Auxiliar 12: Intervalos de Confianza

MA3403 - Probabilidades y Estadistica
Profesor: Vicente Acuña
Auxiliares: Martin Castillo - Felipe Campos
11 de Junio, 2014

Protocolo para la construcción de un intervalo de confianza para un parámetro θ

1. Encontrar un estimador $\hat{\theta}$ de θ. Ojalá insesgado.
2. Encontrar la distribución de $\hat{\theta}$ o de alguna función del estimador.
3. Encontrar un estadístico, digamos Z, tal que dependa de $\hat{\theta}$ y de θ, pero cuya distribución no dependa de θ.
4. Encontrar a y b, relacionados bajo algún criterio, tal que $\mathbb{P}(a < Z < b) = 1 - \alpha$, donde 1_α se denomina, nivel de confianza, que es fijado a priori.
5. En base a los cálculos anteriores construir un intervalo para θ.

P1. Un grupo de partículas tiene velocidad $N(\mu, \sigma^2)$. Se toma una muestra de cinco de ellos y sus velocidades son 10, 12, 9, 8, 13.

a) Construya un I.C. para σ^2, con confianza 0.9, de tipo cota superior (ie, $\sigma^2 <$).

Solución:

Primero, vemos que $\bar{X} = \sum_{i=1}^{5} X_i = 10{,}4$. El estimador que usamos es el que sigue:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n}.$$

Por otro lado observemos lo siguiente:

$$X_i \sim N(\mu, \sigma^2) \Leftrightarrow X_i - \mu \sim N(0, \sigma^2) \Leftrightarrow \frac{X_i - \mu}{\sigma} \sim N(0,1),\ \forall i = 1, \ldots, n.$$

Para encontrar el intervalo imponemos:

$$0{,}9 = \mathbb{P}(0 < \sigma^2 < c)$$

Donde $c = c(X_1, \ldots, X_n)$ depende de los valores observados. Por otro lado usaremos que si tenemos $Z_i \sim N(0,1)$, i.i.d. $\forall i = 1, \ldots, n$. Entonces:

$$Z_1^2 + \cdots + Z_n^2 \sim \chi_n^2 \text{ (Chi cuadrado con n grados de libertad.)}.$$

Luego desarrollamos:

$$0 < \sigma^2 < c \Leftrightarrow 0 < \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \mu)^2} < \frac{c}{\sum_{i=1}^{n}(X_i - \mu)^2} \Leftrightarrow \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{c} < \underbrace{\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2}_{\sim \chi_n^2} < \infty$$

Así,
$$\mathbb{P}(0 < \sigma^2 < c) = 0{,}9 \Leftrightarrow \mathbb{P}\left(\chi_n^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{c}\right) = 0{,}1.$$
Aquí nos gustaría simplemente ocupar la tabla de la χ_n^2 para encontrar el valor de c. Pero no podemos hacerlo pues el parámetro μ es desconocido. Por lo tanto también tenemos que estimarlo. Tomando $\bar{\mu} = \bar{X}$ podemos seguir trabajando pero antes hay que darse cuenta que perdemos un grado de libertad de la χ^2. Esto pues al imponer $\bar{\mu} = \bar{X}$ nos queda:
$$X_1 + \cdots + X_{n-1} + X_n = \bar{\mu} n \in \mathbb{R}.$$
Por lo que se pierde la independencia de una de las variables, pues conociendo el valor de las $n-1$ primeras X_i podemos calcular el valor de X_n. Así la χ^2 pasa de tener n grados de libertad a $n-1$ grados de libertad al imponer $\bar{\mu} = \bar{X}$.
De esta manera tenemos que encontrar el $c \in \mathbb{R}$ tal que:
$$\mathbb{P}\left(\chi_4^2 < \frac{\sum_{i=1}^5 (X_i - \bar{\mu})^2}{c}\right) = 0{,}1.$$
De la tabla de la χ_4^2 y la primera ecuación obtenemos que $\frac{\sum_{i=1}^5 (X_i - \bar{\mu})^2}{c} = 1{,}06$. Por otro lado, $\sum_{i=1}^5 (X_i - \bar{\mu})^2 = 17{,}2$. Luego $c = \frac{17{,}2}{1{,}06} = 16{,}22$, el intervalo buscado corresponde a:
$$I.C. : (0; 16{,}22)$$

b) Si $\sigma^2 = 3{,}5$, determine con qué confianza se construyó el intervalo de confianza para μ dado por $(10; 11{,}5)$.
Solución: Como en este caso estamos estimando el promedio, el estadístico es (aprovechandonos del TCL) $Z = \frac{\bar{\mu}-\mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^n X_i - \mu N}{\sigma\sqrt{n}}$ tenemos que encontrar el valor de $1-\alpha$:
$$\begin{aligned} 1-\alpha &= \mathbb{P}(a \leq Z \leq b) \\ &= \mathbb{P}\left(-b \leq \frac{\mu - \bar{\mu}}{\sigma/\sqrt{n}} < -a\right) \\ &= \mathbb{P}(\underbrace{-b\sigma/\sqrt{n} + \bar{\mu}}_{10} \leq \mu \leq \underbrace{-a\sigma/\sqrt{n} + \bar{\mu}}_{11{,}5}) \end{aligned}$$
Para encontrar los valores de a y b los obtenemos mediante el siguiente sistema
$$\begin{aligned} -b\frac{\sqrt{3{,}5}}{\sqrt{5}} + 10{,}4 &= 10 \\ -a\frac{\sqrt{3{,}5}}{\sqrt{5}} + 10{,}4 &= 11{,}5 \end{aligned}$$
Despejando, obtenemos: $a = -1{,}32,\ b = 0{,}48$. Luego, la confianza del intervalo dado es: $\mathbb{P}(a \leq Z \leq b) = \Phi(b) - \Phi(a) = 0{,}57$.

P2. Considere una marca de ampolletas cuya duración (única característica relevante) es una v.a. $X \sim N(\mu, 25)$. Se toma una m.a.s. de tamaño $n = 25$, obteniéndose $\bar{X} = 170$.

a) Construya un intervalo de confianza de 95 % para μ.
Solución:
El razonamiento es parecido (pero no el mismo) a la parte *b*) de la pregunta 1. Queremos que:
$$\mathbb{P}(b < Z < a) = 0{,}95.$$

Se sabe que para encontrar intervalos más cortos para la $N(0,1)$ es conveniente asumir simetría (con respecto al promedio) del intervalo de confianza. Entonces asumimos $b=-a$. Luego desarrollamos:

$$\mathbb{P}(-a<Z<a)=\mathbb{P}(Z\leq a)-\mathbb{P}(Z\leq -a)=\Phi(a)-\mathbb{P}(Z>a)=Phi(a)-(1-\mathbb{P}(Z\leq a))=2\Phi(a)-1$$

Entonces

$$\mathbb{P}(-a<Z<a)=0{,}95\Leftrightarrow\Phi(a)=\frac{1{,}95}{2}=0{,}975\Leftrightarrow a=1{,}96.$$

b) ¿Con qué nivel de confianza fue construido el siguiente intervalo para μ: (168.5,172)?
Solución:
Queremos saber, con que probabilidad el intevalo dado contiene a μ (la aleatoriedad se encuentra en el intevalo), es decir,

$$\begin{aligned}
\mathbb{P}(168{,}5\leq\mu\leq 172)&=\mathbb{P}(-172\leq-\mu\leq-168{,}5)\\
&=\mathbb{P}\left(\frac{\bar{X}-172}{\sigma/\sqrt{n}}\leq\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\leq\frac{\bar{X}-168{,}5}{\sigma/\sqrt{n}}\right)\\
&=\mathbb{P}(-2\leq Z\leq 1{,}5)\\
&=\Phi(1{,}5)-\mathbb{P}(-2\leq Z)\\
&=\Phi(1{,}5)-(1-\mathbb{P}(Z\leq 2))\\
&=\Phi(1{,}5)+\Phi(2)-1\\
&=0{,}93+0{,}97-1\\
&=0{,}9
\end{aligned}$$

De esta manera, el intervalo dado fue construido con 90 % de confianza.

Momentos La idea de el método de los momentos es el siguiente:

a) Tenemos alguna v.a. X cuya distribución depende de algún parámetro θ que no conocemos. Notar que θ puede tener mas de una componente, ej: $\theta=(\theta_1,\theta_2)$. Solo podemos medir valores .$^{\text{em}}$píricos"de esta variable.

b) La esperanza y la varianza de esta variable deberían depender del parámetro. Así:

$$\mathbb{E}(X)=f(\theta),\ \operatorname{Var}(X)=g(\theta).$$

c) Luego imponemos alguna de las (o ambas) condiciones:

$$\mathbb{E}(X)=\bar{X}=f(\theta),$$

$$Var(X)=\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n}=g(\theta).$$

d) Intentamos despejar θ. Y obtenemos así nuestro estimador.

P3. Se dice que una variable aleatoria X tiene distribución Pareto(c,β) si

$$f(x)=\begin{cases}\beta c^{\beta}x^{-\beta-1} & \text{si } x\geq c\\ 0 & \text{si } x<c\end{cases}$$

Sea $X_1,\ldots,X_n$ muestra aleatoria simple:

a) Suponiendo c conocido, encuentre el estimador de β usando el método de los momentos.
Solución:
Necesitamos calcular $\mathbb{E}(X)$.

$$\begin{aligned}\mathbb{E}(X) &= \int_c^{\infty} x\beta c^{\beta} x^{-\beta-1} dx = \beta c^{\beta} \int_c^{\infty} x^{-\beta} dx \\ &= \beta c^{\beta} \left. \frac{x^{-\beta+1}}{-\beta+1} \right|_c^{\infty} = \frac{-\beta c}{-\beta+1} = \frac{\beta c}{\beta-1}\end{aligned}$$

Luego, imponemos $\mathbb{E}(X) = \overline{X}$ (pues $\mathbb{E}(\hat{X}) = \overline{X}$), y despejamos β, es decir,

$$\overline{X} = \frac{\beta c}{\beta - 1} \Rightarrow \hat{\beta} = \frac{\overline{X}}{\overline{X} - c}$$

b) Determine el estimador máximo verosímil de c y β, denotados por $\hat{c}, \hat{\beta}$.
Solución:

La función de verosimilitud viene dada por:

$$L(\beta, c) = \prod_{i=1}^{n} f(X_i|\beta, c) = \begin{cases} \beta^n c^{n\beta} \prod_{i=1}^{n} X_i^{-(\beta+1)} & , \text{ si } c \leq X_i, \ \forall i = 1, \ldots, n. \\ 0 & , \text{ si } \exists i = 1, \ldots, n \text{ tal que } c \leq X_i \end{cases}$$

Observamos que el valor de c que maximiza f se alcanza cuando $\hat{c} = \min x_i$. Ahora, para β, notamos que f se maximiza en el intervalo en que $c \leq \min x_i$ (puesto que en el complemento es 0) y por lo tanto, en este intervalo buscamos el máximo de la log-verosimilitud. Esto pues sabemos que los problemas $\max L$ y $\max \log(L)$ son equivalentes, en el sentido de que los valores que maximizan L también maximizan $\log(L)$ y visceversa. Al derivar $\log(L(\beta))$ e igualar a 0 se obtiene que el estimador máximo verosímil para β corresponde a:

$$\tilde{\beta} = \frac{n}{\sum_{i=1}^{n} \log\left(\frac{x_i}{\hat{c}}\right)}.$$

c) Determine la densidad de $Y = \ln\left(\frac{x}{c}\right)$ y a partir de ello determine la distribución de $\hat{\lambda} = (1\hat{/}\beta)$ considerando c conocido y n grande.
Solución:

Sea $Y = \ln(X/c)$. Primero notemos que Y queda bien definida sobre $X \geq c$, y en este caso toma valores en $[0, +\infty)$. Calculemos su distribución:

$$\begin{aligned}\mathbb{P}(Y > y) &= \mathbb{P}(\ln(X/c) > y) \\ &= \mathbb{P}(X > ce^y) \\ &= \int_{ce^y}^{+\infty} \beta c^{\beta} x^{-\beta-1} \, dx \\ &= \beta c^{\beta} \left[-\frac{x^{-\beta}}{\beta} \right]_{ce^y}^{+\infty} \\ &= e^{-\beta y}.\end{aligned}$$

Es decir, $Y \sim \text{Exp}(\beta)$. Luego, $\mathbb{E}(Y) = 1/\beta \Rightarrow \hat{\lambda} = \overline{Y}$. Finalmente, para n grande, por TCL, se tiene que

$$\frac{\overline{Y} - \beta^{-1}}{\beta^{-1}/\sqrt{n}} \sim \mathcal{N}(0,1),$$

y por lo tanto $\overline{Y} \sim \mathcal{N}\left(\frac{1}{\beta}, \frac{1}{\beta^2 n}\right)$.

d) Si X representa el ingreso mensual (en miles de \$) de un grupo familiar, $\beta = 3$, $c = 100$ y se toma una muestra de 48 familias de manera independiente. Calcule la probabilidad que el ingreso promedio de las familias supere los 550.000 pesos.
Solución:

Sea v la varianza de X y e la esperanza (calculada en la parte a)). Luego, para calcular la probabilidad se hace uso del TCL, calculando

$$\mathbb{P}\left(\mathcal{N}(0,1) > \frac{550000 - e}{v/\sqrt{48}}\right).$$

Propuesto El cocinero del casino de la facultad preparó masa para hacer 500 empanadas. Ese mismo día, en un grupo de 20 alumnos que almorzaron juntos, alguien propuso contar la cantidad de pasas que cada uno encontrase en su empanada, obteniendo la siguiente distribución:

N° de empanadas	1	3	4	5	4	2	1
N° de pasas	0	1	2	3	4	5	8

Suponiendo que la distribución de la cantidad de pasas X en una empanada sigue una ley de $Poisson$, estime el parámetro λ.
Hint: Use el método de los momentos.