Ejercicios

Fabián Encarnación, Geoconda Molina, Debbie Echanique

2024-02-24

Ejercicio 2

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \in N\left(\mu, \sigma^2\right)$, con $\mu \in \mathbb{R}$ desconocida y $\sigma^2 > 0$ conocida. Encontrar:

1. El ECUMV para $\eta_1 = \mu^4$.
2. El ECUMV para $\eta_2 = P\left(X_1 \leq t\right)$, siendo t un número conocido.
3. El estimador de máxima verosimilitud de η_2. Calcula su distribución límite.

Aproxima mediante simulación (usa 1000 simulaciones con $\mu = 0, \sigma^2 = 1, t = 1$) el error cuadrático medio del ECUMV y del estimador de máxima verosimilitud de η_2. Comprueba (aproximando el sesgo por simulación) que el ECUMV es insesgado. ¿Cuál tiene menor error cuadrático medio? \

Solución.

(i) Sea $\bar{X}$ la media muestral, la cual es completa y suficiente para μ. Dado que

$$0 = E(\bar{X} - \mu)^3 = E\left(\bar{X}^3 - 3\mu\bar{X}^2 + 3\mu^2\bar{X} - \mu^3\right) = E\left(\bar{X}^3\right) - 3\mu\sigma^2/n - \mu^3,$$

obtenemos que

$$E\left[\bar{X}^3 - \left(3\sigma^2/n\right)\bar{X}\right] = E\left(\bar{X}^3\right) - 3\mu\sigma^2/n = \mu^3$$

para todo μ. Por el teorema de Lehmann-Scheffé, el ECUMV de μ^3 es $\bar{X}^3 - \left(3\sigma^2/n\right)\bar{X}$. De manera similar,

$$\begin{aligned}
3\sigma^4 &= E(\bar{X} - \mu)^4 \\
&= E\left[\bar{X}(\bar{X} - \mu)^3\right] \\
&= E\left[\bar{X}^4 - 3\mu\bar{X}^3 + 3\mu^2\bar{X}^2 - \mu^3\bar{X}\right] \\
&= E\left(\bar{X}^4\right) - 3\mu\left(3\mu\sigma^2/n + \mu^3\right) + 3\mu^2\left(\sigma^2/n + \mu^2\right) - \mu^4 \\
&= E\left(\bar{X}^4\right) - 6\mu^2\sigma^2/n - 4\mu^4 \\
&= E\left(\bar{X}^4\right) - \left(6\sigma^2/n\right)E\left(\bar{X}^2 - \sigma^2/n\right) - 4\mu^4.
\end{aligned}$$

Por lo tanto, el ECUMV de μ^4 es $\left[\bar{X}^4 - \left(6\sigma^2/n\right)\left(\bar{X}^2 - \sigma^2/n\right) - 3\sigma^4\right]/4$.

(ii) Dado que $E\left[P\left(X_1 \leq t \mid \bar{X}\right)\right] = P\left(X_1 \leq t\right)$, el ECUMV de $P\left(X_1 \leq t\right)$ es $P\left(X_1 \leq t \mid \bar{X}\right)$. A partir de las propiedades de las distribuciones normales, $\left(X_1, \bar{X}\right)$ es normal bivariada con media (μ, μ) y matriz de covarianza

$$\sigma^2\begin{pmatrix} 1 & n^{-1} \\ n^{-1} & n^{-1} \end{pmatrix}.$$

En consecuencia, la distribución condicional de X_1 dada $\bar{X}$ es la distribución normal $N\left(\bar{X},\left(1-n^{-1}\right)\sigma^2\right)$. Entonces, el ECUMV de $P\left(X_1 \leq t\right)$ es

$$\Phi\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right),$$

donde Φ es la función de distribución acumulada de $N(0,1)$. Por el teorema de convergencia dominada,

$$\frac{d}{dt}P\left(X_1 \leq t\right)=\frac{d}{dt}E\left[\Phi\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right)\right]=E\left[\frac{d}{dt}\Phi\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right)\right].$$

Por lo tanto, el ECUMV de $\frac{d}{dt}P\left(X_1 \leq t\right)$ es

$$\frac{d}{dt}\Phi\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right)=\frac{1}{\sigma\sqrt{1-n^{-1}}}\Phi'\left(\frac{t-\bar{X}}{\sigma\sqrt{1-n^{-1}}}\right).$$

(iii) Del literal anterior, tenemos que ese es el estimador de máxima verosimilitud, además como $\bar{X}$ se distribuye asintóticamente como una $N(\mu,\frac{\sigma^2}{n})$ tenemos que cuando $n\to\infty$, es una distribución $N(0,1)$

```r
set.seed(1)
nsim<-1000
mu<-0
s<-1
t<-1

val<-rnorm(nsim,mu,s)
ecumv<-pnorm(t,val,s)
emv<-ecumv

val2<-pnorm(t,mu,s)
errorcmecumv<-mean((ecumv-val2)^2)
errorcmemv<-errorcmecumv
sesgo<-mean(ecumv)-val2

cat("Error cuadrático medio del ECUMV y Estimador máximo verosimil:", errorcmecumv, '\n')
```

```
## Error cuadrático medio del ECUMV y Estimador máximo verosimil: 0.06484548
```

```r
cat('Sesgo del ECUMV:', sesgo, '\n')
```

```
## Sesgo del ECUMV: -0.08237049
```

Los errores del ECUMV y el Estimador máximo verosimil son los mismos ya que ambos son idénticos, además vemos que el existe sesgo de -0.08, lo cual nos dice que no son insesgados ya que deberían considerar un mayor número de simulaciones.

Ejercicio 3

Sea $X1,...,Xn$ una m.a.s de $X\in N(\mu,\sigma^2)$. Si suponemos que $\mu\neq 0$ calcula el comportamiento limite del estimador de máxima verosimilitud del coeficiente de variación $\tau=\frac{\sigma}{\mu}$.

Solución

Por un lado estudiemos el EMV de μ que es la media muestral:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Por otro lado el EMV de σ^2 es la varianza muestral:

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

Por lo que, si τ se define como $\tau = \sigma/\mu$, sustituyendo las expresiones anteriores se tiene:

$$\tau = \frac{\hat{\sigma}}{\hat{\mu}} = \frac{\sqrt{\hat{\sigma}^2}}{\hat{\mu}}$$

Finalmente, para el comportamiento límite del EMV del coeficiente de variación, $\hat{\mu}$, puede analizarse utilizando la teoría asintótica. Bajo ciertas condiciones regulares, tenemos que:

- $\hat{\mu}$ es asintóticamente normal con media μ y varianza σ^2/n.
- $\hat{\sigma}^2$ tiene una distribución asintótica que permite considerar $\hat{\sigma}$ también asintóticamente normal.

Aplicando el método delta, se puede concluir que $\hat{\tau}$ tendrá una distribución asintótica normal, cuya variabilidad dependerá de las de $\hat{\mu}$ y $\hat{\sigma}$.

Ejercicio 4

Sea $(X1, ..., Xn)$ una m.a.s. de X con densidad $f_\theta(X) = (1+\theta)X^\theta$ si $X \in (0,1)$, 0 si no, donde $\theta > -1$.

1. Calcular el EMV de θ y de $\mu = \mathbb{E}_\theta(X)$
2. Cuál es la distribución en el muestreo de $T = \sum_{i=1}^{n} log(Xi)$
3. Calcular la cota de Cramer-Rao para a estimación insesgada de θ.
4. Contrastar de forma UMP: $H_0 : \theta = 0$ vs. $H_1 : \theta = 1$
5. Obtener el test de Razón de Verosimilitudes para: $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$

Solución

1.

Lafunción de verosimilitud para una muestra aleatoria simple $X1, ..., Xn$ es:

$$L(\theta) = \prod_{i=1}^{n}(1+\theta)X_{i=1}^{\theta} = (1+\theta)^n \prod_{i=1}^{n} X_{i=1}^{\theta}$$

$$\Longrightarrow l(\theta) = nlog(1+\theta) + \theta\sum_{i=1}^{n} log(xi)$$

Derivando:

$$\frac{d}{d\theta}l(\theta) = 0 \Longrightarrow \frac{n}{1+\theta} + \sum_{i=1}^{n} log(xi) = 0 \Longrightarrow \frac{n}{1+\theta} = -\sum_{i=1}^{n} log(xi) \Longrightarrow \frac{-n}{\sum_{i=1}^{n} log(xi)} = 1+\theta \Longrightarrow \theta = -(1+\frac{n}{\sum_{i=1}^{n} log(xi)})$$

Por lo que, el EMV de θ es:

$$\hat{\theta} = -(1 + \frac{n}{\sum_{i=1}^{n} log(xi)})$$

Calculemos $\mu = \mathbb{E}_\theta(X)$, así:

$$\mu = \int_0^1 x(1+\theta)x^\theta dx = (1+\theta)\int_0^1 x^{1+\theta} dx = (1+\theta)\frac{x^{2+\theta}}{\theta+2}|_1 = \frac{\theta+1}{\theta+2}$$

Así, para encontrar el EMV de μ

$$\hat{\mu} = \frac{\hat{\theta}+1}{\hat{\theta}+2} = \frac{2 + \frac{n}{\sum_{i=1}^{n} log(xi)}}{3 + \frac{n}{\sum_{i=1}^{n} log(xi)}} = \frac{2\sum_{i=1}^{n} log(xi) + n}{3\sum_{i=1}^{n} log(xi) + n}$$

2.

Asumiendo que n es grande y bajo condiciones regulares se tiene que $T = \sum_{i=1}^{n} log(xi)$ puede aproximarse por una distribución normal debido al Teorema Central del Límite.

3.

La cota de Cramer-Rao proporcio aun límite inferior para la varianza de cualquier estimador insesgado de θ:

$$Var(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

de donde $I(\theta)$ es la información de Fisher dada por:

$$I(\theta) = \mathbb{E}[\frac{d^2}{d\theta^2} l(\theta)] \Longrightarrow \frac{d^2}{d\theta^2} l(\theta) = \frac{d}{d\theta} l(\theta) = \frac{d}{d\theta}\frac{n}{1+\theta} + \sum_{i=1}^{n} log(xi) = \frac{-n}{(\theta+1)^2}$$

Reemplazando:

$$I(\theta) = \mathbb{E}[\frac{d^2}{d\theta^2} l(\theta)] = \mathbb{E}[\frac{-n}{(\theta+1)^2}] = \frac{n}{(\theta+1)^2}$$

Finalmente

$$Var(\hat{\theta}) \geq \frac{(\theta+1)^2}{n}$$

4.

Usando el lema de Neyman-Pearson cuando se trata de hipótesis simples.

El lema de Neyman-Pearson establece que el test más poderoso para probar H_0 contra H_1 a un nivel de significancia α está basado en la región crítica definida por el cociente de verosimilitudes:

$$\Lambda(x) = \frac{L(\theta = 1|x)}{L(\theta = 0|x)}, \text{se rechaza } H_0 \text{ si } \Lambda(x) > k$$

donde k es una constante que se determina por el nivel de significacia α.

Así, se puede calcular las funciones de verosimilitud para cada caso:

(a) Para $\theta = 0$:

Se tiene que $f_\theta(x) = 2x$ para $x \in (0,1)$, por lo que:

$$L(\theta = 1|x) = 2^n \prod_{i=1}^{n} xi \Longrightarrow \Lambda(x) = \frac{L(\theta = 1|x)}{L(\theta = 0|x)} = 2^n \prod_{i=1}^{n} xi$$

5.

Para obtener el test de Razón de Verosimilitudes para las hipótesis nulas y alternativas dadas ($H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$), primero necesitamos calcular la función de verosimilitud bajo ambas hipótesis y luego compararlas.

Supongamos que tenemos datos observados $x_1, x_2, ..., x_n$ y una distribución conocida parametrizada por θ, entonces la función de verosimilitud bajo H_0 es:

$$L(\theta|x_1, x_2, ..., x_n)$$

La función de verosimilitud bajo la hipótesis alternativa (H_1) implica que estamos restringiendo el parámetro θ a ser mayor que cero. Por lo tanto, la función de verosimilitud bajo H_1 se convierte en:

$$L(\theta|x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i|\theta) \quad \text{para } \theta > 0$$

Donde $f(x_i|\theta)$ es la función de densidad de probabilidad de la distribución que estamos considerando.

Luego, calculamos el logaritmo de la razón de las funciones de verosimilitud:

$$\Lambda = \frac{L(\hat{\theta}_{H_1})}{L(\hat{\theta}_{H_0})}$$

Donde $\hat{\theta}_{H_1}$ y $\hat{\theta}_{H_0}$ son los estimadores de máxima verosimilitud bajo H_1 y H_0, respectivamente.

Si Λ es mayor que un cierto valor crítico (que depende del nivel de significancia deseado y la distribución del estadístico de prueba bajo la hipótesis nula), entonces rechazamos H_0 a favor de H_1.

Es importante tener en cuenta que dependiendo de la distribución de la estadística de prueba, puede que no exista una solución analítica para encontrar el valor crítico, y en ese caso, podría ser necesario recurrir a métodos numéricos o aproximaciones.