The Annals of Statistics
1973, Vol. 1, No. 1, 84–92

COVARIANCE STABILIZING TRANSFORMATIONS

BY PAUL W. HOLLAND[1]

Harvard University[2]

After reviewing the asymptotic variance stabilizing transformations in one dimension, a generalization of these to multivariate cases is discussed. Results are given for the uniqueness of solutions when they exist, but unlike the one-dimensional case, covariance stabilizing transformations need not exist. In the two-dimensional case, a necessary and sufficient condition is given for the existence of solutions. It takes the form of a second-order partial differential equation that the elements of any square root of the inverse of the limiting covariance matrix must satisfy. This condition is applied to three examples with the conclusion that no covariance stabilizing transformation exists for the trinomial distribution. It is conjectured that this non-existence of solutions is true for the general multinomial.

1. Review of the one-dimensional variance stabilizing transformations. Let X_n be a real-valued random variable whose distribution depends upon a real parameter, θ. This parameter varies over the parameter space, D, an open interval in R^1. Suppose that for every $\theta \in D$ the quantity, $n^{\frac{1}{2}}(X_n - \theta)$, converges in distribution to the $N(0, \sigma^2(\theta))$ law, i.e.

$$\mathscr{L}[n^{\frac{1}{2}}(X_n - \theta)] \rightarrow N(0, \sigma^2(\theta)) . \tag{1.1}$$

$\sigma^2(\theta) > 0$ and is continuous for all θ in D. Let X_n denote a random variable and x_n denote the corresponding possible values.

A "variance stabilizing transformation", shall mean a 1–1, continuously differentiable mapping, $f\colon D \rightarrow R^1$ such that $\mathscr{L}[n^{\frac{1}{2}}(f(X_n) - f(\theta))] \rightarrow N(0, 1)$. It would be more precise to call f an "asymptotic" variance stabilizing transformation, but because I deal solely with asymptotic distributions this qualification will be implicit throughout.

Note that X_n converges in probability to θ and thus $X_n \in D$ with a probability that may be as near unity as desired by taking n large enough. Therefore f is defined for the possible values of X_n with a probability that approaches one as $n \rightarrow \infty$, so that we may ignore the fact that f may not be defined for all the possible values of X_n.

One proceeds to find f by assuming it exists and has a differential at each point in D, i.e. if $|x_n - \theta| = O(n^{-\frac{1}{2}})$ then

$$f(x_n) = f(\theta) + (x_n - \theta)\frac{df}{d\theta} + o(n^{-\frac{1}{2}}) . \tag{1.2}$$

But $|X_n - \theta| = O_p(n^{-\frac{1}{2}})$ so that

$$f(X_n) = f(\theta) + (X_n - \theta)\frac{df}{d\theta} + o_p(n^{-\frac{1}{2}}) . \tag{1.3}$$

Received March 3, 1971; revised January 4, 1972.

[1] Supported by NSF Research Contract GS-2044X1 and a Faculty Research Grant from the Social Science Research Council.

[2] Currently at N.B.E.R. Computer Research Center, Cambridge, Mass. 02139.

The reader is referred to Pratt (1959) and Rao (1965) page 321 for a justification of this replacement of $o(n^{-\frac{1}{2}})$ by $o_p(n^{-\frac{1}{2}})$, which is also called the "δ-method". Thus

$$n^{\frac{1}{2}}[f(X_n) - f(\theta)] = n^{\frac{1}{2}}(X_n - \theta)\frac{df}{d\theta} + o_p(1) \tag{1.4}$$

and hence,

$$\mathscr{L}[n^{\frac{1}{2}}(f(X_n) - f(\theta))] \to N\left(0, \sigma^2(\theta)\left(\frac{df}{d\theta}\right)^2\right). \tag{1.5}$$

The problem of finding f is now reduced to solving the differential equation:

$$\left(\frac{df}{d\theta}\right)^2 = \frac{1}{\sigma^2(\theta)}. \tag{1.6}$$

The *non-linear*, first order differential equation, (1.6) is equivalent to any one of these *linear* first-order differential equations:

$$\frac{df}{d\theta} = \frac{\gamma(\theta)}{\sigma(\theta)} \tag{1.7}$$

where $\gamma(\theta)$ is any function such that $\gamma^2(\theta) = 1$ (i.e. $\gamma(\theta)$ only takes on the values $+1$ or -1). Only continuously differentiable solutions to (1.7) are acceptable so that $\gamma(\theta)$ must be continuous and therefore $\gamma(\theta) \equiv 1$ or $\gamma(\theta) \equiv -1$ for all $\theta \in D$. In summary, the one-dimensional variance stabilizing problem always has a 1–1 continuously differentiable solution given by:

$$f(\theta) = f(\theta_0) \pm \int_{\theta_0}^{\theta} (\sigma(t))^{-1}\, dt. \tag{1.8}$$

The solution is unique up to an additive constant and the sign of its derivative. The only requirement is that $\sigma(\theta)$ is a continuous nonzero function of θ in D. $\sigma(\theta)$ need not be 1–1 and in fact may be constant (in which case f may be chosen to be a linear function).

2. A multivariate generalization: stabilizing the entire covariance matrix. I shall consider the following generalization of the univariate setup described in Section 1. The random variable, X_n, of Section 1 is replaced by a p-dimensional random vector, X_n, with coordinates X_{in}. θ becomes a p-dimensional parameter with coordinates, θ_i. This vector parameter, θ, varies over D, an open, simply connected subset of R^p. Finally, assume that $n^{\frac{1}{2}}(X_n - \theta)$ has an asymptotic multivariate Normal distribution with zero mean vector and non-singular covariance matrix, i.e.

$$\mathscr{L}[n^{\frac{1}{2}}(X_n - \theta)] \to N(0, \Sigma(\theta)) \tag{2.1}$$

where $\Sigma(\theta)$ is positive definite for all $\theta \in D$.

A "covariance stabilizing transformation" shall mean a 1–1 continuously differentiable mapping $f\colon D \to R^p$ such that

$$\mathscr{L}[n^{\frac{1}{2}}(f(X_n) - f(\theta))] \to N(0, W) \tag{2.2}$$

where W is a p by p covariance matrix that does not depend on θ. It is easy to show that without loss of generality W may be taken to be the identity matrix, I. We assume $W = I$ throughout the rest of this paper.

The multivariate extension of equation (1.6) is derived by the multivariate version of the δ-method. As this derivation is analogous to the one-dimensional case I omit it and merely report the equation that corresponds to (1.6). Let $(\partial f/\partial\theta)$ denote the Jacobian matrix of partial derivatives of f evaluated at $x = \theta$ i.e. the (i, j)th element of $(\partial f/\partial\theta)$ is $(\partial f_i/\partial x_j)|_{x=\theta}$. f stabilizes the covariance matrix of X_n if and only if f satisfies the matrix differential equation

$$\left(\frac{\partial f}{\partial\theta}\right)'\left(\frac{\partial f}{\partial\theta}\right) = \Sigma^{-1}(\theta)\,. \tag{2.3}$$

Unlike the corresponding equation in one-dimension we shall see that solutions to (2.3) may not exist in D for some important choices of $\Sigma(\theta)$.

3. Uniqueness of solutions. Before dealing with the question of the existence of solutions to (2.3), in this section I shall give some results on the uniqueness of solutions when they exist.

Suppose there are two mappings f and g that satisfy (2.3) then

$$\left(\frac{\partial g}{\partial\theta}\right)'\left(\frac{\partial g}{\partial\theta}\right) = \left(\frac{\partial f}{\partial\theta}\right)'\left(\frac{\partial f}{\partial\theta}\right)\,. \tag{3.1}$$

From (3.1) it follows that

$$\left(\frac{\partial g}{\partial\theta}\right) = \Gamma(\theta)\left(\frac{\partial f}{\partial\theta}\right) \tag{3.2}$$

where $\Gamma(\theta)$ is orthogonal for every $\theta \in D$. Equation (3.2) may be rewritten as

$$\Gamma(\theta) = \left(\frac{\partial g}{\partial\theta}\right)\left(\frac{\partial f}{\partial\theta}\right)^{-1} = \left(\frac{\partial g}{\partial\theta}\right)\left(\frac{\partial\theta}{\partial f}\right) = \left(\frac{\partial g}{\partial f}\right)\,. \tag{3.3}$$

Since the right-hand side of (3.3) does not depend on θ it follows that $\Gamma(\theta)$ is independent of θ, $\Gamma(\theta) \equiv \Gamma_0$. This fact and (3.3) together imply that f and g are related by:

$$g(\theta) = \Gamma_0 f(\theta) + \lambda \tag{3.4}$$

where Γ_0 is constant $p \times p$ orthogonal matrix and λ is a constant p-dimensional vector. In summary then, solutions to the matrix differential equation (2.3) are unique up to an arbitrary additive constant vector λ and an arbitrary constant orthogonal transformation, Γ_0.

4. Existence of solutions: general case. One must now come to grips with the problem of the existence of solutions to (2.3). I have some partial results for the general case and this section summarizes them. The case of $p = 2$ is examined in detail in the next section.

In the case of $p = 1$ a considerable simplification occurs in going from (1.6) to (1.7). To generalize this for $p > 1$ we must recognize that since $\Sigma^{-1}(\theta)$ is a symmetric, positive definite matrix, it has many possible square roots. Let $\Sigma^{-\frac{1}{2}}(\theta)$ denote any one of them—for example it may be taken to be symmetric or triangular. If f is a solution to (2.3) then

$$\left(\frac{\partial f}{\partial \theta}\right)'\left(\frac{\partial f}{\partial \theta}\right) = (\Sigma^{-\frac{1}{2}}(\theta))'\Sigma^{-\frac{1}{2}}(\theta) \tag{4.1}$$

and we see that any solution to (4.1) must also be a solution to

$$\left(\frac{\partial f}{\partial \theta}\right) = \Gamma(\theta)\Sigma^{-\frac{1}{2}}(\theta) \tag{4.2}$$

for some choice of $\Gamma(\theta)$, that is orthogonal for all $\theta \in D$. Therefore the existence of a solution to (2.3) is equivalent to the existence of a $\Gamma(\theta)$, orthogonal, such that $\Gamma(\theta)\Sigma^{-\frac{1}{2}}(\theta)$ is the Jacobian matrix of some mapping, f, of D into R^p.

This leads us to seek conditions that insure that a matrix function of θ, say $M(\theta)$, is a Jacobian matrix. It is well known (assuming that $M(\theta)$ is continuously differentiable) that a necessary and sufficient condition for $M(\theta) = (m_{ij}(\theta))$ to be a Jacobian matrix is that

$$\frac{\partial m_{ij}}{\partial \theta_k} = \frac{\partial m_{ik}}{\partial \theta_j} \tag{4.3}$$

for all i, j, k, (see for example, Courant (1936) page 353). In other words, the Jacobian matrix of each row of M must be symmetric. Therefore, for example, in the case where $p = 2$, the matrix

$$M(\theta) = \begin{pmatrix} 1 & \theta_1\theta_2 \\ \theta_1\theta_2 & 1 \end{pmatrix}$$

violates $\partial m_{11}/\partial\theta_2 = \partial m_{12}/\partial\theta_1$ and hence we have a case where $M(\theta)$ is symmetric, positive definite (if $-1 < \theta_1\theta_2 < 1$) and continuous, but for which the equation

$$\left(\frac{\partial f}{\partial \theta}\right) = M(\theta)$$

does not have a solution.

To find conditions that are necessary and sufficient for the existence of solutions to (4.2) we may proceed by applying (4.3) to the right-hand side of (4.2). This appears to be a very complicated program except when $p = 2$, and I shall not pursue the general case, here. The next section executes this plan for the two-dimensional case.

5. Existence of solutions: case of $p = 2$. When $p = 2$, the orthogonal matrices, $\Gamma(\theta)$, whose elements are continuous functions of θ have one of two forms depending on the value of det $(\Gamma(\theta))$. If det $(\Gamma(\theta)) \equiv 1$, or $\equiv -1$, then

$$\Gamma(\theta) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}, \quad \text{or} \quad = \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix} \tag{5.1}$$

respectively for some continuous real-valued function $\varphi = \varphi(\theta)$. To simplify our notation we shall let $A(\theta) = \Sigma^{-\frac{1}{2}}(\theta)$ be a given square root of $\Sigma^{-1}(\theta)$, and the argument θ will be suppressed whenever possible.

The following theorem solves the problem of the existence of covariance stabilizing transformations when $p = 2$. Observe that the square root of Σ^{-1}, A, is not assumed to be symmetric because the useful square roots of a symmetric matrix need not themselves be symmetric, e.g. the triangular square root is often easier to compute than the symmetric one.

THEOREM 1. *Let $A(\theta)$ be a given 2 by 2 matrix function of θ.*

(a) *An orthogonal matrix, $\Gamma(\theta)$, exists such that $\Gamma(\theta)A(\theta) = (\partial f/\partial\theta)$ for some mapping f: $D^2 \to R^2$ if and only if the elements of A satisfy the following partial differential equation:*

$$\frac{\partial}{\partial\theta_2}\left[\frac{1}{\det(A)}a_{11}\left(\frac{\partial a_{11}}{\partial\theta_2} - \frac{\partial a_{12}}{\partial\theta_1}\right) + a_{21}\left(\frac{\partial a_{21}}{\partial\theta_2} - \frac{\partial a_{22}}{\partial\theta_1}\right)\right] = -\frac{\partial}{\partial\theta_1}\left[\frac{1}{\det(A)}a_{12}\left(\frac{\partial a_{11}}{\partial\theta_2} - \frac{\partial a_{12}}{\partial\theta_1}\right) + a_{22}\left(\frac{\partial a_{21}}{\partial\theta_2} - \frac{\partial a_{22}}{\partial\theta_1}\right)\right]. \tag{5.2}$$

(b) *If $\Gamma(\theta)$ exists, then, $\varphi(\theta)$ may be found by integrating the equations*

$$\begin{bmatrix} \dfrac{\partial\varphi}{\partial\theta_1} \\ \dfrac{\partial\varphi}{\partial\theta_2} \end{bmatrix} = \frac{1}{\det(A)}A'\begin{bmatrix} \left(\dfrac{\partial a_{11}}{\partial\theta_2} - \dfrac{\partial a_{12}}{\partial\theta_1}\right) \\ \left(\dfrac{\partial a_{21}}{\partial\theta_2} - \dfrac{\partial a_{22}}{\partial\theta_1}\right) \end{bmatrix}. \tag{5.3}$$

Before proving Theorem 1 we state a corollary giving this result in a special case needed in the next section.

COROLLARY 1. *If A is diagonal then* (5.2) *becomes*

$$\frac{\partial}{\partial\theta_1}\left[\frac{1}{a_{11}}\frac{\partial a_{22}}{\partial\theta_1}\right] + \frac{\partial}{\partial\theta_2}\left[\frac{1}{a_{22}}\frac{\partial a_{11}}{\partial\theta_2}\right] = 0 \tag{5.4}$$

and (5.3) *becomes*

$$\frac{\partial\varphi}{\partial\theta_1} = \frac{1}{a_{22}}\frac{\partial a_{11}}{\partial\theta_2} \tag{5.5a}$$

and

$$\frac{\partial\varphi}{\partial\theta_2} = -\frac{1}{a_{11}}\frac{\partial a_{22}}{\partial\theta_1}. \tag{5.5b}$$

The remainder of this section is devoted to proving Theorem 1 and may be ignored by the reader only interested in the result. Theorem 1 and Corollary 1 are applied to some examples in the next section. As this proof does not appear to generalize to $p > 2$ we only indicate the essential computations.

PROOF OF THEOREM 1. We assume $\det(\Gamma(\theta)) \equiv 1$; a similar computation works if $\det(\Gamma(\theta)) \equiv -1$. Let $B(\theta) = \Gamma(\theta)A(\theta)$. Now apply the condition (4.3) to each row of B. This sets the θ_2-derivative of the first element of each row of B equal to the θ_1-derivative of the second element of each row of B, and produces two equations which by some algebraic manipulation may be expressed in matrix form as:

$$(5.6)\qquad \Gamma(\theta)\begin{bmatrix}\left(\dfrac{\partial a_{11}}{\partial\theta_2}-\dfrac{\partial a_{12}}{\partial\theta_1}\right)\\ \left(\dfrac{\partial a_{21}}{\partial\theta_2}-\dfrac{\partial a_{22}}{\partial\theta_1}\right)\end{bmatrix} = \Gamma(\theta)\det(A)(A')^{-1}\begin{bmatrix}\dfrac{\partial\varphi}{\partial\theta_1}\\ \dfrac{\partial\varphi}{\partial\theta_2}\end{bmatrix}.$$

Solving (5.6) for $(\partial\varphi/\partial\theta)$ yields (5.3). At this point it is worth summarizing what has been established. First of all, the existence of $\Gamma(\theta)$ is equivalent to the existence of $\varphi(\theta)$ satisfying (5.3). Secondly, if φ exists then it may be found in terms of the elements of A via (5.3). Thirdly, (5.3) has the interpretation that the two components of the vector obtained by carrying out the implied multiplications on the right-hand side of (5.3) are, respectively, the θ_1- and θ_2-partial derivatives of some function, φ. But now apply the condition (4.3) to (5.3) to insure that φ does, indeed, exist. This leads to equation (5.2) and proves the theorem. □

6. Three examples. In this section I examine three examples. The first two illustrate that covariance stabilization can be impossible. The third illustrates how to find a covariance stabilizing transformation when one exists.

EXAMPLE 1. *The sample mean and variance from a normal distribution.* Let $\bar{X}$ and S^2 be the sample mean and variance from a sample of n independent observations from the $N(\mu, \sigma^2)$ distribution. Let $\theta_1 = \mu$ and $\theta_2 = \sigma^2$, $\theta = (\theta_1, \theta_2)'$, $T_n = (\bar{X}, S^2)'$. It is easy to show that

$$\mathscr{L}[n^{\frac{1}{2}}(T_n - \theta)] \to N(0, \Sigma(\theta))$$

where $\Sigma(\theta) = \operatorname{diag}(\theta_2, 2\theta_2^2)$.

Let $A = \Sigma^{-\frac{1}{2}}$, then $A = \operatorname{diag}(\theta_2^{-\frac{1}{2}}, (2\theta_2^2)^{-\frac{1}{2}})$. If these functions are substituted into the left-hand side of (5.4) the result is $(2\theta_2)^{-\frac{3}{2}}$ which is never zero. Thus, this choice of $\Sigma(\theta)$ cannot be stabilized. This is not too surprising, since otherwise we could find *two* asymptotically independent statistics, f_1 and f_2, such that the asymptotic distribution of $f_i(\bar{X}, S^2) - f_i(\mu, \sigma^2)$ is independent of σ^2 (and μ). Log (S^2) is *one* such statistic that does exist.

EXAMPLE 2. *The trinomial distribution.* Let nT_n have a trinomial distribution. The parameter space, D, is given by:

$$D = \{(\theta_1, \theta_2) : \theta_i > 0 \text{ and } \theta_1 + \theta_2 < 1\}.$$

Standard theory implies that

$$\mathscr{L}[n^{\frac{1}{2}}(T_n - \theta)] \to N(0, \Sigma(\theta))$$

where

$$\Sigma(\theta) = \begin{pmatrix}\theta_1(1-\theta_1) & -\theta_1\theta_2\\ -\theta_1\theta_2 & \theta_2(1-\theta_2)\end{pmatrix}.$$

This choice of $\Sigma(\theta)$ may be attacked directly but I choose to diagonalize the problem and apply Corollary 1 as follows. Set

$$Y_{1n} = h_1(T_n) = \sin^{-1}(T_{1n}^{\frac{1}{2}}) + \sin^{-1}(T_{2n}^{\frac{1}{2}}) \quad \text{and} \tag{6.1}$$
$$Y_{2n} = h_2(T_n) = \sin^{-1}(T_{1n}^{\frac{1}{2}}) - \sin^{-1}(T_{2n}^{\frac{1}{2}}) .$$

Let $Y_n = h(T_n) = (h_1(T_n), h_2(T_n))$. Reparametrizing in terms of $\psi = h(\theta)$, the parameter space, D, is transformed to

$$D^* = \{(\psi_1, \psi_2) : 0 < \psi_1 < \pi/2,\ -\psi_1 < \psi_2 < \psi_1\} .$$

The covariance matrix, Σ, is transformed into

$$\Lambda(\psi) = \operatorname{diag}(\sigma^2(\psi), 1 - \sigma^2(\psi)) \tag{6.2}$$

where

$$\sigma^2(\psi) = (\cos\psi_1)/(\cos\psi_1 + \cos\psi_2) . \tag{6.3}$$

Thus, the trinomial problem may be put into the equivalent form of

$$\mathscr{L}[n^{\frac{1}{2}}(Y_n - \psi)] \to N(0, \Lambda(\psi))$$

where Λ is given by (6.2). If $A = \Lambda^{-\frac{1}{2}}$, then A is given by

$$A = \operatorname{diag}(\sigma(\psi)^{-1}, (1 - \sigma^2(\psi))^{-\frac{1}{2}}) .$$

If we now apply Corollary 1, a straightforward calculation reveals that (5.4) is not satisfied for any value of ψ and thus the trinomial distribution does not possess a covariance stabilizing transformation.

EXAMPLE 3. *The joint Poisson-Gamma distribution.* Let X and Y be two jointly distributed random variables. X is Poisson with mean μ and $Y \mid X = x$ is Gamma with parameters λ and x, i.e.

$$f_{Y|X}(y \mid x) = (\Gamma(x))^{-1}\lambda^{-x}y^{x-1}e^{-y/\lambda} \quad \text{if} \quad x \geqq 1\,, \quad y > 0 \quad \text{and} \tag{6.4}$$
$$P\{Y = 0 \mid X = 0\} = 1 .$$

We let $\theta_1 = \mu$ and $\theta_2 = \lambda\mu$. The mean vector of (X, Y) is $\theta = (\theta_1, \theta_2)$ and the covariance matrix is

$$\Sigma(\theta) = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & 2\theta_2^2/\theta_1 \end{pmatrix} . \tag{6.5}$$

The parameter space $D = \{(\theta_1, \theta_2) : \theta_i > 0\}$. Let T_n be the sample mean of n i.i.d. random vectors each distributed as (X, Y). Then $\mathscr{L}[n^{\frac{1}{2}}(T_n - \theta)] \to N(0, \Sigma(\theta))$ where $\Sigma(\theta)$ is given in (6.5). If $A = \Sigma^{-\frac{1}{2}}$ then A may be taken to be

$$A = \begin{pmatrix} \left(\frac{2}{\theta_1}\right)^{\frac{1}{2}} & -\frac{1}{\theta_2}\left(\frac{\theta_1}{2}\right)^{\frac{1}{2}} \\ 0 & \frac{1}{\theta_2}\left(\frac{\theta_1}{2}\right)^{\frac{1}{2}} \end{pmatrix} . \tag{6.6}$$

A straightforward calculation reveals that the differential equation (5.2) with this choice of A becomes

$$\frac{\partial}{\partial\theta_2}[(2\theta_1)^{-1}] = \frac{\partial}{\partial\theta_1}[(-2\theta_2)^{-1}] \tag{6.7}$$

which is true so that A satisfies (5.2) and we may conclude that covariance stabilization is possible in this example. It remains to find the stabilizing transformation. First we solve equation (5.3) to find $\varphi(\theta)$. This yields

$$\varphi(\theta) = L(\theta) + c_1 \quad \text{where} \quad L(\theta) = 2^{-1}\log(\theta_1/\theta_2) \tag{6.8}$$
and c_1 is an arbitrary constant.

Next we use this choice of $\varphi(\theta)$ and solve (4.2) for f_1 and f_2. The equations one obtains for f_1 and f_2 by this procedure may be expressed as

$$\begin{aligned} f_1(\theta) &= 2\theta_1^{\frac{1}{2}}\sin(L(\theta)) \\ f_2(\theta) &= 2\theta_1^{\frac{1}{2}}\cos(L(\theta)) \,. \end{aligned} \tag{6.9}$$

It is easily verified that these choices of f_1 and f_2 do, in fact, satisfy (2.3) and stabilize the covariance matrix of T_n. In terms of the original parameters, μ and λ, the new parameters may be expressed as

$$\begin{aligned} f_1(\mu, \lambda) &= -2\lambda^{\frac{1}{2}}\sin(\log\mu^{\frac{1}{2}}) \\ f_2(\mu, \lambda) &= 2\lambda^{\frac{1}{2}}\cos(\log\mu^{\frac{1}{2}}) \,. \end{aligned} \tag{6.10}$$

7. Discussion. Of the examples in Section 6, the trinomial is the most interesting from a practical point of view. My result states that it is impossible to transform a set of trinomial random variables so that they all have a common known asymptotic covariance matrix for all values of the unknown parameters. Such a transformation would have generalized the arcsine-squareroot transform for sets of binomial proportions and might have proved useful in the multivariate analysis of variance of multiply-indexed sets of trinomial variables.

A negative result for the trinomial is an ominous sign for the possibilities in higher dimensions. I conjecture that covariance stabilization is impossible for the general multinomial distribution, but I do not believe that the elementary techniques I used for $p = 2$ will give this result for $p > 2$.

One attitude towards the trinomial example is that there are three functions of θ—two variances and one covariance—to be stabilized but only two functions—f_1 and f_2—to do the stabilizing. Thus on the surface it looks as if the nonexistence of a solution is due, roughly, to a situation with "too few unknowns and too many equations." However, this intuition must be wrong because the transformation (6.1) produces an entirely equivalent problem with a covariance matrix that is completely determined by a single function of the two parameters (6.3). Now there are "too many unknowns and too few equations." Example 1 finishes out the cases. That situation has two variances to be stabilized and no covariances, i.e., there are the "same number of unknowns and equations." Evidently, the impossibility of covariance stabilization in these cases is due to the way the variances and covariances are related to each other rather than merely to their number.

One important application of the asymptotic solutions to the variance stabilizing problem is to provide a guide for finding useful preasymptotic solutions—i.e. those that approximately stabilize the variance in small samples. For example, see Freeman and Tukey (1950) for a discussion of transforming Binomial and Poisson variates. My negative asymptotic results for the trinomial do not mean that approximate covariance stabilization is impossible in finite samples. However, by not having an asymptotic solution the way towards a reasonably good approximate solution (if it exists at all) for finite samples is still unclear.

8. Acknowledgments. I would like to thank Donald Rubin, Arthur Dempster, Stephen Fienberg, David Hoaglin, Frederick Mosteller, Michael Sutherland, and a referee for their comments on a preliminary draft of this paper.

REFERENCES

COURANT, R. (1936). *Differential and Integral Calculus*, **2**. Interscience Publishers, New York.
FREEMAN, M. F. and TUKEY, J. W. (1950). Transformations related to the angular and the square root. *Ann. Math. Statist.* **21** 607–611.
PRATT, J. (1959). On a general concept of 'In Probability'. *Ann. Math. Statist.* **30** 544–558.
RAO, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS 02138