$$\overline{X}_{n+1} = (n\overline{X}_n + X_{n+1})/(n+1) \qquad (1)$$

$$ns_{n+1}^2 = (n-1)s_n^2 + \frac{n}{(n+1)}(X_{n+1} - \overline{X}_n)^2. \qquad (2)$$

(See below for details.) Now, $\overline{X}_n$ and X_{n+1} are independent, $N(\mu, \sigma^2/n)$ (by the induction hypothesis) and $N(\mu, \sigma^2)$, respectively. Hence $\overline{X}_{n+1}$ is a linear combination of two independent normal random variables, and (a) follows by simply computing $E\overline{X}_{n+1}$ and $V(\overline{X}_{n+1})$. Similarly, it follows that $X_{n+1} - \overline{X}_n$ has a $N(0, ((n+1)/n)\sigma^2)$ distribution, and so $(n/(n+1))(X_{n+1} - \overline{X}_n)^2$ is distributed as the square of a $N(0, \sigma^2)$ random variable. Since X_{n+1} is independent of s_n^2, and $\overline{X}_n$ is also independent of s_n^2 by the induction hypothesis, (b) follows after dividing (2) through by σ^2. Finally, the induction hypothesis (and inspection of (1)) shows that $\overline{X}_{n+1}$ is independent of s_n^2, and (c) follows by noting

$$\text{cov}(n\overline{X}_n + X_{n+1}, X_{n+1} - \overline{X}_n) = \sigma^2 - n \cdot \sigma^2/n = 0.$$

The relationships (1) and (2) are themselves nice exercises in summation notation. (1) is direct, as is the useful consequence $\overline{X}_{n+1} - \overline{X}_n = (X_{n+1} - \overline{X}_n)/(n+1)$. Formula (2) follows by expanding

$$\sum_{i=1}^{n+1}(X_i - \overline{X}_{n+1})^2 = \sum_{i=1}^{n+1}[(X_i - \overline{X}_n) + (\overline{X}_n - \overline{X}_{n+1})]^2$$

$$= \sum_{i=1}^{n+1}(X_i - \overline{X}_n)^2 + 2(\overline{X}_n - \overline{X}_{n+1})$$

$$\times \sum_{i=1}^{n+1}(X_i - \overline{X}_n) + (n+1)(\overline{X}_n - \overline{X}_{n+1})^2.$$

Then, noting that

$$\sum_{i=1}^{n+1}(X_i - \overline{X}_n)^2 = (n-1)s_n^2 + (X_{n+1} - \overline{X}_n)^2,$$

and

$$\sum_{i=1}^{n+1}(X_i - \overline{X}_n) = \sum_{i=1}^{n}(X_i - \overline{X}_n) + (X_{n+1} - \overline{X}_n) = X_{n+1} - \overline{X}_n,$$

(2) follows readily.

One feature of this proof is that it only involves bivariate distributions; hence the independence of $X_1 + X_2$ and $X_1 - X_2$ and of $n\overline{X}_n + X_{n+1}$ and $X_{n+1} - \overline{X}_n$ can be illustrated geometrically at a blackboard. In addition, formulas (1) and (2) are of independent interest as useful algorithmic devices for calculating $\overline{X}$ and s^2 on a computer; see Chan, Golub, and LeVeque (1983) for a discussion and evaluation of these and other algorithms.

Kruskal's original version of this proof did not assume the knowledge that a lack of correlation among bivariate normal random variables implies independence; instead he derived the relevant distributions analytically by transformation.

[Received April 1983. Revised October 1983.]

REFERENCES

BROWNLEE, K.A. (1965), *Statistical Theory and Methodology in Science and Engineering* (2nd ed.), New York: John Wiley.

CHAN, TONY F., GOLUB, GENE A., and LEVEQUE, RANDALL L. (1983), "Algorithms for Computing the Sample Variance: Analysis and Recommendations," *The American Statistician*, 37, 242–247.

HOGG, R.V., and CRAIG, A. (1970), *Introduction to Mathematical Statistics* (3rd ed.), New York: Macmillan.

KRUSKAL, W.H. (1946), "Helmert's Distribution," *American Mathematical Monthly*, 53, 435–438.

RAO, C.R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: John Wiley.

SHUSTER, J. (1973), "A Simple Method of Teaching the Independence of $\overline{X}$ and s^2," *The American Statistician*, 27, 29–30.

The Double Exponential Distribution: Using Calculus to Find a Maximum Likelihood Estimator

ROBERT M. NORTON*

1. INTRODUCTION

When learning how to derive a maximum likelihood estimator (MLE) for an unknown parameter of a known density function, students are tempted to stop after equating the first derivative to zero and solving, particularly if the derivative is of complicated form. Yet students know from calculus that a function may have a maximum, minimum, or neither at a critical number. The purpose of this note is to give a simple maximization argument on one particular exercise encountered by any teacher using the mathematical statistics text by Hogg and Craig. It is an exercise students always find difficult.

*Robert M. Norton is Associate Professor, Department of Mathematics, College of Charleston, Charleston, SC 29424. The author wishes to thank a referee for some insightful suggestions.

2. THE PROBLEM AND SOLUTION

The particular problem is this: Find the MLE for θ when $X_1, X_2, \ldots, X_n$ represents a random sample from the density

$$f(x; \theta) = \tfrac{1}{2}\exp(-|x - \theta|), \quad -\infty < x < \infty, \quad -\infty < \theta < \infty.$$

The likelihood function $L(\theta)$ has as its natural logarithm

$$\psi(\theta) = -n \ln 2 - \sum_{i=1}^{n} |x_i - \theta| = -n \ln 2 - \sum_{i=1}^{n} \{(x_i - \theta)^2\}^{1/2}.$$

The function ψ is continuous everywhere and differentiable except at $\theta = x_1, x_2, \ldots, x_n$. Now when it exists,

$$\psi'(\theta) = \sum_{i=1}^{n} \{(x_i - \theta)^2\}^{-1/2} (x_i - \theta) = \sum_{i=1}^{n} (x_i - \theta)/|x_i - \theta|,$$

a sum of ones and minus ones.

Let $y_1 < y_2 < \cdots < y_n$ denote the order statistics of the sample. For $\theta < y_1$, $\psi'(\theta) = n$ since $(x_i - \theta)/|x_i - \theta| = 1$ for all i. If $y_1 < \theta < y_2$ then $\psi'(\theta) = n - 2$, while if $y_2 < \theta < y_3$ then $\psi'(\theta) = n - 4$, and so forth. The graph of ψ is a continuous polygonal curve. It is now easy to see that for n odd, ψ is strictly increasing on $(-\infty, y_{(n+1)/2}]$ and strictly decreasing on $[y_{(n+1)/2}, \infty)$, so that at $\theta = y_{(n+1)/2}$ the functions ψ and L are maximized. If n is even, the highest points on the graph of L lie on the horizontal segment

$$\{(\theta, L(\theta)) : y_{n/2} \leq \theta \leq y_{(n/2)+1}\}.$$

Hence the sample median $\tilde{x}$ is a MLE for θ. My students have found this argument helpful on this problem.

Students who realize that the derivative is a sum of ones and minus ones sometimes try to equate the derivative to zero, reasoning that there must be as many ones as minus ones, and conclude that the median is an answer. Yet this makes sense only for n even, and then it is known only that $\tilde{x}$ is a critical number for θ. Still more work is required to show that L is maximized at $\theta = \tilde{x}$.

[*Received May 1983. Revised August 1983.*]

REFERENCE

HOGG, R.V., and CRAIG, A.T. (1978), *Introduction to Mathematical Statistics,* New York: Macmillan.