



Soviet-era science, translated into English

N. N. Chentsov

1962

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196201.26832>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

N. N. Chentsov

Estimation of an Unknown Distribution Density from Observations

(Presented by Academician M. V. Keldysh, 21 V 1962)

As is known (see ⁽¹⁾), the deviation of a histogram from the unknown graph of the density decreases, roughly speaking, as $N^{-1/3}$, where N is the number of independent observations from which the histogram is constructed. In this paper a class of methods for estimating an unknown density is considered, generalizing the histogram method and capable of giving greater accuracy. These methods proved useful to the author in solving certain practical problems (see ⁽²⁾).

Let ξ be a random variable (not necessarily numerical), $\mu(dx)$ a measure in the space X of values of ξ , a point function; $p(x) = dP/d\mu$ the sought density of the probability distribution $P(\cdot)$ of the variable ξ with respect to the measure μ .

Let a weight $r(x)$ be given on X , and let the scalar product

$$(\varphi, \psi) = \int_X \varphi(x)\psi(x) \mu(dx) \tag{1}$$

define a Hilbert space $L_2(r)$, with $p(\cdot) \in L_2(r)$.

Consider an arbitrary n -dimensional subspace E_n with orthonormalized basis $\{\varphi_{kn}(x)\}$, $k = 1, \dots, n$. Denote

$$\pi_n(x) = \sum_{k=1}^n a_{kn} \varphi_{kn}(x) = \sum_{k=1}^n (\varphi_{kn}, p) \varphi_{kn}(x)$$

the mean-square approximation of $p(x)$, i.e., the projection of $p(\cdot)$ onto E_n . The coefficients a_{kn} may be interpreted as

$$a_{kn} = (\varphi_{kn}, p) = \int \varphi_{kn}(x) r(x) dP = M\alpha_k(\xi), \tag{2}$$

where $\alpha_k(\xi) = \varphi_{kn}(\xi)r(\xi)$. Let $\xi^{(1)}, \dots, \xi^{(N)}$ be independent observations of the variable ξ . Form the mean

$$a_k^* = \frac{1}{N} \sum_{i=1}^N \alpha_k(\xi^{(i)}),$$

and also the polynomial

$$\pi_{nN}^*(x) = \sum_{k=1}^n a_k^* \varphi_{kn}(x) = \sum_{k=1}^n \left(\varphi_{kn}, \frac{d\Xi_N}{d\mu} \right) \varphi_{kn}(x), \quad (3)$$

i.e., the projection onto E_N of the “derivative” $d\Xi_N/d\mu$ with respect to the measure $\mu(dx)$ of the empirical discrete measure $\Xi_N(dx)$ with atoms of weight $1/N$ at the observed points $\xi^{(i)}$.

By the Pythagorean theorem,

$$\|\pi^*(x) - p(x)\|^2 = \|\pi_n(x) - p(x)\|^2 + \sum_{k=1}^n [a_k^* - a_{kn}]^2. \quad (4)$$

Thus, by choosing a subspace E_n that approximates $p(x)$ sufficiently well and then making a sufficiently large number N of observations,

we can obtain, by formula (3), an approximation to $p(x)$ with an arbitrarily small random error in $L_2(r)$.

Although the density $p(x)$ is unknown, we may have information that $p(x)$ belongs to a set Π of functions with a known rate of approximation (for example, by trigonometric polynomials). Suppose N independent observations of ξ have been made. What dimension n should be chosen in order to use them in the best way for estimating $p(x)$? The rate of approximation of $p(x)$ by a given sequence of subspaces E_n is assumed known. We measure the error $\rho(\|\pi^* - p\|)$ in the approximation of $p(x)$ by the random polynomial $\pi^*(x)$ either by the quantity $\sqrt{M}\|\pi^* - p\|^2$, or by the confidence estimate $Q_\delta(\|\pi^*(x) - p(x)\|)$ —the quantile of order $1 - \delta$ for the corresponding sufficiently small δ .

Theorem 1a. If it is known that $p(\cdot) \in L_2(r)$ and

$$\|\pi_n(x) - p(x)\|^2 \leq Bg(Cn), \quad (5a)$$

and the sequence E_n is such that

$$\int \left\{ \frac{1}{n} \sum_{k=1}^n |\varphi_{kn}(x)|^2 \right\} [r(x)]^2 p(x) \mu(dx) \leq H, \quad (6a)$$

where $g(n) \downarrow 0$; B, C, H are positive constants, then n can be chosen as \max (roughly speaking, $n_N \asymp \Gamma(N)$) so that

$$\rho(\|\pi^* - p\|) = O\left(\sqrt{N^{-1}\Gamma(N)}\right),$$

where $\Gamma(y/g(y)) \equiv y$.

Theorem 1b. If it is known that $p(\cdot) \in L_2(r)$ and

$$\|\pi_n(x) - p(x)\|^2 \geq bg(cn), \quad (5b)$$

$$\int \left\{ \frac{1}{n} \sum_{k=1}^n |\varphi_{kn}(x)|^2 \right\} [r(x)]^2 p(x) \mu(dx) \geq h, \quad (6b)$$

where b, c, h are some positive constants, and, when the deviation is measured by a quantile, also that for sufficiently small δ

$$Q_\delta \left(\sum_{k=1}^n [a_k^* - a_{kn}]^2 \right) \geq lM \left(\sum_{k=1}^n [a_k^* - a_{kn}]^2 \right),$$

then, for any choice of n ,

$$\rho(\|\pi^* - p\|) = \Omega(\sqrt{N^{-1}\Gamma(N)}).$$

Thus, if the conditions of the theorems are satisfied, the norm of the error $\|\pi^* - p\|$ in the best case has order $\sqrt{N^{-1}\Gamma(N)}$. To prove the theorems, one estimates from above and from below the minimum over n of the expectation or quantile of the sum (4). The first term decreases as $g(n)$, the second term behaves as n/N , and their sum will be nearly minimal when the terms are equal. Both of these theorems can be extended to the case of a variable weight $r_n(x)$.

Corollary 1. Let $\pi_{nN}^*(x)$ be a histogram of the random variable ξ , $a \leq \xi \leq b$, with distribution density $p(x)$, constructed from N independent observations. The L_2 -deviation of $\pi^*(x)$ from $p(x)$ in the best case (in particular, for $n \sim \sqrt[3]{N}$ grouping intervals of equal length) has, in probability, order $N^{-1/3}$. (Here $p'(x)$ is assumed to be continuous and not identically equal to zero.)

Corollary 2. The L_2 -deviation of the histogram $\pi_{nN}^*(\vec{x})$ of a vector random variable $\vec{\xi} = \{\xi_m\}$, $a_m \leq \xi_m \leq b_m$; $m = 1, \dots, s$, in the best case has, in probability, order

$$N^{-\frac{1}{s+2}}.$$

Corollary 3. If $\{\varphi_k(x)\}$ is an orthonormal basis in $L_2(r)$ and

$$|(\varphi_k, \rho)| \leq Ak^{-m}, \quad \int [\varphi_k(x)r(x)]^2 \rho(x) \mu(dx) \leq H,$$

then, with the choice

$$n \sim \sqrt[2m]{N},$$

the deviation $\|\pi^* - \rho\|$ has, in probability, order

$$N^{-\frac{1}{2} + \frac{1}{4m}}.$$

Since the rate of approximation of $\rho(x)$ by a segment of the series $\sum a_k \varphi_k(x)$ may be unknown in advance, one can choose n according to the results of the observations, restricting oneself in (3) only to those terms for which the coefficient a_{kN}^* is substantially larger than its experimental root-mean-square error.

Above we considered approximation of $\rho(x)$ in the metric L_2 . In other metrics, Theorems 1a and 1b may give upper or lower estimates. In particular, it follows that the maximum of the absolute deviation of the histogram from the density is a quantity $\Omega(N^{-1/3})$. From the results of N. V. Smirnov ⁽¹⁾ (see also ⁽³⁾) it follows that this estimate can be sharpened only by a logarithmic factor. Unlike the histogram, for $\varphi_{kn}(x)$ of general form the polynomial $\pi_{nN}^*(x)$ may turn out to be nonpositive. To avoid this, one may try to correct it by setting, for example,

$$\tilde{\pi}_{nN}^*(x) = \sum_{k=1}^n \gamma_{kn} a_{kn}^* \varphi_{kn}(x),$$

where γ_{kn} are certain multipliers.

It is clear in advance that the estimate of $\rho(x)$ by formula (3) is not always appropriate. However, if an estimate is needed which is suitable for any unknown density $\rho(x)$ from a substantially infinite-dimensional set Π of densities, then estimate (3) may turn out to be close to optimal (more precisely, with an error only a finite number of times larger than that of the optimal estimate). Let us consider two restrictions on the set Π and the weight $r(x)$:

$$\sup_{\rho \in \Pi, x \in X} \rho(x)r(x) \leq A; \tag{7a}$$

$$\inf_{\rho \in \Pi, x \in X} \rho(x)r(x) \geq a. \tag{7b}$$

If conditions (7a) and (7b) are both satisfied, then all densities $\rho(\cdot) \in \Pi$ are uniformly “Lipschitz-equivalent” with respect to one another:

$$\frac{a}{A} \leq \inf_{\rho, q, x} \frac{\rho(x)}{q(x)} \leq \sup_{\rho, q, x} \frac{\rho(x)}{q(x)} \leq \frac{A}{a}. \tag{8}$$

Denote by $d_n(\Pi)$ the n -dimensional width of Π (see (4)), and by $r_n(\Pi)$ the radius of the largest n -dimensional sphere inscribed in Π .

Theorem 2a. If condition (7a) is satisfied and

$$d_n(\Pi) \leq Bg(Cn), \quad (9a)$$

then one can choose a sequence of subspaces E_n and a dependence of n on N such that

$$\sup_{\rho \in \Pi} \rho(\|\pi_{nN}^* - \rho\|) = O\left(\sqrt{N^{-1}\Gamma(N)}\right).$$

Theorem 2b. If condition (7b) is satisfied and

$$r_n(\Pi) \geq bg(cn), \quad (9b)$$

then, for any estimate $\pi_N^*(x; \xi^{(1)}, \dots, \xi^{(N)})$, whatever the method of its construction,

$$\sup_{\rho \in \Pi} \rho(\|\pi_N^* - \rho\|) = \Omega\left(\sqrt{N^{-1}\Gamma(N)}\right).$$

Theorem 2a follows easily from Theorem 1a. The idea of the proof of Theorem 2b is that, on the one hand,

$$M\|\pi^* - \rho\|^2 \geq \|M\pi^* - \rho\|^2,$$

and, on the other hand, by the Cramér–Rao inequality (see (5)), for any n -dimensional sub-

of the space $E_n \ni p$

$$M\|\pi^* - p\|^2 \geq a \frac{n}{N} \left[\frac{1}{n} \operatorname{div} M\pi_n^* \right]^2,$$

where π_n^* is the projection of π^* onto E_n .

If Π has property (8), and each weight $r(x) \in R$ has properties (7a) and (7b), then all the metrics $L_2(r)$ are equivalent on Π . Under these conditions Theorem 2a can be strengthened by replacing in it

$$\sup_{p \in \Pi} \sup_{r \in R} \rho(\|\pi_{nN}^* - p\|),$$

and in Theorem 2b by

$$\sup_{p \in \Pi} \inf_{r \in R} \rho(\|\pi_N^* - p\|).$$

The most important example is the family of weights $r(x) = 1/p(x)$; at each point $p \in \Pi$ there is its own scalar product and its own metric, so that Π may be interpreted as a Riemannian manifold of infinite dimension. In this case the quality of approximation is determined by the mean square of the relative error. It should be noted, however, that for distributions defined on the entire real line or on the entire Euclidean space, conditions (7a) and (7b), and still more (8), turn out to be excessively restrictive.

In conclusion, the author considers it a pleasant duty to express gratitude to N. V. Smirnov for his constant attention and valuable discussion.

Received
15 V 1962

REFERENCES

- ¹ N. V. Smirnov, DAN, **74**, No. 2, 189 (1950).
- ² I. M. Gel' fand, S. M. Feinberg, A. S. Frolov, N. N. Chentsov, Tr. II International Conf. of the UN on the Peaceful Uses of Atomic Energy, **2**, Moscow, 1959, p. 628.
- ³ B. I. Glivenko, *Course in Probability Theory*, Moscow, 1939.
- ⁴ V. M. Tikhomirov, UMN, **15**, issue 3, 81 (1960).
- ⁵ G. Cramér, *Mathematical Methods of Statistics*, IL, 1948.

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

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.