The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 393–398

BIORTHOGONAL AND DUAL CONFIGURATIONS AND THE RECIPROCAL NORMAL DISTRIBUTION[1]

BY ROBERT H. BERK

University of Michigan

0. Summary. In this note we discuss the notions of biorthogonal and dual configurations and their relevance in certain statistical applications. The first application is to the distribution of a random matrix related to a multi-variate-normal sample matrix. As with the latter, the distribution is preserved by (certain) linear transformations. One consequence of this is the familiar result that if $\mathbf{Q}$ is a non-singular Wishart matrix, then for any non-zero vector α, $1/\alpha'\mathbf{Q}^{-1}\alpha$ is a multiple of a chi-square variable. Application is also made to the Gauss-Markov theorem and to certain estimates of mixing proportions due to Robbins.

1. Biorthogonal and dual configurations. Let $\mathfrak{X}$ be a vector space with an inner-product, denoted by $\langle \cdot , \cdot \rangle$. The configurations (= ordered subsets) $(x_1, \cdots, x_p)$ and $(x_1^*, \cdots, x_p^*)$ are said to be biorthogonal if $\langle x_i, x_j^* \rangle = \delta_{ij}$, the Kroneker delta. Clearly this relation is symmetric. Necessarily, the elements of $\{x_1, \cdots, x_p\}$ (respectively, $\{x_1^*, \cdots, x_p^*\}$) are linearly independent. For if (e.g.) $x_1 \varepsilon \mathcal{V}\{x_2, \cdots, x_p\}$, the subspace spanned by $\{x_2, \cdots, x_p\}$, then $\langle x_i, x_1^* \rangle = 0$, $i = 2, \cdots, p$, implies that $\langle x_1, x_1^* \rangle = 0$. In general, there are many configurations biorthogonal with a given configuration $(x_1, \cdots, x_p)$. One such is distinguished: There is a unique $(y_1, \cdots, y_p) \subset \mathcal{V}\{x_1, \cdots, x_p\}$ which is biorthogonal with $(x_1, \cdots, x_p)$. It is called the configuration dual to $(x_1, \cdots, x_p)$ and is constructed as follows: Let x_i, be the projection of x_i into $\mathcal{V}^{\perp}\{x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_p\}$. ($\mathfrak{U}^{\perp}$ is the orthogonal complement of $\mathfrak{U} \subset \mathfrak{X}$ in $\mathfrak{X}$.) $\langle x_i, x_{i\cdot} \rangle = \langle x_{i\cdot}, x_{i\cdot} \rangle \neq 0$ by linear independence; of course $\langle x_i, x_{j\cdot} \rangle = 0$ if $i \neq j$. Then $y_i = x_{i\cdot}/\langle x_{i\cdot}, x_{i\cdot} \rangle$ gives the configuration dual to $(x_1, \cdots, x_p)$. It readily follows that any configuration $(x_1^*, \cdots, x_p^*)$ biorthogonal with $(x_1, \cdots, x_p)$ has the representation $x_i^* = y_i + \delta_i$, where $\delta_i \varepsilon \mathcal{V}^{\perp}\{x_1, \cdots, x_p\}$.

Since $y_i \varepsilon \mathcal{V}\{x_1, \cdots, x_p\}$, we may write $(y_1, \cdots, y_p) = (x_1, \cdots, x_p)A$, where A is a non-singular $p \times p$ matrix. Letting Q denote the non-singular configuration matrix of $(x_1, \cdots, x_p): q_{ij} = \langle x_i, x_j \rangle$, the duality relations require that $QA = I$. Hence $A = Q^{-1}$, $(y_1, \cdots, y_p) = (x_1, \cdots, x_p)Q^{-1}$ and the dual configuration has configuration matrix Q^{-1}. Thus the dual configuration relation is also symmetric. If $x_1, \cdots, x_p$ are elements of R^p and if $X = ((x_1, \cdots, x_p))$ is the $p \times p$ matrix they generate by their representation as p-tuples, then $Y' = ((y_1, \cdots, y_p))' = X^{-1}$. If $x_1, \cdots, x_p$ are elements of R^n, $n > p$, Y is a pseudo (= one-sided) inverse for X, as is the matrix generated by every other configuration biorthogonal with $(x_1, \cdots, x_p)$.

Received 7 March 1968.

[1] Supported in part by grant GP-6008 from the National Science Foundation.

2. Reciprocal normal distribution. Let $\mathbf{z}, \mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(n)}$, $(n \geqq p)$ independent observations from $N(0, \Sigma)$, the p-dimensional normal distribution with zero mean and covariance matrix Σ. Throughout, Σ will be $p \times p$ non-singular. Let $\mathbf{X} = ((\mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(n)}))$ and let $\mathbf{x}_1, \cdots, \mathbf{x}_p$ denote the rows of $\mathbf{X}$, all random elements of R^n; $\mathbf{X} = ((\mathbf{x}_1, \cdots, \mathbf{x}_p))'$. $\mathbf{X}$ (more properly, $(\mathbf{x}_1, \cdots, \mathbf{x}_p)$) is called the sample configuration. Let $\mathbf{Q} = \mathbf{X}\mathbf{X}'$ be the $p \times p$ sample configuration matrix (of $(\mathbf{x}_1, \cdots, \mathbf{x}_p)$). $\mathbf{Q} \sim W(n, \Sigma)$, the p-dimensional central Wishart distribution based on Σ and having n degrees of freedom. Because Σ is non-singular, wp 1 $\mathbf{x}_1, \cdots, \mathbf{x}_p$ are linearly independent and $\mathbf{Q}$ is non-singular. Let $\mathbf{Y} = \mathbf{Q}^{-1}\mathbf{X}$. Then $(\mathbf{y}_1, \cdots, \mathbf{y}_p)$, the rows of $\mathbf{Y}$, is the configuration dual to $(\mathbf{x}_1, \cdots, \mathbf{x}_p)$. Below we investigate the distribution of the random configuration $\mathbf{Y}$ and show that like $\mathbf{X}$, it has closure properties under (certain) linear transformations. We also cite certain facts about spherically distributed configurations and about the multivariate normal distribution. A fuller discussion of these may be found in [1], [2] and [3].

We begin by noting that the distribution of $\mathbf{Y}$ is spherical (invariant under orthogonal rotations). To see this, we recall that if G is an $n \times n$ orthogonal matrix, $\mathbf{x}_i$ and $G\mathbf{x}_i$ have the same (spherical normal) distribution in R^n. In fact, $\mathbf{X}$ and $\mathbf{X}G$ have the same distribution. Since $\mathbf{X}$ and $\mathbf{X}G$ have the same configuration matrix, $\mathbf{Q}$, it follows that $\mathbf{Y}G = \mathbf{Q}^{-1}\mathbf{X}G$ has the same distribution as $\mathbf{Y}$. If $\mathbf{v} \varepsilon R^n$ is spherically distributed, then $\langle \mathbf{v}, \mathbf{v} \rangle = \mathbf{v}'\mathbf{v}$ and $\mathbf{v}/(\mathbf{v}'\mathbf{v})^{\frac{1}{2}}$ are independent; the latter being uniform over the unit sphere in R^n. Thus the distribution of a spherically distributed vector is characterized by the distribution of $\mathbf{v}'\mathbf{v}$. For $\mathbf{x}_i$, $\mathbf{x}_i'\mathbf{x}_i \sim \sigma_{ii}\chi_n^2$, where $((\sigma_{ij})) = \Sigma$. Similarly, the distribution of a spherically distributed configuration such as $\mathbf{X}$ is characterized by the distribution of its configuration matrix, in this case, $W(n, \Sigma)$. As $\mathbf{Y}$ has configuration matrix $\mathbf{Q}^{-1}$, its distribution is characterized by the fact that $(\mathbf{Y}\mathbf{Y}^1)^{-1} \sim W(n, \Sigma)$.

A random matrix (or configuration) with this spherical distribution will be said to have the reciprocal normal distribution. This is motivated by the fact that for $n = p = 1$, $\mathbf{Y} = ((y_{11})) = ((1/\mathbf{x}_{11}))$ where $\mathbf{X} = ((\mathbf{x}_{11}))$. Propositions 1 and 2 below suggest that the natural parameters of this distribution are $m = n - (p - 1)$ and Σ^{-1}. Accordingly, we write $\mathbf{Y} \sim RN(m, \Sigma^{-1})$ to mean $\mathbf{Y}$ is spherically distributed and $(\mathbf{Y}\mathbf{Y}')^{-1} \sim W(m + p - 1, \Sigma)$. Equivalently, $\mathbf{X}$, the configuration dual to $\mathbf{Y}$, has the distribution of a sample of $m + p - 1$ from $N(0, \Sigma)$. I.e., $\mathbf{X}$ is spherically distributed and $\mathbf{X}\mathbf{X}' \sim W(m + p - 1, \Sigma)$.

We recall some facts about projections of normal vectors. If $\mathbf{x}_1 \varepsilon R^n$ is projected into a fixed k-dimensional subspace $\mathcal{V}$, the resulting vector, $\mathbf{x}_{1*}$, is spherically normally distributed in $\mathcal{V}$ with k degrees of freedom. I.e., $\mathbf{x}_{1*}$ is spherically distributed in $\mathcal{V}$ and $\mathbf{x}_{1*}'x_{1*} \sim \sigma_{11}\chi_k^2$. This result remains true if $\mathcal{V}$ is a random k-dimensional subspace, as long as it is independent of $\mathbf{x}_1$. Similarly if we project $(\mathbf{x}_1, \cdots, \mathbf{x}_p)$ into $\mathcal{V}$, the resulting configuration, $(\mathbf{x}_{1*}, \cdots, \mathbf{x}_{p*})$ is multivariate normal in $\mathcal{V}$ with k degrees of freedom: it is spherically distributed in $\mathcal{V}$ and $\mathbf{X}_*\mathbf{X}_*' \sim W(k, \Sigma)$, where $\mathbf{X}_* = ((\mathbf{x}_{1*}, \cdots, \mathbf{x}_{p*}))'$.

Consider now a partition of the multivariate-normal vector $\mathbf{z}$ into $\mathbf{z}_1$, the first

s and $\mathbf{z}_2$, the remaining $p - s$ coordinates. We obtain a corresponding partition of Σ: Σ_{ii} is the covariance matrix of $\mathbf{z}_i$ and Σ_{12}, the matrix of covariances between the elements of $\mathbf{z}_1$ and $\mathbf{z}_2$. $\Sigma_{11\cdot2}$ denotes the conditional covariance matrix of $\mathbf{z}_1$ given $\mathbf{z}_2$. $\Sigma_{11\cdot2}$ is also the marginal covariance matrix of $\mathbf{z}_{1\cdot2} = \mathbf{z}_1 - B\mathbf{z}_2$, where $B\mathbf{z}_2 = E(\mathbf{z}_1 \mid \mathbf{z}_2)$. (Note that $\mathbf{z}_2$ and $\mathbf{z}_{1\cdot2}$ are independent.) We obtain a corresponding partition of $\mathbf{X}$ (respectively, $\mathbf{Y}$): $\mathbf{X}_1$ (respectively, $\mathbf{Y}_1$) denotes the first s and $\mathbf{X}_2$ (respectively $\mathbf{Y}_2$), the remaining $p - s$ rows of $\mathbf{X}$ (respectively, $\mathbf{Y}$). That is, $\mathbf{X}_1 = ((\mathbf{x}_1, \cdots, \mathbf{x}_s))'$. Ax $\mathbf{X}$ is a sample of n from $N(0, \Sigma)$, $\mathbf{X}_1 - B\mathbf{X}_2$ is a sample of n from $N(0, \Sigma_{11\cdot2})$ and is independent of $\mathbf{X}_2$. If we project the rows of $\mathbf{X}_1 - B\mathbf{X}_2$ into $\mathcal{V}^\perp(\mathbf{X}_2) = \mathcal{V}^\perp\{\mathbf{x}_{s+1}, \cdots, \mathbf{x}_p\}$, we obtain a configuration that is multivariate normal in $\mathcal{V}^\perp(\mathbf{X}_2)$ with covariance matrix $\Sigma_{11\cdot2}$ and having degrees of freedom $=$ dim $\mathcal{V}^\perp(\mathbf{X}_2) = n - (p - s)$. But as the projection of $\mathbf{X}_2$ into $\mathcal{V}^\perp(\mathbf{X}_2)$ is zero, the projection of $\mathbf{X}_1 - B\mathbf{X}_2$ into $\mathcal{V}^\perp(\mathbf{X}_2)$ is just $\mathbf{X}_{1\cdot2}$, the projection of $\mathbf{X}_1$ into $\mathcal{V}^\perp(\mathbf{X}_2)$. Hence $\mathbf{Q}_{11\cdot2} = \mathbf{X}_{1\cdot2}\mathbf{X}_{1\cdot2}' \sim W(n - (p - s), \Sigma_{11\cdot2})$ and is independent of $\mathbf{Q}_{22} = \mathbf{X}_2\mathbf{X}_2' \sim W(n, \Sigma_{22})$.

We discuss now the closure properties of the distribution of $\mathbf{Y}$, considering first $\mathbf{y}_1$ by way of introduction. Since $\mathbf{Y}$ is dual to $\mathbf{X}$, $\mathbf{y}_1 = \mathbf{x}_{1\cdot}/\mathbf{x}_{1\cdot}'\mathbf{x}_{1\cdot}$, where $\mathbf{x}_{1\cdot}$ is the projection of $\mathbf{x}_1$ into $\mathcal{V}^\perp(\mathbf{x}_2, \cdots, \mathbf{x}_p)$. Taking $s = 1$ in the preceding shows that $\mathbf{x}_{1\cdot}$ is spherically normally distributed in $\mathcal{V}^\perp(\mathbf{x}_2, \cdots, \mathbf{x}_p)$ with scale factor $\sigma_{11\cdot2}$ (the only element of $\Sigma_{11\cdot2}$); $\mathbf{x}_{1\cdot}'\mathbf{x}_{1\cdot}/\sigma_{11\cdot2} \sim \chi^2_{n-(p-1)}$. We further note that in general, $(\Sigma_{11\cdot2})^{-1} = (\Sigma^{-1})_{11}$ (this is well known but we present a probabilistic derivation using dual configurations in Section 3). In particular, writing $\Sigma^{-1} = ((\sigma^{ij}))$, $\sigma^{11} = 1/\sigma_{11\cdot2}$. Thus $\sigma^{11}/\mathbf{y}_1'\mathbf{y}_1 \sim \chi^2_{n-(p-1)}$; i.e., $\mathbf{y}_1 \sim RN(n - (p - 1), \sigma^{11})$. More generally, we have

1. Proposition. *If* $\mathbf{Y} \sim RN(m, \Sigma^{-1})$, *then*

(i) $\mathbf{Y}_1 \sim RN(m, (\Sigma^{-1})_{11})$.

(ii) $\mathbf{Y}_{2\cdot1} \sim RN(m + s, (\Sigma^{-1})_{22\cdot1})$ *and is independent of*

$$\mathbf{Y}_1\mathbf{Y}'_1 = (\mathbf{Q}^{-1})_{11} = (\mathbf{Q}_{11\cdot2})^{-1}.$$

Proof. Let $\mathbf{X}$ be the configuration dual to $\mathbf{Y}$. Note that the configuration dual to $\mathbf{X}_{1\cdot2}$ is just $\mathbf{Y}_1$. Dually, the configuration dual to $\mathbf{X}_2$ is $\mathbf{Y}_{2\cdot1}$. By the preceding discussion,

$$\begin{aligned}\mathbf{Y}_1 &= (\mathbf{Q}_{11\cdot2})^{-1}\mathbf{X}_{1\cdot2} \sim RN(n - (p - s) - (s - 1), (\Sigma_{11\cdot2})^{-1}) \\ &= RN(m, (\Sigma^{-1})_{11}).\end{aligned}$$

Moreover, $\mathbf{Q}_{11\cdot2} = \mathbf{X}_{1\cdot2}\mathbf{X}_{1\cdot2}'$ is independent of $\mathbf{X}_2$ and therefore of

$$\begin{aligned}\mathbf{Y}_{2\cdot1} &= (\mathbf{Q}_{22})^{-1}\mathbf{X}_2 \sim RN(n - (p - s - 1), (\Sigma_{22})^{-1}) \\ &= RN(m + s, (\Sigma^{-1})_{22\cdot1}).\end{aligned}$$ □

2. Proposition. *If* $\mathbf{Y} \sim RN(m, \Sigma^{-1})$ *and* C *is* $p \times p$ *non-singular,* $C\mathbf{Y} \sim RN(m, C\Sigma^{-1}C')$.

PROOF. Let $D = C^{-1}$. $D'\mathbf{X}$ is a sample of n from $N(0, D'\Sigma D)$. Hence $C\mathbf{Y} = C\mathbf{Q}^{-1}\mathbf{X} = (D'\mathbf{XX}'D)^{-1}D'\mathbf{X} \sim RN(m, (D'\Sigma D)^{-1}) = RN(m, C\Sigma^{-1}C')$. □

3. COROLLARY. *Let A denote an $s \times p$ matrix of rank $s \leqq p$. Then if $\mathbf{Y} \sim RN(m, \Sigma^{-1})$,*

$$A\mathbf{Y} \sim RN(m, A\Sigma^{-1}A').$$

PROOF. Let C be a $p \times p$ non-singular matrix having A as its first s rows. The corollary follows from Propositions 1 and 2. □

In particular, if α is a non-zero $p \times 1$ vector, $\alpha'\mathbf{Y} \sim RN(m, \alpha'\Sigma^{-1}\alpha)$. Hence $\alpha'\Sigma^{-1}\alpha/\alpha'\mathbf{YY}'\alpha = \alpha'\Sigma^{-1}\alpha/\alpha'\mathbf{Q}^{-1}\alpha \sim \chi^2_{n-(p-1)}$, where $\mathbf{Q} \sim W(n, \Sigma)$. This is a well-known property of the Wishart distribution [5]. A more general consequence of the corollary [1] is that $(A\mathbf{Q}^{-1}A')^{-1} \sim W(n - (p - s), (A\Sigma^{-1}A')^{-1})$. Moreover, if B denotes the remaining $p - s$ rows of C, then again by Propositions 1 and 2,

$$\begin{aligned}[(C\mathbf{Q}^{-1}C')_{22\cdot 1}]^{-1} &= (B\mathbf{Q}^{-1}B' - A\mathbf{Q}^{-1}B'(B\mathbf{Q}^{-1}B')^{-1}B\mathbf{Q}^{-1}A')^{-1} \\ &\sim W(n, [(C\Sigma^{-1}C')_{22\cdot 1}]^{-1}) \\ &= W(n, (B\Sigma^{-1}B' - A\Sigma^{-1}B'(B\Sigma^{-1}B')^{-1}B\Sigma^{-1}A')^{-1})\end{aligned}$$

and is independent of $(A\mathbf{Q}^{-1}A')^{-1}$. (To see this, note that $(C\mathbf{Y})_{2\cdot 1} \sim RN(m + s$' $C\Sigma^{-1}C')$ and $(C\mathbf{Q}^{-1}C')_{22\cdot 1} = (C\mathbf{Y})_{2\cdot 1}(C\mathbf{Y})_{2\cdot 1}'$.)

3. Other statistical applications. Biorthogonal and dual configurations provide interesting interpretations of other statistical phenomena, a few of which we discuss here. We consider first a population analog of the dual sample configurations of Section 2, leading to a probabilistic proof of the well-known fact that $(\Sigma^{-1})_{11} = (\Sigma_{11\cdot 2})^{-1}$.

If $\mathbf{x} \in R^p$ has the $N(0, \Sigma)$ distribution, then the set of linear combinations $\mathfrak{X} = \{\alpha'x: \alpha \in R^p\}$ is a vector space for which covariance is an inner product. The coordinates of $\mathbf{x}$ are a linearly independent configuration in $\mathfrak{X}$ with configuration matrix Σ and we may obtain their dual configuration: $\mathbf{y} = \Sigma^{-1}\mathbf{x} \sim N(0, \Sigma^{-1})$. (Note that the coordinates of $\mathbf{y}$ are elements of $\mathfrak{X}$.) Let $\mathbf{x}_1$ be the first s and $\mathbf{x}_2$, the remaining $p - s$ coordinates of $\mathbf{x}$ and partition $\mathbf{y}$ similarly. Then $\mathbf{x}_{1\cdot 2} = \mathbf{x}_1 - B\mathbf{x}_2 \sim N(0, \Sigma_{11\cdot 2})$ and is independent of $\mathbf{x}_2$. If we dualize the configuration given by the coordinates of $\mathbf{x}_{1\cdot 2}$, we obtain $\mathbf{y}_1$, which therefore has the $N(0, (\Sigma_{11\cdot 2})^{-1})$ distribution. But since $\mathbf{y}_1$ is a partition of $\mathbf{y}$, $\mathbf{y}_1 \sim N(0, (\Sigma^{-1})_{11})$; hence $(\Sigma^{-1})_{11} = (\Sigma_{11\cdot 2})^{-1}$.

The next example serves as a preliminary to the one that follows. Consider the usual set-up of the Gauss-Markov theorem: $\mathbf{y} = \mu + \boldsymbol{\varepsilon}$ is a random element of R^p, where $\mu = \Sigma_1^m\beta_1 u_i$, the u_i being $m < p$ known linearly independent elements of R^p, the β_i are unknown and $E\boldsymbol{\varepsilon} = 0$, $E\boldsymbol{\varepsilon\varepsilon}' = \Sigma$. We interpret the usual derivation of the minimum-variance-linear-unbiased-estimator in terms of biorthogonal configurations. If $A\Sigma^{-1}\mathbf{y}$ is an unbiased estimator of $\beta = (\beta_1, \cdots, \beta_m)'$ for all choices of β, we must have $A\Sigma^{-1}U' = I$, where $U = ((u_1, \cdots, u_m))'$. I.e., if

$A = ((a_1, \cdots, a_m))'$, $(a_1, \cdots, a_m)$ and $(u_1, \cdots, u_m)$ are biorthogonal relative to the Σ^{-1} inner-product on R^p. Hence we may write $a_i = v_i + \delta_i$, where $V = ((v_1, \cdots, v_m))'$ is dual to U and $\delta_i \varepsilon \mathcal{V}^{\perp}\{u_1, \cdots, u_m\}$; both relative to the Σ^{-1} inner-product. Thus $A = V + \Delta$, where $\Delta = ((\delta_1, \cdots, \delta_m))'$ is an arbitrary configuration in $\mathcal{V}^{\perp}\{u_1, \cdots, u_m\}$. cov $(A\Sigma^{-1}\mathbf{y}) = EA\Sigma^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\Sigma^{-1}A' = V\Sigma^{-1}V' + \Delta\Sigma^{-1}\Delta'$ since $V\Sigma^{-1}\Delta' = 0$. It is clear that minimum variance is obtained by choosing $\Delta = 0$, giving the estimator $V\Sigma^{-1}\mathbf{y} = (U\Sigma^{-1}U')^{-1}U\Sigma^{-1}\mathbf{y}$.

The last example has a superficial resemblance to the preceding. We consider an estimator for mixing proportions proposed by Robbins [4]. Let $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \cdots$ be independent observations with distribution F, where it is known that $F = \sum_1^p \alpha_i F_i$, the F_i being known distributions, the α_i, unknown proportions. (Of course $0 \leqq \alpha_i$, $\sum \alpha_i = 1$.) Robbins' ingenious method of estimating the α_i is as follows: One first constructs functions $\phi_1, \cdots, \phi_p$ so that $\int \phi_i \, dF_j = \delta_{ij}$. Then $E_F\phi_i(\mathbf{x}) = \alpha_i$ and the estimator $\boldsymbol{\phi}_{in} = \sum_{k=1}^n \phi_i(\mathbf{x}_k)/n$ is an unbiased and consistent estimator of α_i. (Robbins actually proposes $\boldsymbol{\phi}_{in}^+$ as an estimator.)

Robbins gives a specific construction for finding $\phi_1, \cdots, \phi_p$ on which we elaborate. For this, we need a more general notion of biorthogonal configurations: Let $(x_1, \cdots, x_p)$ be a configuration in a vector space $\mathfrak{X}$ and let $(y_1, \cdots, y_p)$ be a configuration in $\mathfrak{X}^*$, its algebraic adjoint (= linear functionals on $\mathfrak{X}$). Then the two configurations are biorthogonal if $\langle x_i, y_j \rangle = \delta_{ij}$ where $\langle x, y \rangle$ denotes the application of y to x (or x, considered as an element of $\mathfrak{X}^{**}$, to y). Let $\mu = F_1 + \cdots + F_p$. We seek functions $\phi_1, \cdots, \phi_p$ that are elements of $\bigcap_i \{L_1(F_i)\} = \{L_1(\mu)\}$ ($\{L\}$ means the topological vector space L, considered as just a vector space) so that $(F_1, \cdots, F_p)$ and $(\phi_1, \cdots, \phi_p)$ are biorthogonal configurations; the former in $\mathfrak{F} = \{\sum \alpha_i F_i : 0 \leqq \alpha_i, \sum \alpha_i = 1\}$, the latter in $\{L_1(\mu)\}$, which is a representation of $\mathfrak{F}^*$. (If $F \varepsilon \mathfrak{F}$ and $\phi \varepsilon \{L_1(\mu)\}$, $\langle F, \phi \rangle = \int \phi \, dF$.) Let $\mathcal{V}_1^{\perp}\{\mathfrak{F}\} \subset \{L_1(\mu)\}$ be the anihilator of $\mathfrak{F}$. Then if $\delta_i \varepsilon \mathcal{V}_1^{\perp}\{\mathfrak{F}\}$, $i = 1, \cdots, p$, $(\phi_1 + \delta_1, \cdots, \phi_p + \delta_p)$ is also biorthogonal with $(F_1, \cdots, F_p)$. Hence the estimators we seek may be characterized by a specific choice of $(\phi_1, \cdots, \phi_p)$ together with an arbitrary configuration in $\mathcal{V}_1^{\perp}\{\mathfrak{F}\}$. Robbins proposed the following choice: Let $f_i = dF_i/d\mu \; \varepsilon \{L_1(\mu)\}$ and let ϕ_i^* be the unique linear combination of elements of $\{f_1, \cdots, f_p\}$ so that $\langle \phi_i^*, F_j \rangle = \delta_{ij}$. (The natural embedding $F \rightarrow dF/d\mu$ of $\mathfrak{F}$ in $\{L_\infty(\mu)\} \subset \{L_1(\mu)\}$ provides a notion of duality between $(F_1, \cdots, F_p)$ and $(\phi_1^*, \cdots, \phi_p^*)$.)

One might inquire whether there exists an optimal choice of $(\phi_1, \cdots, \phi_p)$. If optimality means small variance, the answer (not surprisingly) is: no choice of $(\phi_1, \cdots, \phi_p)$ gives a uniformly minimum-variance unbiased estimator. To see this, we show that the unbiased estimator having minimum variance at $F^* = \sum \alpha_i^* F_i$ depends on the choice of F^*. Choose F^* so that $\alpha_i^* > 0$, $i = 1, \cdots, p$. (This assumption is not critical, it merely assures that $F_i \ll F^*$, $i = 1, \cdots, p$.) Since variance is to be minimized, we need consider only $\phi_i \varepsilon \{L_2(\mu)\} = \{L_2(F^*)\}$. The natural embedding $F \rightarrow dF/dF^*$ puts everything in $\{L_2(F^*)\}$ which we then treat as the inner-product space $L_2(F^*)$. Letting $f_i^* = dF_i/dF^*$, the problem is to select a configuration in $L_2(F^*)$ which is biorthogonal with $(f_1^*, \cdots, f_p^*)$ and optimal at F^*. There is such a configuration; $(\psi_1, \cdots, \psi_p)$, the dual to $(f_1^*, \cdots, f_p^*)$. For if $\delta_i \varepsilon \mathcal{V}^{\perp}\{f_1^*, \cdots, f_p^*\}$ (note that $\{\mathcal{V}^{\perp}\{f_1^*, \cdots, f_p^*\}\} = \mathcal{V}_2^{\perp}\{\mathfrak{F}\}$, the anihilator of $\mathfrak{F}$ in $\{L_2(F^*)\}$),

$$\text{var}_{F^*}\{\psi_i(\mathbf{x}) + \delta_i(\mathbf{x})\} = \text{var}_{F^*}\psi_i(\mathbf{x}) + E_{F^*}\delta_i^2(\mathbf{x}),$$

since ψ_i and δ_i are orthogonal in $L_2(F^*)$. Clearly the choice $\delta_i \equiv 0$ gives minimum variance at F^*. Thus $(\psi_1, \cdots, \psi_p)$ is optimal at F^* and a uniformly minimum variance estimator does not exist.

4. Acknowledgment. The author wishes to thank Professors A. P. Dempster, I. Olkin, M. Perlman and R. A. Wijsman for their comments on an earlier version of this paper.

REFERENCES

[1] DEMPSTER, A. P. (1969). *Elements of Continuous Multivariate Analysis.* Addison-Wesley, Reading.

[2] JAMES, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.* **25** 40–75.

[3] RAO, C. R. (1966). *Linear Statistical Inference and Its Applications.* Wiley, New York.

[4] ROBBINS, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[5] WIJSMAN, R. A. (1959). Applications of a certain representation of the Wishart matrix. *Ann. Math. Statist.* **30** 597–601.