SAMPLING ENTROPY FOR RANDOM HOMOGENEOUS SYSTEMS WITH COMPLETE CONNECTIONS

By Marius Iosifescu

Academy of the Rumanian People's Republic

Centre of Mathematical Statistics

In this note we derive the asymptotic behaviour of the sampling entropy for random homogeneous systems with complete connections with a finite set of states.

1. Let $X = (i)_{1 \leqq i \leqq r}$ be a finite set and W an arbitrary set. For every $i \varepsilon X$ let u_i be a mapping of W into itself and P_i a real-valued function defined on W such that $P_i \geqq 0$, $\sum_{i=1}^r P_i = 1$. We put $u_{i_1 \cdots i_n} = u_{i_n} \circ \cdots \circ u_{i_1}$ for $i_k \varepsilon X$, $1 \leqq k \leqq n$.

After [6] the mappings u_i and the functions P_i determine a random homogeneous system with complete connections; this concept contains as particular cases the simple or multiple chains with complete connections ([2], [10]), the chains of infinite order ([2], [3]), the stochastic models for learning ([1]) and the random automata ([9]). For every $c \varepsilon W$ there exist [6] a probability space $(\Omega, \mathcal{K}, \mathcal{P}_c)$ and a sequence of random variables $(\xi_n)_{n \varepsilon N^*}$, $N^* = \{1, 2, \cdots\}$, defined on Ω and with values in X, such that

$$\mathcal{P}_c(\xi_1(\omega) = i) = P_i(c)$$

$$\mathcal{P}_c(\xi_{n+1}(\omega) = i \mid \xi_n(\omega) = i_n, \cdots, \xi_1(\omega) = i_1) = P_i(u_{i_1 \ldots i_n}(c))$$

for any $n \varepsilon N^*$, $i \varepsilon X$, $(i_1 \cdots i_n) \varepsilon X^{(n)}$, where $X^{(n)}$ is the nth cartesian product of the set X.

For every $2 \leqq l \varepsilon N^*$, $(i_1 \cdots i_l) \varepsilon X^{(l)}$ let $P_{i_1 \ldots i_l}$ be the function defined on W by the relation

$$P_{i_1 \ldots i_l}(c) = P_{i_1}(c) P_{i_2}(u_{i_1}(c)) \cdots P_{i_l}(u_{i_1 \ldots i_{l-1}}(c)).$$

For every l and $n \varepsilon N^*$, $(i_1 \cdots i_l) \varepsilon X^{(l)}$ let $P_{i_1 \cdots i_l}^{(n)}$ be the function defined on W by the relations

$$P_{i_1 \cdots i_l}^{(n)} = P_{i_1 \cdots i_l}, \qquad \text{if } n = 1,$$

$$P_{i_1 \cdots i_l}^{(n)}(c) = \sum_{i=1}^r P_i(c) P_{i_1 \cdots i_l}^{(n-1)}(u_i(c)), \qquad \text{if } n > 1.$$

We have [6]

$$P_{i_1 \cdots i_l}^{(n)}(c) = \mathcal{P}_c(\xi_n(\omega) = i_1, \cdots, \xi_{n+l-1}(\omega) = i_l).$$

We set

$$a_n = \sup |P_i(u_{i_1 \cdots i_n}(c')) - P_i(u_{i_1 \cdots i_n}(c''))|$$

the upper bound being taken over all $c', c'' \varepsilon W$, $i \varepsilon X$, $(i_1 \cdots i_n) \varepsilon X^{(n)}$.

Received 27 April 1964; revised 20 March 1965.

If $\sum_{n\varepsilon N^*} a_n < \infty$ and there exist $\delta > 0$ such that for any $l \varepsilon N^*$ and any partition $A_1^{(l)} \cup A_2^{(l)}$ of $X^{(l)}$

$$(1) \qquad \sum_{(i_1 \cdots i_l)\varepsilon A_1^{(l)}} P_{i_1 \cdots i_l}(c) > \delta \qquad \text{for any } c \varepsilon W \text{ or}$$
$$\sum_{(i_1 \cdots i_l)\varepsilon A_2^{(l)}} P_{i_1 \cdots i_l}(c) > \delta \qquad \text{for any } c \varepsilon W$$

then for any $l \varepsilon N^*$ there is [6] a probability $P^{\infty}_{i_1 \cdots i_l}$ defined on $X^{(l)}$ such that

$$\left|\sum_{(i_1 \cdots i_l)\varepsilon A^{(l)}} [P^{(n)}_{i_1 \cdots i_l}(c) - P^{\infty}_{i_1 \cdots i_l}]\right| \leqq \inf_{1 \leqq s \leqq n} [(3 \sum_{j \geqq s} a_j/\delta) + (1-\delta)^{(n/s)-1}]$$

for every $l, n \varepsilon N^*$, $c \varepsilon W$, $A^{(l)} \subset X^{(l)}$.

A simpler condition which implies the Condition (1) is the following ([5], [6]): there exists $i_0 \varepsilon X$, $\alpha > 0$ and $k \varepsilon N^*$ such that

$$(2) \qquad P_{i_0}(u_{i_1 \cdots i_k}(c)) > \alpha$$

for any $c \varepsilon W$, $i_\mu \varepsilon X$, $1 \leqq \mu \leqq k$. The Condition (2) implies the Condition (1) with $\delta = \frac{1}{4}\alpha^u$ where u is chosen such that $\sum_{n \geqq u} a_n \leqq \frac{1}{8}$.

In the following we suppose that

$$\sum_{n\varepsilon N^*} \inf_{1 \leqq s \leqq n} (\sum_{j \geqq s} a_j + (1-\delta)^{n/s}) < \infty.$$

2. We put for every $(i_1, \cdots i_l) \varepsilon X^{(l)}$ and $l, n \varepsilon N^*$

$$\nu_{n, i_1 \cdots i_l} = \sum_{k=1}^{n} \chi_{i_1 \cdots i_l}(\xi_k \cdots \xi_{k+l-1})$$

where $\chi_{i_1 \cdots i_l}$ is the indicator of the element $(i_1 \cdots i_l)$ of $X^{(l)}$.

The random variable

$$H_{n,l} = -l^{-1} \sum_{(i_1 \cdots i_l)\varepsilon X^{(l)}} (\nu_{n, i_1 \cdots i_l}/n) lg(\nu_{n, i_1 \cdots i_l}/n)$$

represents a sampling entropy for a sample of size $n + l - 1$ in the sequence $(\xi_n)_{n\varepsilon N^*}$.

We set

$$\sigma_l^2 = l^{-2}[\sum_{(i_1 \cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1 \cdots i_l}\, lg^2\, P^{\infty}_{i_1 \cdots i_l} - (\sum_{(i_1 \cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1 \cdots i_l}\, lg\, P^{\infty}_{i_1 \cdots i_l})^2$$
$$+ 2\sum_{h\varepsilon N^*} \{\sum_{(i_1 \cdots i_{h+l})\varepsilon X^{(h+l)}} P^{\infty}_{i_1 \cdots i_{h+l}}\, lg\, P^{\infty}_{i_1 \cdots i_l}\, lg\, P^{\infty}_{i_{h+1} \cdots i_{h+l}}$$
$$- (\sum_{(i_1 \cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1 \cdots i_l}\, lg\, P^{\infty}_{i_1 \cdots i_l})^2\}]$$

$$H_l = -l^{-1} \sum_{(i_1 \cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1 \cdots i_l}\, lg\, P^{\infty}_{i_1 \cdots i_l}.$$

We shall prove the

THEOREM 1. *For any $c \varepsilon W$, $l \varepsilon N^*$ we have*

$$\lim_{n\to\infty} \mathcal{P}_c\{n^{\frac{1}{2}}(H_{n,l} - H_l)/\sigma_l < \lambda\} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\lambda} e^{-u^2/2}\, du$$

uniformly with respect to λ if $\sigma_l \neq 0$ and

$$\lim_{n\to\infty} \mathcal{P}_c\{n^{\frac{1}{2}}(H_{n,l} - H_l) < \lambda\} = 1 \text{ for } \lambda > 0$$
$$= 0 \text{ for } \lambda \leqq 0$$

if $\sigma_l = 0$.

The proof is based essentially on the central limit theorem of [6], p. 631 (see also [8]) which in our notations may be stated as follows:

Let f be a real-valued function defined on $X^{(l)}$ and

$$\sigma^2(f) = \sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1\cdots i_l} f^2(i_1\cdots i_l) - (\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1\cdots i_l} f(i_1\cdots i_l))^2 + 2\sum_{h\varepsilon N^*}\{\sum_{(i_1\cdots i_{h+l})\varepsilon X^{(h+l)}} P^{\infty}_{i_1\cdots i_{h+l}} f(i_1\cdots i_l) f(i_{h+1}\cdots i_{h+l}) - (\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1\cdots i_l} f(i_1\cdots i_l))^2\}.$$

We have always $0 \leqq \sigma^2(f) < \infty$. For any $c \,\varepsilon\, W$ we have

$$\lim_{n\to\infty} \mathcal{P}_c\{[\sum_{k=1}^n f(\xi_k\cdots\xi_{k+l-1}) - n\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1\cdots i_l} f(i_1\cdots i_l)/\sigma(f)n^{\frac{1}{2}}] < \lambda\} = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\lambda} e^{-u^2/2}\,du$$

uniformly with respect to λ if $\sigma(f) = 0$. If $\sigma(f) = 0$ the random variable

$$\sum_{k=1}^n f(\xi_k\cdots\xi_{k+l-1}) - n\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1\cdots i_l} f(i_1\cdots i_l)/n^{\frac{1}{2}}$$

converges in quadratic mean to zero with respect to $\mathcal{P}_c$ as $n \to \infty$.

We deduce easily that for the theorem we want to prove we may assume that $P^{\infty}_{i_1\cdots i_l} \neq 0$, $(i_1\cdots i_l)\,\varepsilon\, X^{(l)}$. Then, we set $f(i_1\cdots i_l) = -l^{-1}\, lg\, P^{\infty}_{i_1\cdots i_l}$, $(i_1\cdots i_l)\,\varepsilon\, X^{(l)}$. We have

$$n^{-\frac{1}{2}}[\sum_{k=1}^n f(\xi_k\cdots\xi_{k+l-1}) - n\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1\cdots i_l} f(i_1\cdots i_l)] = n^{\frac{1}{2}}[-l^{-1}\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} [\nu_{n,i_1\cdots i_l}/n]\, lg\, P^{\infty}_{i_1\cdots i_l} - H_l].$$

From the theorem which is stated above it follows by choosing f adequately that $\nu_{n,i_1\cdots i_l}/n$ converges in probability with respect to $\mathcal{P}_c$ to $P^{\infty}_{i_1\cdots i_l}$ as $n \to \infty$. Thus

$$n^{\frac{1}{2}}(H_{n,l} - H_l) - n^{\frac{1}{2}}[-l^{-1}\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} (\nu_{n,i_1\cdots i_l}/n)\, lg\, P^{\infty}_{i_1\cdots i_l} - H_l] = l^{-1}\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} (\nu_{n,i_1\cdots i_l}/n^{\frac{1}{2}})(lg\, P^{\infty}_{i_1\cdots i_l} - lg\,(\nu_{n,i_1\cdots i_l}/n))$$

converges in probability with respect to $\mathcal{P}_c$ to zero as $n \to \infty$. (We have even the almost sure convergence ([7], [8]); consequently $H_{n,l}$ converges to H_l almost surely as $n \to \infty$.) The theorem is proved.

REMARK 1. The problem of determining a point estimate for the asymptotic entropy

$$H = -\lim_{l\to\infty} l^{-1}\sum_{(i_1\cdots i_l)\varepsilon X^{(l)}} P^{\infty}_{i_1\cdots i_l}\, lg\, P^{\infty}_{i_1\cdots i_l}$$

remains open in the general case.

For an ergodic finite Markov chain with the transition probabilities p_{ij}, $1 \leqq i, j \leqq r$ we have ([11])

$$H = -\sum_{i,j=1}^r \pi_i p_{ij}\, lg\, p_{ij}$$

where $(\pi_i)_{1\leqq i\leqq r}$ represents the stationary absolute distribution ($\sum_{i=1}^r \pi_i p_{ij} = \pi_j$, $1 \leqq j \leqq r$) of the considered chain and the sampling entropy corresponding to H is

$$\mathbf{H}_n = -\sum_{1,j=1}^r (\nu_{n,ij}/n)\, lg\,(\nu_{n,ij}/\nu_{n,i}).$$

This case may be obtained taking $W = \{\mathbf{p} = (p_i)_{1 \leqq i \leqq r} \mid p_i \geqq 0, 1 \leqq i \leqq r, \sum_{i=1}^{r} p_i = 1\}$, $u_i(\mathbf{p}) = (p_{ij})_{1 \leqq j \leqq r}$, $P_i(\mathbf{p}) = p_i$, $1 \leqq i \leqq r$. The central limit theorem stated above (see also [4]) permits by choosing $f(ij) = -lg\ p_{ij}$ and setting

$$\sigma^2 = \sum_{i,j=1}^{r} \pi_i p_{ij} lg^2 p_{ij} - H^2$$
$$+ 2 \sum_{h \varepsilon N^*} \{ \sum_{(i_1 \cdots i_{h+2}) \varepsilon X^{(h+2)}} \pi_{i_1} p_{i_1 i_2} \cdots p_{i_{h+1} i_{h+2}} lg\ p_{i_1 i_2}\, lg\ p_{i_{h+1} i_{h+2}} - H^2 \}$$

to obtain the

THEOREM 2. *For every initial probability distribution* $\mathbf{p}$ *we have*

$$\lim_{n\to\infty} \mathcal{P}_{\mathbf{p}}\{[n^{\frac{1}{2}}(\mathbf{H}_n - H)/\sigma] < \lambda\} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\lambda} e^{-u^2/2}\, du$$

uniformly with respect to λ *if* $\sigma \neq 0$ *and*

$$\lim_{n\to\infty} \mathcal{P}_{\mathbf{p}}\{n^{\frac{1}{2}}(\mathbf{H}_n - H) < \lambda\} = 1 \quad \text{for} \quad \lambda > 0$$
$$= 0 \quad \text{for} \quad \lambda \leqq 0$$

if $\sigma = 0$.

REMARK 2. The case of independent identically distributed observations (which has been considered also by G. P. Basharin (*Theory Prob. Applications*, 4, 1959, 361–364) can be obtained taking $W = \{\mathbf{p} = (p_i)_{1 \leqq i \leqq r} \mid p_i \geqq 0, 1 \leqq i \leqq r, \sum_{i=1}^{r} p_i = 1\}$, $u_i(\mathbf{p}) = \mathbf{p}$, $P_i(\mathbf{p}) = p_i$, $1 \leqq i \leqq r$. In this case

$$H_l = H = -\sum_{i=1}^{r} p_i\, lg\ p_i$$
$$\sigma_l^2 = \sigma^2 = \sum_{i=1}^{r} p_i\, lg^2\ p_i - H^2, \qquad l\, \varepsilon\, N^*.$$

Particularly, for $r = 2$, $\sigma^2 = p_1 p_2\, lg^2\ p_1/p_2$.

REFERENCES

[1] BUSH, R. R. and MOSTELLER, F. (1955). *Stochastic Models for Learning*. Wiley, New York.

[2] CIUCU, G. and THEODORESCU, R. (1960). *Processes with complete connections*. (In Rumanian.) Acad. R. P. R. Bucharest.

[3] DOEBLIN, W. and FORTET, R. (1937). Sur les chaînes à liaisons complètes. *Bull. Soc. Math. France* **65** 132–148.

[4] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[5] IONESCU TULCEA, C. T. (1959). On a class of operators occurring in the theory of chains of infinite order. *Canad. J. Math.* **11** 112–121.

[6] IOSIFESCU, M. (1963). Random systems with complete connections with an arbitrary set of states. (In Russian.) *Rev. Math. Pures Appl.* **8** 611–645.

[7] IOSIFESCU, M. (1964). Sur la loi forte des grands nombres pour les systèmes aléatoires homogènes à liaisons complètes à un ensemble quelconque d'états. *C. R. Acad. Sci. Paris* **258** 4421–4423.

[8] IOSIFESCU, M. (1965). Quelques propriétés asymptotiques des systèmes aléatoires homogènes à liaisons complètes. *Rev. Roumaine Math. Pures Appl.* **10** (In print).

[9] ONICESCU, O. and GUIASU, S. (1965). Abstract finite random automata. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 3, 279–285.

[10] ONICESCU, O. and MIHOC, G. (1935). Sur les chaînes de variables statistiques. *Bull. Sci. Math.* **59** 174–192.

[11] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27** 379–424, 623–657.