MATHEMATICS

A. M. KAGAN

1963

SovietRxiv

Abstract

Full Text

MATHEMATICS

A. M. KAGAN

ON THE THEORY OF FISHER INFORMATION QUANTITY

(Presented by Academician V. I. Smirnov, 5 February 1963)

In this note an analogue of Fisher' s information quantity ([1]) is constructed for families specified, generally speaking, by densities not differentiable with respect to the parameter. The corresponding generalization of the Rao–Cramér inequality ([1,2]) is also given.

§ 1. W-divergence between two distributions

Let, on an abstract space X of elements x with a distinguished σ-algebra of subsets $\mathfrak{A}$, probability measures P_1 and P_2 be given. One may always assume that they are given by densities $p_1(x) = dP_1/d\mu$, $p_2(x) = dP_2/d\mu$ with respect to some measure μ (as μ one may take, for example, $P_1 + P_2$).

Define the W-divergence between P_1 and P_2 as follows:

$$W(P_1; P_2) = \int_{\{p_1(x)>0\}} \left[1 - \frac{p_2(x)}{p_1(x)}\right]^2 p_1(x)\, d\mu(x) \tag{1}$$

and, analogously,

$$W(P_2; P_1) = \int_{\{p_2(x)>0\}} \left[1 - \frac{p_1(x)}{p_2(x)}\right]^2 p_1(x)\, d\mu(x). \tag{2}$$

Generally speaking, $W(P_1; P_2) \neq W(P_2; P_1)$ (note that the Kullback-Leibler numbers ([3]) have this same property).

The introduced W-divergence has the following properties:

1. $W(P_1; P_2) = 0$ only when $P_1 = P_2$.
2. If $W(P_1; P_2^{(n)}) \to 0$ as $n \to \infty$, then $\mathrm{Var}\,|P_1 - P_2^{(n)}| \to 0$. As examples show, the converse assertion is false.
3. Let $\mathfrak{B}$ be a σ-subalgebra of the algebra $\mathfrak{A}$; let $\widetilde{P}_1$ and $\widetilde{P}_2$ be the restrictions of the measures P_1 and P_2, respectively, to the σ-algebra $\mathfrak{B}$. Then

$W(\widetilde{P}_1; \widetilde{P}_2) \leq W(P_1; P_2)$, with equality if and only if $\mathfrak{B}$ is a sufficient subalgebra for the family $(P_1; P_2)$ ([4]).

4. Suppose that P_1 and P_2 are mutually absolutely continuous. Let $X^n = X \times \cdots \times X$, and let $P_i^{(n)}$ be the direct product of the measure P_i with itself n times, $i = 1, 2$. Then

$$W(P_1^{(n)}; P_2^{(n)}) \geq n\, W(P_1; P_2).$$

§ 2. Parametric families

Let $P = \{p(x|\theta);\ \theta \in \Theta\}$ be a family of distributions on $\{X, \mathfrak{A}\}$, specified with respect to some measure μ by densities $p(x(\theta))$ depending on the parameter θ. The parameter set Θ is assumed to be a finite or infinite interval of the line.

Put

$$W(\theta; \theta + \Delta\theta) = \frac{1}{(\Delta\theta)^2} \int_{\{p(x|\theta)>0\}} \left[1 - \frac{p(x|\theta+\Delta\theta)^2}{p(x|\theta)}\right] p(x|\theta)\, d\mu, \tag{3}$$

$$W(\theta) = \lim_{\Delta\theta \to 0} \inf W(\theta; \theta + \Delta\theta). \tag{4}$$

$W(\theta)$ is an analogue of Fisher' s information quantity $I(\theta)$ in those cases where it does not exist. For sufficiently smooth families $W(\theta) = I(\theta)$. If the family P is assumed homogeneous (i.e., all distributions belonging to it are mutually absolutely continuous), then the following properties of $W(\theta)$ can be established:

1. If

$$W^{(n)}(\theta; \theta + \Delta\theta) = \frac{1}{(\Delta\theta)^2} \int_X \cdots \int_X \left[1 - \frac{p(x_1|\theta+\Delta\theta) \cdots p(x_n|\theta+\Delta\theta)^2}{p(x_1|\theta) \cdots p(x_n|\theta)}\right] \times$$

$$\times\, p(x_1|\theta) \cdots p(x_n|\theta)\, d\mu(x_1), \ldots, d\mu(x_n), \tag{5}$$

$$W^{(n)}(\theta) = \lim_{\Delta\theta \to 0} \inf W^{(n)}(\theta; \theta + \Delta\theta), \tag{6}$$

then

$$W^{(n)}(\theta) = nW(\theta). \tag{7}$$

2. Let $\varphi(x)$ be an unbiased estimate of the parameter θ. Then the following analogue of the Rao-Cramér inequality holds:

$$E(\varphi(x) - \theta)^2 \geq \frac{1}{W(\theta)}. \tag{8}$$

§ 3. Suppose now that the parameter set Θ is an s-dimensional parallelepiped; $\theta = (\theta_1, \dots, \theta_s)$. Put

$$W_{ij}(\theta) = \lim_{|\Delta\theta| \to 0} \inf \frac{1}{\Delta\theta_i \Delta\theta_j} \int_X \left[1 - \frac{p(x|\theta + \Delta\theta_i)}{p(x|\theta)}\right] \times$$

$$\times \left[1 - \frac{p(x|\theta + \Delta\theta_j)}{p(x|\theta)}\right] p(x|\theta)\, d\mu(x), \tag{9}$$

$$W(\theta) = \|W_{ij}(\theta)\|_{i,j=1,\dots,s}. \tag{10}$$

If $B(\theta)$ is the correlation matrix of an unbiased estimate $\varphi(x)$ of the parameter θ, and $W^{-1}(\theta)$ exists, then, in the well-known sense,

$$B(\theta) - W^{-1}(\theta) \geq 0. \tag{11}$$

The proof is carried out by the method of [2].

4. Let Θ be an open subset of a normed space, and let the unbiased estimate $\varphi(x)$ of the parameter θ be Bochner-integrable [5] with respect to the measures $p(x|\theta)\, d\mu$.

$$W(\theta) = \lim_{\|\Delta\theta\| \to 0} \inf \frac{1}{\|\Delta\theta\|^2} \int_X \left[1 - \frac{p(x|\theta + \Delta\theta)^2}{p(x|\theta)}\right] p(x|\theta)\, d\mu. \tag{12}$$

Then

$$E\|\varphi - \theta\|^2 \geq \frac{1}{W(\theta)}. \tag{13}$$

Received
2 II 1963

References Cited

1. H. Cramér, *Mathematical Methods of Statistics*, IL, 1948.

2. O. V. Shalaevsky, Theory of Probability and Its Applications, 6, 3 (1961).

3. S. Kullback, R. Leibler, *Ann. Math. Statistics*, 22, 1 (1951).

4. P. Halmos, L. Savage, *Ann. Math. Statistics*, 20, 1 (1949).

5. E. Hille, *Functional Analysis and Semigroups*, IL, 1951.

Note: Figure translations are in progress. See original paper for figures.