Ann. Inst. Statist. Math.
Vol. 40, No. 1, 93-99 (1988)

CHARACTERIZATION OF CONDITIONAL COVARIANCE AND UNIFIED THEORY IN THE PROBLEM OF ORDERING RANDOM VARIABLES

KENTARO NOMAKUCHI[1] AND TOSHIO SAKATA[2]

[1] *Department of mathematics, Faculty of Science, Kyushu University, Higashi-ku, Fukuoka 812, Japan*
[2] *Faculty of General Education, Kumamoto University, Kumamoto 860, Japan*

(Received May 23, 1986; revised August 28, 1986)

Abstract. Under the assumption that a $(p+q)$-dimensional row vector (Y, X) is elliptically contoured distributed, the conditional covariance of Y given $X=x$ is characterized in the context of correctly ordering the coordinates Y_k's of Y based on X. This is an answer to a conjecture implicit in Portnoy (1982). Moreover some unified theory is presented for the problem of ordering Y_k's based on X. An essential tool is the decreasing in transposition (D. T.) function theory of Hollander *et al.* (1977, *Ann. Statist.*, 5(4), 722–733).

Key words and phrases: Linear predictor, ordering r.v.'s, elliptically contoured distribution.

1. Introduction

Let $Y=(Y_1, Y_2,\dots, Y_p)$ be an unobservable p-dimensional random vector with a mean vector μ and $X=(X_1, X_2,\dots, X_q)$ be an observable q-dimensional random vector with a mean vector ν. If (Y, X) has a joint elliptically contoured distribution, we write $(Y, X)\simeq E_{p+q}((\mu, \nu), \Sigma)$, and if normally distributed we write $(Y, X)\simeq N_{p+q}((\mu, \nu), \Sigma)$. We assume that Σ is non-singular throughout the paper. Let Σ_x, Σ_{xy} and $\Sigma_{y:x}$ denote the covariance matrix of X, and the covariance matrix between X and Y, and the conditional covariance of Y given X respectively. Let e denote the normalized p-vector whose all elements are equal to $1/\sqrt{p}$. Portnoy (1982) has treated the problem of ordering or ranking Y_k's based on X. Assuming that $(Y, X)\simeq N_{p+q}((\mu, \nu), \Sigma)$ and that the conditional covariance of Y given X is of the form

$$\Sigma_{y:x} = \alpha I + \beta e^t e \tag{1.1}$$

with real constants α and β, he showed

(A) the probability of correctly ordering the co-ordinates Y_k's of Y is maximized by ranking according to the order of the best linear predictors, $E(Y|X=x)$.
In this paper we assume that $(Y, X)\simeq E_{p+q}((\mu, \nu), \Sigma)$, and the following four results are presented.

(B_1) If Σ_{xy} is assumed to be of rank p, that the fact (A) holds true for $\mu=\tau e$ for some constant τ implies that $\Sigma_{y\cdot x}$ is of the form

$$\Sigma_{y\cdot x} = \alpha I + a^t e + e^t a\ , \quad \text{where } a \in R^p \text{ and } \alpha \in R\ . \tag{1.2}$$

(B_2) If Σ_{xy} is not assumed to be of full rank p, that the fact (A) holds true for all μ implies that $\Sigma_{y\cdot x}$ is also of the form (1.2).

(B_3) If the conditional density of Y given X is assumed to be unimodal, that $\Sigma_{y\cdot x}$ is of the form (1.2) implies that the fact (A) holds true.

(C) The problem of maximizing the probability of correctly ordering Y_k's based on X may be treated in a unified way in terms of D.T. function in the sense of Hollander *et al.* (1977).

Note that (B_1) and (B_2) are the answers to a conjecture implicit in Portnoy (1982), and (B_3) is a generalization to an elliptically contoured distribution from a normal distribution, and that (C) gives a new point of view to this ranking problem.

Note that the concept of D.T. function is known to be useful to unify the monotonicity theory of many selection problems (see Berger and Proschan (1984)). Our results show that our problem is one of such interesting situations.

2. Terminologies and preliminary results

In this section we define some terminologies and list some preliminary results which are needed in the subsequent sections.

DEFINITION 2.1. If a p-dimensional random vector Y has a density function of the form

$$h(y) = q((y-\mu)\Sigma^{-1}(y-\mu)^t)\ , \tag{2.1}$$

where $q(\cdot)$ is a non-negative function on $(0, \infty)$, Y is said to have an elliptically contoured distribution and it is denoted as $Y\simeq E_p(\mu, \Sigma)$. Moreover if $q(\cdot)$ is non increasing, we say that the distribution of Y is unimodal and denotes as $Y\simeq EM_p(\mu, \Sigma)$.

Note 2.1. When $(Y, X)\simeq E_{p+q}((\mu, \nu), \Sigma)$, we use the terminology $E\{Y|X=x\}=\mu+(x-\nu)\Sigma_x^{-1}\Sigma_{xy}$ and $\Sigma_{y\cdot x}=\Sigma_y-\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}$, even if the expectation and the covariance do not exist. It is well known that the conditional distribution of Y given $X=x$ is $E_p(E\{Y|X=x\}, \Sigma_{y:x})$. In the unimodal case, the unimodality is also preserved. For other properties and notations about an elliptically contoured distribution, see Canbani *et al.* (1981).

Note 2.2. We also need the concept of a decreasing in transposition (D.T.) function. A good reference for this is Hollander *et al.* (1977).

DEFINITION 2.2. Let $S(p)$ be the set of all permutations on the set of integers $\{1,2,\ldots,p\}$. Then for $\sigma \in S(p)$, $C(\sigma)$ denotes the subset of R^p defined as follows,

$$C(\sigma) = \{y \in R^p \,|\, y_{\sigma(1)} \leq y_{\sigma(2)} \leq \cdots \leq y_{\sigma(p)}\} .$$

DEFINITION 2.3. A mapping δ: $x \in R^q \to \delta(x) \in S(p)$ is called a decision (function) for the problem of ranking Y_k's based on X.

DEFINITION 2.4. The decision δ_0 is defined as follows,

$$\delta_0(x) = \sigma \quad \text{if} \quad E\{Y|X = x\} \in C(\sigma) .$$

Note that δ_0 depends on (μ, ν, Σ). Here we cite the fundamental theorem of Portnoy (1982), which will clarify our results, with easier proof than the original.

THEOREM 2.1. *For the ranking problem of Y_k's based on X a decision $\delta(x)$ is optimal (that is, maximizing the probability of correctly ordering Y_k's) if and only if for almost all x*

$$P\{Y \in C(\delta(x))|X = x\} = \max_{\tau \in S(p)} P\{Y \in C(\tau)|X = x\} .$$

PROOF. Let δ' be any decision and δ be a decision satisfying the condition of the theorem. Then we have

$$P\{Y \in C(\delta'(x))|X = x\} \leq P\{Y \in C(\delta(x))|X = x\} \qquad \text{for almost all } x .$$

Taking an expectation of both sides of the above inequality with respect to X, it is clear that δ is optimal in the sense of the theorem. Conversely, any decision not satisfying the condition of the theorem is clearly not optimal since decisions satisfying the condition of the theorem really exist.

3. Characterization of conditional covariance

Before stating the theorems, we give lemmas without proofs (see

Nomakuchi and Sakata (1988)).

LEMMA 3.1. *For $p \geq 3$, let $Y \simeq E_p(0, \Sigma)$. If $P\{Y \in C(\sigma)\} = 1/p!$ for all $\sigma \in S(p)$, then Σ is of the form* (1.2).

LEMMA 3.2. *For $p \geq 3$, let $Y \simeq EM_p(0, \Sigma)$. If Σ is of the form of* (1.2), *then $P\{Y \in C(\sigma)\} = 1/p!$ for all $\sigma \in S(p)$.*

We also need the next lemma for Theorem 3.1.

LEMMA 3.3. *Let $(Y, X) \simeq E_{p+q}((\mu, \nu), \Sigma)$. Let U be any measurable neighbourhood of ν in R^q. Then $P\{X \in U\} > 0$.*

PROOF. Without loss of generality we may assume $\Sigma = I$ and $(\mu, \nu) = (0, 0)$. Set $T = \{(y, x) | x \in U\}$. Then $P\{(Y, X) \in T\} = P\{X \in U\}$. From the sphericity of the joint distribution, the regions obtained by rotations of T have the same probability $P\{(Y, X) \in T\}$. Since R^{p+q} is covered by countable these regions, $P\{X \in U\}$ must be positive.

THEOREM 3.1. *Let $(Y, X) \simeq E_{p+q}((\mu, \nu), \Sigma)$ with $q \geq p \geq 3$. Assume that Σ_{xy} is of rank p. If δ_0 is optimal for $\mu = \tau e$ with some real constant τ, then $\Sigma_{y \cdot x}$ has the form of* (1.2).

PROOF. Without loss of generality we may assume $\nu = 0$. Since $Y - \tau e \in C(\sigma)$ is equivalent to $Y \in C(\sigma)$, we also assume $\mu = 0$. Let δ_0 be optimal. From Theorem 2.1, it holds that if $E\{Y | X = x\} \in C(\sigma)$, $P\{Y \in C(\sigma) | X = x\} = \max_{\tau \in S(p)} P\{Y \in C(\tau) | X = x\}$, which implies $P\{Y \in C(\sigma) | X = x\} \geq 1/p!$. Since (Y, X) has the elliptically contoured distribution, from Lemma 3.3, there exists a sequence $\{x_i\}$ such that $x_i \to 0$, $E\{Y | X = x_i\} = x_i \Sigma_x^{-1} \Sigma_{xy} \in C(\sigma)$, and $P\{Y \in C(\sigma) | x = x_i\} \geq 1/p!$. Since (Y, X) has a density, it is not difficult to check $P\{Y \in C(\sigma) | X = 0\} = \lim P\{Y \in C(\sigma) | X = x_i\}$. Therefore we have $P\{Y \in C(\sigma) | X = 0\} \geq 1/p!$. Since this holds for any $\sigma \in S(p)$, it follows $P\{Y \in C(\sigma) | X = 0\} = 1/p!$ for any $\sigma \in S(p)$. Applying Lemma 3.1, the proof is completed.

Next theorem treats the case where Σ_{xy} is not of rank p. In this case $E\{Y | X = x\}$ may be restricted to the set of Lebesgue measure 0, for example, the boundary of some $C(\sigma)$ even if x take all values of R^q. For this situation, it is obvious that we must introduce the class of randomized decisions. A randomized decision δ is defined as a mapping from the sample space of $X (= R^q)$ to the set of probability measures on $S(p)$. Typically δ_0 is defined as follows: $\delta_0(x) = \sigma$ with probability 1 if $E\{Y | X = x\} \in \mathrm{Int}(C(\sigma))$ and $\delta_0(x) = \sigma_i$ $(i = 1, 2, \ldots, k)$ with probability $1/k$, if $E\{Y | X = x\} \in Bd(C(\sigma_i))$, where $\mathrm{Int}(A)$ and $Bd(A)$ denote the interior and the boundary of the set A respectively.

Moreover, in this degenerate case, to obtain those sequences $\{x_j\}$ in Theorem 3.1, we need to assume that μ is not fixed.

THEOREM 3.2. *Let $(Y, X) \simeq E_{p+q}((\mu, \nu), \Sigma)$. If $\delta_0 = \delta_0((\mu, \nu), \Sigma)$ is optimal for all $\mu \in U$, $\Sigma_{y:x}$ must be of the form* (1.2). *Here U may be an arbitrary small neighbourhood of the origin of R^p.*

PROOF. Under the assumption, the existence of the sequence $\{x_i\}$ in the proof of Theorem 3.1 is clear. Rest of the proof is the same as that of Theorem 3.1.

THEOREM 3.3. *Let $(Y, X) \simeq EM_{p+q}((\mu, \nu), \Sigma)$. If $\Sigma_{y:x}$ has the form of* (1.2), *then δ_0 is optimal.*

PROOF. This is postponed to Example 4.2.

4. Unified theory through D.T. function theory

In this section we present some unified theory for the problem of ranking Y_k's based on X. In Portnoy (1982) and our previous sections the conditions are needed on the conditional covariance of Y given $X=x$ in order that the decision δ_0 is optimal. Here we show that optimal decision functions are determined by the D.T. structure of conditional density function $f(y|x)$.

THEOREM 4.1. *Assume that the conditional density of Y given $X=x$, $f(y|x)$, is represented as $f(y|x)=h(y, g(x))$, where $g(x)=(g_1(x), g_2(x), \dots, g_p(x))$ and $h(y, z)$ is a D.T. function in y and z. Then the probability of correctly ordering Y_k's is maximized if, observing $X=x$, Y_k's are ordered according to the order of $g_k(x)$'s $(k=1, 2, \dots, p)$.*

PROOF. From Theorem 2.1, the maximal probability is realized when, for $X=x$, Y_k's are ordered according to the order represented by the cone $C(\sigma)$, where $P\{Y \in C(\sigma)|X=x\} = \max_{\tau \in S(p)} P\{Y \in C(\tau)|X=x\}$. Now, without loss of generality, we may assume that $g_1(x) < g_2(x) < \cdots < g_p(x)$. Then, since $h(y, z)$ is a D.T. function in y and z, for any $\tau \in S(p)$

$$h(y, g) \le h(y(\tau), g) \quad \text{for all} \quad y \in C(\tau)\ .$$

Integrating both sides of above inequalty on the set $C(\tau)$ with respect to y, we have that

$$\int_{y \in C(\tau)} h(y, z)dy \le \int_{y \in C(\tau)} h(y(\tau), g(x))dy$$

$$= \int_{\{y \mid y_1 < y_2 < \cdots < y_p\}} h((y_1, y_2, \ldots, y_p), g) dy ,$$

which implies that

$$P\{Y \in C(\tau) \mid X = x\} \leq P\{Y \in C(\sigma_0) \mid X = x\} \quad \text{for any} \quad \tau \in S(p) .$$

This completes the proof of the theorem.

Example 4.1. Let assume that $(Y, X) \simeq N_{p+q}((\mu, \nu), \Sigma)$ and $\Sigma_{y \cdot x} = \alpha I + \beta e^t e$. Then the conditional density of given $X = x$ is

$$f(y \mid x) = h(y, g) = c \exp\{-(1/2)(y - g)\Sigma_{y \cdot x}^{-1}(y - g)^t\} ,$$

where $g = E\{Y \mid X = x\}$. Since $\Sigma_{y \cdot x}$ is of equi-correlated structure, h is D.T. (see Hollander *et al.* (1977)). Hence the probability attains its maximum when a decision δ_0 is adopted, which was a main result in Portnoy (1982).

Example 4.2. Let $(Y, X) \simeq EM_{p+q}((\mu, \nu), \Sigma)$ and $\Sigma_{y \cdot x} = \alpha I + e^t a + a^t e$. Then the conditional density is

$$f(y \mid x) = h(y, g) = q((y - g)\Sigma_{y \cdot x}(y - g)^t) ,$$

where $g = E\{Y \mid X = x\}$, and q is a non-increasing function on $[0, \infty)$. Neglecting a constant multiplier we can write $\Sigma_{y \cdot x} = I + \beta e^t e + \gamma(a^t e + e^t a)$ with $|a| = 1$, $ae^t = 0$, where β and $\gamma \in R^1$ and $a \in R^p$. Since $\lambda = 1 + \beta - \gamma^2 = (e - \gamma a)\Sigma_{y \cdot x}(e - \gamma a)^t > 0$, setting $\Gamma = I - (1 + \lambda^{-1/2})e^t e + \gamma\lambda^{-1/2} a^t e$, we have that $Y\Gamma \simeq EM_0(g\Gamma, I)$. Since $z \in C(\sigma)$ is equivalent to $z\Gamma \in C(\sigma)$ for any $z \in R^p$ and $\sigma \in S(p)$, without loss of generality, we may take $\Sigma_{y \cdot x} = I$. Then the quadratic form in $q(\cdot)$ is D.T. and therefore $f(y, g \mid x)$ is D.T. Applying Theorem 4.1, we have δ_0 is optimal. This gives the proof of Theorem 3.3 from a unified point of view.

Example 4.3. A density function $f(x)$ is said to belong to the Furlie-Gumbel-Morgenstein family with four variables if it is represented as follows (see Johnson (1985)),

$$f(x) = \prod_{j=1}^{4} f_j(x_j)\Big[1 + \sum_{j_1 < j_2} \alpha_{j_1 j_2}\{1 - 2F_{j_1}(x_{j_1})\}\{1 - 2F_{j_2}(x_{j_2})\} + \cdots + \alpha_{1234} \prod_{j=1}^{4}\{1 - 2F_j(x_j)\}\Big] ,$$

where F_j is the distribution function of f_j, $j = 1, \ldots, 4$. Here we consider a simple special case where $f_j(x) = h(x)$, $j = 1, \ldots, 4$, $\alpha_{j_1 j_2 j_3} = 0$ $(1 \leq j_1 < j_2 < j_3 \leq 4)$, and $\alpha_{1234} = 0$.

In this case the conditional density of (x_1, x_2) given (x_3, x_4) is represented as follows,

$$
\begin{aligned}
f(x_1, x_2 | x_3, x_4) &= h(x_1)h(x_2)[S_1(x_1, x_2) + S_2(x_3, x_4) \\
&\quad - \{1 - 2H(x_1)\}g_1(x_3, x_4) \\
&\quad - \{1 - 2H(x_2)\}g_2(x_3, x_4)]/\{1 + S_2(x_3, x_4)\} \\
&= h(x_1, x_2, g_1(x_3, x_4), g_2(x_3, x_4) | x_3, x_4) ,
\end{aligned}
$$

where

$$
\begin{aligned}
g_1(x_3, x_4) &= -\alpha_{13}\{1 - 2H(x_3)\} - \alpha_{14}\{1 - 2H(x_4)\} , \\
g_2(x_3, x_4) &= -\alpha_{23}\{1 - 2H(x_3)\} - \alpha_{24}\{1 - 2H(x_4)\} , \\
S_1(x_1, x_2) &= \alpha_{12}\{1 - 2H(x_1)\}\{1 - 2H(x_2)\} , \\
S_2(x_3, x_4) &= \alpha_{34}\{1 - 2H(x_3)\}\{1 - 2H(x_4)\} ,
\end{aligned}
$$

where $H(x)$ is the distribution function of $h(x)$. Routine works ascertain that $h(x_1, x_2, g_1(x_3, x_4), g_2(x_3, x_4) | x_3, x_4)$ is a D.T. function in (x_1, x_2) and (g_1, g_2). Hence it follows that an optimal decision function ranks (x_1, x_2) according to the order of g_1 and g_2. This gives an example other than an elliptically contoured distribution to which Theorem 4.1 is applicable.

Acknowledgements

We are grateful to Professor A. Kudo for his encouragements, and also Mr. S. Sasabuchi for his helpful comments.

REFERENCES

Berger, R. L. and Proschan. F. (1984). Monoticity in selection problems, *Ann. Statist.*, **12** (1), 387–391.

Canbani, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions, *J. Multivariate Anal.*, **11**, 368–385.

Hollander, M., Proschan, F. and Sethuraman, J. (1977). Function decreasing in transposition and their applications in ranking problems, *Ann. Statist.*, **5** (4), 722–733.

Johnson, N. L. (1985). On some generalized Farlie-Gumbel-Morgenstein distributions, *Comm. Statist. A—Theory Methods*, **4**, 415–427.

Nomakuchi, K. and Sakata, T. (1988). Characterization of the form of covariance matrix of elliptically contoured distribution, *Sankhyā Ser. A*. (to appear).

Portnoy, S. (1982). Maximizing the probability of correctly ordering random variables using linear predictors, *J. Multivariate Anal.*, **12**, 256–269.