The Annals of Statistics
1977, Vol. 5, No. 6, 1240–1244

OPTIMUM FISHERIAN INFORMATION FOR MULTIVARIATE DISTRIBUTIONS

BY TAMER BAŞAR

Marmara Scientific and Industrial Research Institute

It is shown that within the class of all multivariate distributions depending on a location parameter (and satisfying certain smoothness conditions) and with a weighted norm constraint on the covariance matrix, the one with minimum Fisherian information is the Gaussian distribution. This result is then used in obtaining a tight upper bound on the error of estimating an unknown random vector observed in additive Gaussian noise under quadratic loss.

1. Introduction. It is known that within the class of distributions of a single variable depending on a location parameter (and satisfying certain regularity conditions) and with a fixed variance, the one with minimum Fisherian information is the Gaussian distribution [2]. In this paper, we generalize this result to n-dimensions, and prove that a similar result still holds. The quantity to be minimized, in this case, is a weighted norm of the Fisher information matrix, and the variance constraint is replaced by a weighted norm constraint on the covariance matrix. We then consider the problem of estimating an unknown random vector in additive Gaussian noise and under quadratic loss, and use the above result to obtain a tight upper bound on the expected value of optimum quadratic error.

2. Main result. Let $\mathscr{A}_\theta$ denote a family of distributions $\{P_\theta, \theta \in R^n\}$ on R^n, with densities $dP_\theta/d\mu = p(x - \theta)$ with respect to the Lebesgue measure, and satisfying the following second-order constraint:

(i) $\operatorname{Tr}[D \operatorname{Cov}(X)] = k^2$, $D > 0$ (an $n \times n$ matrix),

together with the three regularity conditions:

(ii) p is continuously differentiable

(iii) $\int x'xp(x)\,dx < \infty$

(iv) $|x_i|p_i(x_i) \to 0$ as $|x_i| \to \infty$,

where p_i denotes the marginal density function of the ith component of the random vector X.

Then the Fisher information matrix for the above family of distributions is given by

$$I_p(\theta) = \int_{K\theta} [\nabla_\theta \ln p(x-\theta)][\nabla_\theta \ln p(x-\theta)]' p(x-\theta)\,dx \tag{1}$$

Received February 1977; revised April 1977.

AMS 1970 *subject classifications*. 62E10, 62H05, 62B10, 62F10.

Key words and phrases. Fisherian information, multivariate distributions, robust estimation, Gaussian distribution.

where K_θ denotes a region in R^n defined by $\{x \in R^n, p(x - \theta) > 0\}$. Since θ is a location parameter, the above integral is actually independent of θ, and hence the Fisher information matrix can be written as

$$I_p = \int_K [\nabla_x \ln p(x)][\nabla_x \ln p(x)]' p(x)\, dx \tag{2}$$

where $K = \{x \in R^n, p(x) > 0\}$. We now seek to minimize $\mathrm{Tr}\,[CI_p]$ over $\mathscr{A}_0$, where $C > 0$ is an $n \times n$ matrix. The main result is the following:

THEOREM 1. *The minimum of* $J = \mathrm{Tr}\,[CI_p]$ *is attained on* $\mathscr{A}_0$ *uniquely by the Gaussian distribution with covariance*

$$\mathrm{Cov}\,(X) = \{k^2/\mathrm{Tr}\,[F]\} D^{-\frac{1}{2}} F D^{-\frac{1}{2}} \tag{3}$$

where

$$F =_\Delta (D^{\frac{1}{2}} C D^{\frac{1}{2}})^{\frac{1}{2}}\,, \tag{4a}$$

$$D^{-\frac{1}{2}} =_\Delta (D^{\frac{1}{2}})^{-1}\,, \tag{4b}$$

and $A^{\frac{1}{2}}$ *denotes the unique positive definite square-root of* $A > 0$. *The minimum value of* J *is given by*

$$J^* = \{\mathrm{Tr}\,[F]\}^2/k^2\,. \tag{5}$$

PROOF. Without any loss of generality, we may take the mean of X to be zero. Moreover, via the absolutely continuous transformation $Y = D^{\frac{1}{2}}X$, we can rewrite constraint (i) as

(i′) $\mathrm{Tr}\,[\mathrm{Cov}\,(Y)] = k^2$

and I_p as

$$I_f = \int_{\tilde{K}} D^{\frac{1}{2}} \varphi_f(y) \varphi_f'(y) D^{\frac{1}{2}} f(y)\, dy\,, \tag{2′}$$

where $f(y)$ denotes the density function of Y (which exists because of absolute continuity of the transformation), $\varphi_f(\cdot)$ is defined by

$$\varphi_f(y) = \nabla_y \ln f(y) = \nabla_y f(y)/f(y)\,, \tag{6}$$

and $\tilde{K}$ is defined similar to K. Furthermore,

$$J = \mathrm{Tr}\,[CI_p] = \mathrm{Tr}\,[F^2 I_f]\,, \tag{7}$$

and hence we can now seek to minimize (7) under the constraint (i′). To complete the formulation of this new minimization problem, we replace $\mathscr{A}_0$ by $\tilde{\mathscr{A}}_0$, with the definition of the latter being obvious.

For each $f \in \tilde{\mathscr{A}}_0$, we now introduce the inner product

$$\langle \xi, \eta \rangle_f = \int_{\tilde{K}} \xi'(y) \eta(y) f(y)\, dy \tag{8}$$

on the space $L_2(\mu_f)$ of random vectors of dimension n, where μ_f denotes the measure derived from the density f. The natural norm derived from this inner product will be denoted by $\|\cdot\|_f$. Let us now consider the scalar quantity

$$\langle y, F\varphi_f(y) \rangle_f =_\Delta \int_{\tilde{K}} y' F \nabla_y f(y)\, dy = \sum_{i,j} \int_{\tilde{K}} y_i F_{ij} \frac{\partial f(y)}{\partial y_j}\, dy_1 \cdots dy_n \tag{9}$$

where F_{ij} denotes the ijth entry of F. Since $f(y)$ is identically zero outside $\tilde{K}$ and since $f(y)$ is continuous, $\tilde{K}$ is an open subset of R^n and we can, without any loss of generality, take $\partial f(y)/\partial y_i$ to be finite on the boundary of this open set ($i = 1, \cdots, n$) and going to zero in a sufficiently small neigborhood of the boundary. This allows us to replace $\tilde{K}$ above by R^n and apply integration by parts on the right-hand side of (9):

$$F_{ii} \int_{R^n} y_i \frac{\partial f}{\partial y_i} dy_1 \cdots dy_n = F_{ii}[y_i f_{y_i}(y_i)|_{-\infty}^{+\infty} - \int_{R^n} f(y)\, dy] = -F_{ii} \int_{\tilde{K}} f(y)\, dy = -F_{ii} \tag{10a}$$

and for $i \neq j$,

$$\begin{aligned} F_{ij} \int_{R^n} y_i \frac{\partial f}{\partial y_j} dy_1 \cdots dy_n &= F_{ij} \int_{R^{n-1}} y_i f(y)\, dy_1 \cdots dy_k \cdots dy_n |_{y_j=-\infty}^{y_j=+\infty} \quad k \neq j \\ &= 0\,, \end{aligned} \tag{10b}$$

where in writing (10a) and (10b) we have made use of (iv) explicitly. (10a) and (10b) together with (9) now imply that

$$\langle y, F\varphi_f(y)\rangle_f = -\mathrm{Tr}\,[F]\,, \tag{11}$$

and hence this indicates that the inner product is independent of the choice of $f \in \tilde{\mathscr{S}}_0$. Cauchy–Schwarz inequality applied to (11) yields:

$$\{\mathrm{Tr}\,[F]\}^2 = |\langle y, F\varphi_f(y)\rangle_f|^2 \leqq \|y\|_f^2 \|F\varphi_f(y)\|_f^2 = k^2 J\,, \tag{12}$$

and therefore J is bounded above by $\{\mathrm{Tr}\,[F]\}^2/k^2$, which is a tight bound since it is attained when f is chosen as the zero-mean Gaussian density with covariance $(k^2/\mathrm{Tr}\,[F])F$. Moreover, this is the unique minimizing solution, since the inequality sign in the Cauchy–Schwarz inequality can nontrivially be replaced by an equality sign if and only if $F\varphi_f(y) = \lambda y$ for some scalar λ, which in this case is $\lambda = -\{\mathrm{Tr}\,[F]\}/k^2$ because of the constraint (i′) and the regularity condition (iv). To complete the proof of Theorem 1 we simply note that under the linear transformation $X = D^{-\frac{1}{2}}\, Y$, X is still a Gaussian random vector with $\mathrm{Cov}\,(X) = D^{-\frac{1}{2}}\, \mathrm{Cov}\,(Y) D^{-\frac{1}{2}}$. □

REMARK 1. It should be clear from the above, and especially from the relations (12), that if the constraint (i) is replaced by

(i″) $\mathrm{Tr}\,[D\, \mathrm{Cov}\,(X)] \leqq k^2$,

then the statement of Theorem 1 will still be valid.

3. An application. One application of the result presented here would be in the derivation of upper bounds on the estimation error of a vector valued random variable in additive Gaussian noise. To be more specific, let X denote a single observation of an n-dimensional zero-mean random vector Y in independent additive Gaussian noise W, i.e.,

$$X = Y + W \tag{13}$$

where $W \sim N(0, R)$, $R > 0$, and Y is assumed to possess a density function $p(y)$ with respect to the Lebesgue measure and with the norm constraint

$$\int_{R^n} y'yp(y)\,dy \leqq c^2 . \tag{14}$$

We further assume that $p(y)$ satisfies the three regularity conditions (ii)—(iv), and seek to obtain an upper bound on the error of estimating Y using observation (13) and under the quadratic loss function

$$L(\delta, p) = E_p[(\delta(x) - y)'(\delta(x) - y)] , \tag{15}$$

where δ is an estimate for Y and the criterion is minimization of $L(\delta, p)$ over all Borel-measurable δ and for each fixed $p(\cdot)$.

It is well known that (15) is uniquely minimized for each fixed p by the conditional mean

$$\delta_p{}^0(x) = E_p[y \mid x] = \frac{1}{\psi_p(x)} \int y\phi(x - y)p(y)\,dy \tag{16a}$$

where $\phi(\cdot)$ is the density function of W, and

$$\psi_p(x) =_\Delta \int \phi(x - y)p(y)\,dy \tag{16b}$$

is the density function of X and satisfies the three regularity conditions (ii)—(iv). We now note the relation

$$R\nabla_x \ln \psi_p(x) + x = \delta_p{}^0(x) ,$$

and make use of this differential equation in simplifying (15) after substitution of (16a):

$$\begin{aligned} L(\delta_p{}^0, p) = E_p[y'y] - E_p[\delta_p{}^{0\prime}\delta_p{}^0] &= E[W'W] + E_p[x'x] - E_p[\delta_p{}^{0\prime}\delta_p{}^0] \\ &= \mathrm{Tr}\,[R] - \mathrm{Tr}\,[RRI_\psi] , \end{aligned}$$

where I_ψ is defined by (2) with $p(x)$ replaced by $\psi_p(x)$. We thus see that $L(\delta_p{}^0, p)$ depends only on the density function of X, and hence we can now seek to determine the tight upper bound on $L(\delta^0, p)$ when ψ varies over the class of permissible density functions for X, characterized by relation (16b) and the constraint

$$\int_{R^n} x'x\psi(x)\,dx \leqq c^2 + \mathrm{Tr}\,[R] .$$

Ignoring satisfaction of relation (16b) for a moment, we note that the unique solution to the above is given by Theorem 1 to be the Gaussian distribution with covariance

$$\mathrm{Cov}\,(X) = \{(c^2 + \mathrm{Tr}\,[R])/\mathrm{Tr}\,[R]\}R$$

yielding the tight upper bound

$$L(\delta_p{}^0, p) \leqq c^2\,\mathrm{Tr}\,[R]/\{c^2 + \mathrm{Tr}\,[R]\} .$$

The corresponding (least favorable) density function of Y can easily be seen to be Gaussian with covariance

$$\mathrm{Cov}\,(Y) = \{c^2/\mathrm{Tr}\,[R]\}R .$$

We note that a similar result was obtained in [1] under a soft constraint on Cov (Y), which was included in the generalized quadratic loss function. Using an entirely different approach, it was shown in [1] that the tight upper bound is again obtained by taking Y to be Gaussian.

REFERENCES

[1] BAŞAR, T. and MINTZ, M. (1973). On a minimax estimate for the mean of a normal random vector under a generalized quadratic loss function. *Ann. Statist.* **1** 127–134.

[2] KAGAN, A. M., LINNIK, YU. V. and RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.

SCIENTIFIC AND TECHNICAL RESEARCH COUNCIL OF TURKEY
MARMARA SCIENTIFIC AND INDUSTRIAL RESEARCH INSTITUTE
P.O. BOX 21
GEBZE-KOCAELI, TURKEY