$\sqrt{\varepsilon^2 - \sum_{k=0}^{\tau-1} a_k^2}$. However, the accuracy condition (1) for the reproduction $\tilde{\xi}$ of the input process ξ takes the form

$$\mathbf{M}(\xi^{(\tau)}(t) - \tilde{\xi}^{(\tau)}(t))^2 \leqslant \varkappa^2 = \varepsilon^2 - \delta^2 = \varepsilon^2 - \sum_{k=0}^{\tau-1} a_k^2 > 0,$$

where $(\xi^{(\tau)}, \tilde{\xi}^{(\tau)})$ are a stationary normal pair of processes. The nonanticipatory message rate has been obtained in the top row of the inequality (6), and the bottom row of (6) is obtained by substituting in the top row of (6) the value of $f^{(\tau)}(\lambda)$ from (25) for $f(\lambda)$ and $\varkappa^2 = \varepsilon^2 - \sum_{k=0}^{\tau-1} a_k^2$ for ε^2, since the random variable $\tilde{\xi}^{(\tau)}(t)$ is decomposable into the sum of the random variable $\xi(t) - \tilde{\xi}^{(\tau)}(t)$.

LITERATURE CITED

1. A. K. Gorbunov and M. S. Pinsker, "Epsilon-entropy and message rate without anticipation and with prediction," Probl. Peredachi Inform., 9, No. 3, 12-21 (1973).
2. A. K. Gorbunov and M. S. Pinsker, "Epsilon-entropy with prediction for a Gaussian message and a Gaussian source," Probl. Peredachi Inform., 10, No. 2, 5-26 (1974).
3. M. S. Pinsker, "Message sources," Probl. Peredachi Inform., 14, 5-20 (1963).
4. M. S. Pinsker, "Gaussian sources," Probl. Peredachi Inform., 14, 59-100 (1963).
5. Yu. A. Rozanov, Stationary Stochastic Processes [in Russian], Fizmatgiz, Moscow (1963).

SAMPLE ESTIMATE OF THE ENTROPY OF A RANDOM VECTOR

L. F. Kozachenko and N. N. Leonenko

UDC 621.391.1:519.27

We establish conditions for asymptotic unbiasedness and consistency of a simple estimator of the unknown entropy of an absolutely continuous random vector from a sample of independent observations.

Let R^m be the m-dimensional Euclidean space with the metric

$$\rho(x_1, x_2) = \left\{ \sum_{j=1}^{m} (x_1^{(j)} - x_2^{(j)})^2 \right\}^{1/2},$$

where $x_i = (x_i^{(1)}, \ldots, x_i^{(m)}) \in R^m$, i = 1, 2, and $m \geq 1$. Consider the ball $v(y, r) = \{x \in R^m: \rho(x, y) < r\}$ of volume

$$|v(y, r)| = r^m c_1(m), \quad c_1(m) = \pi^{m/2}/\Gamma(m/2+1).$$

Assume that the random vector has an unknown probability density $f(x)$, $x \in R_m$. Our problem is to estimate, from the independent observations $X_1, \ldots, X_N$, $N \geq 2$, of the vector ξ, its entropy

Translated from Problemy Peredachi Informatsii, Vol. 23, No. 2, pp. 9-16, April-June, 1987. Original article submitted June 10, 1985.

0032-9460/87/2302-0095$12.50

$$H=H(f)=-\int_{R^m} f(x)\ln f(x)\,dx<\infty. \tag{1}$$

The functional (1) is usually estimated by replacing f(x) and the corresponding distribution function with their empirical counterparts. This approach has been applied, in particular, in [1-3], where the history of the subject is surveyed. However, the consistency conditions of such estimators are highly restrictive [1, 3], while the methods of [2] are quite difficult to apply to estimators of the functional (1). This suggests that we should look for "simple" estimators of the functional (1). In this paper, we investigate one such simple estimator (1), which has been proposed in [4] on the basis of Dobrushin's ideas [5]. Note that [4] offers no mathematical propositions concerning the entropy estimator (2) defined below. In [6], as in [4], the entropy estimator is used in tests for normality of a random vector.

From the sample $X_1, \ldots, X_N$, $N \geq 2$, compute $\rho_i = \min\{\rho(X_i, X_j), j \in \{1, 2, \ldots, N\} \setminus \{i\}\}$ and let $\bar\rho=\left\{\prod_{i=1}^{N}\rho_i\right\}^{1/N}$. We estimate the entropy (1) by

$$H_N=m\ln\bar\rho+\ln c_1(m)+\ln\gamma+\ln(N-1), \tag{2}$$

where $\ln\gamma = c_2$ is the Euler constant ($c_2 \simeq 0.5772$), i.e., $\gamma=\exp\left\{-\int_0^\infty e^{-t}\ln t\,dt\right\}$.

We now state two main theorems.

THEOREM 1. Form some $\varepsilon > 0$, let

$$\int_{R^m}|\ln f(x)|^{1+\varepsilon}f(x)\,dx<\infty, \tag{3}$$

$$\int_{R^m}\int_{R^m}|\ln\rho(x,y)|^{1+\varepsilon}f(x)f(y)\,dx\,dy<\infty. \tag{4}$$

Then $\lim MH_N = H$.

THEOREM 2. For some $\varepsilon > 0$, let

$$\int_{R^m}|\ln f(x)|^{2+\varepsilon}f(x)\,dx<\infty, \tag{5}$$

$$\int_{R^m}\int_{R^m}|\ln\rho(x,y)|^{2+\varepsilon}f(x)f(y)\,dx\,dy<\infty. \tag{6}$$

Then H_N for $N \to \infty$ is a consistent estimator of H.

Before proceeding to prove Theorem 1, recall Lebesgue's theorem.

THEOREM 3 [7]. If $f(x) \in L_1(R^m)$, then for almost all $x \in R^m$ and any sequences of open balls $v(x, r_k)$, of radius $r_k \to 0$,

$$\lim_{k\to\infty}\frac{1}{|v(x,r_k)|}\int_{v(x,r_k)}f(y)\,dy=f(x). \tag{7}$$

Proof of Theorem 1. Represent H^N in the form

$$H_N=\frac{1}{N}\sum_{i=1}^{N}\zeta_i,$$

where $\zeta_i = \ln\{\rho_i^m\gamma(N-1)c_1(m)\}$ are identically distributed random variables. Then $MH_N = M\zeta_i$ for any $i \in \{1, \ldots, N\}$. Let $r_N(u) = \{u/[c_1(m)\gamma(N-1)]\}^{1/m}$, $u \in R^1$. Then for $y \in R^m$ and $N \to \infty$,

$$|v(y,r_N(u))|=\{u/[\gamma(N-1)]\}\to 0.$$

Using the definition of conditional probabilities and Theorem 3, we obtain for almost all x for $N \to \infty$,

$$F_{N,x}(u)=P\{\exp(\zeta_i)<u\,|\,X_i=x\}=1-P\{\bigcap_{\substack{j=1\\ j\neq i}}^{N}\{X_j\notin v(x,r_N(u))\}\}=$$

$$=1-\Big(1-\int\limits_{v(x,r_N(u))} f(y)\,dy\Big)^{N-1}\to 1-\exp\{-f(x)u/\gamma\}. \tag{8}$$

We will show that for almost all x,

$$\lim_{N\to\infty}\mathbf{M}\{\zeta_i\,|\,X_i=x\}=-\ln f(x). \tag{9}$$

Remark 1. In (9) and below, the expression "for almost all x" implies that the corresponding propositions hold for almost all $x \in R^m$ with respect to the probability measure induced by the density f(x).

Let $\xi_{N,x}$ be a random variable with the distribution function $F_{N,x}(u)$ and ξ_x a random variable with the distribution function $F_x(u) = 1 - \exp\{-f(x)u/\gamma\}$ [see (8)]. Since for $f(x) > 0$,

$$\mathbf{M}\ln\xi_x=\int\limits_0^\infty \ln u\exp\{-f(x)u/\gamma\}[f(x)/\gamma]\,du=\int\limits_0^\infty \ln\{t\gamma/f(x)\}e^{-t}\,dt=$$

$$=\ln\gamma-\ln f(x)+\int\limits_0^\infty \ln t e^{-t}\,dt=-\ln f(x),$$

and $\mathbf{M}\ln\xi_{N,x} = \mathbf{M}\{\ln\zeta_i\,|\,X_i = x\}$, then (9) holds if for almost all x

$$\lim_{N\to\infty}\mathbf{M}\ln\xi_{N,x}=\mathbf{M}\xi_x. \tag{10}$$

We will show that for some $\varepsilon > 0$ and $C > 0$, we have $\mathbf{M}|\ln\xi_{N,x}|^{1+\varepsilon} < C$ for $x \in R^m$ such that $f(x) > 0$, (7) holds at the point x, and

$$\int\limits_0^\infty |\ln|u||^{1+\varepsilon}\,dF_{2x}(u)<\infty. \tag{11}$$

Remark 2. The conditions (7), (11) and the condition $f(x) > 0$ are satisfied simultaneously for almost all $x \in R^m$ with respect to the probability measure induced by the density f(x).

Indeed, the condition (7) and $f(x) > 0$ are satisfied for almost all x, the condition (11) is satisfied for almost all x since

$$\int\limits_0^\infty |\ln|u||^{1+\varepsilon}\,dF_{2x}(u)=\mathbf{M}|\ln\xi_{2x}|^{1+\varepsilon},$$

and

$$\mathbf{M}\,|\ln\xi_{2x}/[\gamma c_1(m)]|^{1+\varepsilon}m^{1+\varepsilon}=\mathbf{M}\,|\ln\rho(x,\xi)|^{1+\varepsilon}=\int\limits_{R^m}|\ln\rho(x,y)|^{1+\varepsilon}f(y)\,dy.$$

Note that the relationship

$$\int\limits_{R^m}|\ln\rho(x,y)|^{1+\varepsilon}f(y)\,dy<\infty$$

for almost all x follows from (4). In the sequel, we will need

LEMMA 1. Let F(u) be a distribution function. For $\alpha \geq 1$ we have

$$\int\limits_1^\infty (\ln u)^\alpha\,dF(u)<\infty \tag{12}$$

if and only if

$$\int_1^\infty (\ln u)^{\alpha-1}u^{-1}(1-F(u))\,du<\infty. \tag{13}$$

Here

$$\int_1^\infty (\ln u)^\alpha\, dF(u)=\alpha\int_1^\infty (\ln u)^{\alpha-1}u^{-1}(1-F(u))\,du. \tag{14}$$

The proof of Lemma 1 is quite standard are therefore omitted.

Let us now examine $M|\ln \xi_{N,x}|^{1+\varepsilon}$ for some $\varepsilon > 0$ and x such that $f(x) > 0$ and (7), (11) hold:

$$\mathbf{M}|\ln \xi_{N,x}|^{1+\varepsilon}=\int_0^\infty |\ln u|^{1+\varepsilon}\, dF_{N,x}(u)\ \int_0^1 |\ln u|^{1+\varepsilon}\, dF_{N,x}(u)+\int_1^\infty (\ln u)^{1+\varepsilon}\, dF_{N,x}(u). \tag{15}$$

By Lemma 1,

$$\int_1^\infty (\ln u)^{1+\varepsilon}\, dF_{N,x}(u)=(1+\varepsilon)\int_1^\infty (\ln u)^{\varepsilon}u^{-1}(1-F_{N,x}(u))\,du= \tag{16}$$

$$=(1+\varepsilon)\int_1^{\sqrt{N-1}} (\ln u)^{\varepsilon}u^{-1}(1-F_{N,x}(u))\,du\ (1+\varepsilon)\int_{\sqrt{N-1}}^\infty (\ln u)^{\varepsilon}u^{-1}(1-F_{N,x}(u))\,du=I_1(N)+I_2(N).$$

Let us estimate the two integrals separately:

$$I_2(N)=(1+\varepsilon)\int_{\sqrt{N-1}}^\infty (\ln u)^{\varepsilon}u^{-1}\Big(1-\int_{v(x,r_N(u))} f(z)\,dz\Big)^{N-1} du\leqslant$$

$$\leqslant(1+\varepsilon)\Big(1-\int_{v(x,r_N(\sqrt{N}))} f(z)\,dz\Big)^{N-2}\int_{\sqrt{N-1}}^\infty (\ln u)^{\varepsilon}u^{-1}\Big(1-\int_{v(x,r_N(u))} f(z)\,dz\Big)du. \tag{17}$$

Since $\frac{\gamma(N-1)}{\sqrt{N}}\int_{v(x,r_N(\sqrt{N}))} f(z)\,dz\to f(x)$ for $N \to \infty$, then for $\delta > 0$ such that $f(x) - \delta > 0$ and sufficiently large N,

$$\Big(1-\int_{v(x,r_N(\sqrt{N}))} f(z)\,dz\Big)^{N-2}\leqslant\Big(1-\frac{f(x)-\delta}{\sqrt{N-1}\,\gamma}\Big)^{N-2}\ \exp\Big\{-\frac{(f(x)-\delta)\,(N-2)}{\sqrt{N-1}\,\gamma}\Big\}. \tag{18}$$

Now,

$$\int_{\sqrt{N-1}}^\infty (\ln u)^{\varepsilon}u^{-1}\Big(1-\int_{v(x,r_N(u))} f(z)\,dz\Big)du$$

$$=\Big[\int_{1/\sqrt{N-1}}^1+\int_1^\infty\Big]\{(\ln(u(N-1)))^{\varepsilon}u^{-1}(1-F_{2x}(u))\,du\}\leqslant \tag{19}$$

$$\leqslant\sqrt{N-1}(\ln(N-1))^{\varepsilon}+C_\varepsilon[\ln(N-1)]^{\varepsilon}\int_1^\infty (1-F_{2x}(u))\,u^{-1}\,du+$$

$$+C_\varepsilon\int_1^\infty(\ln u)^{\varepsilon}u^{-1}(1-F_{2x}(u))\,du,$$

where $C_\varepsilon > 0$. The integrals

$$\int_1^\infty (\ln u)^\varepsilon u^{-1}(1-F_{2x}(u))\,du, \qquad \int_1^\infty (1-F_{2x}(u))\,u^{-1}\,du$$

converge by (11) and Lemma 1. Therefore, (19), (18), (17) imply that $I_2(N) \to 0$ for $N \to \infty$. Now,

$$I_1(N)=(1+\varepsilon)\int_1^{\sqrt{N-1}} (\ln u)^\varepsilon u^{-1}\Big(1-\int_{v(x,r_N(u))} f(z)\,dz\Big)^{N-1} du \leqslant$$

$$\leqslant (1+\varepsilon)\int_1^{\sqrt{N-1}} (\ln u)^\varepsilon u^{-1}\exp\Big\{-(N-1)\int_{v(x,r_N(u))} f(z)\,dz\Big\}du.$$

Since for $u < \sqrt{N}$ we have $r_N(u) \le r_N(\sqrt{N-1}) \to 0$ for $N \to \infty$, then noting that (7) is satisfied for x when N is sufficiently large, we obtain for all $u < \sqrt{N-1}$ simultaneously

$$\int_{v(x,r_N(u))} f(z)\,dz \geqslant \frac{f(x)-\delta}{\gamma}\,\frac{u}{\sqrt{N-1}}.$$

Therefore, for sufficiently large N,

$$I_1(N)\leqslant(1+\varepsilon)\int_1^{\sqrt{N-1}} (\ln u)^\varepsilon u^{-1}\exp\Big\{-\frac{f(x)-\delta}{\gamma}u\Big\}du \leqslant$$

$$\leqslant(1+\varepsilon)\int_1^{\infty} (\ln u)^\varepsilon u^{-1}\exp\Big\{-\frac{f(x)-\delta}{\gamma}u\Big\}du < \infty.$$

Noting that $I_2(N) \to 0$, the last relationship and (16) imply that for $N \to \infty$ there exists a constant C_1 such that for all sufficiently large N,

$$\int_1^\infty (\ln u)^{1+\varepsilon}\,dF_{N,x}(u)<C_1. \tag{20}$$

Now,

$$\int_0^1 |\ln u|^{1+\varepsilon}\,dF_{N,x}(u)$$

$$=(1+\varepsilon)\int_0^1 (-\ln u)^{1+\varepsilon}u^{-1}\Big(1-\Big(1-\int_{v(x,r_N(u))} f(z)\,dz\Big)^{N-1}\Big)du.$$

By (7), for sufficiently large N, for all z and some $\delta > 0$ we have

$$\Big(1-\int_{v(x,r_N(u))} f(z)\,dz\Big)^{N-1} \geqslant \exp\Big\{-\frac{f(x)+\delta}{\gamma}u\Big\}.$$

Therefore, for sufficiently large N,

$$\int_0^1 |\ln u|^{1+\varepsilon}\,dF_{N,x}(u)\leqslant$$

$$\leqslant(1+\varepsilon)\int_0^1 (-\ln u)^\varepsilon u^{-1}\Big(1-\exp\Big\{-\frac{f(x)+\delta}{\gamma}u\Big\}\Big)du<\infty.$$

The last inequality, combined with (20), shows that there exists a constant C_2 such that for all N [see (15)]

$$\mathbf{M}|\ln \xi_{N,x}|^{1+\varepsilon}<C_2.$$

The last inequality leads (see [8, p. 290, d]) to (10) and hence (9). To complete the proof of Theorem 1, it suffices to show that for N → ∞,

$$\mathbf{M}\zeta_i = \int_{R^m} \mathbf{M}(\zeta_i \mid X_i = x) f(x)\,dx \to \int_{R^m} (-\ln f(x)) f(x)\,dx. \tag{21}$$

By Fatou's lemma and condition (3), we have

$$\limsup_{N\to\infty} \int_{R^m} |\mathbf{M}(\zeta_i \mid X_i = x)|^{1+\varepsilon} f(x)\,dx \leqslant$$

$$\leqslant \int_{R^m} |\limsup_{N\to\infty} \mathbf{M}(\zeta_i \mid X_i = x)|^{1+\varepsilon} f(x)\,dx = \int_{R^m} |\ln f(x)|^{1+\varepsilon} f(x)\,dx < \infty.$$

Hence follows (21) (see [9, p. 176]).

Proof of Theorem 2. We omit the proof of the passages to the limit, since it is quite cumbersome and virtually identical to that in Theorem 1. Clearly,

$$\mathbf{D}H_N = \Big[\sum_{i=1}^{N} \mathbf{D}\zeta_i + \sum_{i\neq j} \mathrm{cov}(\zeta_i, \zeta_j)\Big] \Big/ N^2 = \tag{22}$$

$$= \mathbf{D}\zeta_i / N + 2\sum_{i<j} \mathrm{cov}(\zeta_i, \zeta_j)/N^2.$$

From (8), as in the proof in Theorem 1, we obtain

$$\lim_{N\to\infty} \mathbf{M}\{\zeta_i^2 \mid X_i = y\} = \int_0^\infty \ln^2 u \exp\Big\{-\frac{f(y)u}{\gamma}\Big\} \frac{f(y)}{\gamma}\,du =$$

$$= \int_0^\infty (\ln t + \ln\gamma - \ln f(y))^2 e^{-t}\,dt = c_2 + \ln^2 f(y),$$

where

$$c_2 = \int_0^\infty \ln^2 t e^{-t}\,dt - \ln^2\gamma.$$

Then using Theorem 1, we obtain

$$\lim_{N\to\infty} \{\mathbf{D}\zeta_i\} = c_2 + I_1 - H^2, \tag{23}$$

where

$$I_1 = \int_{R^m} f(x)(\ln f(x))^2\,dx.$$

Note that for normal density f(x), $I_1 - H^2 = m/2$. Substituting (23) in (22), we see that for N → ∞, the first term in (22) goes to zero. To complete the proof, it remains to show that for N → ∞ and i ≠ j,

$$\mathrm{cov}(\zeta_i, \zeta_j) = \mathbf{M}\zeta_i\zeta_j - (\mathbf{M}\zeta_i)^2 \to 0. \tag{24}$$

For i ≠ j; u, w ∈ R^1; x, y ∈ R^m, we have

$$P\{e^{\zeta_i} < u, e^{\zeta_j} < w \mid X_i = x, X_j = y\} = 1 - P\{\min_{k\neq j} \rho(y, X_k) \geqslant r_N(w)\} -$$

$$-P\{\min_{k\neq i} \rho(x, X_k) \geqslant r_N(u)\} + P\{\min_{k\neq i} \rho(x, X_k) \geqslant r_N(u),\ \min_{k\neq j} \rho(y, X_k) \geqslant r_N(w)\}. \tag{25}$$

As before, for N → ∞

$$P\{\min_{k\neq i} \rho(x, X_k) \geqslant r_N(u)\} \to \exp\{-f(x)u/\gamma\}, \tag{26}$$

$$P\{\min_{k\neq j} \rho(y, X_k) \geqslant r_N(w)\} \to \exp\{-f(y)w/\gamma\}.$$

The last term in (25) is equal to

$$P\{\bigcap_{k\neq i\neq j}[X_k\notin v(x,r_N(u))\cup v(y,r_N(w))]\}=\left[1-\int\limits_{v(x,r_N(u))\cup v(y,r_N(w))}f(z)\,dz\right]^{N-2}. \tag{27}$$

For $N\to\infty$ we have $r_N(u)\to 0$, $r_N(w)\to 0$. Therefore, by the separation theorem, there exists $N_0>0$ such that $v(x, r_N(u))\cap v(y, r_N(w))=\emptyset$ for $N>N_0$, and so for $N\to\infty$ (11) implies

$$\begin{aligned}&P\{\min_{k\neq i}\rho(x,X_k)\geqslant r_N(u),\ \min_{k\neq j}\rho(y,X_k)\geqslant r_N(w)\}\to\\&\to\exp\{-(uf(x)+wf(y))/\gamma\}.\end{aligned} \tag{28}$$

Substituting (26)-(28) in (25), we obtain

$$\begin{aligned}&\lim_{N\to\infty}P\{e^{\zeta_i}<u,e^{\zeta_j}<w\,|\,X_i=x,\ X_j=y\}=1-\exp\{-uf(x)/\gamma\}-\\&-\exp\{-wf(y)/\gamma\}+\exp\{-(uf(x)+wf(y))/\gamma\}.\end{aligned}$$

The last relationship gives

$$\begin{aligned}&\lim_{N\to\infty}\mathbf{M}\{\zeta_i\zeta_j\,|\,X_i=x,X_j=y\}=\int_0^\infty\int_0^\infty\ln u\ln w\times\\&\times e^{-uf(x)/\gamma}e^{-wf(y)/\gamma}\left[\frac{f(x)}{\gamma}\frac{f(y)}{\gamma}\right]du\,dw=\ln f(x)\ln f(y).\end{aligned}$$

Then

$$\lim_{N\to\infty}\mathbf{M}\zeta_i\zeta_j=\int\limits_{R^m}\int\limits_{R^m}f(x)f(y)\ln f(x)\ln f(y)\,dx\,dy. \tag{29}$$

The right-hand side of (29) coincides with $\lim(\mathbf{M}\zeta_i)^2$. Thus, (24) holds.

LITERATURE CITED

1. A. V. Ivanov and M. N. Rozhkova, "On properties of sample estimator of entropy of a random vector with probability density," Probl. Peredachi Inform., 17, No. 3, 34-43 (1981).
2. B. Ya. Levit, "Asymptotically efficient estimation of nonlinear functionals," Probl. Peredachi Inform., 14, No. 3, 65-72 (1978).
3. T. A. Azlarov and Sh. A. Khashimov, "On limit distribution of estimators of some non-linear functionals of probability density," in: Limit Theorems for Stochastic Processes and Related Topics [in Russian], Fan, Tashkent (1982), pp. 3-15.
4. S. N. Podkorytov, "On estimators of the entropy of distribution density and testing the fit of a sample to multivariate normal distribution," Proc. 7th All-Union Symposium on Methods of Representations and Hardware Analysis of Stochastic Processes and Fields [in Russian], Leningrad (1974), pp. 75-78.
5. R. L. Dobrushin, "A simple method of empirical estimation of the entropy of a stationary sequence," Teor. Veroyatn. Ee Primen., 3, No. 4, 462-464 (1958).
6. O. Vasicek, "A test for normality on sample entropy," J. R. Stat. Soc., B38, No. 1, 54-59 (1976).
7. H. Lebesgue, "Sur l'intégration des fonctions discontinues," Ann. Ecole Norm., No. 27, 361-450 (1910).
8. W. Feller, An Introduction to Probability Theory and Its Applications [Russian translation], Vol. 2, Mir, Moscow (1984).
9. M. Loeve, Probability Theory, 4th edn., Springer-Verlag (1977).