Handout 9

PROPERTIES OF ESTIMATORS: EFFICIENCY, SUFFICIENCY

1. Supposse you want to estimate the proportion of people who smoke marihuana. Compare the method of randomized response introduced in the previous handout ("Go next room and throw a regular coin: if it falls heads, answer 'Yes', otherwise answer the truth.") and the method of asking the question plainly, in the case everyone would tell the truth. What is the relative efficiency of both methods? What is the price to pay for protecting the privacy of respondents?

2. In a previous handout the moment and MLE estimator of θ of the Rayleigh distribution,

$$\begin{aligned} F_X(x) &= 1 - e^{-x^2/2\theta} \\ f_X(x) &= \frac{x}{\theta} e^{-x^2/2\theta} \end{aligned}$$

were found to be:

$$\begin{aligned} \hat{\theta}_M &= 2\overline{X}^2/\pi \\ \hat{\theta}_{MLE} &= \sum_{i=1}^{n} x_i^2/2n \end{aligned}$$

Check that:

(a) $\hat{\theta}_{MLE}$ in unbiased.

(b) $\hat{\theta}_M$ is biased but the bias decreases to zero as the sample size increases.

(c) The Fisher information for one observation is $1/\theta^2$.

(d) $\hat{\theta}_{MLE}$ is efficient.

HINT: You may want to use the fact that the first, second and fourth moments of this distribution are given by

$$\begin{aligned} E[X] &= \sqrt{\pi\theta/2} \\ E[X^2] &= 2\theta \\ E[X^4] &= 8\theta^2. \end{aligned}$$

3. Consider the MLE estimator of θ in a $U(0,\theta)$ distribution, based on n independent observations $X_1, \ldots, X_n$. We have seen $\hat{\theta}_{MLE} = X_{(n)} = \max\{X_1, \ldots, X_n\}$.

(a) Show that $X_{(n)}$ is sufficient for θ.

(b) Find the relative efficiency of $\hat{\theta}_M = 2\overline{X}_n$ as compared with $\hat{\theta}_U$ (both unbiased) for sample sizes $n = 10$, $n = 20$ and $n \to \infty$.

(c) Could we use the Cramér-Rao result to establish that $\hat{\theta}_U$ is optimal among all unbiased estimators?

4. Using the factorization theorem, show sufficiency in the following cases:

   (a) $\overline{X}_n$ for μ in the $N(\mu, \sigma = 1)$.

   (b) $\overline{X}_n$ and $\sum_{i=1}^n X_i^2$ jointly for μ and σ^2 in the $N(\mu, \sigma)$.

   (c) $\overline{X}_n$ for θ in the geometric distribution, $P(x) = \theta(1-\theta)^x$, $x = 0, 1, \ldots$, $0 < \theta < 1$.

   (d) $X_{(1)}$ (the smallest observation) is sufficient for c in the shifted exponential distribucion, $f(x; c) = e^{-(x-c)}$ for $x > c$, zero otherwise.

   (e) $\max |X_1|, \ldots, |X_n|$ for θ in a $U(-\theta, \theta)$.

5. Find the Cramér-Rao lower bound in the case of the exponential distribution, $f(x) = \lambda e^{-\lambda x}$, $x > 0$.

For more examples and help with those above you may turn to the books in the references (note though that Lehmann (1983) is much more advanced than the level of this course). You also may turn to sources in Internet, notable Wikipedia and searchable questions and answers in `https://stats.stackexchange.com`.

References

P. H. Garthwaite, I. T. Jolliffe, and B. Jones. *Statistical Inference.* Prentice Hall, London, 1995.

E. L. Lehmann. *Theory of Point Estimation.* Wiley, New York, 1983.

S. M. Ross. *Introductory Statistics, Third Edition.* Academic Press, 2010.

L. Ruíz-Maya and F. Martín-Pliego. *Fundamentos de Inferencia Estadística.* Thomson - Paraninfo, 2005. Signatura: AL-519.23 RUI.

Respuestas abreviadas

1. Una “solución” incorrecta es:

> *Vimos en el handout previo que $2(s/n) - 1$ es un estimador insesgado de p cuando preguntamos con respuesta aleatorizada. Su varianza es $4\,Var(s/n)$. Si preguntáramos sin hacer uso de respuesta aleatorizada, nuestro estimador de p sería (s/n). Por tanto, la eficiencia del estimador con respuesta aleatorizada es* $1/4$.

Pero al escribirlo me pasó desapercibido que s/n es la frecuencia binomial *de una diferente binomial* en el caso de la respuesta aleatorizada. La probabilidad de “sí” en el estimador de respuesta aleatorizada no es p, la verdadera proporción de gente que fuma marihuana, sino $\frac{1}{2} + \frac{1}{2}p$. Por lo tanto, la varianza del estimador de respuesta aleatorizada es:

$$4 \times \frac{(\frac{1}{2} + \frac{1}{2}p)(1 - \frac{1}{2} - \frac{1}{2}p)}{n} = \frac{(1+p)(1-p)}{n}$$

y la eficiencia en comparación al estimador ordinario, que supone que todo el mundo dice la verdad, es:

$$\frac{p(1-p)/n}{(1+p)(1-p)/n} = \frac{p}{1+p}$$

2. (a) $\hat{\theta}_{MLE}$ in unbiased. En efecto,

$$E[\hat{\theta}_{MLE}] = E\left[\sum_{i=1}^{n} x_i^2/2n\right] = \frac{1}{2n} nE[X_i^2] = \frac{2n\theta}{2n} = \theta \qquad (1)$$

(b) $\hat{\theta}_M$ is biased but the bias decreases to zero as the sample size increases. En efecto,

$$\begin{aligned}
E[\hat{\theta}_M] &= E\left[2\overline{X}^2/\pi\right] \\
&= \frac{2}{\pi}E\left[\left(\frac{X_1 + \ldots + X_n}{n}\right)^2\right] \\
&= \frac{2}{\pi n^2}(nE[X_i^2] + n(n-1)E[X_i]E[X_j]) \\
&= \frac{2}{\pi n^2}(2n\theta + n(n-1)\pi\theta/2) \\
&= \frac{2}{\pi n^2}(n^2\pi\theta/2) + O(1/n) \\
&= \theta + O(1/n)
\end{aligned}$$

(c) The Fisher information for one observation is $1/\theta^2$. Tomando el logaritmo de la densidad de una observación, tenemos:

$$\log f_X(X) = \log(X) - \log(\theta) - \frac{X^2}{2\theta}.$$

Derivando respecto a θ,

$$\frac{\partial \log f_X(X)}{\partial \theta} = -\frac{1}{\theta} + \frac{X^2}{2\theta^2}$$

La información de Fisher será así:

$$\begin{aligned}
E\left(\frac{x^2}{2\theta^2} - \frac{1}{\theta}\right)^2 &= E\left(\frac{X^2 - 2\theta}{2\theta^2}\right)^2 \\
&= E\left(\frac{X^4 + 4\theta^2 - 4X^2\theta}{4\theta^4}\right) \\
&= E\left(\frac{8\theta^2 + 4\theta^2 - 4\theta \times 2\theta}{4\theta^4}\right) \\
&= \frac{1}{\theta^2}
\end{aligned}$$

(d) La mínima varianza alcanzable por un estimador insesgado es $1/nI(\theta) = \theta^2/n$. La varianza de $\hat{\theta}_{MLE}$ es:

$$\begin{aligned}
\operatorname{Var}(\hat{\theta}_{MLE}) &= \operatorname{Var}(\sum_{i=1}^{n} X_i^2/2n) \\
&= \frac{1}{4n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i^2) \\
&= \frac{1}{4n^2}\sum_{i=1}^{n}\left[E[X_i^4] - (E[X_i^2])^2\right] \\
&= \frac{1}{4n^2} n \left[8\theta^2 - (2\theta)^2\right] \\
&= \theta^2/n
\end{aligned}$$

Por consiguiente, $\hat{\theta}_{MLE}$ es eficiente.

En este caso, contra lo que es más habitual, el estimador MLE es insesgado y el estimador por momenots no lo es.

3. (a) Podemos escribir la verosimilitud de una muestra de tamaño n así:

$$f(\vec{x};\theta) = \prod_{i=1}^{n}\left(\frac{1}{\theta}\right)^n \delta(\theta > x_{(n)})$$

La función $\delta(\theta > x_{(n)})$ se define como 1 si $\theta > x_{(n)}$ y 0 en otro caso: tiene por objeto hacer la verosimilitud cero para valores de θ menores que $x_{(n)}$ (que no pueden presentarse). Si comparamos esa expresión con el teorema de factorización, vemos que $g(s(\vec{x}),\theta) = (1/\theta)^n\delta(\theta > x_{(n)})$ y $h(\vec{x}) = 1$. Por tanto, $x_{(n)}$, la única función de los datos que se integra en $g(s(\vec{x}),\theta)$ es estadístico suficiente.

(b) En el handout previo constatamos que

$$\operatorname{Var}(\hat{\theta}_U) = \frac{\theta^2}{n(n+2)};$$

Por otra parte, la varianza de $\hat{\theta}_M = 2\overline{X}_n$ es $\theta^2/3n$. La eficiencia relativa de éste último respecto a $\hat{\theta}_U$ es:

$$\frac{\theta^2/n(n+2)}{\theta^2/3n} = \frac{3n}{n(n+2)}$$

Para los valores de n indicados tenemos que tal eficiencia es:

```
> n <- 10
> (3*n) / (n*(n+2))
[1] 0.25
> n <- 20
> (3*n) / (n*(n+2))
[1] 0.1363636
```

La eficiencia del estimador de momentos va disminuyendo al crecer n. Cuando $n \to \infty$, la eficiencia de $\hat{\theta}_M$ respecto a $\hat{\theta}_U$ tiende a cero; el segundo es abrumadoramente mejor que el primero. $\hat{\theta}_M$ es $\sqrt{n}$-consistente, $\hat{\theta}_U$ es n-consistente, algo que nunca ocurre entre estimadores regulares.

(c) No podemos hacer uso de Cramér-Rao aquí para probar optimalidad de $\hat{\theta}_U$ entre los estimadores insesgados[1], pues falla la condición de regularidad que hace el soporte de la distribución (= conjunto de valores que puede tomar X) independiente del parámetro.

4. Todos ejemplos fáciles de manipulación. En el último (exponencial trasladada), basta escribir la verosimilitud así:

$$\begin{aligned} f(\vec{x};c) &= \prod_{i=1}^{n} e^{-(x-c)} \\ &= e^{-\sum_{i=1}^{n}(x_i-c)} \\ &= e^{-\sum_{i=1}^{n}(x_i-x_{(1)}+x_{(1)}-c)} \\ &= e^{-n(x_{(1)}-c)} \times e^{-\sum_{i=1}^{n}(x_i-x_{(1)})} \end{aligned}$$

5. Muy simple, la cota es λ^2/n.

Orientaciones docentes

En relación a el apartado 3b, se les puede hacer notar como la eficiencia puede variar con el tamaño muestral. Para diferentes valores de n un estimador puede variar mucho su eficiencia respecto a otro.

[1] No obstante, sí se puede demostrar que $\hat{\theta}_U$ es insesgado óptimo, si alguien preguntara (teorema de Lehmann-Scheffé).