Handout 7.

Method of moments and maximum likelihood estimation

1. Let X be distributed with density function $f_X(x;\theta) = (1+\theta)x^\theta$ for $0 \le x \le 1$ and $\theta > -1$. Obtain the method of moments estimator and the MLE based on a sample of n independent observations.

2. Consider the so-called Rayleigh distribution, whose density is given by:

$$f_X(x;\theta) = \left(\frac{x}{\theta}\right)\exp(-x^2/2\theta) \qquad (0 \le x < \infty)$$

   (a) Find the method of moments estimator of θ based in a sample on n independent observations. (Hint: You can make use of the fact that

$$\int_0^\infty \frac{x^2}{\theta}\exp(-x^2/2\theta)dx = \sqrt{\frac{\theta\pi}{2}}.$$

   (b) Find the MLE of θ.

3. Let X be a random variable with the following probability distribution:

$P(x)$	$\frac{2}{10}$	$\frac{2-\theta}{10}$	$\frac{\theta}{10}$	$\frac{6}{10}$
x	1	2	3	4

   (a) What are the feasible values of θ?

   (b) Consider a random sample of independent observations from this distribution, whose values happen to be:

$$1, 3, 4, 4, 1, 2, 4, 4, 4, 4, 4, 4, 4$$

   What is the method of moments estimator of θ?

   (c) What is, with the same sample, the MLE of θ?

4. A battery fires continuously shells on targets, all of them of the same size and at the same distance. A bored soldier wants to estimate the probability p that a shell hits its target. The following sequence of questions, that you must answer in his place, suggests one way of doing it.

   (a) What is the probability of hitting the target with the first shot? With the second (we assume that the first missed the target)? With the third? With the n-th shot?

   (b) If a given target, repeatedly fired upon, is hit with the 3rd. shot, can you work the method of moments estimate of p?

   (c) What would be the MLE of p?

(d) After the first target is hit, the soldier witnesses the destruction of another two, of the same size and at the same distance, requiring respectively 2 and 4 shots. What would be the new MLE of p encompassing all information?

5. The negative binomial distribution can be seen as the general case of the preceding problem. If we have a probability p of success in one trial, the number X of trials needed to achieve k successes is negative binomial. Obtain the probability function of X (which depends on k) and the maximum likelihood estimator of p for given k given a single observation X from a negative binomial.

6. Consider the Cauchy distribution (Student's t with 1 degree of freedom) shifted by θ. The density is given by:

$$f_X(x;\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2} \qquad (-\infty < x < \infty)$$

   Can we find a moment estimator of θ? Find out what would be the maximum likelihood estimator of θ based on a sample of size 2.

   (a) What happens when n grows large? (Hint: Here you have an example of an intractable MLE).

   (b) Suggest a possible estimator of θ.

Aside from the references below, looking in the Wikipedia for the name of a distribution will provide you with such basic information as the moments, characteristic function, etc. and is a good source to further references.

References

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

A. Jeffrey. *Handbook of Mathematical Formulas and Integrals.* Academic Press, 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.

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

Bosquejos de respuesta, pautas docentes

Los ejemplos en las transparencias son de mejor rendimiento que estos, y debieran hacerse en primer lugar. A continuación pueden hacerse los que siguen u otros similares, hasta donde el tiempo permita.

Conceptualmente son todos muy simples; se hace uso sin embargo de algunos resultados (integrales, suma de series) que puede ser un poco largo justificar, y no creo que convenga hacerlo.

1. La media de la distribución proporcionada es:

$$\int_0^1 (1+\theta)x^{(\theta+1)}dx = \left[\frac{\theta+1}{\theta+2}x^{(\theta+2)}\right]_0^1 = \frac{\theta+1}{\theta+2}$$

   Para obtener el estimador por el método de momentos, igualamos esto al primer momento muestral $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ y obtenemos:

$$\frac{\theta+1}{\theta+2} = \overline{X}_n$$

   y despejando $\hat{\theta}_M = (2\overline{X}_n - 1)/(1-\overline{X}_n)$.

   Para obtener el estimador máximo-verosímil, basta escribir la verosimilitud

$$f(\vec{x};\theta) = (1+\theta)^n \left(\prod_{i=1}^n x_i^\theta\right)$$

   cuyo logaritmo es

$$\log f(\vec{x};\theta) = n\log(1+\theta) + \theta\sum_{i=1}^n \log(x_i).$$

   Derivando respecto a θ e igualando a cero tenemos:

$$\hat{\theta}_{MLE} = \frac{-n}{\sum_{i=1}^n \log(x_i)} - 1.$$

   En este caso, ambos estimadores son bien diferentes.

   Hacedles notar que el estimador MLE sólo depende de la muestra a través de $\sum_{i=1}^n \log(x_i)$. Más tarde, cuando introduzcamos estadísticos suficientes, es útil volver sobre estos ejemplos y recordarles que en multitud de casos en que hemos hecho estimación máximo-verosímil los estadísticos suficientes han aparecido por sí solos.

2. No creo que merezca la pena invertir tiempo en hacer la integral (2a) por partes. Basta decirles el resultado. Alternativamente, con un poco de ingenio, pueden reconducir tal integral a una conocida:

$$\begin{aligned}\int_0^\infty \frac{x^2}{\theta}\exp(-x^2/2\theta)dx &= \sqrt{\frac{2\pi}{\theta}}\int_0^\infty \frac{x^2}{\sqrt{2\pi\theta}}\exp(-x^2/2\theta)dx \\ &= \sqrt{\frac{2\pi}{\theta}} \times \frac{\theta}{2} \\ &= \sqrt{\frac{\pi\theta}{2}}\end{aligned}$$

El truco consiste en ver que la última integral sería la mitad de la varianza de una $N(0,\sigma^2=\theta)$, y por tanto toma el valor $\theta/2$.

Haciendo uso del resultado, el estimador de por el método de momentos es resultado de resolver

$$\sqrt{\frac{\pi\theta}{2}} = \overline{X}_n$$

de donde obtenemos $\hat{\theta}_M = 2\overline{X}_n^2/\pi$. Por su parte, el estimador máximo verosímil resulta de maximizar la verosimilitud

$$f(\vec{x};\theta) = \prod_{i=1}^n (x_i/\theta)\exp\left(-\sum_{i=1}^n x_i^2/2\theta\right)$$

cuyo logaritmo es:

$$\log f(\vec{x};\theta) = -n\log(\theta) + \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n x_i^2/2\theta.$$

Derivando respecto a θ e igualando a cero,

$$\frac{\partial \log f(\vec{x};\theta)}{\partial\theta} = -n/\theta - \left(\sum_{i=1}^n x_i^2/2\right)\left(-\frac{1}{\theta^2}\right) = 0,$$

de donde despejamos $\hat{\theta}_{MLE} = \sum_{i=1}^n x_i^2/2n$.

Hacedles notar que el estimador MLE sólo depende de la muestra a través de $\sum x_i^2$.

3. (a) Para que las probabilidades estén entre 0 y 1, θ no puede se menor que 0 ni mayor que 2.

   (b) La media de la distribución es

   $$\frac{2}{10} + 2\times\frac{2-\theta}{10} + 3\times\frac{\theta}{10} + 4\times\frac{6}{10} = \frac{\theta}{10} + 3$$

   Por tanto, el estimador por el método de momentos resulta de resolver $\overline{X}_n = \frac{\theta}{10} + 3$, lo que da $\hat{\theta}_M = 10\overline{X}_n - 30$. En nuestro caso, $\overline{X}_n = 3.307692$ y $\hat{\theta}_M = 3.07692$. Hay que notar que el estimador queda fuera del rango factible.

   (c) La verosimilitud es

   $$\left(\frac{2}{10}\right)^2 \times \left(\frac{2-\theta}{10}\right) \times \left(\frac{\theta}{10}\right) \times \left(\frac{6}{10}\right)^9.$$

   Tomando logaritmos, derivando respecto de θ e igualando a cero obtenemos: $\hat{\theta}_{MLE} = 1$.

4. (a) Con el primero, p. Con el segundo, $(1-p)p$. Con el tercero, $(1-p)^2p$. Con el n-simo, $(1-p)^{n-1}p$.

(b) Conceptualmente, no es difícil. Pero el valor medio de la distribución sería:

$$m = 1 \times p + 2 \times (1-p)p + 3 \times (1-p)^2 p + \dots$$

Esta es una serie aritmético-geométrica que se puede sumar (Jeffrey (1995), p. 65, formula 1.8.3.1; el valor en este caso es $1/p$), pero es engorrosa de utilizar. Contra lo que es usual, en este caso el estimador máximo verosímil es más simple.

(c) El estimador máximo verosímil resulta de maximizar:

$$f(x;p) = (1-p)^2 p$$

o su logaritmo $2\log(1-p) + \log(p)$ con respecto a p. Derivando esta última expresión e igualando a cero obtenemos:

$$\frac{-2}{1-p} + \frac{1}{p} = 0$$

de donde $\hat{p}_{MLE} = 1/3$. (El estimador de momentos hubiera resultado de igualar $1/p$ a 3 y habría tomado el mismo valor.)

(d) La verosimilitud conjunta de las tres observaciones es ahora:

$$f(\vec{x};p) = (1-p)^2 p \times (1-p)p \times (1-p)^3 p = (1-p)^6 p^3$$

y derivando el logaritmo e igualando a cero como antes,

$$\frac{-6p}{p(1-p)} + \frac{3-3p}{p(1-p)} = 0,$$

de donde obtenemos de nuevo $\hat{p}_{MLE} = 1/3$.

5. La función de probabilidad pedida es:

$$P(X = x) = \binom{x-1}{k-1} p^{k-1}(1-p)^{x-k} \times p$$

("probabilidad de obtener exactamente $k-1$ éxitos en los primeros $x-1$ ensayos y un nuevo éxito en el x-ésimo").

Dada una observación x habríamos de maximizar respecto a p la expresión anterior o su logaritmo:

$$\log \binom{x-1}{k-1} + k\log(p) + (x-k)\log(1-p).$$

Derivando e igualando a cero obtenemos

$$\frac{k}{p} - (x-k)\frac{1}{1-p} = 0$$

de donde resulta $\hat{p}_{MLE} = k/x$. (Acontece, aunque es laborioso de demostrar, que el valor medio de la binomial negativa es $m = k/p$; el método de momentos nos hubiera dado el mismo resultado.)

6. (a) Para sólo dos observaciones, habríamos de maximizar algo como:

$$\frac{1}{\pi^2}\frac{1}{1+(x_1-\theta)^2}\frac{1}{1+(x_2-\theta)^2}$$

que es bien factible de realizar, al menos numéricamente (las derivadas se hacen laboriosas).

(b) Al crecer n, la expresión anterior se hace cada vez más compleja al involucrar en el denominador un polinomio de grado creciente en θ y la maximización se torna rápidamente impracticable.

(c) Dado que la distribución es simétrica en torno a θ, la mediana parece un estimador razonable.

Este ejemplo es discutido ampliamente en el Seminario 2.