Ann. Inst. Statist. Math.
Vol. 54, No. 3, 621–625 (2002)

A MONOTONICITY OF MOMENTS CONCERNED WITH ORDER RESTRICTED STATISTICAL INFERENCE

KENTARO NOMAKUCHI

Department of Mathematics and Information Science, Kochi University, Kochi 780-8520, Japan

(Received December 27, 1999; revised March 29, 2001)

Abstract. In the order restricted statistical inference problem, moments of the distance between a true parameter and the least square estimate are non-decreasing when the true parameter moves along a half line from an initial point in the null space. This follows from "stochastically larger" property of the distance.

Key words and phrases: Order restricted statistical inference, least square estimate, maximum likelihood estimate, monotonicity, moments, stochastically larger.

1. Introduction

The purpose of this paper is to show a monotonicity of moments concerned with what is called an order restricted statistical inference problem (Barlow *et al.* (1972) or Robertson *et al.* (1988)). We treat parameters of a location family. In such a statistical problem the parameter space C becomes a closed convex cone in R^n. Especially in a testing problem, the null space is usually a linear subspace L contained in C, and the alternative is $C - L$. In order to see the performance of a statistical procedure, it is important to examine its behavior when the true parameter moves along a half line in $\Theta_1 = C - L$ from an initial point of the null space $\Theta_0 = L$. We will show that, in such a case, the distance between the true parameter and the least square estimate (LSE) becomes "stochastically larger." Hence we have the monotonicity of its moments.

Anraku's Order-Restricted Information Criterion (ORIC) motivates us to write the paper. Anraku (1999) applies an Akaike Information Criterion (AIC) approach to simple order restricted inference and proposes ORIC. AIC is defined by (the log-likelihood) + (a penalty) and used as a criterion in many model-selection problems, where the penalty usually depends on the model itself, but not on each parameter in the model. Anraku obtains such a similar quantity in the simple order restricted model about the normal distribution. The penalty, however, depends on the parameters. Hence, he proposes its minimum value as a penalty of the model and names the criterion as ORIC. Our monotonicity property plays a key role in finding the minimum penalty, since the maximum likelihood estimate (MLE) is LSE in his problem. He treats it in the simple order restriction, but we will consider it under more general conditions. Hence the result would extend the region where his ORIC approach is applicable, for example, to simple tree order restriction or umbrella order restriction.

There are several similar results on monotonicity. It is well known that the distance between the sample vector X and the MLE has a "stochastically larger" property (see Perlman (1969) or Robertson *et al.* (1988)), from which the monotonicity of its moments follows. The monotonicity of the power function of the likelihood ratio test is also well known (see Nomakuchi (1983) or Mukerjee *et al.* (1986)).

2. Theorems

Let X be a random vector in R^n. We assume that

(1) the $\theta \in R^n$ is a location parameter, that is,

$$X \sim F(x - \theta), \qquad x \in R^n,$$

where F is a distribution function;

(2) the parameter space C of θ is a closed convex cone in R^n;

Under these assumptions, we consider a least square estimate (LSE) $\hat{x}$ of θ, that is, the $\hat{x}$ is the orthogonal projection of $x \in R^n$ onto C which is defined as a unique point $\hat{x} \in C$ satisfying

$$|x - \hat{x}| = \min_{y \in C} |x - y|,$$

where $|\cdot|$ is a norm induced by an inner product $(\cdot, \cdot)$ in R^n. Note that maximum likelihood estimate (MLE) is equal to LSE in the estimation of the mean of multivariate normal distribution. Especially when the convex cone is given by the order restrictions, MLE is also equal to LSE in many statistical problems (see Barlow *et al.* (1972)).

Let L be a linear subspace contained in C. For any non zero vector $\lambda \in R^n$ and any $a \in L$, we define a ray such that

$$\theta(t) = t\lambda + a, \qquad t \geq 0.$$

We also prepare a parallel ray with $\theta(t)$ as follows

$$x(t) = t\lambda + b, \qquad t \geq 0,$$

where b is any vector in R^n. Let $\hat{x}(t) \in C$ be the orthogonal projection of $x(t)$ onto C. Setting a function as follows

$$\phi(t) = |\theta(t) - \hat{x}(t)|, \qquad t \geq 0,$$

we have the following theorem.

THEOREM 2.1. *$\phi(t)$ is a non-decreasing function in $t \geq 0$.*

PROOF. At first we assume that C is a closed convex polyhedral cone. According to the notations and the definitions in Nomakuchi (1983), the cone C can be decomposed into faces F_i, $i = 1, \ldots, m$, that is,

$$C = \sum_{i=1}^{m} F_i$$

where these faces are mutually disjoint, hence we use a summation sign instead of a union sign. Note that the definition of faces is slightly different from the one of Rockafellar (1970), in which the faces are closed, but ours are relatively open and so disjoint. Hence, there exists a unique face F_i which contains the projection $\hat{x}(t)$. Let M_i be the smallest linear subspace containing this F_i. It is well known that $\hat{x}(t)$ is obtained by projecting

$x(t)$ onto M_i (see Theorem 1 in Nomakuchi (1983) for details). Since this projection onto M_i is linear, let P_i be the corresponding projection matrix, that is,

$$\hat{x}(t) = P_i x(t) \quad \text{if} \quad \hat{x}(t) \in F_i.$$

Since $a \in L \subset M_i$, clearly $P_i a = a$. Setting $\hat{\theta}(t) = P_i \theta(t)$, we have

$$\begin{aligned}
\phi^2(t) &= |\theta(t) - \hat{x}(t)|^2 \\
&= |\theta(t) - \hat{\theta}(t) + \hat{\theta}(t) - \hat{x}(t)|^2 \\
&= |\theta(t) - \hat{\theta}(t)|^2 + |\hat{\theta}(t) - \hat{x}(t)|^2 \\
&= |t\lambda + a - P_i(t\lambda + a)|^2 + |P_i(t\lambda + a) - P_i(t\lambda + b)|^2 \\
&= t^2|\lambda - P_i\lambda|^2 + |a - P_i b|^2.
\end{aligned}$$

Hence, when the projection $\hat{x}(t)$ is in F_i, the function $\phi(t)$ is non-decreasing in t. On the other hand, the projection $x(t)$ onto a closed convex cone is continuous, therefore $\phi(t)$ is non-decreasing in $t \geq 0$.

Since any convex set can be approximated by a polyhedral one, we can take a sequence of polyhedral closed convex cones C_k such that

$$C_1 \supset C_2 \supset \cdots \supset C = \cap_{k=1}^{\infty} C_k.$$

Since the linear subspace L is contained in all C_k, the ray $\theta(t)$, $t \geq 0$, clearly satisfies the assumption of the theorem with C_k. Therefore, if $\hat{x}_k(t)$ be the projection of $x(t)$ onto C_k, the function

$$\phi_k(t) = |\theta(t) - \hat{x}_k(t)|, \qquad t \geq 0$$

is non-decreasing. Since the ϕ_k continuously converges to ϕ, we have a proof of the theorem.

Remark 2.1. The result is also valid in Hilbert space though we proved it in R^n.

Remark 2.2. We do not assume $\lambda \in C$. Hence the result holds true without it. It is easy to see that, when $\lambda \notin C$, the $\phi(t)$ goes non-decreasingly to ∞. On the other hand, it is also clear that, when $\lambda \in C^o$, the interior of C, the $\phi(t)$ approaches non-decreasingly to $|a - b|$ because $t\lambda + b \in C^o$ for sufficiently large t and its projection onto C is itself.

Remark 2.3. The assumption, $a \in L$, is essential. Without this one, the theorem does not hold. It is easy to construct a counterexample. Let us consider a following example in R^2;

$$C = \{(\theta_1, \theta_2)' : \theta_2 \geq \max(-\theta_1, 0)\}, \qquad \lambda = (1,0)', \quad a = (-1,1)', \ \ b = (-1,-1)'.$$

When $0 \leq t \leq 1$, clearly $\hat{x}(t) = 0$, hence the $\phi(t) = \sqrt{(t-1)^2 + 1}$ becomes decreasing in this interval. This also implies that, unless $a \in L$, the theorem does not necessarily hold true even when the entire of $\theta(t)$, $t \geq 0$, is contained in C.

Assuming $X_t \sim F(x - \theta(t))$, we have the following theorems.

THEOREM 2.2. *When t moves from 0 to ∞, $|\hat{X}_t - \theta(t)|$ becomes "stochastically larger."*

PROOF. We can set $X_t = X_0 + \theta(t)$, where $X_0 \sim F$. If we fix X_0 as a constant, we have from Theorem 2.1 that $\phi(t) = |\hat{X}_t - \theta(t)|$ is non-decreasing in $t \geq 0$, from which the proof follows.

We have easily the following theorem from the "stochastically larger" property of $|\hat{X}_t - \theta(t)|$.

THEOREM 2.3. *Assume that $r > 0$ and $E|X|^r < \infty$. Then, the absolute moment $E|\hat{X}_t - \theta(t)|^r$ is non-decreasing in $t \geq 0$.*

Combining this with Remark 2.2, we have the following corollary.

COROLLARY 2.1. *Let $\lambda \notin L$. If t moves from $-\infty$ to ∞ then $E|\hat{X}_t - \theta(t)|^r$ is non-increasing in $t \leq 0$ and non-decreasing in $t \geq 0$.*

COROLLARY 2.2. *When $X \sim F(x - \theta)$, then the minimum value of $E|\hat{X} - \theta|^r$ is attained at $\theta = 0$.*

Remark 2.4. Let $\psi(\theta) = E(|\hat{X} - \theta|^r)$. When $\lambda \in C^o$, $\psi(\theta(t))$ goes to the r-th moment $E(|X - \theta|^r)$, which does not depend on the θ or λ. On the other hand, $\psi(\theta(t))$ goes to ∞ when $\lambda \in C^c$. Hence, the function $\psi(\theta)$ is not convex in θ. But, we have a conjecture that the level sets

$$D_a = \{\theta \in R^n : \psi(\theta) \leq a\}, \qquad a > 0,$$

would be convex. We have not obtained the proof.

Example 2.1. Normal distribution.

If we assume multivariate normal distribution with mean 0 and variance matrix I as the underlying distribution F, the distribution of $|\hat{X}|^2$ with $\theta = 0$ becomes a mixture of χ^2 distributions, that is,

$$P(|\hat{X}|^2 \leq x) = \sum_{i=0}^{n} p_i P(\chi_i^2 \leq x),$$

where the weight p_i is a probability of the event that i is equal to the dimension of F_j such that $\hat{X} \in F_j$, and χ_i^2 is a random variable with χ^2 distribution with i degrees of freedom. Hence, we have from Corollary 2.2

$$\mathrm{Min}_\theta E(|\hat{X} - \theta|^r) = \sum_{i=1}^{n} \frac{p_i \Gamma((r+i)/2) 2^{r/2}}{\Gamma(i/2)}.$$

The minimum of the mean square error of $\hat{X}$ as an estimate of θ is especially given when $\theta = 0$, that is,

$$\mathrm{Min}_\theta E(|\hat{X} - \theta|^2) = \sum_{i=1}^{n} i p_i,$$

which is used by Anraku (1999) to propose ORIC in the simple order restriction problem. These results, however, are applicable in a much more general setting.

Acknowledgements

The author thanks two referees for their valuable comments and advices.

References

Anraku, K. (1999). An information criterion for parameters under a simple order restriction, *Biometrika*, **86**, 141–152.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*, Wiley, London.

Mukerjee, H., Robertson, T. and Wright, F. T. (1986). A probability inequality for elliptically contoured densities with applications in order restricted inference, *Ann. Statist.*, **14**, 1548–1554.

Nomakuchi, K. (1983). The likelihood ratio test of normal mean with hypothesis determined by a convex polyhedral cone and the monotonicity of its power function, *Mem. Fac. Sci. Kyushu Univ. Ser. A*, **37**, 195–209.

Perlman, M. D. (1969). One-sided problems in multivariate analysis, *Ann. Math. Statist.*, **40**, 549–567.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, Chichester.

Rockafellar, R. T. (1970). *Convex Analysis*, Princeton University Press, New Jersey.