

UNBIASED ESTIMATION OF SOME MULTIVARIATE PROBABILITY DENSITIES AND RELATED FUNCTIONS¹

BY S. G. GHURYE² AND INGRAM OLKIN

Indiana University and Stanford University

1. Introduction and summary. A problem which has received considerable attention in recent years is that of estimating the tail probability in a distribution belonging to a specified family. Kolmogorov (1950), Lieberman and Resnikoff (1955), and Healy (1956) obtained the unique minimum variance unbiased (UMVU) estimator of $P\{x > c\}$, where x has a normal distribution with unknown mean and variance; they used the Rao-Blackwell theorem for this purpose. Washio, Morimoto and Ikeda (1956) used integral transform theory to study the Koopman-Pitman family. Barton (1961) provided UMVU estimators of the normal, Poisson and binomial distribution functions. The most extensive study is that by Tate (1959) who obtained the UMVU estimators for the cdf (and other functionals) for several probability densities; this was accomplished by using transform theory. In the context of reliability theory, Glasser (1962) estimated the tail of the exponential distribution and Basu (1964) that of the gamma distribution; other related work in the area of reliability theory is that of Rutenfranz (1966), and Zacks and Even (1966). Non-parametric estimation of probability densities has been treated by Rosenblatt (1956), Parzen (1962), Leadbetter and Watson (1963), and Weiss and Wolfowitz (1967).

The more general problem of estimating $Ef(x)$, where f is any given function, was considered by Neyman and Scott (1960) and Schmetterer (1961) in the case that x has a normal distribution with unknown mean and variance. Schmetterer's solution really consists in obtaining the UMVU estimator of the density and then integrating the product of this estimate and f . That this procedure is of wider applicability is seen from the following result.

LEMMA 1. *Let $\{p_\theta, \theta \in \Theta\}$ be a family of densities relative to a σ -finite measure M . If T is a statistic such that there exists an unbiased estimator $\tilde{p}(T, u)$ of $\{p_\theta(u), \theta \in \Theta\}$, for almost every u , then $\int h(u)\tilde{p}(T, u) dM(u)$ is an unbiased estimator of $\varphi_\theta = \int h(u)p_\theta(u) dM(u)$.*

It might, therefore, be of interest to obtain UMVU estimators of the probability density for various families of distributions, which is, in part, the motivation for this paper.

REMARK. It is to be noted that this procedure for estimating φ_θ is not necessarily better than the direct approach, nor is it always feasible; for it can happen that there exists an unbiased estimator of φ_θ , but not of p_θ . Thus, in cases where there is an essential discontinuity in p (such as the family of uniform distribu-

Received 19 August 1968; revised 11 February 1969.

¹ Research supported in part by NSF Grant GP-6681 at Stanford University.

² Now at the University of Alberta, Edmonton, Canada.

tions on finite sub-intervals of the line), p is not estimable, although there are many estimable functions.

Suppose now that (X_1, \dots, X_n) is a sample of independent observations on a population with density $p_\theta(x)$, $\theta \in \Theta$, and $S(X_1, \dots, X_n)$ is a complete sufficient statistic based on this sample. If $\{p_\theta(x), \theta \in \Theta\}$ has an unbiased estimator, then the conditional density (if it exists) of $X_1 | S(X_1, \dots, X_n)$ evaluated at x is the UMVU estimator of $\{p_\theta(x), \theta \in \Theta\}$. Two methods of computation have been used in this context:

(1) the conditional density of $X_1 | S$ may be obtained directly from the joint density of the sample;

(2) alternatively, under suitable conditions, it may be obtained from the conditional cdf by differentiation, $(\partial/\partial x)P_\theta\{X_1 \leq x | S\}$.

Both procedures involve the computation of conditional distributions, which might be cumbersome. E. L. Lehmann (private communication) has shown how the computation of the conditional density is facilitated in many cases by the use of a result due to D. Basu (1955). However, it can happen in some problems that there exist methods specifically suited to those problems which are simpler than the general procedures indicated above. In this paper, we treat a class of such problems, and incidentally are able to estimate some parametric functions which are not densities.

In these problems, we are concerned with location-and-scale-parameter families; we note that a simplification results from breaking up each problem into two sub-problems: first, the problem of estimation assuming the scale-parameter fixed, followed by the problem of estimating an appropriate function with the scale-parameter unknown. This is a consequence of the following result whose proof is simple.

LEMMA 2. *Let S and T be independent statistics with a joint cdf $F(t; \sigma, \tau)G(s; \sigma)$ depending on the parameters $(\sigma, \tau) \in \Theta_1 \times \Theta_2$. Let $b(\sigma, T)$ be an unbiased estimator of $\{h(\sigma, \tau), \tau \in \Theta_2\}$ for each $\sigma \in \Theta_1$ and $a(S, \tau)$ be an unbiased estimator of $\{b(\sigma, \tau), \sigma \in \Theta_1\}$ for each $\tau \in \Theta_2$. Then $a(S, T)$ is an unbiased estimator of $\{h(\sigma, \tau), (\sigma, \tau) \in \Theta_1 \times \Theta_2\}$.*

In the problems we consider, the "scale-parameter" is the covariance matrix of a family of normal distributions; we therefore deal with the gamma and Wishart distributions in Section 2 and use these results to estimate multivariate normal densities with known means. In Section 3, we estimate the multivariate normal density with unknown mean vector and unknown or partially known covariance matrix. Section 4 is concerned with the multivariate normal family with a linearly restricted mean, including the univariate linear model as a special case. Finally, in Section 5, we obtain the UMVU estimators of the entropy of a multivariate normal distribution and of the Kullback-Leibler information number for a pair of multivariate normal distributions.

NOTATION. The determinant of a matrix A is denoted by $|A|$; $A > 0 (\geq 0)$ means that A is a positive (non-negative) definite matrix. The function Ψ of a

symmetric matrix is defined by

$$(1.1) \quad \Psi(A) = |A|, \quad \text{if } A > 0, \\ = 0, \quad \text{otherwise.}$$

Vectors are row-vectors unless transposed (with a prime). The vector $(1, \dots, 1)$ is denoted by e . For a square matrix A , $\text{etr } A$ denotes $\exp(\text{tr } A)$. By $\mathcal{L}(x) = N(k; \xi, \Sigma)$ we mean that x is a random vector with a k -dimensional normal distribution whose density is denoted by

$$(1.2) \quad n(k; x, \xi, \Sigma) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \text{etr} [-\Sigma^{-1}(x - \xi)'(x - \xi)/2].$$

By $\mathcal{L}(S) = W(k, \nu; \Sigma)$ we mean that S is a random $k \times k$ symmetric matrix having a Wishart distribution whose density, relative to $\{k(k+1)/2\}$ -dimensional Lebesgue measure, is denoted by

$$(1.3) \quad w(k, \nu; S, \Sigma) = c(k, \nu) |\Sigma|^{-\nu/2} [\Psi(S)]^{(\nu-k-1)/2} \text{etr} (-\Sigma^{-1}S/2),$$

where $c(k, \nu) = [2^{k\nu/2} \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma(\frac{1}{2}(\nu - i + 1))]^{-1}$. Note that $W(1, \nu; \sigma^2) = \mathcal{L}(\sigma^2 \chi_\nu^2)$.

2. Gamma and Wishart distributions; multivariate normal distribution with a known mean. In this section, the UMVU estimator of the Wishart density is obtained in three cases: (i) Σ is known up to a scalar multiple, (ii) Σ has the form $\sigma^2[(1 - \rho)I + \rho e'e]$, σ and ρ unknown (intra-class correlation model), (iii) Σ is completely unknown. In practice, we typically have a sample from a multivariate normal population whose parameters are partially known, and we construct from the sample an appropriate sufficient statistic. The component of the sufficient statistic which pertains to the covariance matrix has a Wishart distribution, and the canonical problem is to find the UMVU estimator of another Wishart density with a covariance matrix of the same form. This problem is dealt with in the following

THEOREM 1. For any $A \geq 0$ and any real $m < \nu - k + 1$, let

$$(2.1) \quad \gamma(k, m; A, \Sigma) = |\Sigma|^{-m/2} \text{etr} (-\Sigma^{-1}A/2).$$

If $\mathcal{L}(S) = W(k, \nu; \Sigma)$, then the UMVU estimator of $\gamma(k, m; A, \Sigma)$ is

$$(2.2) \quad g_\nu(k, m; A, S) = [c(k, \nu - m)/c(k, \nu)] |S|^{-(\nu-k-1)/2} [\Psi(S - A)]^{(\nu-m-k-1)/2}.$$

PROOF. Unbiasedness follows by direct integration, since

$$g_\nu(k, m; A, S) w(k, \nu; S, \Sigma) = \gamma(k, m; A, \Sigma) w(k, \nu - m; S - A, \Sigma).$$

Completeness of the Wishart distribution then establishes that the estimator is UMVU.

COROLLARY 1. If $\mathcal{L}(S) = W(k, \nu; \sigma^2 \Sigma_0)$, where $\Sigma_0(>0)$ is known and σ is arbitrary, then the UMVU estimator of $w(k, m; A, \sigma^2 \Sigma_0)$, $m < \nu$, is

$$c(k, m) |\Sigma_0|^{-m/2} [\Psi(A)]^{(-k-1)/2} g_{\nu k}(1, mk; a, s),$$

where $a = \text{tr } \Sigma_0^{-1}A$, $s = \text{tr } \Sigma_0^{-1}S$.

COROLLARY 2. If $\mathcal{L}(S) = W(k, \nu; \Sigma)$ with $\Sigma = \sigma^2[(1 - \rho)I + \rho e' e]$, where σ and $\rho(- (k - 1)^{-1} < \rho < 1)$ are unknown, then the UMVU estimator of $w(k, m; A, \Sigma)$, $m < \nu$, is

$$c(k, m)[\Psi(A)]^{(m-k-1)/2} g_\nu(1, m; a, u) g_{k\nu-\nu}(1, km - m; b, v),$$

where $a = eAe'/k$, $u = eSe'/k$, $b = \text{tr } A - a$, $v = \text{tr } S - u$.

COROLLARY 3. If $\mathcal{L}(S) = W(k, \nu; \Sigma)$ and $\Sigma(>0)$ is arbitrary, then the UMVU estimator of $w(k, m; A, \Sigma)$, $m < \nu - k + 1$, is

$$c(k, m)[\Psi(A)]^{(m-k-1)/2} g_\nu(k, m; A, S).$$

REMARK. It has been pointed out to us by W. J. Hall that this result may also be obtained (when $m > k - 1$) using conditional densities as follows: We may write $S = S_1 + S_2$, where $\mathcal{L}(S_1) = W(k, m; \Sigma)$, $\mathcal{L}(S_2) = W(k, \nu - m; \Sigma)$, and S_1 and S_2 are independent. Then the conditional density of $S | S_1 + S_2$ at $S_1 = A$ is the UMVU estimator of $W(k, m; A, \Sigma)$. This provides an interpretation of the estimator in Corollary 3 (when $m > k - 1$): it is the conditional density.

PROOFS OF COROLLARIES. In Corollary 1, since $\Sigma = \sigma^2 \Sigma_0$, where Σ_0 is known, the usual estimator S of Σ can be further reduced to $s = \text{tr } \Sigma_0^{-1} S$, which is sufficient and is complete, since $\mathcal{L}(s) = W(1, k\nu; \sigma^2)$. The density to be estimated is

$$w(k, m; A, \sigma^2 \Sigma_0) = c(k, m)[\Psi(A)]^{(m-k-1)/2} |\Sigma_0|^{-m/2} \gamma(1, mk; a, \sigma^2).$$

The result now follows from Theorem 1 with the association $(k, m, A, \Sigma, s, \nu) = (1, mk, a, \sigma^2, s, \nu k)$.

In Corollary 2, if we set $\alpha^{-1} = \sigma^2(1 + (k - 1)\rho)$ and $\beta^{-1} = \sigma^2(1 - \rho)$, it is well known that (u, v) is a complete sufficient statistic, u and v are independent with $\mathcal{L}(u) = W(1, \nu; \alpha^{-1})$ and $\mathcal{L}(v) = W[1, \nu(k - 1); \beta^{-1}]$, $\Sigma^{-1} = \alpha e' e/k + \beta(I - e' e/k)$ and $|\Sigma|^{-1} = \alpha \beta^{k-1}$. The density to be estimated is

$$w(k, m; A, \Sigma) = c(k, m)[\Psi(A)]^{(m-k-1)/2} \gamma(1, m; a, \alpha^{-1}) \gamma(1, m[k - 1]; b, \beta^{-1}).$$

The result now follows from (2.2) with the associations

$$(k, m, A, \Sigma, S, \nu) = (1, m, a, \alpha^{-1}, u, \nu), \quad \text{and}$$

$$(k, m, A, \Sigma, S, \nu) = (1, m(k - 1), b, \beta^{-1}, v, \nu(k - 1)).$$

Corollary 3 is an immediate consequence of Theorem 1.

REMARK. These corollaries have an immediate application to the problem of estimating a multivariate normal density with a known mean. If x_1, \dots, x_n is a sample from a $N(k; \xi_0, \Sigma)$ population with a known mean vector ξ_0 and an unknown or partially known covariance matrix Σ , and it is desired to estimate $n(k; c, \xi_0, \Sigma)$, the results follow immediately from Corollaries 1-3 by substituting $S = \sum_i (x_i - \xi_0)'(x_i - \xi_0)$, and noting that $n(k; c, \xi_0, \Sigma) = (2\pi)^{-k/2} \gamma(k, 1; (c - \xi_0)'(c - \xi_0), \Sigma)$.

3. Multivariate normal distribution with an unknown mean. Let x_1, \dots, x_n be a sample from a $N(k; \xi, \Sigma)$ population with an unknown mean vector ξ and an

unknown or partially known covariance matrix Σ . It is desired to estimate $n(k; c, \xi, \Sigma)$ at a prescribed point $c \in R_k$, under different assumptions concerning Σ . We consider four cases: (1) Σ is known, (2) Σ is known up to a scalar multiple, (3) Σ has the form $\sigma^2[(1 - \rho)I + \rho e'e]$, (4) Σ is unknown. In each case we find the estimator of

$$\gamma(k, m; c, \xi, \Sigma) = |\Sigma|^{-m/2} \text{etr}[-\Sigma^{-1}(c - \xi)'(c - \xi)/2],$$

which is $(2\pi)^{k/2} n(k; c, \xi, \Sigma)$ when $m = 1$. A complete sufficient statistic is provided in each case by $t = \sum_i x_i/n$ and a suitable function of $S = \sum_i x_i' x_i - nt't$. The results for the four cases are stated first, and are then followed by the proofs. Note that the sufficient statistics t, S are independent, with $\mathcal{L}(t) = N(k; \xi, r\Sigma)$, $\mathcal{L}(S) = W(k, \nu; \Sigma)$, and r is a known positive proper fraction. Normally, t and S come from the same sample of size n , in which case $r = 1/n$ and $\nu = n - 1$; however, such statistics might also arise in other ways.

3.1. When $\Sigma = \Sigma_0$ is known, the UMVU estimator of $\gamma(k, m; c, \xi, \Sigma_0)$ is given by

$$|\Sigma_0|^{-m/2} (1 - r)^{-k/2} \exp\{-[(c - t)\Sigma_0^{-1}(c - t)'/2(1 - r)]\}.$$

3.2. When $\Sigma = \sigma^2 \Sigma_0$, with Σ_0 known, σ unknown, let $s = \text{tr } \Sigma_0^{-1} S$; the UMVU estimator of $\gamma(k, m; c, \xi, \sigma^2 \Sigma_0)$, $m < \nu$, is given by

$$|\Sigma_0|^{-m/2} \frac{2^{mk/2} \Gamma(\nu k/2)}{\Gamma[(\nu - m)k/2](1 - r)^{k/2}} s^{-(\nu k - 2)/2} \cdot \left(\Psi \left[s - \frac{(c - t)\Sigma_0^{-1}(c - t)'}{(1 - r)} \right] \right)^{[(\nu - m)k - 2]/2}.$$

3.3. When $\Sigma = \sigma^2[(1 - \rho)I + \rho e'e]$, let $u = eSe'/k = e[\sum_i x_i' x_i - nt't]e'/k$, $v = \sum_i x_i x_i' - nt't - u$; then the UMVU estimator of $\gamma(k, m; c, \xi, \Sigma)$, $m < \nu$, is given by

$$\frac{2^{mk/2} \Gamma(\nu/2) \Gamma[\nu(k - 1)/2] (1 - r)^{-k/2}}{\Gamma[(\nu - m)/2] \Gamma[(\nu - m)(k - 1)/2]} u^{-(\nu - 2)/2} v^{-[\nu(k - 1) - 2]/2} \cdot [\Psi(u - a)]^{(\nu - m - 2)/2} [\Psi(v - b)]^{[(\nu - m)(k - 1) - 2]/2},$$

where $a = [(c - t)e']^2/k(1 - r)$, $b = (c - t)(c - t)'/(1 - r) - a$.

3.4. When Σ is unknown, the UMVU estimator of $\gamma(k, m; c, \xi, \Sigma)$, $m < \nu - k + 1$, is given by

$$c(k, \nu - m)/\{c(k, \nu)(1 - r)^{k/2}\} |S|^{-(\nu - k - 1)/2} \cdot \{\Psi[S - (c - t)'(c - t)/(1 - r)]\}^{(\nu - k - m - 1)/2}.$$

PROOFS. The proof of 3.1 consists of a direct computation from the fact that $\mathcal{L}(t) = N(k; \xi, r\Sigma_0)$. Note that when $r = 1/n$, where n is an integer, the result can be obtained by considering a sample $\{x_1, \dots, x_n\}$ from a $N(k; \xi, \Sigma_0)$ popu-

lation. The conditional density of $x_1 \mid \Sigma x_i/n$ evaluated at c is then the UMVU estimator of $n(k; c, \xi, \Sigma_0)$.

In 3.2, we have unknown parameters (ξ, σ^2) and complete sufficient statistics (t, s) which are independent with $\mathcal{L}(t) = N(k; \xi, r\sigma^2\Sigma_0)$ and $\mathcal{L}(s) = W(1, \nu k; \sigma^2)$. We use Lemma 2, first estimating γ for fixed σ^2 . From 3.1, we obtain the estimator $|\Sigma_0|^{-m/2} (1-r)^{-k/2} \sigma^{-mk} \exp\{-a/2\sigma^2\}$, where $a = (c-t)\Sigma_0^{-1}(c-t)'/(1-r)$, assuming σ^2 known. An application of Theorem 1, with the association $(k, m, A, \Sigma, S, \nu) = (s, mk, a, \sigma^2, s, \nu k)$ yields the final estimator

$$|\Sigma_0|^{-m/2} (1-r)^{-k/2} [c(1, [\nu-m]k)/c(1, \nu k)] s^{-(\nu k-2)/2} [\Psi(s-a)]^{[(\nu-m)k-2]/2}.$$

In 3.3, we have unknown parameters (ξ, σ^2, ρ) and complete sufficient statistics (t, u, v) which are independent with $\mathcal{L}(t) = N(k; \xi, r\Sigma)$, $\mathcal{L}(u) = W(1, \nu; \alpha^{-1})$, $\mathcal{L}(v) = W(1, \nu(k-1); \beta^{-1})$. Again we use Lemma 2, first estimating

$$\gamma = \alpha^{m/2} \beta^{m(k-1)/2} \exp[-(c-\xi)\{\alpha e'e/k + \beta(kI - e'e)/k\}(c-\xi)'/2]$$

for fixed α, β . From 3.1, we obtain the estimator

$$(1-r)^{-k/2} [\alpha^{m/2} \exp(-a\alpha/2)] [\beta^{m(k-1)/2} \exp(-b\beta/2)].$$

The last step is to estimate the two factors in terms of the statistics u and v , using Theorem 1 with the associations

$$(k, m, A, \Sigma, S, \nu) = (1, m, a, \alpha^{-1}, u, \nu),$$

$$(k, m, A, \Sigma, S, \nu) = (1, m(k-1), b, \beta^{-1}, v, \nu(k-1)),$$

respectively, and obtain the final estimate

$$(1-r)^{-\frac{1}{2}k} \frac{c(1, \nu-m)}{c(1, \nu)} u^{-\frac{1}{2}(\nu-2)} [\Psi(u-a)]^{\frac{1}{2}(\nu-m-2)} \cdot \frac{c(1, [\nu-m][k-1])}{c(1, \nu[k-1])} v^{-\frac{1}{2}(\nu k - \nu - 2)} [\Psi(v-b)]^{\frac{1}{2}(\nu-m)(k-1)-1}.$$

Finally, 3.4 follows similarly by the use of Lemma 2 with Σ fixed and Theorem 1.

REMARK ON A GENERALIZATION. Suppose that t and S are independent statistics with $\mathcal{L}(S) = W(k, \nu; \Sigma)$ and $\mathcal{L}(t) = N(k; \xi, A'\Sigma A)$, where A is a known non-singular $k \times k$ matrix. This is a generalization of Case 3.2, where $A = r^{\frac{1}{2}}I$. The problem is to find a function of t and S which is an unbiased estimator of $\gamma(k, m; c, \xi, \Sigma)$. It is somewhat surprising that no such estimator exists unless A is a scalar multiple of the identity—as in Case 3.2.

PROOF. Suppose there exists a function $g(c; t, S)$ such that

$$(3.5) \quad \int g(c; t, S) n(k; t, \xi, A'\Sigma A) w(k, \nu; S, \Sigma) dt dS \equiv_{\xi, \Sigma} \gamma(k, m; c, \xi, \Sigma).$$

We assert that (3.5) implies $\Sigma - A'\Sigma A \geq 0$ for all $\Sigma > 0$, and then show that this implies $A = aI$.

Suppose there exists a $\Sigma_0 > 0$ such that $\Sigma_0 - A'\Sigma_0 A$ is not positive semi-definite; then there is a non-vanishing vector ℓ and a positive number b^2 such that

$\ell(\Sigma_0 - A'\Sigma_0A)\ell' = -b^2$. In (3.5), set $\Sigma = \Sigma_0$ and $\xi = \eta\ell$ with $\eta \in R_1$. Upon simplification, we obtain

$$(3.6) \quad \int_{t \in R_k, S > 0} g(c; t, S) \operatorname{etr} \{-\Sigma_0^{-1}[(A')^{-1}t' t A^{-1} + S]/2\} \\ \cdot \exp [tA^{-1}\Sigma_0^{-1}(A')^{-1}t'\eta] dt dS \equiv_{\eta} (\text{constant}) \exp (-b^2\eta^2).$$

On making a non-singular linear transformation from (t_1, \dots, t_k) to (y_1, \dots, y_k) with $y_1 = tA^{-1}\Sigma_0^{-1}(A')^{-1}\ell'$, and integrating over $S > 0$ and $(y_2, \dots, y_k) \in R_{k-1}$, (3.6) reduces to

$$(3.7) \quad \int_{-\infty}^{\infty} f(y) \exp (y\eta) dy \equiv_{\eta} (\text{constant}) \exp (-b^2\eta^2),$$

where f is some function on R_1 .

But the right-hand side of (3.7) is not a Laplace transform (e.g., see Widder (1941), Theorem 7a). Consequently, the assumption that there exists a $\Sigma_0 > 0$ such that $\Sigma_0 - A'\Sigma_0A$ is not positive definite leads to a contradiction. Hence, $\Sigma - A'\Sigma A \geq 0$ for all $\Sigma > 0$.

LEMMA 3. *If A is a non-singular $k \times k$ matrix such that $\Sigma - A'\Sigma A \geq 0$ for all $\Sigma > 0$, then $A = aI$ with $|a| \leq 1$.*

PROOF. By choosing Σ to be a diagonal matrix $D_\sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_k)$, we see that $\sigma_j \geq \sum_i \sigma_i a_{ij}^2$ for all $\sigma_j > 0$, which implies that $a_{ij} = 0$ for $i \neq j$, and $a_{ii}^2 \leq 1$. Thus A is a diagonal matrix $D_a = \operatorname{diag}(a_1, \dots, a_k)$. Now choose a general Σ and consider any 2×2 principal minor of $\Sigma - D_a \Sigma D_a$. Since this minor is non-negative, we have $\sigma_{ii}(1 - a_i^2)\sigma_{jj}(1 - a_j^2) \geq \sigma_{ij}^2(1 - a_i a_j)^2$ for all $\sigma_{ii} > 0$, $\sigma_{ii}\sigma_{jj} > \sigma_{ij}^2$. This implies $(1 - a_i^2)(1 - a_j^2) \geq (1 - a_i a_j)^2$ and hence $a_i = a_j$, $i, j = 1, \dots, k$.

4. Multivariate normal distribution with a linearly restricted mean. We next consider a situation which is, in some respects, intermediate between those of the previous two sections, namely one in which the k -dimensional mean vector is known to be restricted to a subspace of less than k dimensions. The problem can be formulated in some generality as follows. Let x_1, \dots, x_n be independent k -dimensional random vectors with $\mathcal{L}(x_i) = N(k; \theta A_i, \Sigma)$, $i = 1, \dots, n$, where θ is an unknown ℓ -dimensional vector whose range is not contained in an $(\ell - 1)$ -dimensional subspace, and the A_i are known $\ell \times k$ matrices such that (A_1, \dots, A_n) has rank ℓ . It is desired to estimate $\prod_1^r n(k; c_j, \theta B_j, \Sigma)$ in terms of $x = (x_1, \dots, x_n)$, where B_1, \dots, B_r are given $\ell \times k$ matrices. We again consider the three cases with different assumptions on Σ : (1) Σ known, (2) Σ known up to a scalar multiple, (3) the intraclass correlation model. In each case, we find the UMVU estimator of

$$\gamma(k, m, r; \mathbf{c}, \theta, \Sigma) = |\Sigma|^{-m/2} \exp [-\sum_{j=1}^r (c_j - \theta B_j) \Sigma^{-1} (c_j - \theta B_j)'/2],$$

from which the UMVU estimator of

$$\prod_1^r n(k; c_j, \theta B_j, \Sigma) = (2\pi)^{-kr/2} \gamma(k, r, r; \mathbf{c}, \theta, \Sigma)$$

is immediate.

4.1. $\Sigma = \Sigma_0$ known. Define $t = \sum_i x_i \Sigma_0^{-1} A_i'$ and $A = (A_1 \Sigma_0^{-\frac{1}{2}}, \dots, A_n \Sigma_0^{-\frac{1}{2}})$, so that $\mathcal{L}(t) = N(k; \theta A A', A A')$. The UMVU estimator of $\gamma(k, m, r; \mathbf{c}, \theta, \Sigma)$ is given by

$$|\Sigma_0|^{-m/2} |I - B'(A A')^{-1} B|^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} [c - t(A A')^{-1} B] [I - B'(A A')^{-1} B]^{-1} [c - t(A A')^{-1} B'] \right\},$$

if and only if $A A' > B B'$, where

$$B = (B_1 \Sigma_0^{-\frac{1}{2}}, \dots, B_r \Sigma_0^{-\frac{1}{2}}), \quad c = (c_1 \Sigma_0^{-\frac{1}{2}}, \dots, c_r \Sigma_0^{-\frac{1}{2}}).$$

Note that the alternative form of the UMVU estimator

$$|\Sigma_0|^{-m/2} |A A'|^{-\frac{1}{2}} |A A' - B B'|^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2} [c c' - t(A A')^{-1} t' + (t - c B')(A A' - B B')^{-1} (t - c B')'] \right\},$$

may be more convenient for computational purposes.

4.2. $\Sigma = \sigma^2 \Sigma_0$, Σ_0 known, σ^2 unknown. Using the notation of 4.1, define $s = \sum_i x_i \Sigma_0^{-1} x_i' - t(A A')^{-1} t'$, so that $\mathcal{L}(s) = W(1, \nu; \sigma^2)$. The UMVU estimator of $\gamma(k, m, r; \mathbf{c}, \theta, \sigma^2 \Sigma_0)$ is given by

$$2^{mk/2} \Gamma(\frac{1}{2}\nu) |\Sigma_0|^{-m/2} [\Gamma(\frac{1}{2}(\nu - mk))]^{-1} s^{-(\nu-2)/2} |I - B'(A A')^{-1} B|^{-\frac{1}{2}} \cdot \{\Psi[s - (c - t(A A')^{-1} B)(I - B'(A A')^{-1} B)^{-1} (c - t(A A')^{-1} B')]\}^{(\nu-mk-2)/2}.$$

Normally, when t and s arise from a sample of size n , $\nu = nk - \ell$. When $k = 1$, $\Sigma_0 = 1$, the model is that of univariate linear regression.

4.3. *Intraclass correlation model*, $\Sigma = \sigma^2[(1 - \rho)I + \rho e' e]$. We give here the explicit solution to two relatively common cases.

Case (i). The matrices A_i and B_j have the special form $A_i = a_i' e$, $B_j = b_j' e$, $i = 1, \dots, n; j = 1, \dots, r$, where the a 's and b 's are ℓ -dimensional vectors with rank $(a_1', \dots, a_n') = \ell$. Define $A = (A_1, \dots, A_n)$, $A_0 = A A' = k \sum_i a_i' a_i$, and $t = x A' A_0^{-\frac{1}{2}}$, $u_1 = k^{-1} \sum_i (x_i e')^2 - t t'$, $u_2 = x(I - e' e/k) x'$, so that $\mathcal{L}(t) = N(\ell; \theta A_0^{\frac{1}{2}}, \alpha^{-1} I)$, $\mathcal{L}(u_1) = W(1, \eta_1; \alpha^{-1})$, $\mathcal{L}(u_2) = W(1, \eta_2; \beta^{-1})$. The UMVU estimator of $\gamma(k, m, r; \mathbf{c}, \theta, \Sigma)$ is given by $\tilde{\gamma} = \tilde{\gamma}_1 \tilde{\gamma}_2 |I - B' A_0^{-1} B|^{-\frac{1}{2}}$, where $B = (B_1, \dots, B_r)$, and $\tilde{\gamma}_j = k(\delta_j, \eta_j) u_j^{-(\eta_j-2)/2} [\Psi(u_j - q_j)]^{(\eta_j-\delta_j-2)/2}$, $j = 1, 2$, and where $k(\delta, \eta) = 2^{\delta/2} \Gamma(\eta/2) / \Gamma[(\eta - \delta)/2]$, $\delta_1 = m$, $\delta_2 = m(k - 1)$, $\eta_1 = n - \ell$, $\eta_2 = n(k - 1)$, $q_1 = (\bar{c} - t A_0^{-\frac{1}{2}} B)(I - B' A_0^{-1} B)^{-1} (\bar{c} - t A_0^{-\frac{1}{2}} B)'$, $q_2 = c(I - e' e/k) c'$, $\bar{c} = (c_1 e' e, \dots, c_r e' e)/k$.

Case (ii). The matrices A_i and B_j have the property that $A_i e' = B_j e' = 0$, $i = 1, \dots, n; j = 1, \dots, r$, where $A = (A_1, \dots, A_n)$ is of rank ℓ . Define $A_0 = A A'$, $M = (I - e' e/k)$, $t = \sum_i x_i M A_i' A_0^{-\frac{1}{2}}$, $u_1 = \sum_i x_i M x_i' - t t'$, $u_2 = \sum_i (x_i e')^2 / k$, so that $\mathcal{L}(t) = N(\ell; \theta A_0^{\frac{1}{2}}, \beta^{-1} I)$, $\mathcal{L}(u_1) = W(1, \eta_1; \beta^{-1})$, $\mathcal{L}(u_2) = W(1, \eta_2; \alpha^{-1})$. The UMVU estimator of $\gamma(k, m, r; \mathbf{c}, \theta, \Sigma)$ is given by $\tilde{\gamma}$ as in (i) with the association $\delta_1 = m(k - 1)$, $\delta_2 = m$, $\eta_1 = n(k - 1) - \ell$, $\eta_2 = n$,

$q_1 = (\bar{c} - tA^{-\frac{1}{2}}B)(I - B'A_0^{-\frac{1}{2}}B)^{-1}(\bar{c} - tA_0^{-\frac{1}{2}}B)'$, $q_2 = \sum_j c_j e' e c_j' / k$, where $\bar{c} = (c_1 M, \dots, c_r M)$.

PROOFS. We start with a result which is a version of 4.1 in canonical form.

4.4. Let t be a random vector with $\mathcal{L}(t) = N(\ell; \theta, I)$, where θ is unknown, and let H be a $k \times k$ non-negative definite matrix. Then there exists a function of t , say $\tilde{\gamma}_0$, which is an unbiased estimator of

$$\gamma_0 = \exp[-\sum_{j=1}^r (c_j - \theta B_j)H(c_j - \theta B_j)'/2],$$

if and only if $I - \sum_j B_j H B_j' > 0$, in which case

$$\tilde{\gamma}_0 = |I - B'B|^{-\frac{1}{2}} \exp[-(c - tB)(I - B'B)^{-1}(c - tB)'/2],$$

where $B = (B_1 H^{\frac{1}{2}}, \dots, B_r H^{\frac{1}{2}})$, $c = (c_1 H^{\frac{1}{2}}, \dots, c_r H^{\frac{1}{2}})$. The unbiasedness of $\tilde{\gamma}_0$ may be verified in a direct manner by writing out the identity, $Eg \equiv \gamma_0$, to be satisfied by an unbiased estimator, g , of γ_0 ; viz., the identity

$$(2\pi)^{-\frac{1}{2}} \int g(t) \exp[-(t - \theta)(t - \theta)'/2] \equiv \exp[-\sum_j (c_j - \theta B_j)H(c_j - \theta B_j)'/2],$$

which can be rewritten in the form

$$(2\pi)^{-\frac{1}{2}} \int \{g(t) \exp[-(t' - c'c)/2] \exp[(t - cB')\theta'] dt \equiv \exp[\theta(I - BB')\theta'/2].$$

The left-hand side is a Laplace transform, whereas the right-hand side is one if and only if $I > BB'$, in which case 4.4 is obtained by inversion.

In 4.1, unbiasedness follows directly from 4.4 with the association $[t, H, B_j] = [t(AA')^{-\frac{1}{2}}, \Sigma_0^{-1}, (AA')^{-\frac{1}{2}}B_j]$. The result follows from the completeness of t .

In 4.2, we have independent statistics s and t , with $\mathcal{L}(s) = W(1, nk - \ell; \sigma^2)$ and $\mathcal{L}(t) = N(\ell; \theta AA', \sigma^2 AA')$. Unbiasedness follows from Lemma 2, Theorem 1 and 4.4 using the association $[t, H, B_j] = [t(\sigma^2 AA')^{-\frac{1}{2}}, (\sigma^2 \Sigma_0)^{-1}, (\sigma^2 AA')^{-\frac{1}{2}}B_j]$. The result then follows from the completeness of (t, s) .

In 4.3, we first investigate the existence of a complete sufficient statistic and then reduce the problem to canonical form. The explicit solution is then obtained for the two special cases.

Let Γ be a $k \times k$ orthonormal matrix with first column $e'k^{-\frac{1}{2}}$. Define $x_i \Gamma = (y_i, z_i)$, $A_i \Gamma = (g_i', G_i')$, where $y_i = x_i e'k^{-\frac{1}{2}}$, $g_i' = A_i e'k^{-\frac{1}{2}}$. Then $y_1, \dots, y_n, z_1, \dots, z_n$ are mutually independent with $\mathcal{L}(y_i) = N(1; \theta g_i', \alpha^{-1})$ and $\mathcal{L}(z_i) = N(k - 1; \theta G_i', \beta^{-1}I)$. Hence, $(\sum_i y_i^2, \sum_i z_i z_i', \sum_i y_i g_i, \sum_i z_i G_i')$ is a minimal sufficient statistic for (α, β, θ) (see Lehmann and Scheffé (1950)). In terms of the original variables, a sufficient statistic is $(\sum_i (x_i e')^2, \sum_i x_i M x_i', \sum_i x_i A_i')$. Completeness may then be shown by using a necessary and sufficient condition obtained by Wijsman (1958, Theorem 2).

The joint density of the y_i and z_i has the form

$$f(\alpha, \beta, \theta) \exp[-\sum_i (y_i^2 \alpha + z_i z_i' \beta - 2y_i g_i \alpha \theta' - 2z_i G_i \beta \theta')/2].$$

The minimal sufficient statistic is complete if and only if, among the 2ℓ functions occurring as coefficients of $\alpha \theta'$ and $\beta \theta'$, exactly ℓ are linearly independent. Since $\text{rank}(g_1', \dots, g_n', G_1', \dots, G_n') = \ell$, this is the case if and only if $\text{rank}(g_1', \dots, g_n') + \text{rank}(G_1', \dots, G_n') = \ell$. In Case (i), $(g_1', \dots, g_n') = k^{\frac{1}{2}}(a_1', \dots,$

$a_n')$, which is of rank ℓ , and $(G_1', \dots, G_n') = 0$. In Case (ii), $(g_1', \dots, g_n') = 0$ and, $(G_1', \dots, G_n') = (A_1\Gamma_2, \dots, A_n\Gamma_2)$, where $\Gamma = (e'k^{-1}\Gamma_2)$. By hypothesis, $A = (A_1, \dots, A_n)$ is of rank ℓ , so that (G_1', \dots, G_n') is of rank ℓ . Hence, in each case, the minimal sufficient statistic is complete.

The unbiasedness follows by a direct computation using Lemma 2, Theorem 1 and 4.1.

5. Entropy and Kullback-Leibler information numbers for the normal distribution. The entropy and Kullback-Leibler information numbers for the normal distribution are defined by

$$(5.1) \quad \int n(k; x, \xi, \Sigma) [\log n(k; x, \xi, \Sigma)] dx = -\frac{1}{2}[k + k \log(2\pi) - \log |\Sigma|],$$

$$(5.2) \quad \int n(k; x, \xi_1, \Sigma_1) [\log n(k; x, \xi_1, \Sigma_1) - \log n(k; x, \xi_2, \Sigma_2)] dx \\ = \frac{1}{2}[\log |\Sigma_2| - \log |\Sigma_1| + \text{tr } \Sigma_1 \Sigma_2^{-1} - k + (\xi_1 - \xi_2) \Sigma_2^{-1} (\xi_1 - \xi_2)'].$$

Since both quantities play a role in statistical and communication theory, it is of interest to obtain the UMVU estimators. In this connection, the maximum likelihood estimator of the entropy of a discrete distribution has been discussed by Miller and Madow (1954).

Because they involve the parameters the integrals in (5.1) and (5.2) do not have the same form as that in Lemma 1, so that UMVU estimators of $\log |\Sigma|$, $\text{tr } \Sigma_1 \Sigma_2^{-1}$ and $(\xi_1 - \xi_2) \Sigma_2^{-1} (\xi_1 - \xi_2)'$ must be found directly. The result then follows from the additivity of unbiased estimators.

LEMMA 4. If $\mathcal{L}(S) = W(k, \nu; \Sigma)$, then

$$(5.3) \quad E \log |S| = \log |\Sigma| + k \log 2 + \sum_1^k \frac{d \log \Gamma(\alpha)}{d\alpha} \Big|_{\alpha=(\nu-i+1)/2}.$$

PROOF. From the fact that $\mathcal{L}(|\Sigma^{-1}S|) = \mathcal{L}(\prod_1^k z_i)$, where $\mathcal{L}(z_i) = \chi_{n-i+1}^2$, we have $E \log |S| = \log |\Sigma| + \sum_1^k E \log z_i$. To obtain $E \log z_i$, we note that $\int_0^\infty z^{\alpha-1} e^{-\frac{1}{2}z} dz = 2^\alpha \Gamma(\alpha)$ may be differentiated under the integral sign with respect to α , to yield $E \log z = \log 2 + d \log \Gamma(\alpha)/d\alpha$.

To estimate $\text{tr } \Sigma_1 \Sigma_2^{-1}$, suppose $\mathcal{L}(S_j) = W(k, \nu_j; \Sigma_j)$, $j = 1, 2$, with S_1 and S_2 independently distributed. Since $ES_1/\nu_1 = \Sigma_1$, $ES_2^{-1}/(\nu_2 - k - 1) = \Sigma_2^{-1}$, and the trace is a linear function, we have

$$(5.4) \quad E \text{tr } (S_1 S_2^{-1})/\nu_1(\nu_2 - k - 1) = \text{tr } \Sigma_1 \Sigma_2^{-1}.$$

When ξ_1 and ξ_2 are known, we have

$$(5.5) \quad E(\xi_1 - \xi) S_2^{-1} (\xi_1 - \xi_2)' / (\nu_2 - k - 1) = (\xi_1 - \xi_2) \Sigma_2^{-1} (\xi_1 - \xi_2)'.$$

When ξ_1 and ξ_2 are unknown, and we have available independent statistics x, y, S_1, S_2 , with $\mathcal{L}(x) = N(k; \xi_1, \Sigma_1/N)$ and $\mathcal{L}(y) = N(k; \xi_2, \Sigma_2/M)$, then

$$(5.6) \quad E \{ \text{tr } S_2^{-1} [(x - y)'(x - y) - S_1/(\nu_1 N)] / (\nu_2 - k - 1) \} - k/M \\ = \text{tr } \Sigma_2^{-1} \{ [\Sigma_1/N + \Sigma_2/M + (\xi_1 - \xi_2)'(\xi_1 - \xi_2)] - \Sigma_1/N \} - k/M \\ = (\xi_1 - \xi_2) \Sigma_2^{-1} (\xi_1 - \xi_2)'.$$

Combining (5.4) with either (5.5) or (5.6) yields the UMVU estimator of the entropy or the Kullback-Leibler information numbers.

Acknowledgment. The authors are indebted to M. L. Eaton, W. J. Hall, E. L. Lehmann and M. Perlman for comments and suggestions.

REFERENCES

- [1] BARTON, D. E. (1961). Unbiased estimation of a set of probabilities. *Biometrika* **48** 227–229.
- [2] BASU, A. P. (1964). Estimates of reliability for some distributions useful in life-testing. *Technometrics* **6** 215–219.
- [3] BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15** 377–380.
- [4] GHURYE, S. G. and OLKIN, INGRAM (1961). Unbiased estimation of probability densities (abstract). *Ann. Math. Statist.* **32** 1346 (Correction: **33** 833).
- [5] GLASSER, GERALD J. (1962). Minimum variance unbiased estimator for Poisson probabilities. *Technometrics* **4** 409–418.
- [6] HEALY, W. C., JR. (1956). Unbiased estimation of the normal distribution function. (abstract) *Ann. Math. Statist.* **27** 873.
- [7] KOLMOGOROV, A. N. (1950). Unbiased estimators. *Izv. Akad. Nauk SSSR. Ser. Mat.* **14** 303–326 (English translation in *Amer. Math. Soc. Transl.* **98**).
- [8] LEADBETTER, M. R. and WATSON, G. S. (1963). On the estimation of the probability density, I. *Ann. Math. Statist.* **34** 480–491.
- [9] LEHMANN, E. L. AND SCHEFFÉ, HENRY (1950). Completeness, similar regions and unbiased estimation—Part 1. *Sankhyā* **10** 305–340.
- [10] LIEBERMAN, GERALD J. and RESNIKOFF, GEORGE J. (1955). Sampling plans for inspection by variables. *J. Amer. Statist. Assoc.* **50** 457–516.
- [11] MILLER, GEORGE A. and MADOW, WILLIAM D. (1954). On the maximum likelihood estimate of the Shannon-Wiener measure of information. Technical Report No. 54-75, Air Force Cambridge Research Center.
- [12] NEYMAN, J. and SCOTT, E. L. (1960). Correction for bias introduced by a transformation of variables. *Ann. Math. Statist.* **31** 643–655.
- [13] PARZEN, E. (1962). On the estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.
- [14] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.
- [15] RUTEMILLER, H. C. (1966). Point estimation of reliability of a system comprised of k elements from the same exponential distribution. *J. Amer. Statist. Assoc.* **61** 1029–1032.
- [16] SCHMETTERER, L. (1961). On a problem of J. Neyman and E. Scott. *Ann. Math. Statist.* **31** 656–661.
- [17] TATE, R. F. (1959). Unbiased estimation: functions of location and scale parameters. *Ann. Math. Statist.* **30** 341–366.
- [18] WASHIO, YASUTOSHI, MORIMOTO, HARUKI AND IKEDA, NOBUYAKI (1956). Unbiased estimation based on sufficient statistics. *Bull. Math. Statist.* **6** 69–93.
- [19] WEISS, L. and WOLFOWITZ, J. (1967). Estimation of a density function at a point. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7** 327–335.
- [20] WIDDER, DAVID VERNON (1941). *The Laplace Transform*. Princeton Univ. Press.
- [21] WIJSMAN, ROBERT A. (1958). Incomplete sufficient statistics and similar tests. *Ann. Math. Statist.* **29** 1028–1045.
- [22] ZACKS, S. and EVEN, M. (1966). The efficiencies in small samples of the maximum likelihood and best unbiased estimators of reliability functions. *J. Amer. Statist. Assoc.* **61** 1033–1051.
- [23] ZACKS, S. and EVEN, M. (1966). Minimum variance unbiased and maximum likelihood estimators of reliability functions for systems in series and in parallel. *J. Amer. Statist. Assoc.* **61** 1052–1062.