KYBERNETIKA — VOLUME 31 (1995), NUMBER 4, PAGES 331 – 335

MINIMUM ENTROPY OF ERROR ESTIMATE FOR MULTI–DIMENSIONAL PARAMETER AND FINITE–STATE–SPACE OBSERVATIONS[1]

Antonín Otáhal

The minimum entropy of error estimate (MEEE) is studied for a finite mixture of probability densities on a finite-dimensional Euclidean space. It is proved that the MEEE coincides with the conditional expectation in case all the densities in the mixture are isotropic and unimodal; further a counter-example is given which shows that the result cannot be generalized for symmetric non-isotropic densities.

1. INTRODUCTION

The minimum entropy of error principle was introduced by Weidemann and Stear [6, 7] and the idea has been further pursued by Janžura, Koski and Otáhal [2, 3]. The principle consists in that one random variable (parameter) is estimated by means of another random variable (observation), so that the (differential) entropy of the estimation error is minimized. The principle is intuitively plausible, though its application is, due to problems with differential entropy, somewhat technically involved – cf. also Ikeda [1], Otáhal [4], Vajda [5],(10.20).

One of the main results of [2] states that the minimum entropy of error estimate (MEEE) is the same as the conditional expectation in case the state space of the observation is finite, the parameter space is (a subset of) the real line and all the conditional densities (of the parameter given the observation value) are symmetric and unimodal. The present paper studies a possibility of generalizing this result for a multi-dimensional parameter.

2. BASIC NOTIONS

For convenience we define the real function Φ on $[0, +\infty)$ as

$$\Phi(t) = -t \cdot \log(t)$$

with the usual convention $\Phi(0) = 0$.

[1]Supported by the Grant Agency of the Czech Republic under Grant No. 201/94/0321.

Suppose there is given an m-dimensional random vector U whose distribution is absolutely continuous with respect to the m-dimensional Lebesgue measure with the corresponding density f_U. Further there is given another random variable Z (observation, or data) which has a finite range of possible values $\boldsymbol{Z} = \{z_1, \ldots, z_n\}$. For $j = 1, \ldots, n$ we denote $f_j(u) = f_{U|Z}(u|z_j)$ and $\alpha_j = P\{Z = z_j\}$. The MEE estimate is defined as the mapping G from $\boldsymbol{Z}$ to the m-dimensional Euclidean space R^m minimizing the entropy of the error $e = U - G(Z)$. In other words, denoting $t_j = G(z_j)$ we can express the problem of finding the MEEE as the problem of minimizing, for $\boldsymbol{t} = (t_1, \ldots, t_n)$ and $f_{\boldsymbol{t}}(x) = \sum_{j=1}^n \alpha_j f_j(x + t_j)$, the value of

$$H(\boldsymbol{t}) = \int_{R^m} \Phi(f_{\boldsymbol{t}}(x))\,\mathrm{d}x$$

with respect to the shifts $t_1, \ldots, t_n$. We have to suppose $\int \Phi(f_j(x))\,\mathrm{d}x < +\infty$. In [2] we can find details of this construction, as well as further results: H is a continuous and bounded function of $\boldsymbol{t}$ which takes on its minimum on $(R^m)^n$. For further reference we point out the following result:

2.1. Theorem. If $m = 1$ and all the densities $f_1, \ldots, f_n$ are symmetric unimodal, then the minimum of H takes place at $\boldsymbol{t} = \boldsymbol{0}$.

Proof. Cf. [2], Proposition 3.12. □

3. ISOTROPIC UNIMODAL DENSITIES

If we want to generalize Theorem 2.1 to the case $m > 1$ we have to decide which generalization of symmetry is the 'proper' one to ensure that a similar result will hold. We will show that under an assumption of isotropic (rotation invariant) densities the result can be generalized and, by means of a simple example, we will establish that a more general notion of symmetric (i. e. even) densities is not sufficient.

For $x \in R^m$ we denote by $|x|$ the usual Euclidean norm of x. A real function g on R^m is *isotropic* if the value of $g(x)$ depends only on $|x|$. An isotropic function g is *unimodal* if $g(x)$ is a non-increasing function of $|x|$.

Before coming to the main result we have to go through an auxiliary technical one.

3.1. Lemma. Let V be a proper linear subspace of R^d and W be its orthogonal complement. Suppose that there is given a real function g on R^d such that

$$\int_V g(v + w)\,\mathrm{d}v \geq 0$$

for every $w \in W$.

Then $\int_{R^d} g(x)\,\mathrm{d}x \geq 0$.

Proof is an immediate consequence of the Fubini theorem. □

3.2. Theorem. Let, in the notations of Part 2, the conditional densities $f_1, \dots, f_n$ be isotropic and unimodal. Then the minimum of H takes place at $\boldsymbol{t} = \mathbf{0}$, in other words, the MEEE of U by means of Z is the same as the conditional expectation $E\{U|Z\}$.

P r o o f will be carried out by induction in the dimension m.

1. For $m = 1$ the assertion is the same as that of Theorem 2.1, since on the real line the notions of isotropy and symmetry are the same.
2. We will assume that the assertion holds for the dimension equal to $m - 1$ and will prove it for m.

Fix an $(m-1)$-dimensional subspace A in R^m, denote by π the orthogonal projection onto A and put $\tau_j = \pi(t_j)$ for $j = 1, \dots, n$.

Now we first prove the inequality

$$H(\tau_1, \dots, \tau_n) \geq H(0, \dots, 0). \tag{1}$$

In fact, denote, for a hyperplane B parallel to A and for $j = 1, \dots, n$, by $\tilde{f}_j$ the restriction of f_j to B. Since obviously all $\tilde{f}_j$'s are isotropic and unimodal in B (with $\tau(\mathbf{0})$ playing the role of the origin), by the induction assumption we conclude that

$$\int_B \Phi\left(\sum_j \alpha_j \tilde{f}_j(y + \tau_j)\right) \mathrm{d}y \geq \int_B \Phi\left(\sum_j \alpha_j \tilde{f}_j(y)\right) \mathrm{d}y$$

and (1) follows by Lemma 3.1.

In the second step we prove

$$H(t_1, \dots, t_n) \geq H(\tau_1, \dots, \tau_n). \tag{2}$$

For this we choose a unit vector a (of either orientation) in R^m orthogonal to A and define, for $j = 1, \dots, n$, the real r_j by the relation $t_j = \tau_j + r_j a$. For a straight line p orthogonal to A and $j = 1, \dots, n$ we consider the restriction $\hat{f}_j$ of f_j to p, which is obviously a symmetric unimodal function on p. Hence by Theorem 2.1 the inequality

$$\int_p \Phi\left(\sum_j \alpha_j \hat{f}_j(\xi + r_j)\right) \mathrm{d}\xi \geq \int_p \Phi\left(\sum_j \alpha_j \hat{f}_j(\xi)\right) \mathrm{d}\xi$$

is true and (2) again follows by Lemma 3.1.

Putting together (1) and (2) we complete the proof. □

4. SYMMETRIC DENSITIES

In this section we will present an example of even unimodal densities for which the MEEE differs from the conditional expectation.

Let us, for α, β, $\gamma \geq 0$, denote

$$\Psi(\alpha, \beta, \gamma) = \Phi(\alpha + \beta + \gamma) - \Phi(\alpha + \beta) - \Phi(\alpha + \gamma) - \Phi(\beta + \gamma) + \Phi(\alpha) + \Phi(\beta) + \Phi(\gamma).$$

4.1. Lemma. For all α, β, $\gamma \geq 0$ it holds

$$\Psi(\alpha, \beta, \gamma) \geq 0.$$

Proof. For fixed β and γ we define $\psi(\alpha) = \Psi(\alpha, \beta, \gamma)$. Then $\psi(0) = 0$ and the derivative of ψ is given as

$$\psi'(\alpha) = \log \frac{(\alpha + \beta)(\alpha + \gamma)}{\alpha(\alpha + \beta + \gamma)}$$

which is obviously non-negative. □

4.2. Example. Suppose $m = 2$, i.e. the example takes place in the plane; we write λ for the two-dimensional Lebesgue measure. By K we denote a rectangle whose center is at the coordinate origin, whose width is w and whose length is ℓ; we take $\ell \gg w$. The orientation of K is such that its vertices are given by coordinates $(\pm\ell/2, \pm w/2)$. For $\eta > 0$ and the rotation ρ which rotates by the angle $2\pi/3$ we take the sets

$$\begin{aligned} K_1(\eta) &= K + (0, \eta), \\ K_2(\eta) &= \rho(K + (0, \eta)), \\ K_3(\eta) &= \rho^2(K + (0, \eta)). \end{aligned}$$

For positive reals α, β, γ such that $(\alpha + \beta + \gamma) \cdot \lambda(K) = 1$ we put

$$h(\eta) = \int_{R^2} (\alpha \cdot 1_{K_1(\eta)} + \beta \cdot 1_{K_2(\eta)} + \gamma \cdot 1_{K_2(\eta)}) \, \mathrm{d}\lambda,$$

where 1_M stands for the indicator function of a set M.

Let us for the sake of brevity write

$$\begin{aligned} \psi_1 &= \psi_1(\alpha, \beta, \gamma) = \Phi(\alpha) + \Phi(\beta) + \Phi(\gamma), \\ \psi_2 &= \psi_2(\alpha, \beta, \gamma) = \Phi(\alpha + \beta) + \Phi(\alpha + \gamma) + \Phi(\beta + \gamma), \\ \psi_3 &= \psi_3(\alpha, \beta, \gamma) = \Phi(\alpha + \beta + \gamma). \end{aligned}$$

Then we can calculate for $\eta \in [0, w/6]$

$$\begin{aligned} h(\eta) \quad = \quad & \Big(\tfrac{6}{4\sqrt{3}} w^2 - 6\sqrt{3}\eta^2) \, \psi_3 + \\ & +(\tfrac{2}{4\sqrt{3}} w^2 + 6\sqrt{3}\eta^2) \, \psi_2 + \\ & +(\lambda(K) - \tfrac{10}{4\sqrt{3}} w^2 - 6\sqrt{3}\eta^2\Big) \, \psi_1 \end{aligned}$$

and for $\eta \in [w/6, w/2]$

$$\begin{aligned} h(\eta) \quad = \quad & \Big(\tfrac{9}{4\sqrt{3}} w^2 - 3\sqrt{3}(w\eta - \eta^2)\Big) \, \psi_3 + \\ & + \Big(-\tfrac{1}{4\sqrt{3}} w^2 + 3\sqrt{3}(w\eta - \eta^2)\Big) \, \psi_2 + \\ & + \Big(\lambda(K) - \tfrac{7}{4\sqrt{3}} w^2 - 3\sqrt{3}(w\eta - \eta^2)\Big) \, \psi_1. \end{aligned}$$

It is easy to see that the function h of η has a local maximum at 0, because Lemma 4.1 ensures that on the interval $[0, w/6]$ the function h is decreasing. Since $\eta = 0$ corresponds to zero shifts of symmetric sets $K_1(0)$, $K_2(0)$ and $K_3(0)$, we have really proved that for conditional densities given by their indicator functions the MEEE differs from the conditional expectation.

The example is based on the fact that minimizing the differential entropy of a mixture of shifted indicator functions of the sets K_1, K_2, K_3 we seek for a small volume of the intersection $K_1 \cap K_2 \cap K_3$ while maximizing the pairwise intersections $K_1 \cap K_2$, $K_2 \cap K_3$ and $K_1 \cap K_3$.

Just for the completeness of the analysis of the example let us mention that for η large enough the sets $K_1(\eta)$, $K_2(\eta)$ and $K_3(\eta)$ are pairwise disjoint, $h(\eta) = \lambda(K)\,\psi_1$ and $h(\eta) > h(0)$. That is, the local maximum of h at 0 is not global. □

This example seems to indicate (the exact calculations would be rather complicated) that even for Gaussian densities, if they are suitably chosen (i. e. if their level ellipses 'copy' the shape of the sets considered in Example 4.2) the MEEE will not coincide with the conditional expectation, which is a rather surprising statement.

(Received August 25, 1994.)

REFERENCES

[1] S. Ikeda: Continuity and characterization of Shannon–Wiener information measure for continuous probability distributions. Ann. Inst. Statist. Math. *11* (1959), 131–144.

[2] M. Janžura, T. Koski and A. Otáhal: Minimum entropy of error principle in estimation. Inform. Sci., to appear.

[3] M. Janžura, T. Koski and A. Otáhal: Minimum entropy of error principle in estimation: a short survey. In: Proceedings of 6th Joint Swedish–Russian Internat. Workshop on Inform. Theory, Moelle 1993, pp. 429–431.

[4] A. Otáhal: Finiteness and Continuity of Differential Entropy. Asymptotic Statistics. In: Proceedings of 5th Prague Symposium on Asymptotic Statistics (P. Mandl and M. Hušková, eds.), Physica–Verlag, Heidelberg 1993, pp. 415–419.

[5] I. Vajda: Theory of Statistical Inference and Information. Kluwer, Dodrecht – Boston – London 1989.

[6] H. L. Weidemann and E. B. Stear: Entropy analysis of parameter estimation. Inform. and Control *14* (1969), 493–506.

[7] H. L. Weidemann and E. B. Stear: Entropy analysis of estimation systems. IEEE Trans. Inform. Theory *16* (1970), 264–270.

RNDr. Antonín Otáhal, CSc. Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic), Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.