The Annals of Statistics
1989, Vol. 17, No. 2, 711–721

ON THE DETERMINANTS OF MOMENT MATRICES[1]

BY BRUCE G. LINDSAY

The Pennsylvania State University

An investigation is carried out in the behavior of the determinants of certain moment matrices, for which the (i, j) entry is the $(i+j)$th moment of a distribution F. The determinant can be represented as the expected value of a U-statistic type kernel. The structure of the kernel illustrates how the determinant carries information about the number of support points of the distribution F. The kernel representation can be extended to the determinant of a matrix of moment generating function derivatives, where the (i, j) entry is the $i+j$th derivative of the moment generating function of F. When done, this reveals that this determinant is itself, as a function of t, a moment generating function. When this somewhat surprising result is applied to members of the quadratic variance exponential family, one obtains the result that they are closed under this two-step operation of taking derivatives, then computing determinants. This results in an elementary recursion for the values of the moment determinants. The final result gives the convergence of the moment determinants to the normal theory values under central limit theorem conditions.

1. Introduction. Let F be an arbitrary univariate distribution function with moments $m_0 = 1$, $m_1 = E[X]$, $m_2 = E[X^2], \dots$ and moment generating function $m(t)$, here assumed to be finite on an interval containing 0. The objective of this article is to analyze the structure of the determinants of the *moment matrices*

$$\mathbf{M}_p := \begin{bmatrix} 1 & m_1 & m_2 & \cdots & m_p \\ m_1 & m_2 & m_3 & & \cdot \\ m_2 & m_3 & m_4 & & \cdot \\ \vdots & & & & \cdot \\ m_p & \cdot & \cdot & \cdot & m_{2p} \end{bmatrix} = (m_{i+j}), \tag{1.1}$$

$$i = 0, \dots, p, \ j = 0, \dots, p,$$

and the corresponding Hankel matrices $\mathbf{M}_p(t) = (m^{(i+j)}(t))$, whose entries are derivatives of the moment generating function.

The key motivation for this analysis is the importance of these matrices in describing certain fundamental properties of the underlying distribution [e.g., Uspensky (1937), Widder (1947) and Karlin (1968)]. Moment matrices also arise in the study of optimal design for polynomial regression [e.g., Hoel (1958)]. In this article we develop several new properties of the determinants of these matrices, some of which are used in a sequel article [Lindsay (1989)] to describe

Received January 1987; revised May 1988.
[1]This research was supported by National Science Foundation Grant DMS-88-01514.
AMS 1980 *subject classifications*. Primary 62E10, 62G05; secondary 62H05.
Key words and phrases. Moment matrix, Hankel determinant, moment generating function, quadratic variance exponential family.

the behavior of moment estimators of mixing distributions. One particular application in this article is evaluation of $\det \mathbf{M}_p$ for all members of the quadratic variance exponential families. This result is achieved by establishing several intriguing relationships for the Hankel determinants.

The first result, in Section 2, is a suggestive representation of $\det \mathbf{M}_p$. Let $X_0, X_1, \ldots, X_p$ be a random sample from F. Then it will be shown in Corollary 2B that

$$\det \mathbf{M}_p = \frac{1}{(p+1)!} E\left[\prod_{j>k} (X_j - X_k)^2\right], \tag{1.2}$$

thereby extending to higher-order moments the relationship

$$\det \mathbf{M}_1 = \operatorname{Var} X = \tfrac{1}{2} E[X_1 - X_0]^2.$$

The latter is the well-known kernel representation of the variance used in the theory of U-statistics [e.g., Serfling (1980)] and (1.2) is one natural extension. [See Good (1975) for a different extension, via representation, to the higher-order cumulants.] The method of constructing this representation is important as well, as in this and the sequel article it will be used repeatedly to create useful and informative representations of related matrices.

The representation (1.2) makes clear two other important attributes of $\det \mathbf{M}_p$: (1) $\det \mathbf{M}_p$ is invariant under location changes in F; and (2) $\det \mathbf{M}_p$ is nonnegative. Moreover, it equals 0 if and only if F is a discrete distribution with p or fewer points of support as then, and only then, for every realization $x_0, \ldots, x_p$ (on a set of probability 1) there will exist at least one pair of indices (j, k) with $x_j = x_k$.

The second property indicates that there is a sense in which $\det \mathbf{M}_p$ generalizes the idea of variance from "variability about one point" to "variability about p points." Hereafter we will use the notation v_p for $\det \mathbf{M}_p$ to suggest this aspect, with the variance thereby being v_1. In the spirit of this aspect, it will be shown in Section 2 that there exists a set of p points $\{r_1, \ldots, r_p\}$ such that the following generalized Tchebycheff inequality holds:

$$P\left\{\inf_i \{|X - r_i|\} > \varepsilon\right\} \leq v_p / v_{p-1} \varepsilon^{2p}. \tag{1.3}$$

That is, when the ratio v_p / v_{p-1} is small, we can be sure the distribution is concentrated near the set of p-points. The Tchebycheff inequality $(p = 1)$ is useful for showing convergence in probability when the variance goes to 0; inequality (1.3) can be used in a similar fashion to show weak convergence to a p-point distribution when v_p goes to 0 and the low-order moments converge. This will be used in the sequel article to show consistency of a method-of-moments technique.

Being a determinant, v_p appears rather cumbersome to evaluate and awkward to manipulate. The following results concerning moment generating functions are useful in this regard, but also in themselves reveal a somewhat surprising structure to the moment problem.

Next, consider the moment generating function $m(t)$, hereafter assumed to be finite on some nonempty domain (a, b) containing 0. Let D^k denote k-fold differentiation with respect to t. It is well known that $D^2 \log m(t) \ge 0$; that is, the moment generating function is log-convex. In particular, this implies that $mm'' - m'^2 \ge 0$, with equality only for degenerate (one support point) distributions. To generalize, let

$$\mathbf{M}_p(t) := \begin{bmatrix} m(t) & m'(t) & m''(t) & \cdots & m^{(p)}(t) \\ m'(t) & m''(t) & & & \\ m''(t) & & & & \vdots \\ \vdots & & & & \\ m^{(p)}(t) & \cdots & & & m^{(2p)}(t) \end{bmatrix}. \tag{1.4}$$

Observe that $\mathbf{M}_p(0) = \mathbf{M}_p$ and that $\det \mathbf{M}_1(t) = m(t)m''(t) - m'^2(t)$.

Karlin (1968), page 73, has shown that $V_p(t) := \det \mathbf{M}_p(t)$ is a nonnegative function, being identically 0 if and only if F has p or fewer points of support. Theorem 3A of this article presents the stronger result that $V_p(t)$ *itself* has the form $\int \exp(tx)\, d\mu(x)$ for a finite nonnegative measure μ. We will call such an integral a moment generating function or mgf; it is also known as a Laplace–Stieltjes integral. The total mass of μ is $V_p(0) = v_p$. Thus the mapping Γ_p: $m(t) \to V_p(t)$ takes mgf into mgf, with a kernel corresponding to all p-or-fewer point distributions. Moreover, it will be shown that Γ_p takes the $n(0, \sigma^2)$ distribution into a measure which, when normalized to mass 1, is $n(0, \sigma^2(p+1))$.

Having thus identified $V_p(t)$ as a potentially important accessory to the problem, a second important structural result, Lemma 3C, establishes a simple recursion relationship for use in the construction of $V_p(t)$ from $\{V_q(t)\colon q < p\}$. From the lemma one can deduce that the second ratios of the determinant sequence, that is,

$$\Lambda_p := v_p v_{p-2}/v_{p-1}^2 \tag{1.5}$$

are basic elements in the analysis of the moment sequence. In fact, Λ_p is simply the variance of the distribution corresponding to the normalized mgf $V_{p-1}(t)/v_{p-1}$. In the $n(0, \sigma^2)$ model, therefore, Λ_p is simply $\sigma^2 p$.

We note as an aside regarding the interpretation of Λ_p that it can be shown that

$$v_p/v_{p-1} = \inf_{\boldsymbol{\beta}} \left\{ \operatorname{Var}\left[X^p - \sum_{i=0}^{p-1} \beta_i X^i \right] \right\}. \tag{1.6}$$

That is, the ratio of consecutive moment matrix determinants is the residual variance in the linear prediction of X^p by $1, X, X^2, \ldots, X^{p-1}$. Thus, as a ratio of ratios, Λ_p represents the growth in the linear unpredictability of the pth moment.

In Section 4, as a first application, the recursion methods of Section 3 are employed upon the quadratic variance class of exponential family distributions [Morris (1982, 1983)]. In particular, if F is any such distribution, then the distribution corresponding to $V_p(t)$ is itself a member of the same quadratic variance family. Moreover, the sequence $\{\Lambda_p\}$ is linear in p for the normal and Poisson distributions, and quadratic for the gamma, binomial and negative binomial (Corollary 4C).

Next, let F_n be the distribution function of $\sqrt{n}\,\overline{X}$, where $\overline{X}$ is the mean of n i.i.d. observations from F. A natural sequel to the last result is to consider the convergence of $\Lambda_p(F_n)$ to the normal theory value of $p\sigma^2$ under these central limit theorem conditions. In Theorem 5A it is shown that the expansion in n of $\Lambda_p(F_n)$ is

$$\Lambda_p(F_n)/p\sigma^2 = 1 + (p-1)\big(\kappa_2\kappa_4 - \kappa_3^2\big)/\kappa_2^3 n + O\big(1/n^2\big), \tag{1.7}$$

where κ_j is the jth cumulant of F. The simplicity of the $O(1/n)$ term in the expansion is somewhat surprising in light of the seemingly complicated function under investigation. In the quadratic variance exponential family the order $1/n^2$ term is 0.

2. Representations of the moment determinants. The representation result (1.2) is a special case of the following theorem, for which we must first develop some notation. Let $\mathbf{X}^t = (X_0, X_1, X_2, \dots, X_p)$ be a vector of random variables from a $(p+1)$-variate distribution F. The second moment matrix of F is the matrix $\mathbf{C} := E\{\mathbf{X}\mathbf{X}^t\}$. Let $\mathbf{X}_0, \dots, \mathbf{X}_p$ be $p+1$ independent replicates from F. Define the matrix $\mathbf{A}$ to be the $(p+1) \times (p+1)$ matrix with jth column $\mathbf{X}_j$. We can interpret $W := \det \mathbf{A}$ to be the volume (signed) of the parallelepiped formed by the replicates of X. The next theorem indicates that the determinant of the second moment matrix is a measure of the variability of that volume.

THEOREM 2A. *Suppose that F has second-moment matrix* $\mathbf{C}$. *The random variable W defined above has mean* 0 *and variance* $(p+1)!\det \mathbf{C}$.

This result can be found in Wilks (1960). Since the method of proof will be repeatedly used to create representations, it is here provided in the Appendix. The main device in the proof is to represent the moments in the matrix $\mathbf{C}$ as expectations using independent replicates in each column. This enables one to commute the E and det operations. Once this is done, we simply average over the $(p+1)!$ possible permutations of the indices on the replicates.

REMARK. The result could also be written as $\det \mathbf{C} = E(\det \mathbf{A}\mathbf{A}^t)/(p+1)!$, or, letting $\hat{\mathbf{C}}$ be the matrix of second moments of the sample of $p+1$ replicates, we have

$$\det \mathbf{C} = (p+1)^{p+1} E \det \hat{\mathbf{C}}/(p+1)!.$$

This formula can be extended via the Cauchy–Binet identity to a sample of n replicates,

$$\det \mathbf{C} = E\big[\det \hat{\mathbf{C}}\big]\frac{(n-1)\cdots(n-p)}{n\cdots n}. \tag{2.1}$$

The formula can also be shown to hold if $\mathbf{C}$ and $\hat{\mathbf{C}}$ are the theoretical and sample covariance matrices, and therefore identifies the bias correction for $\det \hat{\mathbf{C}}$ as an estimator of $\det \mathbf{C}$.

COROLLARY 2B. *If* $\mathbf{M}_p$ *is the moment matrix for a distribution* F, *then representation* (1.2) *holds.*

PROOF. Simply use $\mathbf{X}^t = (1, X, X^2, \dots, X^p)$ in the theorem. In this case $\det \mathbf{A}$ is simply the Vandermonde determinant. □

REMARKS. The corollary could also be derived from a "basic composition result" in Karlin (1968), page 17, (2.5), as extended by averaging over the permutations of the indices. For the special case when the distribution F has exactly $p + 1$ points of support, the result in the corollary can be found in Hoel (1958) and the result in the theorem in Karlin and Studden (1966).

Formula (2.1) indicates the bias adjustment in the use of the sample moments for the true moments in the estimation of $v_p = \det \mathbf{M}_p$.

To conclude this section, we prove the generalized Tchebycheff inequality (1.3). We will need the following information: If the sequence of numbers $\{1, m_1, \dots, m_{2p-1}\}$ are the initial moment sequence of some distribution with p or more points, then there exists a unique p-point distribution F_p with the same initial moment sequence. This result, found in Uspensky (1937), is discussed in greater detail in the sequel article.

THEOREM 2C. *Suppose that* F *is a distribution with* p *or more points of support. Let* F_p *be the* p*-point distribution with the same initial* $2p - 1$ *moments, and let* $\{r_1, \dots, r_p\}$ *be its points of support. Then inequality* (1.3) *holds.*

PROOF. We start with the equality (1.6). If we let β^* be the coefficients which minimize the given expectation, and $\rho(t) = t^p - \sum \beta_i^* t^i$ be the corresponding polynomial, then we will show that $\rho(t) = \prod(t - r_i)$. We note that the minimization problem is completely determined by the first $2p - 1$ moments of X. Thus without loss of generality we can solve for the β^* values by using the distribution F_p. However, in this case we can make the given expectation 0 if and only if we use the polynomial $\rho(X)$. Thus $E[\rho^2(X)] = v_p/v_{p-1}$. Finally, we have the simple inequality

$$P\Big[\inf_i \big\{(X - r_i)^2\big\} > \varepsilon^2\Big] \le P\Big[\prod (X - r_i)^2 > \varepsilon^{2p}\Big].$$

The proof is completed by applying the Markov inequality to the right-hand side. □

3. Moment generating functions and Hankel determinants. Following (1.3) the function $V_p(t)$ was defined to be $\det \mathbf{M}_p(t)$, where $\mathbf{M}_p(t)$ is the Hankel matrix of the moment generating function $m(t)$. This section develops some basic results concerning $V_p(t)$. The first theorem presents the result that $V_p(t)$ is itself a moment generating function for a finite measure.

Theorem 3A. *Suppose that the moment generating function $m(t)$ exists in a neighborhood of* 0. *If the distribution F has fewer than $p+1$ points of support, then $V_p(t)=0$ for all $t \in R$. Otherwise $V_p(t)$ is the moment generating function for the measure μ_p which has the density*

$$g(z)=E\left[\prod_{j>k}(X_j-X_k)^2 \mid Z=z\right]$$

with respect to the distribution of the $(p+1)$*-fold convolution* $Z=X_0+X_1+\cdots+X_p$. *The measure μ_p has total mass $V_p(0)=v_p$.*

Proof. Using Theorem 2A with vector $X^t=(1,X,X^2,\dots,X^p)\exp(tX/2)$ gives

$$V_p(t)=E\left[\prod_{j>k}(X_j-X_k)^2 e^{t(X_0+\cdots+X_p)}\right]\Big/(p+1)!.$$

Conditioning on Z yields the result. □

Although this result might strike one as surprising, it may not yet be clear how the moment generating function property can be useful. The answer is that we are going to be able to identify the distributions corresponding to $V(t)$ for important classes of initial moment generating functions. The first one is easy.

Example. For the normal distribution $n(\mu,\sigma^2)$ the differences X_j-X_k are independent of the sum $X_0+\cdots+X_p$ and so the distribution corresponding to $V_p(t)/v_p$ is the convolution of $p+1$ normals, hence $n(\mu(p+1),\sigma^2(p+1))$. (The normalizing constant v_p will be derived shortly.)

To tackle other distributions, we need better tools; the key is to develop a recursion that can be used to derive $V_p(t)$ from $V_0(t)=m(t),\dots,V_{p-1}(t)$. To proceed, we first need to transform to a system of variables with simpler recursive relationships. First, define the (unstandardized) cumulant generating function corresponding to $V_p(t)$,

$$\kappa_p(t) := \log\det \mathbf{M}_p(t)=\log V_p(t).$$

Then form the sequence of second differences of the $\{\kappa_p\}$ sequence,

$$\Delta_p(t) := \kappa_{p-2}(t)-2\kappa_{p-1}(t)+\kappa_p(t), \qquad p=0,1,2,\dots, \tag{3.1}$$

where $\kappa_{-2}(t)$ and $\kappa_{-1}(t)$ are defined to be 0. We note that the inverse transformation is

$$\kappa_p=(p+1)\Delta_0+p\Delta_1+\cdots+\Delta_p. \tag{3.2}$$

The following equality is a special case of Sylvester's identity [e.g., Karlin (1968), page 72]:

$$V_{p+1}(t)V_{p-1}(t)=V_p(t)V_p''(t)-V_p'^2(t). \tag{3.3}$$

By dividing by V_p^2 we may rewrite (3.3), provided $V_p(t)$ is not 0, as

$$\exp(\Delta_{p+1}) = D^2\kappa_p. \tag{3.4}$$

This identity has an important implication. Evaluation at 0 proves the following.

COROLLARY 3B. *The variance of the distribution corresponding to mgf $V_p(t)$ is Λ_{p+1}.*

In particular, this implies that $\Lambda_p = p\sigma^2$ for the $n(\mu, \sigma^2)$ distribution. By using appropriate initial values, one can then derive v_p,

$$\Lambda_p = p\sigma^2 \Rightarrow v_p/v_{p-1} = p!\sigma^{2p} \Rightarrow v_p = \prod_{j=1}^{p} j!\sigma^{2p(p+1)}.$$

Finally, putting (3.4) together with (3.2) implies the following recursive relationship. This will soon be used to generalize the results for the normal distribution to the entire quadratic variance exponential family.

LEMMA 3C. *Provided that F has more than p points of support,*

$$\exp(\Delta_{p+1}) = D^2\left[(p+1)\Delta_0 + p\Delta_1 + \cdots + \Delta_p\right].$$

4. Quadratic variance exponential families. Sufficient tools have now been developed to derive the behavior of $V_p(t)$ and hence $\{\det \mathbf{M}_p\}$ for a very important class of exponential family distributions.

Let $\kappa(t) := \log[m(t)]$ be the cumulant generating function for F and suppose that for some a, b and c the function $\kappa(t)$ satisfies

$$\kappa''(t) = a + b[\kappa'(t)] + c[\kappa'(t)]^2. \tag{4.1}$$

If we construct the *exponential tilt* of F, a family of distributions for X defined by

$$dF_\theta(x) = e^{\theta x}\, dF(x)\exp(-\kappa(\theta)),$$

then it is an exponential family of distributions and (4.1) indicates that this family has the "quadratic variance property." That is, the variance of X in this family is a quadratic function of the mean value parameter [Morris (1982, 1983)]. As will be seen, several important distributions fall in this class, and we can easily determine $V_p(t)$ for them. However, a rather differently appearing differential equation than (4.1) turns up in a natural way in the proof, so we first note its equivalence to (4.1) in Lemma 4A, due to Morris (1982), (3.7).

LEMMA 4A. *Suppose $\kappa(t)$ is a cumulant generating function which satisfies one of the following differential equations:*

(i) $\log \kappa''(t) = a + bt + 2c\kappa(t)$, *some* a, b, c.
(ii) $\kappa''(t) = a + b\kappa'(t) + c[\kappa'(t)]^2$, *some* a, b, c.

Then it satisfies the other (with the same b and c).

Note that in evaluation at $t = 0$ in part (i) shows that $a = \log \sigma^2$, where σ^2 is the variance of F.

Morris catalogued all possible quadratic variance distributional families (modulo certain transformations). The basic list is included here for reference, together with our main result concerning Λ_p (Corollary 4C):

	b	c	$\Lambda_p/p\sigma^2$
Normal (μ, σ^2)	0	0	1
Poisson (λ)	1	0	1
Gamma (n, λ)	0	$1/n$	$1 + \frac{1}{n}(p-1)$
Binomial (n, q)	0	$-1/n$	$1 - \frac{1}{n}(p-1)$
Negative binomial (n, q)	1	$1/n$	$1 + \frac{1}{n}(p-1)$
NEF–GHS (n, λ)	0	$1/n$	$1 + \frac{1}{n}(p-1)$

We are now prepared to prove the main result for these families.

THEOREM 4B. *Let F be a distribution with cgf $\kappa(t)$ existing on an interval about* 0. *Suppose that* $\log \kappa(t) = a + bt + 2c\kappa(t)$, *for some a, b and c. Suppose $p <$ number of support points of F. Then*

(i) $$\Delta_p(t) = \log\left[p + c\frac{p(p-1)}{2}\right] + a + bt + 2c\kappa(t) \quad \textit{for } p \geq 2;$$

(ii) $$\kappa_p(t) - \kappa_p(0) = (p+1)\kappa(t) + \frac{p(p+1)}{2}[bt + 2c\kappa(t)] \quad \textit{for } p \geq 0.$$

PROOF. We start with the algebraic identity

$$\det \mathbf{M}_1(t) = m^2(t) D^2 \log m(t).$$

This implies, by using the definition of $\Delta_1(t)$ and then the assumption on κ'', that

$$\Delta_1(t) = \log \kappa''(t) = a + bt + 2c\kappa(t).$$

The recursion of Theorem 3B then gives

$$\begin{aligned}\Delta_2(t) &= \log D^2\{2\kappa(t) + bt + c\kappa(t)\}\\ &= \log[(2+c)\kappa''(t)] = \log(2+c) + (a + bt + 2c\kappa(t)).\end{aligned}$$

Continuing in a like fashion, one obtains by induction

$$\Delta_p = \log\{[p + c[(p-1) + (p-2) + \cdots + 1]]\kappa''\}.$$

This in turn implies result (i). Result (ii) follows from (3.2). □

COROLLARY 4C. *Suppose that $\kappa(t)$ is the cumulant generating function of a distribution F satisfying* (4.1). *Then*

$$\Lambda_p = v_p v_{p-2}/v_{p-1}^2 = \sigma^2 p[1 + c(p-1)],$$

where $\sigma^2 = \exp(a) = \operatorname{Var} X$.

PROOF. Evaluate the formula for $\Delta_p(t)$ at $t = 0$. □

5. Limiting behavior. The results of Section 4 suggest that the behavior of Λ_p under convolution of the basic distribution F might have a relatively simply structure. We illustrate this in the context of the central limit theorem. Let $X_1, X_2, \ldots, X_n$ be a random sample from distribution F. It is known that if the moments of F exist, then the moments of the normalized mean $Y_n = \sqrt{n}(\overline{X} - \mu)$ converge to the appropriate normal theory values. Let F_n be the distribution of Y_n. If $E[X]^{2p}$ is finite, then this implies that $\Lambda_p(F_n) \to p\sigma^2$. The following theorem specifies the first-order error term in n, for p fixed.

We first define

$$m(t; \alpha) := [m(t/\alpha)]^{\alpha^2},$$

which for $\alpha = \sqrt{n}$ represents the moment generating function of $\sqrt{n}\,\overline{X}$. Moreover, let $\Delta_p(t; \alpha)$ be the corresponding $\Delta_p(t)$.

THEOREM 5A. *Suppose that F is a nondegenerate distribution. Let p be a positive integer. Then as $\alpha = \sqrt{n} \to \infty$,*

$$\exp[\Delta_p(t; \alpha)] = p\Delta_0''(t/\alpha) + \alpha^{-2}p(p-1)\Delta_1''(t/\alpha) + \alpha^{-4}R_\alpha(t/\alpha), \tag{5.1}$$

where $R_\alpha(s)$ is $O(1)$ as a function of α; and so

$$\Lambda_p(F_n)/p\sigma^2 = 1 + (p-1)\gamma/n + O(1/n^2),$$

where $\gamma = [\kappa_2\kappa_4 - \kappa_3^2]/\kappa_2^3$.

PROOF. The proof will be by induction. To begin with

$$\Delta_0(t; \alpha) = \alpha^2\Delta_0(t/\alpha)$$

and so

$$\Delta_1(t; \alpha) = \log D^2[\Delta_0(t; \alpha)] = \log \Delta_0''(t/\alpha) = \log \Delta_1(t/\alpha).$$

This verifies (5.1) for $p = 1$.

Next consider the implications of the induction hypothesis (5.1). We start by writing

$$\Delta_p(t; \alpha) = \log p\Delta_0''(t/\alpha)$$
$$+ \log[1 + \alpha^{-2}(p-1)\Delta_1''(t/\alpha)/\Delta_0''(t/\alpha) + \alpha^{-4}R_\alpha^*(t/\alpha)]$$

with new remainder term R_α^*. Differentiating twice with respect to t and using

$\log \Delta_0'' = \Delta_1$ gives

$$\Delta_p''(t;\,\alpha) = \alpha^{-2}\Delta_1''(t/\alpha) + \alpha^{-4}R_\alpha^{**}(t/\alpha).$$

The rest of the proof follows from Lemma 3C. □

APPENDIX

PROOF OF THEOREM 2A. First, $E[W] = E[\det \mathbf{A}] = E[\det \mathbf{A}(e)]$, where we use the symbol $\mathbf{A}(e)$ to denote the matrix $(\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_p)$. For any permutation σ of the indices $(0, 1, 2, \dots, p)$, let $\mathbf{A}(\sigma)$ be the matrix with ith column $\mathbf{X}_{\sigma(i)}$. We then have the relationship

$$\det \mathbf{A}(\sigma) = \begin{cases} \det \mathbf{A}(e) & \text{if } \sigma \text{ is an even permutation,} \\ -\det \mathbf{A}(e) & \text{if } \sigma \text{ is an odd permutation.} \end{cases} \tag{A.1}$$

Since $\mathbf{A}(e)$ and $\mathbf{A}(\sigma)$, with σ odd, are identically distributed, $\det \mathbf{A}(e)$ has a symmetric distribution about 0. Since the following calculation shows that W has a finite second moment, the mean exists and is 0.

Next, we may write, letting the second subscript denote coordinate within replicate,

$$\det \mathbf{C} = \det E\left[X_{00}\mathbf{X}_0, \dots, X_{pp}\mathbf{X}_p\right].$$

That is, we construct the moments in the kth column strictly from the kth replicate of $\mathbf{X}$. The key now is that "E" and "det" commute because the determinant is a sum of products, each of which has exactly one term from each column, and hence from each independent replicate. This gives

$$\det \mathbf{C} = E\left[X_{00} \dots X_{pp} \det \mathbf{A}(e)\right].$$

A permutation σ over the subscripts merely changes the labels of the replicates, so gives the same expectation. That is, we have

$$\det \mathbf{C} = E\left[X_{\sigma(0)0} \dots X_{\sigma(p)p} \det \mathbf{A}(\sigma)\right].$$

Now, using (A.1), average over all permutations on the right-hand side, noticing that the coefficients of $\det \mathbf{A}(e)$ are themselves summands in the determinantal expansion of $\det \mathbf{A}(e)$. The result:

$$\det \mathbf{C} = E\left[\det \mathbf{A}(e) \det \mathbf{A}(e)\right] / (p+1)!. \qquad \square$$

Acknowledgments. The author is very grateful for the assistance of those involved in the editorial process.

REFERENCES

GOOD, I. J. (1975). A new formula for cumulants. *Math. Proc. Cambridge Philos. Soc.* **78** 333–337.
HOEL, P. G. (1958). Efficiency problems in polynomial estimation. *Ann. Math. Statist.* **29** 1134–1146.
KARLIN, S. (1968). *Total Positivity.* Stanford Univ. Press, Stanford, Calif.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley Interscience, New York.

LINDSAY, B. G. (1989). Moment matrices: Applications in mixtures. *Ann. Statist.* **17** 722–740.

MORRIS, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10** 65–80.

MORRIS, C. N. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *Ann. Statist.* **11** 515–529.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill, New York.

WIDDER, D. V. (1947). *The Laplace Transform*. Princeton Univ. Press, Princeton, N.J.

WILKS, S. S. (1960). Multidimensional statistical scatter. In *Contributions to Probability and Statistics* (I. Olkin et al., eds.) 486–503. Stanford Univ. Press, Stanford, Calif.

DEPARTMENT OF STATISTICS
THE PENNSYLVANIA STATE UNIVERSITY
219 POND LABORATORY
UNIVERSITY PARK, PENNSYLVANIA 16802