1981, pp. 201–203). This n-dimensional space and its two-dimensional subspace are the ones to which we direct attention.

Each variable is a vector lying in the observation space of n dimensions. Linearly independent variables are those with vectors that do not fall along the same line; that is, there is no multiplicative constant that will expand, contract, or reflect one vector onto the other. Orthogonal variables are a special case of linearly independent variables. Not only do their vectors not fall along the same line, but they also fall perfectly at right angles to one another (or, equivalently, the cosine of the angle between them is zero). The relationship between "linear independence" and "orthogonality" is thus straightforward and simple.

Uncorrelated variables are a bit more complex. To say variables are uncorrelated indicates nothing about the raw variables themselves. Rather, "uncorrelated" implies that once each variable is centered (i.e., the mean of each vector is subtracted from the elements of that vector), then the vectors are perpendicular. The key to appreciating this distinction is recognizing that centering each variable can and often will change the angle between the two vectors. Thus, orthogonal denotes that the *raw* variables are perpendicular. Uncorrelated denotes that the *centered* variables are perpendicular.

Each of the following situations can occur: Two variables that are perpendicular can become oblique once they are centered; these are orthogonal but not uncorrelated. Two variables not perpendicular (oblique) can become perpendicular once they are centered; these are uncorrelated but not orthogonal. And finally, two variables may be both orthogonal and uncorrelated if centering does not change the angle between their vectors. In each case, of course, the variables are linearly independent. Figure 1 gives a pictorial portrayal of the relationships among these three terms. Examples of sets of variables that correspond to each possible situation are shown.

[*Received April 1983. Revised October 1983.*]

REFERENCES

BOX, J.F. (1978), *R.A. Fisher: The Life of a Scientist*, New York: John Wiley.

DRAPER, N., and SMITH, H. (1981), *Applied Regression Analysis*, New York: John Wiley.

GIBBONS, J.D. (1968), "Mutually Exclusive Events, Independence, and Zero Correlation," *The American Statistician*, 22, 31–32.

HUNTER, J.J. (1972), "Independence, Conditional Expectation, and Zero Covariance," *The American Statistician*, 26, 22–24.

POLLAK, E. (1971), "A Comment on Zero Correlation and Independence," *The American Statistician*, 25, 53.

SHIH, W. (1971), "More on Zero Correlation and Independence," *The American Statistician*, 25, 62.

Kruskal's Proof of the Joint Distribution of $\bar{X}$ and s^2

STEPHEN M. STIGLER*

In introductory courses in mathematical statistics, the proof that the sample mean $\bar{X}$ and sample variance s^2 are independent when one is sampling from normal populations is commonly deferred until substantial mathematical machinery has been developed. The proof may be based on Helmert's transformation (Brownlee 1965, p. 271; Rao 1973, p. 182), or it may use properties of moment-generating functions (Hogg and Craig 1970, p. 163; Shuster 1973). The purpose of this note is to observe that a simple proof, essentially due to Kruskal (1946), can be given early in a statistics course; the proof requires no matrix algebra, moment-generating functions, or characteristic functions. All that is needed are two minimal facts about the bivariate normal distribution: Two linear combinations of a pair of independent normally distributed random variables are themselves bivariate normal, and hence if they are uncorrelated, they are independent.

*Stephen M. Stigler is Professor, Department of Statistics, University of Chicago, Chicago, IL 60637.

Let $X_1, \ldots, X_n$ be independent, identically distributed $N(\mu, \sigma^2)$. Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,\ s_n^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

We suppose that the chi-squared distribution $\chi^2(k)$ has been defined as the distribution of $U_1^2 + \cdots + U_k^2$, where the U_i are independent $N(0, 1)$.

Theorem. (a) $\bar{X}_n$ has a $N(\mu, \sigma^2/n)$ distribution. (b) $(n-1)s_n^2/\sigma^2$ has a $\chi^2(n-1)$ distribution. (c) $\bar{X}_n$ and s_n^2 are independent.

Proof. The proof is by induction. First consider the case $n = 2$. Here $\bar{X}_2 = (X_1 + X_2)/2$ and, after a little algebra, $s_2^2 = (X_1 - X_2)^2/2$. Part (a) is an immediate consequence of the assumed knowledge of normal distributions, and since $(X_1 - X_2)/\sqrt{2}$ is $N(0, 1)$, (b) follows too, from the definition of $\chi^2(1)$. Finally, since $\text{cov}(X_1 - X_2, X_1 + X_2) = 0$, $X_1 - X_2$ and $X_1 + X_2$ are independent and (c) follows.

Now assume the conclusion holds for a sample of size n. We prove it holds for a sample of size $n + 1$. First establish the two relationships

$$\bar{X}_{n+1} = (n\bar{X}_n + X_{n+1})/(n+1) \quad (1)$$

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

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

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

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

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

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

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

Then, noting that

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

and

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

(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 $n\bar{X}_n + X_{n+1}$ and $X_{n+1} - \bar{X}_n$ can be illustrated geometrically at a blackboard. In addition, formulas (1) and (2) are of independent interest as useful algorithmic devices for calculating $\bar{X}$ 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 $\bar{X}$ 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.

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

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

*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.