Inequality (3.6) reduces to (1.1) and (2.3) for $l = 1$ and $m = 1$; it reduces to (3.4) for $l = 1$, $m = 2$. Because inequality (3.6) is a multivariate version it permits multiple comparisons for contrasts.

[*Received July 1990. Revised September 1991.*]

REFERENCES

Arnold, B. C. (1974), "Schwarz, Regression, and Extreme Deviance," *The American Statistician*, 28, 22–23.

——— (1988), "Bounds on the Expected Maximum," *Communications in Statistics, A—Theory and Methods*, 17, 2135–2150.

Arnold, B. C., and Balakrishnan, N. (1989), "Relations, Bounds and Approximations for Order Statistics" (Lecture Notes in Statistics, No. 53), New York: Springer-Verlag.

Arnold, B. C., and Groeneveld, R. A. (1974), "Bounds for Deviations Between Sample and Population Statistics," *Biometrika*, 61, 387–389.

——— (1978), "Bounds on Deviations of Estimates Arising in Finite Population Regression Models," *Communications in Statistics, A*, 7, 1173–1179.

——— (1979), "Bounds on Expectations of Linear Systematic Statistics Based on Dependent Samples," *The Annals of Statistics*, 7, 220–223.

——— (1981), "Maximal Deviation Between Sample and Population Means in Finite Populations," *Journal of the American Statistical Association*, 76, 443–445.

Beesack, P. R. (1973), "On Bounds for the Range of Ordered Values," *Univerzitet u Beogradu, Publikacije Elektrotechničkog Fakulteeta, Serija Matematika i Fizika*, No. 412–460, 93–96.

——— (1976), "On Bounds for the Range of Ordered Variates II," *Aequationes Mathematicae*, 14, 293–301.

Boyd, A. V. (1971), "Bounds for Order Statistics," *Univerzitet u Beogradu, Publikacije Elektrotehničkog Fakulteeta, Serija Matematika i Fizika*, No. 357–380, 31–32.

Brunk, H. D. (1959), "Note on Two Papers of K. R. Nair," *Journal of the Indian Society of Agricultural Statistics*, 11, 186–189.

David, H. A. (1988), "General Bounds and Inequalities in Order Statistics," *Communications in Statistics, A—Theory and Methods*, 17, 2119–2134.

——— (1990), "Problem 90-4," *Bulletin of the Institute of Mathematical Statistics*, 19, 608.

David, H. A., Hartley, H. O., and Pearson, E. S. (1954), "The Distribution of the Ratio, in a Single Normal Sample, of Range to Standard Deviation," *Biometrika*, 41, 482–493.

Dwass, M. (1975), "The Extreme Deviations Inequality," *The American Statistician*, 28, 108.

Farnum, N. R. (1989), "An Alternative Proof of Samuelson's Inequality and Its Extensions," *The American Statistician*, 43, 46–47.

Hawkins, D. M. (1971), "On the Bounds of the Range of Order Statistics," *Journal of the American Statistical Association*, 66, 644–645.

Hotelling, H., and Solomons, L. M. (1932), "The Limits of a Measure of Skewness," *The Annals of Mathematical Statistics*, 3, 141–142.

Kabe, D. G. (1980), "On Extensions of Samuelson's Inequality," *The American Statistician*, 34, 249.

Loynes, R. M. (1979), "A Note on Prescott's Upper Bound for Normed Residuals," *Biometrika*, 66, 387–389.

Mallows, C. L., and Richter, D. (1969), "Inequalities of Chebyshev Type Involving Conditional Expectations," *The Annals of Mathematical Statistics*, 40, 1922–1932.

Murty, V. N. (1989), "A Note on Samuelson's Inequality," *The American Statistician*, 44, 64–65.

Nair, K. R. (1948), "Certain Symmetrical Properties of Unbiased Estimates of Variance and Covariance," *Journal of the Indian Society of Agricultural Statistics*, 1, 162–172.

——— (1956), "A Note on the Estimation of Mean Change," *Journal of the Indian Society of Agricultural Statistics*, 8, 122–124.

——— (1959), "A Tail-Piece to Brunk's Paper," *Journal of the Indian Society of Agricultural Statistics*, 11, 189–190.

O'Cinneide, C. A. (1990), "The Mean Is Within One Standard Deviation of Any Median," *The American Statistician*, 44, 292–293.

Pearson, E. S., assisted by Chandra Sekar, C. (1936), "The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations," *Biometrika*, 28, 308–320.

Prescott, P. (1977), "An Upper Bound for Any Linear Function of Normed Residuals," *Communications in Statistics, B—Simulation and Computation*, 6, 83–88.

Quesenberry, C. P. (1974), "The Extreme Deviations Inequality," *The American Statistician*, 28, 112.

Samuelson, P. A. (1968), "How Deviant Can You Be? *Journal of the American Statistical Association*, 63, 1522–1525.

Shapiro, S. S., and Wilk, M. B. (1965), "An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, 52, 591–611.

Stephansky, W. (1971), "Rejecting Outliers by Maximum Normed Residual," *The Annals of Mathematical Statistics*, 42, 35–45.

Stephansky, W. (1972), "Rejecting Outliers in Factorial Designs," *Technometrics*, 14, 469–479.

Thompson, W. R. (1935), "On a Criterion for the Rejection of Observations and the Distribution of the Ratio of Deviation to Sample Standard Deviation," *The Annals of Mathematical Statistics*, 6, 214–219.

Thomson, G. W. (1955), "Bounds for the Ratio of Range to Standard Deviation," *Biometrika*, 42, 268–269.

Wolkowicz, H., and Styan, G. P. H. (1979), "Extensions of Samuelson's Inequality," *The American Statistician*, 62, 634–637.

Wolkowicz, H., and Styan, G. P. H. (1980), "Reply," *The American Statistician*, 34, 250.

On Invariance and Maximum Likelihood Estimation

NABENDU PAL and J. CALVIN BERRY*

The invariance of maximum likelihood estimators to non–one-to-one parameter transformations is considered. Three approaches to maximum likelihood estimation are discussed, and it is shown that invariance depends on the approach adopted. A theorem is provided that gives sufficient conditions for invariance to hold under all three approaches.

KEY WORDS: Nuisance parameter; One-to-one function; Partial sufficiency; Specific sufficiency.

In many estimation problems the parameter or parameters of interest do not appear explicitly in the original parameterization of the model. Let $\boldsymbol{\theta}$ denote the

*Nabendu Pal and J. Calvin Berry are Assistant Professors, Department of Statistics, University of Southwestern Louisiana, Lafayette, LA 70504-1006. The authors thank an associate editor and a referee for their helpful comments and suggestions.

parameters of interest and let $\boldsymbol{\omega}$ denote the original parameters of the model. It is well known that the method of maximum likelihood estimation is parameterization invariant. That is, if $\hat{\boldsymbol{\omega}}$ is a maximum likelihood estimator (MLE) of $\boldsymbol{\omega}$ and if there is a one-to-one correspondence between $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, say $\boldsymbol{\theta} = g(\boldsymbol{\omega})$ with g a one-to-one function, then $\hat{\boldsymbol{\theta}} = g(\hat{\boldsymbol{\omega}})$ is a MLE of $\boldsymbol{\theta}$. If the correspondence between $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ is not one-to-one, then there are several possible definitions of a MLE of $\boldsymbol{\theta}$. The invariance property mentioned above does not necessarily hold for some of these definitions of a MLE of $\boldsymbol{\theta}$. Three specific approaches to maximum likelihood estimation of $\boldsymbol{\theta} = g(\boldsymbol{\omega})$ are considered in this article and a theorem is proved providing conditions under which all three approaches yield the same estimate.

Many authors have considered the problem of selecting an appropriate likelihood function and an appropriate definition of a MLE for the situation under consideration here. The results presented in Basu (1977, 1978), Edwards (1972), and Kalbfleisch and Sprott (1970) are particularly relevant and provide a good starting point for the interested reader. The omission of references to other works should not be interpreted as a negative comment on their relevance. This article provides a simple exposition and summary of three specific approaches to maximum likelihood estimation with no claim of priority.

Let $\mathbf{X}$ denote a p-variate random vector with density function $f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\omega})$, where $\boldsymbol{\omega}$ is a q-variate parameter vector, with $p \geq q \geq 1$. Let $\boldsymbol{\theta} = g(\boldsymbol{\omega})$ denote the parameter vector of interest and let $\boldsymbol{\phi}$ be defined so that there is a one-to-one correspondence between $\boldsymbol{\omega}$ and $(\boldsymbol{\theta}, \boldsymbol{\phi})$. Since there is a one-to-one correspondence between $\boldsymbol{\omega}$ and $(\boldsymbol{\theta}, \boldsymbol{\phi})$ there is no loss of generality in taking $\boldsymbol{\omega} = (\boldsymbol{\theta}, \boldsymbol{\phi})$; this identification will be made for the remainder of this article. Let Ω, Θ, and Φ denote the range of values of $\boldsymbol{\omega}$, $\boldsymbol{\theta}$, and $\boldsymbol{\phi}$, then $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are said to be variation independent if $\Omega = \Theta \times \Phi$. For the purposes of this discussion it is necessary that $\boldsymbol{\phi}$ be chosen so that $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are variation independent. This requirement of variation independence is needed for the definition of partial sufficiency and does not appear to be overly restrictive.

If g is not one-to-one, then the most common approach to maximum likelihood estimation of $\boldsymbol{\theta} = g(\boldsymbol{\omega})$ is that of Zehna (1966). This approach is based on the likelihood function induced by g which is given by

$$L_2(\boldsymbol{\theta}|\mathbf{x}) = \sup_{\{\omega: g(\omega) = \theta\}} L_1(\boldsymbol{\omega}|\mathbf{x}) = \sup_{\{\phi \in \Phi\}} L_1(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{x}),$$

where $L_1(\boldsymbol{\omega}|\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\omega})$. In the terminology of Barndorff–Nielson (1988) L_4 is known as the profile likelihood function. Define an induced MLE of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, as a value of $\boldsymbol{\theta}$ for which $L_2(\boldsymbol{\theta}|\mathbf{x})$ is maximized. It is well known and readily verified that $\hat{\boldsymbol{\theta}} = g(\hat{\boldsymbol{\omega}})$.

As a first alternative approach to maximum likelihood estimation of $\boldsymbol{\theta}$, suppose that there is a statistic $\mathbf{Z}$, based on $\mathbf{X}$, such that the density function of $\mathbf{Z}$ depends on $\boldsymbol{\omega}$ only through $\boldsymbol{\theta}$. In the terminology of Basu (1977, 1978) such a statistic is said to be $\boldsymbol{\theta}$ oriented. Instead of maximizing the induced likelihood function, one might maximize the marginal likelihood function

$$L_3(\boldsymbol{\theta}|\mathbf{z}) = f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}),$$

where $f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta})$ is the marginal density function of $\mathbf{Z}$. Define a marginal MLE of $\boldsymbol{\theta}$, denoted by $\tilde{\boldsymbol{\theta}}$, as a value of $\boldsymbol{\theta}$ for which L_3 is maximized.

A second alternative approach to maximum likelihood estimation of $\boldsymbol{\theta}$ is based on the elimination of $\boldsymbol{\phi}$ by integration of summation. Let $\pi(\boldsymbol{\phi})$ denote a density function, possibly improper, for $\boldsymbol{\phi}$ and define the integrated likelihood function as

$$L_4(\boldsymbol{\theta}|\mathbf{x}) = \int_\Phi f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\phi})\pi(\boldsymbol{\phi})\, d\boldsymbol{\phi},$$

where the integral, or sum, is over all possible values of $\boldsymbol{\phi}$. Define an integrated MLE of $\boldsymbol{\theta}$ with respect to π, denoted by $\hat{\boldsymbol{\theta}}_\pi$, as a value of $\boldsymbol{\theta}$ for whith L_4 is maximized. The density function $\pi(\boldsymbol{\phi})$ can be viewed as a prior distribution for the "nuisance" parameter $\boldsymbol{\phi}$. In the terminology of Edwards (1972), a density π for which $L_4(\boldsymbol{\theta}|\mathbf{x}) \propto L_3(\boldsymbol{\theta}|\mathbf{z})$ for some statistic $\mathbf{Z}$ is said to be a neutral prior density for $\boldsymbol{\theta}$ with respect to $\boldsymbol{\phi}$. If π is neutral for $\boldsymbol{\theta}$ with respect to $\mathbf{Z}$, then $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_\pi$.

The question of interest is, under what conditions on $\mathbf{Z}$ are the three MLE's, $\hat{\boldsymbol{\theta}} = g(\hat{\boldsymbol{\omega}})$, $\tilde{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\theta}}_\pi$, equal? Since interest centers on the comparison of the different MLE's, all MLE's will be assumed to exist. A partial answer is provided in the theorem that follows. In the terminology of Basu (1977, 1978), the statistic $\mathbf{Z}$ is said to be partially sufficient for $\boldsymbol{\theta}$ if $\mathbf{Z}$ is $\boldsymbol{\theta}$ oriented and $\mathbf{Z}$ is specific sufficient for $\boldsymbol{\theta}$. The statistic $\mathbf{Z}$ is said to be specific sufficient for $\boldsymbol{\theta}$ if it is sufficient in the usual sense when the value of $\boldsymbol{\phi}$ is fixed and known. The statistic $\mathbf{Z}$ is partially sufficient for $\boldsymbol{\theta}$ iff the density function of $\mathbf{X}$ factors as

$$f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\omega}) = f_{\mathbf{X}}(\mathbf{z}|\boldsymbol{\theta}) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}, \boldsymbol{\phi}). \tag{1}$$

Theorem. If $\mathbf{Z}$ is partially sufficient for $\boldsymbol{\theta}$, then the three likelihood functions $L_2(\boldsymbol{\theta}|\mathbf{z})$, $L_3(\boldsymbol{\theta}|\mathbf{z})$, and $L_4(\boldsymbol{\theta}|\mathbf{z})$ are proportional as functions of $\boldsymbol{\theta}$. Hence, if $\mathbf{Z}$ is partially sufficient for $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_\pi$, for any $\pi(\boldsymbol{\phi})$ for which the integral in the definition of L_4 converges.

Proof. From the factorization of Equation (1) the three likelihood functions are given by

$$L_2(\boldsymbol{\theta}|\mathbf{x}) = \left[\sup_{\{\phi \in \Phi\}} f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}, \boldsymbol{\phi})\right] f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}),$$

$$L_3(\boldsymbol{\theta}|\mathbf{z}) = f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}),$$

and

$$L_4(\boldsymbol{\theta}|\mathbf{x}) = \left[\int_\Phi f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}, \boldsymbol{\phi})\pi(\boldsymbol{\phi})\, d\boldsymbol{\phi}\right] f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}).$$

Since $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are assumed to be variation independent, the terms in square brackets in the expressions for L_2 and L_4 do not depend on $\boldsymbol{\theta}$ and the theorem is proved.

Notice that the theorem shows that if $\mathbf{Z}$ is partially sufficient for $\boldsymbol{\theta}$, then the three likelihood functions are proportional; hence, any likelihood based procedure will yield the same results under all three approaches. For example, when $\mathbf{Z}$ is partially sufficient for $\boldsymbol{\theta}$, the observed information, obtained as minus the second derivative of the logarithm of the likelihood function (or the corresponding Hessian matrix), is the same under each of the three approaches. Similarly, the expected information, that is, the expected value of the observed information with respect to $\mathbf{Z}$, is the same under each approach.

If $\mathbf{Z}$ is not partially sufficient for $\boldsymbol{\theta}$, then there may be some loss of information when the problem is reexpressed in terms of a likelihood function which depends on $\boldsymbol{\theta}$ alone. The question of loss of information will not be addressed here, but the interested reader is encouraged to consult the articles referenced previously for an introduction to this topic.

The following six examples are given to clarify the applicability of the theorem. The first three examples provide situations where there is a partially sufficient statistic and the theorem applies. The last three examples provide situations where the theorem does not apply. A partial counter example to the converse of the theorem is provided in Example 6.

Example 1. Let $X_1 \sim N(\mu_1, 1)$ independently of $X_2 \sim N(\mu_2, 1)$ be given. (This is equivalent to the independent sample problem with equal samples sizes and sample means X_1 and X_2.) $\theta = \mu_1 + \mu_2$ and $\phi = \mu_1 - \mu_2$ are variation independent. The statistic $Z = X_1 + X_2$ is partially sufficient for θ, hence the theorem applies with all three likelihood functions proportional to $\exp\{-(z - \theta)^2/4\}$, where the neutral prior $\pi(\theta) = 1$ is used to compute L_4. The estimators are given by $\hat{\theta} = \tilde{\theta} = \hat{\theta}_\pi = Z$.

Example 2. Let $X_1 \sim \text{Poisson}(\lambda_1)$ independently of $X_2 \sim \text{Poisson}(\lambda_2)$ be given. (This is equivalent to the independent samples problem with arbitrary sample sizes and sample sums X_1 and X_2.) $\theta = \lambda_1 + \lambda_2$ and $\phi = \lambda_1/\lambda_2$ are variation independent. The statistic $Z = X_1 + X_2$ is partially sufficient for θ, hence the theorem applies with all three likelihood functions proportional to $\theta^z \exp\{-\theta\}$, where the neutral prior $\pi(\theta) = 1$ is used to compute L_4. The estimators are given by $\hat{\theta} = \tilde{\theta} = \hat{\theta}_\pi = Z$.

Example 3. [Example 5 of Basu (1977)]. Let $X_1, \ldots, X_n$ be iid with density function

$$\begin{aligned} f_X(x|\phi) &= (1-\theta)\phi e^{\phi x} && \text{for } x \le 0, \\ &= \theta\phi e^{-\phi x} && \text{for } x > 0, \end{aligned}$$

where $\theta \in (0, 1)$ and $\phi \in (0, \infty)$ are variation independent. The statistic $Z = \Sigma I_{(0,\infty)}(X_i)$, that is, the number of positive observations, is partially sufficient for θ, hence the theorem applies with all three likelihood functions proportional to $\theta^z(1 - \theta)^{n-z}$, where the neutral prior $\pi(\phi) = 1$ is used to compute L_4. The estimators are given by $\hat{\theta} = \tilde{\theta} = \hat{\theta}_\pi = Z/n$.

Example 4. Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$. (a) $\theta = \sigma^2$ and $\phi = \mu$ are variation independent. The statistic $Z = \Sigma(X_i - \bar{X})^2$ is θ oriented but not specific sufficient, hence, Z is not partially sufficient for θ. In this situation the three approaches yield two distinct likelihood functions with

$$L_2(\theta|z) \propto \theta^{-(n/2)} \exp\left\{-\frac{z}{2\theta}\right\}$$

and

$$L_3(\theta|z) \propto L_4(\theta|z) \propto \theta^{-(n-1/2)} \exp\left(-\frac{z}{2\theta}\right),$$

where $\pi(\phi) = 1$ is used to compute L_4. The estimators are $\hat{\theta} = Z/n$ and $\tilde{\theta} = \hat{\theta}_\pi = Z/(n-1)$. (b) $\theta = \mu$ and $\phi = \sigma^2$ are variation independent. The statistic $Z = \bar{X}$ is not θ oriented and there does not appear to be any θ-oriented statistic in this situation.

Example 5. Let $X_1 \sim \text{Uniform}[0, \omega_1]$ independently of $X_2 \sim \text{Uniform}[0, \omega_2]$ be given. Let ϕ be defined by $\phi = 1$ if $\omega_1 \le \omega_2$ and $\phi = -1$ if $\omega_1 > \omega_2$. $\boldsymbol{\theta} = (\min\{\omega_1, \omega_2\}, \max\{\omega_1, \omega_2\})$ and ϕ are variation independent. The statistic $\mathbf{Z} = (Z_1, Z_2) = (\min\{X_1, X_2\}, \text{mas}\{X_1, X_2\})$ is $\boldsymbol{\theta}$ oriented but not specific sufficient, hence $\mathbf{Z}$ is not partially sufficient for $\boldsymbol{\theta}$. Using the invariance property of the first approach yields $\hat{\theta} = \mathbf{Z} = (Z_1, Z_2)$, since $(\hat{\omega}_1, \hat{\omega}) = (X_1, X_2)$. The likelihood functions for the second and third approaches are given by

$$\begin{aligned} L_3(\boldsymbol{\theta}|\mathbf{z}) &= 2L_4(\boldsymbol{\theta}|\mathbf{z}) \\ &= \frac{1}{\theta_1\theta_2} \qquad \text{if } 0 \le z_1 \le \theta_1 < z_2 \le \theta_2 \\ &= \frac{2}{\theta_1\theta_2} \qquad \text{if } 0 \le z_1 \le z_2 \le \theta_1 \le \theta_2, \end{aligned}$$

where $\pi(\phi) = \frac{1}{2}I_{\{-1,1\}}(\phi)$ is used to compute L_4. The corresponding estimator is given by

$$\begin{aligned} \tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_\pi &= (Z_1, Z_2) && \text{if } Z_2 \ge 2Z_1 \\ &= (Z_2, Z_2) && \text{if } Z_2 < 2Z_1. \end{aligned}$$

In this situation $P(\hat{\boldsymbol{\theta}} \neq \tilde{\boldsymbol{\theta}}) = P(Z_2 < 2Z_1) > 0$.

Example 6. Let $X_1, \ldots, X_m$ be iid $N(\theta, 1)$ independently of $Y_1, \ldots, Y_n$ iid $N(\phi\theta, 1)$, where θ is real and $\phi \in [0, \infty)$. θ and ϕ are variation independent. The statistic $Z = \bar{X}$ is θ oriented but not specific sufficient, hence $\bar{X}$ is not partially sufficient for θ. The first two approaches yield likelihood functions that are proportional to $\exp\{-m(\bar{x} - \theta)^2/2\}$. The corresponding estimators are $\hat{\theta} = \tilde{\theta} = Z$. Taking $\pi(\phi) = 1$ for $\phi \ge 0$ yields an unbounded integrated likelihood function in this situation. This example shows that the converse of the theorem with respect to the first two approaches and for a particular statistic Z is false, that is, it is possible to have $\hat{\theta} = \tilde{\theta}$ even when Z is not partially sufficient for θ.

[*Received September 1990. Revised July 1991.*]

REFERENCES

Barndorff-Nielson, O. E. (1988), *Parametric Statistical Models and Likelihood*, New York: Springer-Verlag.

Basu, D. (1977), "On the Elimination of Nuisance Parameters," *Journal of the American Statistical Association*, 72, 355–366.
——— (1978), "On Partial Sufficiency: A Review," *Journal of Statistical Planning and Inference*, 2, 1–13.
Edwards, A. W. F. (1972), *Likelihood*, London: Cambridge University Press.
Kalbfleisch, J. D., and D. A. Sprott (1970), "Application of Likelihood Methods to Models Involving Large Numbers of Parameters," *Journal of the Royal Statistical Society*, Ser. B, 32, 175–208.
Zehna, P. W. (1966), "Invariance of Maximum Likelihood," *The Annals of Mathematical Statistics*, 37, 744.

Powers of Two-Sample Rank Tests Under the Lehmann Alternatives

SHASHIKALA SUKHATME*

This note presents a method of finding the power of a two-sample rank test under the Lehmann alternatives. The method considers expressing a rank statistic in terms of the empirical distribution functions of the samples and uses properties of order statistics. The author finds that the students in an introductory course on nonparametric statistics feel more comfortable with this than with the elegant approach presented by Lehmann.

KEY WORDS: Empirical distribution functions; Mathisen median test; Rosenbaum location test.

1. INTRODUCTION

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples from continuous distributions F and G with probability densities f and g. In a course on nonparametric statistics, the class of rank tests for testing H_0: $F = G$ receives considerable attention. The exact distributions of the test statistics, in principle, can be obtained using the uniform distribution of a vector of ranks on the set of permutations of $\{1, 2, \ldots, m + n\}$. In some cases, the recurrence relations for the probability function of a rank statistic are handy to do the job. However, the task of finding the exact distribution of a rank statistic, under most of the alternatives of interest, which is necessary to find power, is quite difficult. It is relatively easy to find the distribution of a two-sample rank statistic under the Lehmann alternatives, that is, when $G(x) = w(F(x))$, where $w(x)$ is a continuous nondecreasing function on (0, 1), $w(0) = 0, w(1) = 1$. Lehmann (1953) uses Hoeffding's theorem to get the probability distribution of $(R_{m+1}, \ldots, R_{m+n})$, the vector of ranks of $Y_1, \ldots, Y_n$ in the pooled sample. This is useful to obtain the power of a rank statistic for the Lehmann alternatives under consideration in the following manner: List the sets $(r_{m+1}, \ldots, r_{m+n})$ forming the critical region and sum the probabilities associated with these. Several textbooks on nonparametric statistics, for example Randles and Wolfe (1979), discuss Lehmann's method. Lehmann's method, although elegant, is difficult for students in an introductory course in nonparametric methods to grasp.

In this note, I give an elementary method of finding the distribution of a two-sample rank statistic under the Lehmann alternatives. The students in my class like this direct approach. For the sake of illustration, I consider powers of the Mathisen test and the Rosenbaum test.

Let $X_{(1)} \leqq X_{(2)} \leqq \ldots \leqq X_{(m)}$ and $Y_{(1)} \leqq Y_{(2)} \leqq \ldots, \leqq Y_{(n)}$ denote the order statistics of the two samples, and $F_m(x)$ and $G_n(x)$ be the empirical distribution functions (edf) of the samples. Define $S_j = F_m(Y_j)$, $S_{(j)} = F_m(Y_{(j)})$, $S_j^* = G_n(X_j)$, $S_{(j)}^* = G_n(X_{(j)})$; then $S_{(1)} \leqq \ldots \leqq S_{(n)}$ and $S_{(1)}^* \leqq S_{(2)}^* \leqq \ldots \leqq S_{(m)}^*$. The main idea is to express a two-sample rank statistic in terms of the $S_{(j)}$s and/or $S_{(j)}^*$s, find the joint distributions of the $S_{(j)}$s under the Lehmann alternatives of interest, and use them to get the distribution of the statistic. Here I consider the Lehmann alternatives: (a) $G(x) = F^h(x)$, $h > 0$; (b) $F(x) = 1 - \exp(-x)$, $x > 0$, $G(x) = 1 - \exp(-(x - \theta))$, $x \geqq \theta$, $\theta > 0$. In Section 2 I derive the necessary distributions of S_j's and $S_{(j)}$'s and in Section 3 illustrate computation of the powers of the Mathisen (1943) two-sample median test, and the Rosenbaum (1953) two-sample location test.

2. DISTRIBUTIONS UNDER THE LEHMANN ALTERNATIVES

Suppose R_i denotes the rank of $Y_{(i)}$ in the pooled sample and R_j^* denotes the rank of $X_{(j)}$ in the pooled sample, then $R_i = i + mS_{(i)}$, $i = 1, \ldots, n$, and $R_j^* = j + nS_{(j)}^*$, $j = 1, 2, \ldots, m$. Therefore, every two-sample rank statistic can be written as a function of the $S_{(j)}$'s and/or $S_{(j)}^*$'s. For example, the Mann–Whitney–Wilcoxon test statistic is, apart from an additive constant, equal to $U = n^{-1} \sum_{j=1}^{n} S_j$, the Mathisen two-sample median test statistic, when n is odd, is $T_1 = mS_{((n+1)/2)}$, the Rosenbaum two-sample location statistic is $T_2 = m(1 - S_{(n)})$, the Rosenbaum (1954) two-sample scale statistic is $T_3 = m(1 - S_{(n)} + S_{(1)})$, and the Kamat (1956) two-sample scale statistic is $T_4 = n(S_{(m)}^* - S_{(1)}^*) - m(S_{(n)} - S_{(1)}) + m$.

*Shashikala Sukhatme is Associate Professor, Department of Statistics, Iowa State University, Ames, IA 50011. This research was supported in part by a research grant funded by ISU Achievement Foundation.

