SORT 30 (1) January-June 2006, 91-100

About one problem of Bernoulli and Euler from the theory of statistical estimation

Mikhaïl Nikulin

Université Bordeaux

Abstract

We consider some results by D. Bernoulli and L. Euler on the method of maximum likelihood in parametric estimation. The statistical analysis is made by considering a parametric family with a shift parameter.

MSC: 62H03, 62F10

Keywords: Asymptotic normality, sample mean, Bernoulli's estimator, Euler's estimator, maximum likelihood, unbiased estimator.

1 Introduction

Kendall (1961) published a paper on Daniel Bernoulli and the maximum likelihood. This paper quotes two papers: Bernoulli (1961) and Euler (1961). The paper of D. Bernoulli and the commentary by Euler appeared in Latin (1777). An interesting discussion about this problem can be found in Stigler (1997).

We shall consider here one contribution of D. Bernoulli and L. Euler in the estimation of parameters, in particular on the method of maximum likelihood. For more aspects about the principle of maximum likelihood (ML) in estimation, see, for example, Huber and Nikulin (1997).

Address for correspondence: Mikhaïl Nikulin. Université Bordeaux
Received: April 2006
Accepted: May 2006

Bernoulli and Euler considered the problem of statistical estimation of the parameter θ of the probability density

$$p(x;\theta) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-(x-\theta)^2} & \text{if} \quad |x-\theta| \le 1, \quad |\theta| < \infty. \\ 0 & \text{otherwise.} \end{cases}$$

Bernoulli proposed to estimate θ by the ML method. Euler agreed with Bernoulli, but he provided a different estimator. Who was right? This question was posed by L.N. Bolshev in 1969. We shall consider here both approaches under a more general case.

Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)^T$ be a sample, where $X_1, X_2, \cdots, X_n$ are independent identically distributed random variables with density

$$p_k(x;\theta) = \begin{cases} \dfrac{\Gamma(2k)}{\Gamma(k)\Gamma(k)}\left(\dfrac{1}{2}\right)^{2k-1}\left[1-(x-\theta)^2\right]^{k-1} & \text{if} \quad |x-\theta| \le 1, \\ 0 & \text{if} \quad |x-\theta| > 1. \end{cases}$$

We have to estimate the shift parameter θ, where k is given, with $k \in [1,2]$.

The family $\{p_k(x;\theta)\}$ is quite rich. In particular, if $k = 1$ it contains the uniform distribution with support

$$\theta - 1 \le x \le \theta + 1, \quad |\theta| < \infty.$$

If $k = 1.5$ (the case of Bernoulli) the graph of $p_{1.5}(x;\theta) = p(x;\theta)$ is a half ellipse with parameters $a = 1$ and $b = 2/\pi$, and two tangents to the extremes of the curve, $x = \theta - 1$ and $x = \theta + 1$, orthogonal to the axis OX.

L. N. Bolshev proposed to find the ML estimate of θ, as Bernoulli did for the case $k = 1.5$.

Let us denote

$$L(\theta) = \prod_{i=1}^{n} p_k(X_i;\theta)$$

the likelihood function, obtained with the data $\mathbf{X}$, and let $\hat{\theta}_n$ be the value of θ that maximises $L(\theta)$:

$$L(\hat{\theta}_n) = \max_{|\theta|<\infty} L(\theta),$$

with the constraint $\max_i |X_i - \theta| \le 1$, i.e., $0 \le X_{(n)} - X_{(1)} \le 2$, where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ is the ordinal statistic.

As is well known, it is more convenient to consider $\ln L(\theta)$. In our case we have

$$\ln L(\theta) = (k-1)\sum_{i=1}^{n} \ln\left[1-(X_i-\theta)^2\right] + n\ln\frac{\Gamma(2k)}{\Gamma(k)\Gamma(k)2^{2k-1}}. \tag{1.1}$$

To find $\hat{\theta}$ we must solve the ML equation:

$$\frac{\partial}{\partial\theta}\ln L(\theta) = 0. \tag{1.2}$$

From (1.1) and (1.2) we obtain:

$$\sum_{i=1}^{n}\frac{X_i-\theta}{1-(X_i-\theta)^2} = 0. \tag{1.3}$$

Following Euler, this equation can be expressed as:

$$\sum_{i=1}^{n}\frac{1}{1+X_i-\theta} = \sum_{i=1}^{n}\frac{1}{1-X_i+\theta}.$$

It is worth noting that (1.3) does not depend on k. A solution $\hat{\theta}_n$ is the ML estimator, which was proposed by Bernoulli. One can verify that $\hat{\theta}_n$ satisfies the equation

$$\hat{\theta}_n = \frac{\sum_{i=1}^{n}\left\{\frac{1}{1-(X_i-\hat{\theta}_n)^2}X_i\right\}}{\sum_{i=1}^{n}\frac{1}{1-(X_i-\hat{\theta}_n)^2}}. \tag{1.4}$$

More exactly, we can say that (1.4) is (1.3) "solved" with respect to θ. We can find $\hat{\theta}_n$ by using iterative procedures.

The same problem was considered by Euler, who knew Bernoulli's result. Euler proposed the estimator

$$\theta_n^* = \frac{\sum_{i=1}^{n}\left\{\left[1-(X_i-\theta_n^*)^2\right]X_i\right\}}{\sum_{i=1}^{n}\left[1-(X_i-\theta_n^*)^2\right]}. \tag{1.5}$$

Note the difference between θ_n^* and $\hat{\theta}_n$, as in (1.4) and (1.5) the observations X_i have different weights.

Clearly, to estimate θ we can also take the arithmetic mean

$$\bar{X}_n = \frac{1}{n}(X_1+\cdots+X_n).$$

Our aim is to compare these three estimators $\hat{\theta}_n$, θ_n^* and $\overline{X}_n$. We can suppose $\theta = 0$, i.e. we can consider that $\hat{\theta}_n, \theta_n^*$ and $\overline{X}_n$ are estimators of zero, as they are invariant under translation when the loss is quadratic.

2 Arithmetic mean

Since $\mathrm{E}(X_i) = 0$, we obtain that $\overline{X}_n$ is an unbiased estimator of 0, and $\mathrm{Var}(X_i) = \mathrm{E}(X_i)^2$. To compute $\mathrm{Var}(X_i)$ let us find the moments. It is evident that

$$\mathrm{E}(X_i)^{2m+1} = 0, \quad m = 0, 1, 2, \cdots \tag{2.1}$$

On the other hand we have ($\theta = 0$):

$$\begin{aligned} \mathrm{E}(X_i)^{2m} &= \frac{\Gamma(2k)}{\Gamma(k)\Gamma(k)} \left(\frac{1}{2}\right)^{2k-1} \hat{E} \int_{-1}^{1} x^{2m}(1-x^2)^{k-1}dx \\ &= \frac{\Gamma(2k)}{\Gamma(k)\Gamma(k)} \left(\frac{1}{2}\right)^{2k-1} B\left(m+\frac{1}{2}, k\right) \\ &= \frac{\Gamma(2k)\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(k)\Gamma\left(m+k+\frac{1}{2}\right)} \left(\frac{1}{2}\right)^{2k-1}. \end{aligned} \tag{2.2}$$

Table 1

$k \backslash m$	1	2	3
1	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{7}$
$\frac{3}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{64}$
2	$\frac{1}{5}$	$\frac{3}{35}$	$\frac{1}{21}$

From (2.2) we obtain the Table 1 for some values of $\mathrm{E}(X_i)^{2m}$.

From (2.1), (2.2) and Table 1 it follows that

$$\mathrm{Var}(X_i) = \mathrm{E}(X_i)^2 = \frac{\sqrt{\pi}}{2^{2k}} \frac{\Gamma(2k)}{\Gamma(k)\Gamma\left(k+\frac{3}{2}\right)}. \tag{2.3}$$

(In the particular case $k = 1.5$, i.e., in the cases of Bernoulli and Euler we have $\mathrm{Var}(X_i) = 1/4$). From (2.3) we obtain

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n}\frac{\sqrt{\pi}}{2^{2k}}\frac{\Gamma(2k)}{\Gamma(k)\Gamma\left(k+\frac{3}{2}\right)}, \tag{2.4}$$

and hence we have

$$\mathrm{Var}(\bar{X}_n) = \begin{cases} \dfrac{1}{3n} & \text{if } k = 1, \\[2ex] \dfrac{1}{4n} & \text{if } k = \frac{3}{2}, \\[2ex] \dfrac{1}{5n} & \text{if } k = 2. \end{cases} \tag{2.5}$$

Furtheremore, from the central limit theorem for any $k \in [1, 2]$ and any $x \in \boldsymbol{R}$

$$\lim_{n\to\infty} \mathrm{P}\left(\left\{\bar{X}_n \le x\sqrt{\frac{1}{n}\frac{\sqrt{\pi}}{2^{2k}}\frac{\Gamma(2k)}{\Gamma(k)\Gamma\left(k+\frac{2}{3}\right)}}\right\}\right) = \Phi(x), \tag{2.6}$$

where $\Phi(x)$ is the cdf of the $N(0, 1)$ distribution.

3 Euler's estimator θ_n^*

Recall that $\theta = 0$. From (1.5) θ_n^* is the solution of the equation

$$\theta_n^* \sum_{i=1}^{n}\left[\left(1 - X_i^2\right) + 2X_i\theta_n^* - (\theta_n^*)^2\right] = \sum_{i=1}^{n}\left[\left(1 - X_i^2\right) + 2X_i\theta_n^* - (\theta_n^*)^2\right] X_i. \tag{3.1}$$

Let us denote $\bar{X}_n^k = \frac{1}{n}\sum_{i=1}^n X_i^k$. It is evident that $\{\bar{X}_n\}$ converges in probability to 0, i.e., $\mathrm{P}\left(\lim_{n\to\infty}\bar{X}_n\right) = 0$, and similarly $\mathrm{P}\left(\lim_{n\to\infty}\bar{X}_n^3\right) = 0$. Furthermore, we have

$$\mathrm{P}\left(\lim_{n\to\infty}\bar{X}_n^2\right) = \mathrm{E}(X_i)^2 = \begin{cases} \dfrac{1}{3} & \text{if } k = 1, \\[2ex] \dfrac{1}{4} & \text{if } k = \frac{3}{2}, \\[2ex] \dfrac{1}{5} & \text{if } k = 2. \end{cases} \tag{3.2}$$

Next, let us consider equation (3.1) in terms of $\overline{X}_n$

$$\theta_n^*(1 - 3\overline{X}_n^2) = \overline{X}_n - \overline{X}_n^3 - 3\overline{X}_n(\theta_n^*)^2 + (\theta_n^*)^3, \tag{3.3}$$

which is an equation of third degree and hence there exists at least one real root. From (3.2) we have

$$\mathrm{P}\left(\lim_{n\to\infty}(1 - 3\overline{X}_n^2)\right) = 1 - 3\mathrm{E}(X_i)^2 \geq 0,$$

so, by taking limits in both sides of (3.3), we get the equation

$$(1 - 3\mathrm{E}(X_i)^2)\theta^* = (\theta^*)^3, \tag{3.4}$$

whose three roots

$$\theta_1^* = 0, \qquad \theta_2^* = \sqrt{1 - 3\mathrm{E}(X_i)^2}, \qquad \theta_3^* = -\sqrt{1 - 3\mathrm{E}(X_i)^2}, \tag{3.5}$$

are not random, where $\mathrm{P}\,(\lim_{n\to\infty}\theta_n^*) = \theta^*$. Clearly, the roots θ_i^* of (3.5) are very close to the roots θ_{ni}^* of (3.4).

Now, if we consider once again equation (3.3) we can write

$$\sqrt{n}\theta_{n1}^* = \frac{\sqrt{n}\left[\left(\overline{X}_n - \overline{X}_n^3\right) - 3\overline{X}_n\theta_{n1}^{*2}\right]}{1 - 3\overline{X}_n^2 - \theta_{n1}^{*2}}. \tag{3.6}$$

It is evident that the numerator of (3.6) is asypmtotically normal distributed with parameters

$$\mu_n = 0, \quad \sigma_n^2 = \mathrm{E}(X_i - X_i^3)^2.$$

We also have

$$\mathrm{P}\left(\lim_{n\to\infty}(1 - 3\overline{X}_n^2 - \theta_{n1}^{*2})\right) = 1 - 3\mathrm{E}(X_i)^2 = \begin{cases} 0 & \text{if } k = 1, \\ \dfrac{1}{4} & \text{if } k = \frac{3}{2}, \\ \dfrac{2}{5} & \text{if } k = 2. \end{cases} \tag{3.7}$$

What do this result mean? If $k = 1$, then

$$\theta_{n1}^* = \sqrt[3]{-\overline{X}_n + \overline{X}_n^3 + 3\overline{X}_n\theta_{n1}^{*2} + \theta_{n1}^{*2}(1 - 3\overline{X}_n^2)},$$

so

$$n^{1/6}\theta^*_{n1} = \sqrt[3]{-\sqrt{n}(\overline{X}_n - \overline{X^3_n}) + 3\hat{E}\theta^{*2}_{n1}\sqrt{n}\overline{X}_n + \sqrt{n}(1 - 3\overline{X^2_n})\theta^*_{n1}}.$$

On the other hand, since

$$\mathrm{P}\left(\lim_{n\to\infty} \theta^*_{n1}\right) = 0 \quad \text{and} \quad \mathrm{P}\left(\lim_{n\to\infty} \theta^{*2}_{n1}\right) = 0,$$

we obtain

$$\mathrm{P}\left(\lim_{n\to\infty} \theta^{*2}_{n1}\sqrt{n}\overline{X}_n\right) = 0, \quad \mathrm{P}\left(\lim_{n\to\infty} \sqrt{n}(1 - 3\overline{X^2_n})\theta^*_{n1}\right) = 0$$

and hence we have

$$\begin{aligned} \lim_{n\to\infty} \mathrm{P}\left(\left\{n^{1/6}\theta^*_{n1} < x\right\}\right) &= \lim_{n\to\infty} \mathrm{P}\left(\left\{\sqrt{n}\theta^{*3}_{n1} < x^3\right\}\right) \\ &= \lim_{n\to\infty} \mathrm{P}\left(\left\{-\sqrt{n}(\overline{X}_n - \overline{X^3_n}) < x^3\right\}\right) \\ &= \Phi\left(x^3 / \sqrt{\mathrm{E}(X_i - X^3_i)^2}\right). \end{aligned} \tag{3.8}$$

Since

$$\mathrm{E}(X_i - X^3_i)^2 = \mathrm{E}(X_i)^2 - 2\mathrm{E}(X_i)^4 + \mathrm{E}(X_i)^6,$$

from (2.2) we find that

$$\mathrm{E}(X_i - X^3_i)^2 = \begin{cases} \dfrac{8}{105} & \text{if } k = 1, \\ \dfrac{5}{64} & \text{if } k = 1.5, \\ \dfrac{8}{105} & \text{if } k = 2. \end{cases} \tag{3.9}$$

Hence from (3.8) and (3.9) it follows that if $k = 1$ then the sequence $\{n^{1/6}\theta^*_{n1}\}$ converges in distribution as $n \to \infty$ to a random variable $Z^{1/3}$, where Z is normal $N(0, 8/105)$. Thus

$$\lim_{n\to\infty} \mathrm{P}\left(\{n^{1/6}\theta^*_{n1} \leq x)\}\right) = \Phi\left(x^3 / \sqrt{\mathrm{E}(X_i - X^3_i)^2}\right).$$

With the help of (3.8) and (3.9) it follows that

$$E(n^{1/6}\theta^*_{n1})^2 = n^{1/3}E(\theta^{*2}_{n1}) \cong 3\sqrt{\frac{105}{8}}\int_{-\infty}^{\infty}\varphi\left(\sqrt{\frac{105}{8}}x^3\right)x^4dx < \infty,$$

where $\varphi(x) = \Phi'(x)$, i.e., when $k = 1$

$$E(\theta^{*2}_{n1}) = O\left(\frac{1}{n^{1/3}}\right). \tag{3.10}$$

On the other side if $1 < k \leq 2$, then $E(1 - 3X_i^2) > 0$ and hence the sequence $\{\sqrt{n}\theta^*_{n1}\}$ converges in distribution as $n \to \infty$ to a random variable Z normal $N(0, \sigma^2)$, where $\sigma^2 = [E(X_i - X_i^3)^2]/[E(1 - 3X_i^2)]^2$. In particular, from (3.7) and (3.9) (see also Table 1) we obtain

$$\frac{E(X_i - X_i^3)^2}{\left[E(1 - 3X_i^2)\right]^2} = \begin{cases} \frac{5}{64}/\left(\frac{1}{4}\right)^2 = \frac{5}{4} & \text{if } k = 1.5, \\ \frac{8}{105}/\left(\frac{2}{5}\right)^2 = \frac{10}{21} & \text{if } k = 2. \end{cases} \tag{3.11}$$

A simple comparison of (3.10), (3.11) and (2.5) shows that $\bar{X}_n$ is better than Euler's estimator. Finally, note the different rates of convergence of the estimators $\hat{\theta}_n$ and θ^*_{n1}.

4 Bernoulli's (ML) estimator

Let us consider the statistic

$$T = \frac{X_{(1)} + X_{(n)}}{2},$$

as was done, for example, by Voinov and Nikulin (1993, pp. 47-51). Clearly $E(T = 0)$ and from (2.2) it follows that

$$\text{Var}(T = 2)\left[\frac{\Gamma(k)\Gamma(k+1)}{\Gamma(2k)}\right]^{2/k}\left[\Gamma\left(\frac{2}{k}+1\right) - \Gamma^2\left(\frac{1}{k}+1\right)\right]\frac{1}{n^{2/k}}(1 + o(1)). \tag{4.1}$$

From (4.1) with large values of n we have

$$\mathrm{Var}(T) \sim \begin{cases} \dfrac{2}{n^2} < \frac{1}{n} & \text{if } k = 1, \\ \dfrac{1}{18}\left(\dfrac{3\pi}{2}\right)^{4/3}\left[\Gamma\left(\dfrac{1}{3}\right) - \Gamma^2\left(\dfrac{2}{3}\right)\right]\dfrac{1}{n^{4/3}} < \dfrac{1}{4n} & \text{if } k = 3/2, \\ \dfrac{2}{3}\left(1 - \dfrac{\pi}{4}\right)\dfrac{1}{n} < \dfrac{1}{5n} & \text{if } k = 2. \end{cases} \tag{4.2}$$

Since $\mathrm{Var}(X_{(n)}) = \mathrm{Var}(X_{(1)}) = O(n^{-2/k})$, we also have

$$\mathrm{Var}(\hat{\theta}_n) = O(n^{-2/k}). \tag{4.3}$$

Now it is clear (compare (4.2), (4.3), and (2.5)), that Bernoulli's estimator $\hat{\theta}_n$ is better than $\overline{X}_n$ and θ^*_{n1}. It is also clear that, in practice, for $1 \leq k \leq 2$ it is reasonable to use the above statistic T to estimate θ.

Acknowledgements

The author would like to thank Professors T. Smith (Queen's University, Kingston, Canada), C. M. Cuadras (University of Barcelona, Spain), and V. Solev (Steklov Mathematical Institute, St Petersburg, Russia) for helpful discussions and remarks.

Bibliography

Bernoulli, D. (1777). Memoirs of the Academy of St. Petersburg, *Acta Acad. Petrop.*, 3-33.

Bernoulli, D. (1961). The most probable choice between several discrepant observations and the formation therefrom of the most likely induction, *Biometrika*, 1961, 48, 3-13.

Euler, L. (1961). Observation on the foregoing dissertation of Bernoulli, *Biometrika*, 1961, 48, 13-18.

Huber, C., Nikulin, M. (1997). Remarques sur le maximum de vraisemblance. *Qüestiió*, 21, 37-58.

Kendall, M. G. (1961). Studies in the history of probability and statistics XI. Daniel Bernoulli and maximum likelihood, *Biometrika*, 48, 1-18.

Stigler, S. M. (1997). Daniel Bernoulli, Leonhard Euler, and maximum likelihood (including a new translation of a paper by D. Bernoulli). In: *Festschrift for Lucien Le Cam*, (Eds. David Pollard, Eric Torgensen, Grace L. Yang), Springer. New York, pp. 345-357.

Voinov, V., Nikulin, M. (1993). *Unbiased Estimators and their Applications*. Vol. 1: *Univariate Case*, Kluwer Academic Publishers. Dordrecht.