with zero boundary conditions; and therefore will be finite for any bounded domain U, and thus for unbounded U if $c(x) \leqq \alpha < 0$. Now, applying Lemma 1 and (2), we finally obtain

$$\pi_V(x, \Gamma) = \pi'_V(x, \Gamma),$$
$$m_V(x, \Gamma) = m'_V(x, \Gamma),$$

whence follows the assertion of the theorem.

It would be very interesting to find a direct probability theoretic proof of the theorem. In conclusion I am deeply indebted to E. B. Dynkin, who has made a series of valuable suggestions for the writing of this paper.

Submitted to the editors
May 20, 1959.

REFERENCES

[1] I. N. VEKUA, *Concerning complete systems of harmonic polynomials in space*, Dokl. Akad. Nauk SSSR, 90, 4, 1953, pp. 495. (In Russian.)

[2] E. B. DYNKIN, *Infinitesimal operators of Markov processes*, Theory Prob. Applications, 1, 1956, pp. 34–54. (Translation in English, this journal.)

[3] E. B. DYNKIN and A. A. YUSHKEVICH, *Strong Markov processes*, Theory Prob. Applications, 1, 1956, pp. 134–139. (Translation in English, this journal.)

[4] E. B. DYNKIN, *Basic Notions in the Theory of Markov Processes*, Fizmatgiz, 1959. (In Russian.)

[5] E. M. LANDIS, *Concerning certain properties of the solutions of elliptic equations*, Dokl. Akad. Nauk SSSR, 107, 5, 1956, p. 640. (In Russian.)

[6] K. MIRANDA, *Elliptic Partial Differential Equations*, IIL, 1957. (In Russian.)

[7] W. FELLER, *The general diffusion operator and positivity preserving semigroups in one dimension*, Ann. of Math., 60, 3, 1954, pp. 417–436.

[8] W. FELLER, *On second order differential operators*, Ann. of Math., 61, 1, 1955, pp. 90–105.

ON A PROPERTY OF NON-DEGENERATE DIFFUSION PROCESSES

I. V. GIRSANOV

(*Summary*)

Let X be a non-degenerate diffusion process in n-dimensional Euclidean space R^n, U a domain, τ_U the time at which the boundary of U is first attained.

In the present article it is proved that the course of the process X inside U is determined by the functions

$$m(x) = M_x \tau_u \quad \text{and} \quad \pi(x, \Gamma) = \mathrm{P}_x\{x_{\tau_u} \in \Gamma\}.$$

The proof is based on an analytic method of the theory of elliptic differential operators.

ON A STATISTICAL ESTIMATE FOR THE ENTROPY OF A SEQUENCE OF INDEPENDENT RANDOM VARIABLES

G. P. BASHARIN

(*Translated by Newcomb Greenleaf*)

1. We consider the sequence of mutually independent random variables $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ each of which takes on the value E_i with probability p_i $(0 < p_i < 1,\ i = 1, 2, \cdots, s,\ \sum_1^s p_i = 1)$. To get an estimate of the amount of entropy (see [1])

$$H(p_1, p_2, \cdots, p_s) = -\sum_1^s p_i \log_2 p_i \tag{1}$$

of this sequence on the basis of a given sampling, the unknown *a priori* probability p_i in formula (1) is generally replaced by an estimate $\hat{p}_i = m_i/N$ $(i = 1, \cdots, s)$, where m_i is the frequency of occurrence of E_i in the sample of size N. Such a statistical estimate of entropy is often made in various applications. For example, in [4] a statistical estimate of the entropy of a sequence of printed letters of English words is made, and in [5] the same is done for the entropy of a sequence of measurements of values of brightness of a television signal. In this connection, then, the study of sample distribution of the random variable

$$\hat{H} = H(\hat{p}_1, \cdots, \hat{p}_s) = -\sum_1^s \hat{p}_i \log_2 \hat{p}_i. \tag{2}$$

is of interest.

It will be shown below that $\hat{H}$ is a biased, consistent, asymptotically normal estimate of the entropy H, and that the mean value and variance of this estimate have the following values:

$$\mathbf{E}\hat{H} = H - \frac{s-1}{2N}\log_2 e + O\left(\frac{1}{N^2}\right), \tag{3}$$

$$\mathbf{D}\hat{H} = \frac{1}{N}\left[\sum_1^s p_i \log_2{}^2 p_i - H^2\right] + O\left(\frac{1}{N^2}\right), \tag{4}$$

where $O(1/N^a)$ designates a quantity of order N^{-a} $(a > 0)$. With the help of formula (3) it is possible to estimate the amount of the bias of the estimate, and, when necessary, to make a correction of this amount in the sample value of $\hat{H}$. Using formula (4) and the asymptotic normality of the distribution of $\hat{H}$, it is possible to compute a confidence bound for sample estimates of entropy. It is interesting to note that for $p_1 = \cdots = p_s = 1/s$, the main term in (4) vanishes, i.e., for a fixed size of the sample, the estimate of entropy will be most precise in the case of equiprobable outcomes.

The sample distribution for the entropy of a sequence of random variables, connected with Markov chains, is not examined in this paper.

2. In this and in the following section, by the symbol ln will be meant the natural logarithm, but in the final formulation the natural logarithm will be replaced by the logarithm with base two. For simplicity of writing, the entropy, $-\sum_1^s p_i \ln p_i$ will also be designated by H.

We will expand the function $H(\hat{p}_1, \cdots, \hat{p}_s)$ in a Taylor series about the point $(p_1, \cdots, p_s)$, confining ourselves to derivatives of fourth order. The value of the derivative of the function H at the point $(p_1, \cdots, p_s)$ is the following:

$$\frac{\partial H}{\partial p_i} = -1 - \ln p_i, \qquad \frac{\partial^k H}{\partial p_i{}^k} = (-1)^{(k-1)}(k-2)!\,p_i^{-k+1}, \qquad i = 1, 2, \cdots, s, \quad k \geqq 2. \tag{5}$$

Mixed derivatives of all orders of H vanish at $(p_1, \cdots, p_s)$. It is evident from the condition $0 < p_i < 1$, $i = 1, \cdots, s$, that the derivative of any order of the function H is continuous in a neighborhood of the point $(p_1, \cdots, p_s)$, and consequently it is possible to expand H in a convergent Taylor series (see [3], p. 185). We can write down the expansion with Lagrange's remainder term in the following form:

$$\begin{aligned} H(\hat{p}_1, \cdots, \hat{p}_s) = H(p_1, \cdots, p_s) &- \sum_1^s (\hat{p}_i - p_i)(1+\ln p_i) - \frac{1}{2}\sum_1^s \frac{(\hat{p}_i - p_i)^2}{p_i} \\ &+ \frac{1}{6}\sum_1^s \frac{(\hat{p}_i - p_i)^3}{p_i^2} - \frac{1}{12}\sum_1^s \frac{(\hat{p}_i - p_i)^4}{[p_i + \theta(\hat{p}_i - p_i)]^3}, \end{aligned} \tag{6}$$

$f(b) = f(a) + f'(a)(b-a) + \frac{f''(a)}{2}(b-a)^2 + \cdots$

where $0 < \theta < 1$.

3. In order to determine the moments of the estimate $\hat{H}$ with the help of expansion (6), it is necessary to compute the central moments of the random variables $\hat{p}_i$ $(i = 1, \cdots, s)$. Carrying out elementary calculations, we obtain the following values for the first few moments, which will be required in what follows $(i, j, u = 1, \cdots, s)$:

$$\begin{aligned} &\mathbf{E}\hat{p}_i = p_i, \quad \mathbf{E}(\hat{p}_i - p_i)^2 = \frac{p_i(1-p_i)}{N}, \\ &\mathbf{E}(\hat{p}_i - p_i)(\hat{p}_j - p_j) = -\frac{p_i p_j}{N}, && i \neq j \\ &\mathbf{E}(\hat{p}_i - p_i)^3 = \frac{2p_i^2 - 3p_i^2 + p_i}{N^2} \\ &\mathbf{E}(\hat{p}_i - p_i)^2(\hat{p}_j - p_j) = -\frac{p_i p_j(1-2p_i)}{N^2}, && i \neq j, \\ &\mathbf{E}(\hat{p}_i - p_i)(\hat{p}_j - p_j)(\hat{p}_u - p_u) = O\left(\frac{1}{N^2}\right), && i \neq j \neq u, \end{aligned} \tag{7}$$

$$\mathbf{E}(\hat{p}_i-p_i)^4 = O\left(\frac{1}{N^2}\right),$$

$$\mathbf{E}(\hat{p}_i-p_i)^2(\hat{p}_j-p_j)^2 = O\left(\frac{1}{N^2}\right), \qquad i \neq j,$$

$$\mathbf{E}(\hat{p}_i-p_i)^3(\hat{p}_j-p_j) = O\left(\frac{1}{N^2}\right), \qquad i \neq j, \tag{7}$$

$$\mathbf{E}(\hat{p}_i-p_i)^2(\hat{p}_j-p_j)(\hat{p}_u-p_u) = O\left(\frac{1}{N^2}\right), \qquad i \neq j \neq u,$$

$$\mathbf{E}(\hat{p}_i-p_i)^6 = O\left(\frac{1}{N^3}\right).$$

The remaining moments have order of magnitude not greater than N^{-2}.

4. With the help of formula (7) it is possible to compute the first four moments of the random variable $\hat{H}$. For the mean of $\hat{H}$ we obtain the following expression:

$$\mathbf{E}\hat{H} = H(p_1, \cdots, p_s) - \frac{s-1}{2N} + \frac{2-3s+\sum_1^s \frac{1}{p_i}}{6N^2} - \frac{1}{12}\sum_1^s \mathbf{E}\,\frac{(\hat{p}_i-p_i)^4}{[p_i(1-\theta)+\theta\hat{p}_i]^3}.$$

In order to estimate the remainder term, we note that

$$\frac{(\hat{p}_i-p_i)^4}{[p_i(1-\theta)+\theta\hat{p}_i]^3} \leqq \frac{(\hat{p}_i-p_i)^4}{p_i^3(1-\theta)^3}.$$

Hence it follows that

$$\mathbf{E}\,\frac{(\hat{p}_i-p_i)^4}{[p_i(1-\theta)+\theta\hat{p}_i]^3}$$

is of order of magnitude less than N^{-2} and that

$$\mathbf{E}\hat{H} = H - \frac{s-1}{2N} + O\left(\frac{1}{N^2}\right). \tag{3A}$$

In order to obtain formula (3) from this formula it is necessary to pass from natural logarithms to logarithms with base two. For this it suffices to multiply both sides of formula (3A) by $\log_2 e$.

5. For the calculation of the variance of $\hat{H}$ we make use of the following relation:

$$\mathbf{D}\hat{H} = \mathbf{E}\left\{\hat{H}-H+\frac{s-1}{2N}+O\left(\frac{1}{N^2}\right)\right\}^2. \tag{8}$$

Using (6), restricted to third order derivatives, we obtain

$$\mathbf{D}\hat{H} = \mathbf{E}\left\{-\sum_1^s(\hat{p}_i-p_i)(1+\ln p_i)+\frac{s-1}{2N}-\frac{1}{2}\sum_1^s\frac{(\hat{p}_i-p_i)^2}{p_i}\right.$$
$$\left.+\frac{1}{6}\sum_1^s\frac{(\hat{p}_i-p_i)}{[p_i(1-\theta)+\theta\hat{p}_i]^2}+O\left(\frac{1}{N^2}\right)\right\}^2.$$

Removing the brackets and neglecting those terms whose mathematical expectation, according to formula (7), is of order of magnitude less than or equal to N^{-2}, we obtain the following expression for the variance:

$$\mathbf{D}\hat{H} = \mathbf{E}[\sum_1^s(\hat{p}_i-p_i)(1+\ln p_i)]^2-\frac{1}{3}\mathbf{E}\sum_{i,j=1}^s\frac{(\hat{p}_i-p_i)(\hat{p}_j-p_j)^3(1+\ln p_i)}{[p_j(1-\theta)+\theta\hat{p}_j]^2}$$
$$+\frac{s-1}{6N}\mathbf{E}\sum_1^s\frac{(\hat{p}_i-p_i)^3}{[p_i(1-\theta)+\theta\hat{p}_i]^2}-\frac{1}{6}\mathbf{E}\sum_{i,j=0}^s\frac{(\hat{p}_i-p_i)^2}{p_i}\frac{(\hat{p}_j-p_j)^3}{[p_j(1-\theta)+\theta\hat{p}_j]^2}+O\left(\frac{1}{N^2}\right). \tag{9}$$

The first term is of order of magnitude N^{-1}:

$$\mathbf{E}[\sum_1^s(\hat{p}_i-p_i)(1+\ln p_i)]^2 = \frac{1}{N}\sum_1^s p_i(1-p_i)(1+\ln p_i)^2$$
$$-\frac{1}{N}\sum_{\substack{i,j=1\\ i\neq j}}^s p_ip_j(1+\ln p_i)(1+\ln p_j) = \frac{1}{N}[\sum_1^s p_i\ln^2 p_i - H^2].$$

We will show that all remaining terms have order of magnitude less than or equal to N^{-2}. For these evaluations we make use of Schwarz's inequality (see [2], 9. 5. 1.). Applying that inequality to the mathematical expectation under the summation sign in the second term, we obtain

$$\mathbf{E}\,\frac{(\hat{p}_i-p_i)(\hat{p}_j-p_j)^3}{[p_j(1-\theta)+\theta\hat{p}_j]^2} \leq \frac{\mathbf{E}|\hat{p}_i-p_i||\hat{p}_j-p_j|^3}{p_j^2(1-\theta)^2} \leq \frac{\{\mathbf{E}(\hat{p}_i-p_i)^2 E(\hat{p}_j-p_j)^6\}^{1/2}}{p_j^2(1-\theta)^2} = O\left(\frac{1}{N^2}\right).$$

Consequently the second term is of order of magnitude N^{-2}. In similar fashion it is shown that the third and fourth terms on the right side of formula (9) have order $N^{-5/2}$. Thus we have proved that

$$\text{(4A)} \qquad \mathbf{D}\hat{H} = \frac{1}{N}\left[\sum_1^s p_i \ln^2 p_i - H^2\right] + O\left(\frac{1}{N^2}\right).$$

In order to pass from formula (4A) to formula (4), it suffices to multiply both sides of formula (4A) by $\log_2{}^2 e$.

In similar fashion it is possible to show that the coefficients of asymmetry and excess of the random variable $\hat{H}$ have order N^{-2}.

6. In order to prove that as $N \to \infty$ the distribution $\hat{H}$ converges to the normal distribution with mean value (3) and variance (4), we write expansion (6) in the form

$$\sqrt{N}(\hat{H}-H) = -\sqrt{N}\sum_1^s (\hat{p}_i-p_i)(1+\ln p_i) - \frac{\sqrt{N}}{2}\sum_1^s \frac{(\hat{p}_i-p_i)^2}{p_i(1-\theta)+\theta\hat{p}_i}$$

and apply Theorem 28.4 of [2] to the random variable $\sqrt{N}(\hat{H}-H)$. Consequently, $\hat{H}$ is an asymptotically normal estimate of entropy.

Submitted to the editors
February 16, 1959.

REFERENCES

[1] C. E. Shannon. *The Mathematical Theory of Communication*, The Univ. of Illinois Press, 1949, pp. 3–89.
[2] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1945.
[3] G. M. Fikhtengol'ts, *Course of Differential and Integral Calculus*, Vol. 1, 1947. (In Russian.)
[4] C. E. Shannon, *Prediction and entropy of printed English*, Bell Syst. Tech. J., 30, 1, 1951.
[5] D. S. Lebedev and E. I. Piil', *Statistical investigation of television communication*, Report LPS Academy of Sciences SSSR, 1957. (In Russian.)

ON A STATISTICAL ESTIMATE FOR THE ENTROPY OF A SEQUENCE OF INDEPENDENT RANDOM VARIABLES

G. P. BASHARIN *(MOSCOW)*

(Summary)

The mean value and variance are computed for a statistical estimate for the entropy of a sequence of mutually independent random variables having a similar distribution. The estimate is shown to be biased, consistent and asymptotically normal.

REVIEWS AND BIBLIOGRAPHIES

(Translated by Newcomb Greenleaf)

Marek Fisz, *Rachunek Prawdopodobieństwa i Statistika Matematyczna*, Warsaw, Państwowe Wydawnictwo Naukowe, 1958, 530 pp. (Polish.) Simultaneously published as: *Wahrscheinlichkeitsrechnung und Mathematische Statistik*, Berlin, VEB Deutscher Verlag der Wissenschaften, 1958, 528 pp. (German.)

A second edition of a text on the theory of probability and mathematical statistics by Professor M. Fisz of Warsaw University appeared recently as No. 18 of the well-known Polish series "Biblioteka Matematiczna". The book has undergone considerable revisions in comparison with the first edition published by the same publisher four years ago. It will suffice now to say that more than 150 pages have been added; there are three new chapters (Markov chains, Stochastic