir a temperaturas por debajo de los 5 °C. Para la pregunta de abajo se utilizan porciones de carne de 75 gr. Se mide el número de bacterias psicrófilas en la carne después de 9 días de almacenamiento a 4ºC. La variable respuesta o medición fue el logaritmo del número de bacterias por cm^2. (Estas bacterias se encuentran en la superficie de la carne y aparecen cuando la carne se echó a perder). En el laboratorio A, el logaritmo del número de bacterias por cm^2 mínimo diario X tiene distribución del valor mínimo tipo II de Fréchet entonces $f(x)=\dfrac{\alpha}{\mu}\left(\dfrac{x}{\mu}\right)^{\alpha-1}\exp\left[-\left(\dfrac{x}{\mu}\right)^{\alpha}\right]I_{(0,\infty)}(x)$. Halle e interprete un intervalo de confianza de 90% para el logaritmo modal mínimo diario. Asuma en esta pregunta que el parámetro $\alpha=4$ y que en una muestra aleatoria de 4 porciones de 75 gr de carne se hallaron los siguientes logaritmos 7.4, 8.2, 6.8 y 7.6.

21. El investigador Diego Ollero Carmona analizó 3 exoesqueletos de langostino blanco obteniendo los siguientes porcentajes de quitosano extraído con respecto al peso inicial del exoesqueleto: 48.8, 49.7, 47.6. Ollero tomó muestras aleatorias simples de tamaño 2, de esa población. Obtenga los intervalos de confianza de por lo menos 98% de confianza para la variancia poblacional utilizando $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2$ y después $S_1^2=\dfrac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2$. Diga qué estimador es más preciso.

22. De 100 perritas a las que se les hizo un seguimiento de su último parto resultó que con cada una de 30 perritas resultaron 2 perros callejeros, con cada una de 20 resultaron

3 perros callejeros y con 50 cuatro callejeros. Si de estas 100 perritas se toman muestras aleatorias de tamaño 2, determine los intervalos de por lo menos 98% de confianza para el mínimo paramétrico de perros callejeros que resultaron. Utilice el mínimo muestral y la media geométrica. Diga cuál de estos estimadores es más preciso y por qué.