MAXIMUM LIKELIHOOD ESTIMATION FOR DENSITY AS AN INFINITE-DIMENSIONAL GAUSSIAN SHIFT

S. G. Bobkov

A new approach is suggested for the nonparametric estimation of the unknown distribution density under the assumption of the bounded variation of the true density. As an estimator there occurs a statistic based on the application of the maximum likelihood method to the estimation of an infinite-dimensional shift of a Gaussian process with a known correlation function. The quality of the obtained estimate is investigated.

Let $x_1, \ldots, x_n$ be a sample, extracted from a distribution F, continuous on the real axis **R**. It is known that for the sample distribution F_n the "normalized" random functions $\xi_n(t) = n^{1/2}(F_n(t) - F(t))$ converge in distribution to the Brownian bridge W^0, more exactly, to the distribution $\mathscr{L}(\xi)$ of the Gaussian random process $\xi(t) = W^0(F(t))$. Therefore, for large n, the function ξ_n can be considered as a one-element sample from $\mathscr{L}(\xi)$ or, equivalently, the sample function F_n can be considered as a one-element sample from the distribution of the process $n^{-1/2}\xi + F$, and, consequently, with respect to one "observation" F_n the unknown $F = E\{n^{-1/2}\xi + F\}$ can be estimated as a shift of a centralized Gaussian process. A somewhat informal (since the correlation function depends on F) application of the maximum likelihood method to the estimation of an infinite-dimensional shift of a Gaussian process with a known correlation function (see [1]) leads to the following definition.

Let V be an arbitrary family of densities on **R**. From a sample $x_1, \ldots, x_n \in \mathbf{R}$ we define $\hat{p}_n \in V$ by the equality

$$\hat{p}_n = \arg\max_{p \in V} \left\{ \frac{p(x_1) + \ldots + p(x_n)}{n} - \frac{1}{2}\int_{-\infty}^{+\infty} p(t)^2\, dt \right\}. \tag{1}$$

Definition (1) has to be understood in the following sense: if a maximum point in the right-hand side of (1) exists, then the density $\hat{p}_n \in V$ is one (any) of these points. In the next statement V is such that $\hat{p}_n$ is defined for almost all values of the sample $x_1, \ldots, x_n$ from the distribution F, whose density is $p \in V$.

THEOREM. We assume that each density from V is a function of bounded variation and, moreover,

$$C = \sup_{q \in V} \operatorname{var} q < \infty. \tag{2}$$

Then a. s.

$$\int_{-\infty} |\hat{p}_n(t) - p(t)|^2\, dt \leq 4C \max_{|t|} |F_n(t) - F(t)|. \tag{3}$$

Remarks. 1. The distribution of the random variable in the right-hand side of (3) does not depend on F and, therefore, for the true p with a prescribed significance level one can construct a "confidence" ball in $L^2(\mathbf{R}, dx)$ with center at $\hat{p}_n$ and radius of order const$\cdot n^{-1/4}$. 2. By virtue of (2) we have $\int p^2(t)\, dt < \infty$ for all $p \in V$ and, in addition, the right-hand side in (1) is bounded.

The proof of the theorem is based on the lemma given below. Let $V \subset H \subset E$, E being a normed linear space with norm $\|\cdot\|_E$, H is a linear subspace of E, and there is defined a bilinear form $\langle \eta, x\rangle$ on $H \times E$ with the following properties: 1) $\forall \eta_1, \eta_2 \in H$ we have $\langle \eta_1, \eta_2\rangle = \langle \eta_2, \eta_1\rangle$, 2) $\forall \eta \in H$ we have $<\eta, \eta> \geq 0$, 3) $\forall \eta \in V$ we have $|<\eta, x>| \leq C \cdot \|x\|_E$ for any $x \in E$.

Translated from Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta im. V. A. Steklova Akademii Nauk SSSR, Vol. 177, pp. 6-7, 1989.

0090-4104/92/6101-1825$12.50

Properties 1) - 2) define on H $\times$ H an inner product (possibly without the separation property), which generates in H a Euclidean distance $\|\eta - \theta\|_H$. As in (1), we define

$$\hat{\theta}(x) = \arg\max_{\eta \in V} \langle \eta, x \rangle - \frac{1}{2}\|\eta\|_H^2, \tag{4}$$

under the condition that the maximum in (4) exists and $\hat{\theta}(x)$ is any of its maximum points.

LEMMA. For all $\theta \in V$ and all $x \in E$, for which $\hat{\theta}(x)$ is defined, we have $\|\hat{\theta}(x) - \theta\|_H^2 \leqslant 4C\|x - \theta\|_E$.

In order to apply the lemma to the theorem we have to take E to be the space of finite measures on the line (functions of bounded variation) with the uniform norm $\|x\|_E = \sup_t |x(t)|$, H , consists of those $\eta \in E$ for which there exists a density η' of bounded variation, and

$$\langle \eta, x \rangle - \int_{-\infty}^{+\infty} \eta'(t)\, dx(t) = -\int_{-\infty}^{+\infty} x(t)\, d\eta'(t).$$

Proof of the Lemma. $\|\hat{\theta}(x) - \theta\|_H \geq \varepsilon \Rightarrow \exists \eta \in V$, $\|\eta - \theta\|_H \geq \varepsilon$ such that $\langle \eta, x \rangle - \frac{1}{2}\|\eta\|_H^2 = \sup_{\eta \in V}\{\langle \eta, x \rangle - \frac{1}{2}\|\eta\|_H^2\} \geqslant \langle \theta, x \rangle - \frac{1}{2}\|\theta\|_H^2 \Rightarrow \exists \eta \in V$, $\|\eta - \theta\|_H \geqslant \varepsilon$ such that $\langle \eta - \theta, x \rangle \geqslant \frac{1}{2}(\|\eta\|_H^2 - \|\theta\|_H^2) \Rightarrow \exists \eta \in V$, $\|\eta - \theta\|_H \geqslant \varepsilon$ such that $\langle \eta - \theta, x - \theta \rangle \geqslant \frac{1}{2}(\|\eta\|_H^2 - \|\theta\|_H^2) - \langle \eta - \theta, \theta \rangle = \frac{1}{2}\|\eta - \theta\|_H^2 \geqslant \frac{\varepsilon^2}{2}$. But $\langle \eta - \theta, x - \theta \rangle \leq 2C\|x - \theta\|_E$ on the strength of 3). Therefore $\|\hat{\theta}(x) - \theta\|_H \geq \varepsilon \Rightarrow 2C\|x - \theta\|_E \geq \varepsilon^2/2$. It remains to set $\varepsilon = \|\hat{\theta}(x) - \eta\|_H$.

We give an example of a nonparametric family. Assume that V_C consists of densities, equal to zero on $(-\infty, 0]$, nonincreasing and left-continuous on $(0, +\infty)$, and the left limit at zero is $\leq C$. In this case the density $\hat{p}_n$ has the form: $\hat{p}_n = \alpha_1$ on $(0, x_1]$, $\hat{p}_n = \alpha_2$ on $(x_1', x_2]$,..., $\hat{p}_n = \alpha_n$ on (x_{n-1}', x_n'), $\hat{p}_n = 0$ on $(x_n', +\infty)$, where $x_1' < \ldots < x_n'$ is the variational series, constructed from the sample, while the constants α_k realize the maximum $(\Delta_k = x_k' - x_{k-1}',\ \Delta_1 = x_1')$

$$\max_{\substack{C \geqslant \alpha_1 \geqslant \ldots \geqslant \alpha_n \geqslant 0 \\ \alpha_1\Delta_1 + \ldots + \alpha_n\Delta_n = 1}} \left\{\frac{\alpha_1 + \ldots + \alpha_n}{n} - \frac{1}{2}(\alpha_1^2\Delta_1 + \ldots + \alpha_n^2\Delta_n)\right\}.$$

LITERATURE CITED

1. B. S. Tsirel'son, "A geometric approach to maximum likelihood estimation for an infinite-dimensional Gaussian location. I," Teor. Veroyatnost. Primenen., 27, No. 2, 388-395 (1982).