The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://www.tandfonline.com/loi/utas20

# Examples of Nonunique Maximum Likelihood Estimators

Sudhakar Dharmadhikari & Kumar Joag-Dev

**To cite this article:** Sudhakar Dharmadhikari & Kumar Joag-Dev (1985) Examples of Nonunique Maximum Likelihood Estimators, The American Statistician, 39:3, 199-200

**To link to this article:** http://dx.doi.org/10.1080/00031305.1985.10479427

Published online: 12 Mar 2012.

# TEACHER'S CORNER

In this department *The American Statistician* publishes articles, reviews, and notes of interest to teachers of the first mathematical statistics course and of applied statistics courses. The department includes the Accent on Teaching Materials section; suitable contents for the section are described under the section heading. Articles and notes for the department, but not intended specifically for the section, should be useful to a substantial number of teachers of the indicated types of courses or should have the potential for fundamentally affecting the way in which a course is taught.

# Examples of Nonunique Maximum Likelihood Estimators

SUDHAKAR DHARMADHIKARI and KUMAR JOAG-DEV*

This article presents a large class of probability densities $f(x, \theta)$ for which, with positive probability, the maximum likelihood estimator $\hat{\theta}$ based on a sample of size 2 is non unique, and the possible values of $\hat{\theta}$ do not form an interval. Such a density $f(x, \theta)$ can even be chosen to be unimodal, and one such example is the Cauchy density with a location parameter. A discrete version of the argument gives examples in which the nonuniqueness of the maximum likelihood estimator persists for samples of arbitrary size.

KEY WORDS: Unimodal distribution; Log concave distribution; Pólya frequency function.



## 1. INTRODUCTION

Let $(X_1, \ldots, X_n)$ be a random sample from a distribution with density $f(x, \theta)$. The usual example in which the maximum likelihood estimator (MLE) $\hat{\theta}$ of $\theta$ is not unique takes $f(x, \theta)$ to be the uniform density on, say, $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$; see Bickel and Doksum (1977, p. 111) or Hogg and Craig (1978, p. 207). In addition to being somewhat artificial, this example has the feature that the possible MLE's form an interval. In this article we construct a large class of examples of $f(x, \theta)$ for which, with positive probability, the MLE based on a sample of size 2 is nonunique in an essential way, in that the possible choices of $\hat{\theta}$ do not form an interval. The densities $f(x, \theta)$ are somewhat natural in the sense that they can be chosen to be unimodal so that the MLE based on a single observation is always unique. The Cauchy density with a location parameter is such an example. The arguments can be made discrete to illustrate the essential nonuniqueness of the MLE for samples of arbitrary size. For a discussion of the problem of constructing efficient estimators in the presence of nonunique MLE's, see Lehmann (1980).

*Sudhakar Dharmadhikari is Professor, Department of Mathematics, Southern Illinois University, Carbondale, IL 62901. Kumar Joag-dev is Professor, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801. This work was carried out while the first author was on sabbatical leave at Virginia Polytechnic Institute and State University. Work on this article was supported by U.S. Air Force Grant AFOSR-84-0208.

## 2. THE NONUNIQUENESS OF THE MLE FOR A CLASS OF DENSITIES

Let $g$ be a probability density on $R$ satisfying the following three conditions:

1. $g$ is continuous, symmetric about 0, and positive everywhere.
2. $g$ is twice continuously differentiable everywhere except perhaps at 0.
3. If we write $h = \log g$, then $h''(y) > 0$ for some nonzero $y$.

The ease with which such a density can be constructed is discussed later in this section.

Let $(X_1, X_2)$ be a random sample from the distribution with density $f(x, \theta) = g(x - \theta)$, where $x \in R$ and $\theta \in R$. Let $x_1$ and $x_2$ be the observed values of $X_1$ and $X_2$ and write $\bar{x} = (x_1 + x_2)/2$ and $\Delta = (x_1 - x_2)/2$. The likelihood function is given by

$$\begin{aligned} L(\theta|\mathbf{x}) &= g(x_1 - \theta)g(x_2 - \theta) \\ &= g(\bar{x} + \Delta - \theta)g(\bar{x} - \Delta - \theta). \end{aligned}$$

Assumption 1 shows that for every $k \geq 0$,

$$L(\bar{x} + k|\mathbf{x}) = L(\bar{x} - k|\mathbf{x}) = g(\Delta + k)g(\Delta - k). \quad (1)$$

Thus $L(\cdot|\mathbf{x})$ is symmetric about $\bar{x}$. Therefore either $\hat{\theta} = \bar{x}$ or $\hat{\theta}$ is not unique.

By assumptions 2 and 3, there is an interval $(a, b)$ such that for every $y \in (a, b)$, there exists a $\delta > 0$ such that

$$h(y + \delta) - h(y) > h(y) - h(y - \delta),$$

or

$$g(y + \delta)g(y - \delta) > [g(y)]^2. \quad (2)$$

Suppose that $x_1$ and $x_2$ are such that $\Delta \in (a, b)$. Then (2) and (1) show that there is a $\delta > 0$ such that

$$\begin{aligned} L(\bar{x} \pm \delta|\mathbf{x}) &= g(\Delta + \delta)g(\Delta - \delta) \\ &> [g(\Delta)]^2 = L(\bar{x}|\mathbf{x}). \end{aligned}$$

Therefore $\hat{\theta} \neq \bar{x}$, $\hat{\theta}$ is not unique, and the possible choices of $\hat{\theta}$ do not form an interval. The probability that $\hat{\theta}$ is not unique is at least as large as the probability that $|X_1 - X_2|/2$ belongs to $(a, b)$. The latter probability is positive because $g$ is positive everywhere.

We now show that there is a large supply of densities $g$ satisfying the required conditions. Conditions 1 and 2 are easy to meet. Suppose that, contrary to 3, $h''(y) \le 0$ for all nonzero $y$. Then it is easy to show that $g(x)$ tends to zero at an exponential rate as $x \to \pm\infty$, and so $g$ has a moment generating function. In particular, $g$ must have finite moments of all orders if $h'' \le 0$ everywhere. Thus if $g$ satisfies 1 and 2 and has an infinite moment of some order, then $g$ will satisfy 3 automatically. It is also clear that we can choose $g$ to be unimodal with a unique mode so that for a sample of size 1, the MLE is unique.

Densities $g$ that satisfy the condition

$$g^2(y) \ge g(y + \delta)g(y - \delta) \qquad (3)$$

for all $y \in R$ and $\delta \in R$ are called *log concave*. If $g$ is log concave, then $f(x, \theta) = g(x - \theta)$ is a *Pólya frequency function* of type 2 [see Lehmann (1959, p. 115)]. If strict inequality holds in (3) whenever $\delta \ne 0$, one can easily see that the MLE of $\theta$ is unique for samples of any size. Thus our approach to achieving nonuniqueness of $\hat\theta$ is to require that log $g$ should be strictly convex over some interval.

## 3. EXAMPLES

In this section we present two specific examples illustrating the result of Section 2. We recall that $h = \log g$, $f(x, \theta) = g(x - \theta)$, and $\hat\theta$ is the MLE of $\theta$ based on a sample of size 2. The two observations $x_1$ and $x_2$ are written as $\bar{x} \pm \Delta$.

*Example 1*. Let $g(y) = [\pi(1 + y^2)]^{-1}$. Then $h''(y) = 2(y^2 - 1)/[(1 + y^2)^2]$. Therefore $h''(y) > 0$ whenever $|y| > 1$. It follows that $\hat\theta$ is nonunique as soon as $|x_1 - x_2| > 2$. This can also be verified directly as follows. Since $h'(y) = -2y(1 + y^2)^{-1}$, we have

$$\frac{1}{2}\frac{\partial \log L}{\partial\theta} = \frac{(\bar{x} + \Delta - \theta)}{[1 + (\bar{x} + \Delta - \theta)^2]} + \frac{(\bar{x} - \Delta - \theta)}{[1 + (\bar{x} - \Delta - \theta)^2]}.$$

Routine algebra yields

$$\frac{1}{2}\frac{\partial \log L}{\partial\theta} = \frac{2(\bar{x} - \theta)[(\bar{x} - \theta)^2 - \Delta^2 + 1]}{[1 + (\bar{x} + \Delta - \theta)^2][1 + (\bar{x} - \Delta - \theta)^2]}.$$

Suppose that $|\Delta| > 1$. Then the likelihood equation has three roots, namely, $\theta = \bar{x}$ and $\theta = \bar{x} \pm (\Delta^2 - 1)^{1/2}$. Furthermore, as $\theta$ increases from $-\infty$ to $\infty$, the preceding derivative is, in turn, positive, negative, positive, and negative. It follows that $\hat\theta = \bar{x} \pm (\Delta^2 - 1)^{1/2}$ as soon as $|\Delta| > 1$. Of course $\hat\theta = \bar{x}$ if $|\Delta| \le 1$. This corrects problem 23(b) on page 114 of Bickel and Doksum (1977).

*Example 2*. This example shows that $\hat\theta$ can be nonunique with probability 1. Let $g(y) = C(1 + |y|)^{-\alpha}$, where $\alpha > 1$ is a known constant and $C$ is a normalizing constant. Then $h''(y) = \alpha(1 + |y|)^{-2} > 0$ for all $y \ne 0$. Consequently, the MLE is unique only if $\Delta = 0$. Since the event $X_1 = X_2$ has probability 0, the MLE is nonunique with probability 1.

## 4. NONUNIQUENESS OF THE MLE FOR SAMPLES OF ARBITRARY SIZE

In Section 2 the fact that $x_1$ and $x_2$ are situated symmetrically with respect to $\bar{x}$ was used in an essential way. For this reason there are some difficulties in extending the result of that section to samples of arbitrary size. The argument can be made discrete, however, to show the nonuniqueness of the MLE for samples of any size $\ge 2$. This extension is sketched in this section.

Let $I$ denote the set of all integers. Let $\{q(i), i \in I\}$ be a distribution on $I$ such that (a) $q(i) > 0$ and $q(i) = q(-i)$ for all $i \in I$; (b) $q$ is unimodal with a unique mode at 0; and (c) if $\eta = q(1)/q(0)$, then for some $N \ge 2$,

$$\eta q(N + 1)q(N - 1) > [q(N)]^2. \qquad (4)$$

To see that condition (c) can hold, set $q(1) = \eta q(0)$ and $q(2) = q(3) = \xi\eta q(0)$, where $0 < \xi < \eta < 1$. Then (4) holds for $N = 2$. We note that in view of condition (b), the inequality (4) cannot hold for $N = 0$ or 1. Let $X_1, \ldots, X_n$ be independent observations such that $P_\theta(X_j = i) = q(i - \theta)$ for $\theta \in I$, $i \in I$, and $j = 1, \ldots, n$. Suppose $x_j$ is the observed value of $X_j$. Condition (b) shows that the MLE $\hat\theta$ is unique and equals $X_1$ when $n = 1$. Now suppose that $n \ge 2$. Let $n$ be odd and $n = 2m + 1$ with $m \ge 1$. Suppose that $x_j = a$, $x_{m+j} = a + 2N$ for $j = 1, \ldots, m$, and $x_{2m+1} = a + N$. Then

$$L(\theta|\mathbf{x}) = [q(a - \theta)q(a + 2N - \theta)]^m \cdot q(a + N - \theta).$$

Therefore, for $k \in I$,

$$L(a + N + k|\mathbf{x}) = [q(N + k)q(N - k)]^m \cdot q(k).$$

Thus $L$ is symmetric about $a + N$. Furthermore, we use the conditions $\eta < 1$ and (4) to get

$$\begin{aligned} L(a + N \pm 1|\mathbf{x}) &= [q(N + 1)q(N - 1)]^m \cdot q(1) \\ &= [q(N + 1)q(N - 1)]^m \cdot \eta q(0) \\ &\ge [q(N + 1)q(N - 1)]^m \eta^m \cdot q(0) \\ &> [q(N)]^{2m} \cdot q(0) = L(a + N|\mathbf{x}). \end{aligned}$$

Thus $\hat\theta \ne a + N$, the MLE is nonunique, and the possible choices of $\hat\theta$ do not form a discrete interval. We note that the event $X_1 = \cdots = X_m = a$, $X_{m+1} = \cdots = X_{2m} = a + 2N$, and $X_{2m+1} = a + N$ has positive probability. The computations for the case $n = 2m$ are similar, and here it suffices to have a weakened form of (4), where the factor $\eta$ is absent from the left side.

[*Received December 1984. Revised February 1985.*]

## REFERENCES


Bickel, Peter J., and Doksum, Kjell A. (1977), *Mathematical Statistics*, San Francisco: Holden-Day.

Hogg, Robert V., and Craig, Alan T. (1978), *Introduction to Mathematical Statistics* (4th ed.), New York: Macmillan.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley.

——— (1980), "Efficient Likelihood Estimators," *The American Statistician*, 34, 233–235.