Handout 8

Properties of estimators: unbiasedness, consistency

1. A random sample made of two independent observations from a $N(\alpha, 1)$ distribution is used to estimate α. The following two estimators ares considered:

$$\begin{aligned}\alpha^* &= \frac{2}{3}X_1 + \frac{1}{3}X_2 \\ \alpha^{**} &= \frac{2}{5}X_1 + \frac{4}{5}X_2\end{aligned}$$

   (a) Which (if any) of those two estimators are unbiased?

   (b) What are their respective variances?

   (c) Can you obtain an unbiased estimator of α whose variance is less than that of α^*? If your answer is yes, provide an example; if your answer is no, explain why not.

2. Let X be distributed with density function $f_X(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$, for $x > 0$. (Remark: This is the usual exponential distribution parameterized in terms of the mean θ; an alternative parameterization is in terms of $\lambda = 1/\theta$, in which case the mean is $1/\lambda$ and the variance $1/\lambda^2$). In a previous problem we obtained the MLE of θ based on n independent observations.

   (a) Is it: i) Consistent? ii) Unbiased?

   (b) Is the MLE estimator of λ unbiased?

3. A common method to ask sensitive questions in surveys is to use randomized responses: "Do you smoke marihuana? Go next room and throw a regular coin: if it falls heads, answer 'Yes', otherwise answer the truth." This method of asking questions ensures that no one can ever be criminalized: an answer 'Yes' can be the outcome of the coin falling heads. If you ask n persons and obtain s affirmative answers, provide an unbiased estimator of the proportion of people which smoke marihuana.

4. When we take a sample of n independent observations $X_1, \ldots, X_n$ from a $b(p)$ distribution, we have found that the MLE is $\overline{X}$, the number of "ones" divided by the number of throws. Show that it is consistent. At what speed does the standard deviation converge to zero? As fast as $1/n$? As fast as $1/\sqrt{n}$ ("square root consistency")? As fast or faster than $1/n^2$?

5. *(long and somewhat involved, but worthwile to work in detail)* Consider the MLE estimator of θ in a $U(0, \theta)$ distribution, based on n independent observations $X_1, \ldots, X_n$. We already know $\hat{\theta}_{MLE} = X_{(n)} = \max\{X_1, \ldots, X_n\}$.

   (a) Show that it is consistent. (Hint: You can follow a direct approach here. Fix a neighborhood of θ of arbitrary width ϵ and show that for sufficiently large n, the probability of being there is as close to 1 as desired.)

(b) Find the distribution, then the density of $\hat{\theta}_{MLE} = X_{(n)}$. (HINT: $F_{\hat{\theta}_{MLE}}(x) = P(\hat{\theta}_{MLE} \leq x) = P(\cap_{i=1}^{n} X_i \leq x)$.)

(c) Find the mean of $\hat{\theta}_{MLE}$. Find the bias. Obtain an unbiased estimator $\hat{\theta}_U$.

(d) Find the variance of $\hat{\theta}_U$. (HINT: Make use of the fact that $\sigma^2 = \alpha_2 - m^2$; you have computed the mean m in the previous step, so you only need α_2.)

(e) What is the convergence rate to zero of the standard deviation you have found?

6. Show that in a $N(m, \sigma^2)$, the median Med $= \overline{X}_{(0.50)}$ is a consistent estimator of m. (HINT: Assume Med not consistent. Then, Med $\notin (m - \epsilon, m + \epsilon)$ with probability $\eta > 0$ no matter how large n. Show that this cannot happen.)

References

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

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.

A. F. Trocóniz. *Probabilidades. Estadística. Muestreo.* Tebar-Flores, Madrid, 1987.

Respuestas abreviadas, pautas docentes.

Alguno de estos ejemplos o muy similares habréis hecho ya.

Interesa en particular el problema 3 (pues lo explotaremos en el siguiente handout sobre eficiencia) y el 5, que es bastante completo y podéis incluso ampliar comparando $\hat{\theta}_{MLE}$, $\hat{\theta}_M$ y $\hat{\theta}_U$ (o dos cualesquiera de ellos, uno sesgado y otro insesgado) en términos de error cuadrático medio; pero lleva bastante tiempo hacerlo.

El ejercicio 6 muestra lo requerido suponiendo lo contrario y comprobando que ello es imposible; un tipo de razonamiento que los alumnos suelen encontrar difícil y que podéis sustituir por otro más diáfano si lo encontráis.

1. (a) α^* es insesgado, α^{**} no lo es.
   (b) Respectivamente, $(2/3)^2 + (1/3)^2$ y $(2/5)^2 + (4/5)^2$.
   (c) Siendo las observaciones iguales en varianza, sin ningún recurso todavía a la noción de eficiencia, el sentido común sugiere hacerles "igual caso", sin privilegiar a una sobre otra. Si ponderamos con 1/2 cada una de ellas, obtenemos efectivamente un estimador de menor varianza.

2. (a) El parámetro θ es la media de la distribución (a diferencia de lo que ocurre con λ, que es el inverso de la media). El estimador MLE de θ es $\overline{X}$, que fácilmente se ve que es insesgado. Es también consistente, pues además de insesgado, su varianza tiende a cero.
   (b) El estimador MLE de λ es $1/\overline{X}$, que no es insesgado: el valor medio de una función no lineal de $\overline{X}$, tal como $g(\overline{X}) = 1/\overline{X}$ **no** es en general igual a la función del valor medio, $g(E[\overline{X}]) = 1/E[\overline{X}] = 1/\theta = \lambda$.

3. La probabilidad de obtener una respuesta afirmativa es la probabilidad de obtener "cara" (C) más la probabilidad de obtener "cruz" (+) y que la persona interrogada fume marihuana. Si la proporción de personas fumando marihuana es p tenemos:

   $$P(\text{'Sí'}) = P(C) + P(+) \times p.$$

   Estamos interesados en estimar insesgadamente p. Un estimador insesgado de $P(\text{'Sí'})$ es simplemente s/n. Por tanto,

   $$E[s/n] = \frac{1}{2} + \frac{1}{2}p$$

   de donde resulta que:

   $$E\left[\frac{s/n - \frac{1}{2}}{\frac{1}{2}}\right] = p.$$

   Por tanto,

   $$\frac{s/n - \frac{1}{2}}{\frac{1}{2}} = 2(s/n) - 1$$

   es un estimador insesgado de p.

4. La consistencia deriva del hecho de ser $\overline{X}_n$ insesgado y con varianza que tiende a cero. La varianza es pq/n y tiende a cero a la misma velocidad que $1/n$; la desviación típica converge a cero a la misma velocidad que $1/\sqrt{n}$.

   En general, en especial para examinar convergencia de varios estimadores al hacerse n muy grande, se recurre a comparar expresiones como $\sqrt{n}(\hat{\theta}_n - \theta)$. $\sqrt{n}$, o el orden de convergencia que rija en cada caso, provee la "amplificación" suficiente para que la varianza de $\sqrt{n}(\hat{\theta}_n - \theta)$ converja a una constante. Podemos comparar así dos estimadores consistentes

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &\rightarrow N(0, \sigma^2) \\ \sqrt{n}(\hat{\theta}_{*n} - \theta) &\rightarrow N(0, \sigma_*^2) \end{aligned}$$

   comparando σ^2 y σ_*^2. Sin la "amplificación", las varianzas convergerían en ambos casos a cero y la distribución asintótica sería en ambos casos causal, lo que dejaría en la penumbra el hecho de que uno de los dos estimadores puede ser mucho mejor que el otro.

5. (a) Haciendo uso de la sugerencia, construyamos un entorno de θ de radio $\epsilon > 0$, tan pequeño como queramos. Para probar consistencia de $X_{(n)}$ basta mostrar $X_{(n)}$ estará en dicho intervalo con probabilidd que podemos aproximar a 1 tanto como deseemos sin más que aumentar lo suficiente n. En efecto,

$$\begin{aligned} P(X_{(n)} \in (\theta - \epsilon, \theta]) &= 1 - P(X_{(n)} \leq \theta - \epsilon) \\ &= 1 - P(\cap_{i=1}^n X_i \leq \theta - \epsilon) \\ &= 1 - \left(\frac{\theta - \epsilon}{\theta}\right)^n; \end{aligned}$$

   y es claro que la última expresión converge a 1.

   (b) Haciendo uso de la sugerencia,

$$F_{\hat{\theta}_{MLE}}(x) = P(\hat{\theta}_{MLE} \leq x) = P(\cap_{i=1}^n X_i \leq x) = \left(\frac{x}{\theta}\right)^n$$

   Podemos obtener la función de densidad derivando la distribución recién obtenida:

$$f_{\hat{\theta}_{MLE}}(x) = \frac{nx^{n-1}}{\theta^n}$$

   (c) Tenemos que

$$E[\hat{\theta}_{MLE}] = \int_0^\theta \frac{nx^n}{\theta^n} dx = \left[\frac{nx^{n+1}}{\theta^n(n+1)}\right]_0^\theta = \frac{n\theta}{n+1}$$

   El sesgo es la diferencia entre este valor y el verdadero valor θ. Para cancelarlo, bastaría multiplicar $\hat{\theta}_{MLE}$ por $(n+1)/n$:

$$\hat{\theta}_U = \frac{n+1}{n}\hat{\theta}_{MLE} = \frac{n+1}{n}X_{(n)}.$$

(d) Haciendo uso de la sugerencia, basta que calculemos el segundo momento de $\hat{\theta}_U$ y restemos su media (θ) al cuadrado:

$$\begin{aligned}
\alpha_2 &= \left(\frac{n+1}{n}\right)^2 \int_0^\theta \frac{nx^{n-1}}{\theta^n} x^2 dx \\
&= \left(\frac{n+1}{n}\right)^2 \left[\frac{nx^{n+2}}{(n+2)\theta^n}\right]_0^\theta \\
&= \frac{(n+1)^2}{n(n+2)}\theta^2
\end{aligned}$$

La varianza de $\hat{\theta}_U$ es por tanto:

$$\mathrm{Var}(\hat{\theta}_U) = \frac{(n+1)^2}{n(n+2)}\theta^2 - \theta^2 = \frac{\theta^2}{n(n+2)}$$

(e) Podría compararse este estimador de θ con el estimador (también insesgado) obtenible por el método de momentos, $\hat{\theta}_M = 2\overline{X}_n$, cuya varianza es $\theta^2/3n$. Ambos son consistentes, pero $\hat{\theta}_U$ es n-consistente mientras que $\hat{\theta}_M$ es $\sqrt{n}$-consistente.

6. Supongamos que con probabilidad $\eta > 0$ Med no está en el entorno $(m-\epsilon, m+\epsilon)$, sea cual fuere el valor de n. Entonces, con dicha probabilidad η ha de estar a la izquierda de $m-\epsilon$ o a la derecha de $m+\epsilon$; esto nos conduce a un absurdo.

   En efecto, la probabilidad de estar a la izquierda de $m-\epsilon$ es $p < 0.5$. La frecuencia relativa o proporción de observaciones a la izquierda de $m-\epsilon$ converge en probabilidad a $p < 0.5$; ello excluye que Med, permanezca en dicha región al crecer n, porque de hacerlo la frecuencia relativa no podría bajar de 0.5.

   El mismo razonamiento es de aplicación a la región a la derecha de $m+\epsilon$: la frecuencia relativa de observaciones en dicha región converge en probabilidad a $p < 0.5$, y sin embargo la permanencia de Med en la misma supondría que dicha frecuencia relativa es al menos 0.5. Por consiguiente, la probabilidad de que Med esté en la unión de ambas regiones converge a 0, contra la hipótesis de que es $\eta > 0$.