The Annals of Probability
1980, Vol. 8, No. 1, 142–147

A CONDITIONAL LAW OF LARGE NUMBERS

BY OLDRICH ALFONSO VASICEK

Wells Fargo Bank

It is shown that, when conditional on a set of given average values, the frequency distribution of a series of independent random variables with a common finite distribution converges in probability to the distribution which has the maximum relative entropy for the given mean values.

1. Introduction. In statistical mechanics and other areas of physics, empirical distributions in the phase space conform in many circumstances to the distribution maximizing the entropy of the system subject to its constraints. The constraints are typically in the form of specified mean values of some functions of phase. If $p = (p_1, p_2, \cdots, p_k)$ denotes the probability distribution over the state space, the constraints on p take the form

$$\sum_{i=1}^{k} a_{ji} p_i = c_j, \qquad j = 1, 2, \cdots, r,$$

and the maximum entropy distribution is the one that maximizes the entropy function

$$-\sum_{i=1}^{k} p_i \log p_i$$

subject to the constraints.

A principle stating that the empirical distribution possesses the maximum entropy within the restrictions of the system is due to Gibbs (1902). As a special case, he proposed the so-called canonical distribution as a description of systems subject to a single constraint that the average energy has a fixed value,

$$\sum_{i=1}^{k} a_i p_i = c,$$

where $a_1, a_2, \cdots, a_k$ are the energy levels of each state. In this case, the maximum entropy distribution has the form

$$p_i = \exp(\nu + \lambda a_i), \qquad i = 1, 2, \cdots, k,$$

which is the form that Gibbs called canonical.

Gibbs offered no justification for the canonical distribution, and the principle of maximum entropy in general. In spite of its apparent arbitrariness, however, the maximum entropy principle has since found a number of successful applications in a wide range of situations, and has led to many new developments in physics. For an informed discussion, see Jaynes (1967).

In a subsequent paper, Jaynes (1968) presented a demonstration that the distribution with the maximum entropy "can be realized experimentally in overwhelm-

Received February 28, 1978.

AMS 1970 *subject classifications.* Primary 60F05, 60F99; secondary 82A05.

Key words and phrases. Limit theorems, laws of large numbers, conditional convergence, entropy, maximum entropy principle.

ingly more ways than can any other." Therefore, for large physical systems, the empirical distribution should, indeed, agree with the maximum entropy distribution.

In this note, a limit theorem is given which provides a foundation for the above physical principle in the same sense in which the law of large numbers justifies interpretation of limiting frequencies as probabilities. Informally stated, the theorem asserts that in the equiprobable case, the frequencies conditional on given constraints converge in probability to the distribution that has the maximum entropy subject to these constraints.

A generalization of this result is also given, which relaxes the assumption of all states being equally likely. In the general case, the frequencies conditional upon a set of conditions converge to the distribution that maximizes the entropy relative to the underlying distribution.

2. The limit theorems. Let $\mathcal{X} = (x_1, x_2, \cdots, x_k)$ be a finite set of k elements and consider a series $X_1, X_2, \cdots$ of independent identically distributed random variables with values on $\mathcal{X}$, such that

$$(1) \qquad P[X_1 = x_i] = 1/k, \qquad i = 1, 2, \cdots, k.$$

Denote by $f_n = (f_{n1}, f_{n2}, \cdots, f_{nk})$, $n = 1, 2, \cdots$ the frequency distribution of $X_1, X_2, \cdots, X_n$,

$$f_{ni} = \frac{1}{n}\Sigma_{m=1}^{n} I[X_m = x_i], \qquad i = 1, 2, \cdots, k,$$

where I is the characteristic function. Let (a_{ji}) be a given $r \times k$ matrix and $(c_1, c_2, \cdots c_r)$ a given vector. Put $p = (p_1, p_2, \cdots, p_k)$ and define

$$(2) \qquad D_0 = [p : p \in S, \Sigma_{i=1}^{k} a_{ji} p_i = c_j, \ \ j = 1, 2, \cdots, r]$$

where S is the set of probability distributions on $\mathcal{X}$,

$$S = [p : p_i \geqslant 0, \ \ i = 1, 2, \cdots, k, \Sigma_{i=1}^{k} p_i = 1].$$

Assume that $D_0 \neq \varnothing$. Define the entropy of a distribution in S by

$$(3) \qquad H(p) = -\Sigma_{i=1}^{k} p_i \log p_i, \qquad p \in S,$$

with the convention $0 \cdot \log 0 = 0$. Denote by $p_0 = (p_{01}, p_{02}, \cdots, p_{0k})$ the maximum point of H on D_0,

$$(4) \qquad \max_{p \in D_0} H(p) = H(p_0).$$

Since H is continuous on S and D_0 is compact, the maximum exists. Moreover, it is unique by virtue of strict concavity of H on S and convexity of the set D_0.

THEOREM 1. *For every* $\varepsilon > 0$, *there exists* $\delta(\varepsilon) > 0$ *such that for every* δ, $0 < \delta \leqslant \delta(\varepsilon)$,

(5)

$$P[|f_{ni} - p_{oi}| \leqslant \varepsilon, i = 1, 2, \cdots, k \,|\, |\Sigma_{i=1}^{k} a_{ji} f_{ni} - c_j| \leqslant \delta, j = 1, 2, \cdots, r] \to 1$$

as $n \to \infty$, *where* $p_0 = (p_{01}, p_{02}, \cdots, p_{0k})$ *is the maximum entropy distribution*, $\max_{p \in D_0} H(p) = H(p_0)$.

This theorem is a special case of the more general conditional law of large numbers, which will now be stated.

Replace the assumption (1) of the equiprobable case by a general assumption that

$$(6) \qquad P[X_1 = x_i] = q_i, \qquad i = 1, 2, \cdots, k$$

where $q = (q_1, q_2, \cdots, q_k) \in S$ is a given distribution. Assume, without loss of generality, that $q_i > 0$, $i = 1, 2, \cdots, k$. Define the entropy H_q of a distribution in S relative to the distribution q by

$$(7) \qquad H_q(p) = -\Sigma_{i=1}^k p_i \log(p_i/q_i), \qquad p \in S.$$

Again let D_0 be the set in (2), $D_0 \neq \varnothing$, and replace the definition (4) of p_0 by the definition

$$(8) \qquad \max_{p \in D_0} H_q(p) = H_q(p_0).$$

Again, the maximum relative entropy point p_0 exists and is unique.

THEOREM 2. *For every* $\varepsilon > 0$, *there exists* $\delta(\varepsilon) > 0$ *such that for every* δ, $0 < \delta \leqslant \delta(\varepsilon)$,

$$(9) \quad P\big[\,|f_{ni} - p_{oi}| \leqslant \varepsilon, i = 1, 2, \cdots, k \,\big|\, |\Sigma_{i=1}^k a_{ji} f_{ni} - c_j| \leqslant \delta, j = 1, 2, \cdots, r\big] \to 1$$

as $n \to \infty$, *where* $p_0 = (p_{01}, p_{02}, \cdots, p_{0k})$ *is the distribution with the maximum entropy relative to* q,

$$\max_{p \in D_0} H_q(p) = H_q(p_0).$$

The maximum relative entropy distribution p_0 is easy to find. It is given by

$$p_{0i} = q_i \exp(\nu + \Sigma_{j=1}^r \lambda_j a_{ji}), \qquad i = 1, 2, \cdots, k,$$

where the constants ν, λ_j, $j = 1, 2, \cdots, r$ are determined by the condition $p_0 \in D_0$.

3. Proof of the theorems. Theorem 1 follows immediately from Theorem 2, since for $q = (1/k, 1/k, \cdots, 1/k)$

$$H_q(p) = H(p) - \log k$$

so that the maximum points in (4) and (8) coincide.

PROOF OF THEOREM 2. Let $\varepsilon > 0$ be fixed, and put

$$(10) \qquad V = [p : p \in S, |p_i - p_{0i}| \leqslant \varepsilon, i = 1, 2, \cdots, k]$$

where p_0 is given by (8). For each $\delta > 0$, define

$$(11) \qquad D_\delta = \big[p : p \in S, |\Sigma_{i=1}^k a_{ji} p_i - c_j| \leqslant \delta, j = 1, 2, \cdots, r\big].$$

Define uniquely a point p_δ by

$$(12) \qquad \max_{p \in D_\delta} H_q(p) = H_q(p_\delta).$$

Introduce a topology on S by the metric

$$d(u, v) = \max_{1 \leqslant i \leqslant k} |u_i - v_i|, \qquad u, v \in S.$$

We will first prove that

$$\lim_{\delta \to 0^+} p_\delta = p_0. \tag{13}$$

Let the set $\{p_\delta, \delta > 0\}$ be directed by the relation $\delta_1 \prec \delta_2$ if $\delta_1 \geqslant \delta_2$. Since S is compact, the directed set $\{p_\delta\}$ has at least one limit point. Let p^* be one such limit point. Choose an arbitrary $\delta > 0$ and put

$$\eta = \tfrac{1}{2}\delta / \Sigma_{j=1}^r \Sigma_{i=1}^k |a_{ji}|, \quad \delta' = \tfrac{1}{2}\delta.$$

There exists δ'', $0 < \delta'' < \delta'$ such that

$$\max_{1 \leqslant i \leqslant k} |p_{\delta'' i} - p_i^*| \leqslant \eta.$$

Then

$$\left|\Sigma_{i=1}^k a_{ji} p_i^* - c_j\right| \leqslant \left|\Sigma_{i=1}^k a_{ji}(p_i^* - p_{\delta'' i})\right| + \left|\Sigma_{i=1}^k a_{ji} p_{\delta'' i} - c_j\right|$$

$$\leqslant \eta \Sigma_{i=1}^k |a_{ji}| + \delta'' \leqslant \delta, \qquad j = 1, \cdots, r$$

and therefore, $p^* \in D_\delta$. Since this is true for every $\delta > 0$, it follows that

$$p^* \in D_0 = \cap_{\delta > 0} D_\delta.$$

Now $H_q(p_\delta) \geqslant H_q(p_0)$ for every $\delta > 0$. Since H_q is a continuous function, the same is true for the limiting point,

$$H_q(p^*) \geqslant H_q(p_0).$$

But p_0 is the unique maximum point of H_q on D_0, and therefore $p^* = p_0$. Thus, p_0 is the only limit point of $\{p_\delta\}$, which proves (13).

It follows that there exists $\delta(\varepsilon) > 0$ such that for every δ, $0 < \delta \leqslant \delta(\varepsilon)$,

$$|p_{\delta i} - p_{0i}| \leqslant \tfrac{1}{2}\varepsilon, \qquad i = 1, 2, \cdots, k. \tag{14}$$

Let δ be selected arbitrarily from $0 < \delta \leqslant \delta(\varepsilon)$ and fixed. Put $W = S - V$ where V is given by (10), and denote the adherence of W by $\overline{W}$. Put

$$h = \max_{p \in \overline{W} \cap D_\delta} H_q(p).$$

Since

$$\max_{p \in D_\delta} H_q(p) = H_q(p_\delta),$$

and $p_\delta \notin \overline{W}$ by virtue of (14), it follows that

$$h < H_q(p_\delta).$$

Put

$$h' = \tfrac{1}{2}(h + H_q(p_\delta))$$

so that

$$h < h', \tag{15}$$

and define

$$R = [p : p \in S, H_q(p) \geqslant h'].$$

Let

$$B = R \cap V \cap D_\delta.$$

We will now show that B contains an open set.

Let $0 < \delta' < \delta$ and put

$$s_\lambda = (1-\lambda)p_\delta + \lambda p_{\delta'}, \qquad 0 < \lambda < 1.$$

The point s_λ is an interior point of D_δ for every $0 < \lambda < 1$. To prove that, choose

$$\eta = \lambda(\delta - \delta') \Big/ \Sigma_{j=1}^{r} \Sigma_{i=1}^{k} |a_{ji}|.$$

For every $p = (p_1, p_2, \cdots, p_k)$ such that

$$|p_i - s_{\lambda i}| \leqslant \eta, \qquad i = 1, 2, \cdots, k$$

it is true that

$$\begin{aligned} \left|\Sigma_{i=1}^{k} a_{ji} p_i - c_j\right| &\leqslant \eta \Sigma_{i=1}^{k} |a_{ji}| + (1-\lambda)\left|\Sigma_{i=1}^{k} a_{ji} p_{\delta i} - c_j\right| \\ &\quad + \lambda \left|\Sigma_{i=1}^{k} a_{ji} p_{\delta' i} - c_j\right| \\ &\leqslant \lambda(\delta - \delta') + (1-\lambda)\delta + \lambda\delta' = \delta, \ j = 1, 2, \cdots, r \end{aligned}$$

so that $p \in D_\delta$. Thus, s_λ belongs to the interior of D_δ for every λ, $0 < \lambda < 1$. Since p_δ is an interior point of V and, by continuity of H_q, also of R, the point s_λ will be in the interior of both V and R if λ is sufficiently small. Thus, such s_λ is an interior point of B, and consequently B contains an open set, say C.

To summarize our results so far, we have proven that there exists an open set C such that

$$C \subset V \cap D_\delta,$$

$$H_q(p) \geqslant h' \qquad \text{for every} \quad p \in C,$$

and

$$H_q(p) \leqslant h < h' \qquad \text{for every} \quad p \in W \cap D_\delta.$$

Now

$$P\left[|f_{ni} - p_{oi}| \leqslant \varepsilon, i = 1, 2, \cdots, k \middle| |\Sigma_{i=1}^{k} a_{ji} f_{ni} - c_j| \leqslant \delta, j = 1, 2, \cdots, r\right] = \frac{1}{1 + g_n}$$

where

$$g_n = \Sigma_{f_n \in W \cap D_\delta} \frac{n!}{(nf_{n1})!\,(nf_{n2})! \cdots (nf_{nk})!} q_1^{nf_{n1}} q_2^{nf_{n2}} \cdots q_k^{nf_{nk}}$$

$$\Big/ \Sigma_{f_n \in V \cap D_\delta} \frac{n!}{(nf_{n1})!\,(nf_{n2})! \cdots (nf_{nk})!} q_1^{nf_{n1}} q_2^{nf_{n2}} \cdots q_k^{nf_{nk}}.$$

We will make use of the inequality

$$n^n e^{-n} \leqslant n! \leqslant 3(n+1)^{\frac{1}{2}} n^n e^{-n} \tag{16}$$

valid for $n \geqslant 0$, where we define $0^0 = 1$ in agreement with the earlier convention $0.\log 0 = 0$. The inequality (16) is easily established from the Stirling formula. Then

$$\begin{aligned} g_n &\leqslant \Sigma_{f_n \in W \cap D_\delta} 3(n+1)^{\frac{1}{2}} n^n (nf_{n1})^{-nf_{n1}} (nf_{n2})^{-nf_{n2}} \cdots (nf_{nk})^{-nf_{nk}} \cdot q_1^{nf_{n1}} q_2^{nf_{n2}} \cdots q_k^{nf_{nk}} \\ &\quad \Big/ \Sigma_{f_n \in V \cap D_\delta} 3^{-k} (nf_{n1}+1)^{-\frac{1}{2}} (nf_{n2}+1)^{-\frac{1}{2}} \cdots (nf_{nk}+1)^{-\frac{1}{2}} \\ &\quad \cdot n^n (nf_{n1})^{-nf_{n1}} (nf_{n2})^{-nf_{n2}} \cdots (nf_{nk})^{-nf_{nk}} q_1^{nf_{n1}} q_2^{nf_{n2}} \cdots q_k^{nf_{nk}} \\ &\leqslant 3^{k+1}(n+1)^{(k+1)/2} \Sigma_{f_n \in W \cap D_\delta} \exp(nH_q(f_n)) \Big/ \Sigma_{f_n \in V \cap D_\delta} \exp(nH_q(f_n)) \\ &\leqslant 3^{k+1}(n+1)^{(k+1)/2} \Sigma_{f_n \in W \cap D_\delta} \exp(nH_q(f_n)) \Big/ \Sigma_{f_n \in C} \exp(nH_q(f_n)) \\ &\leqslant 3^{k+1}(n+1)^{(k+1)/2} \exp(-n(h'-h)) \frac{\#[f_n : f_n \in W \cap D_\delta]}{\#[f_n : f_n \in C]}, \end{aligned}$$

and therefore

$$g_n \leqslant 3^{k+1}(n+1)^{(k+1)/2} \exp(-n(h'-h)) \frac{\#[f_n : f_n \in S]}{\#[f_n : f_n \in C]} \tag{17}$$

where $\#[Z]$ denotes the number of elements of a finite set Z. Now

$$\frac{\#[f_n : f_n \in S]}{\#[f_n : f_n \in C]}$$

converges with $n \to \infty$ to a finite limit $\mu(S)/\mu(C)$ where $\mu(S)$, $\mu(C)$ are the volumes of S, C, respectively, by $(k-1)$-dimensional Lebesgue measure, and $\mu(C) > 0$. Since $h' - h > 0$, the right-hand side of (17) converges to zero as $n \to \infty$, and consequently

$$P\Big[|f_{ni} - p_{0i}| \leqslant \varepsilon, i = 1, 2, \cdots, k \big| |\Sigma_{i=1}^k a_{ji} f_{ni} - c_j| \leqslant \delta, j = 1, 2, \cdots, r\Big] \to 1,$$

which completes the proof.

REFERENCES

GIBBS, J. W. (1902). *Elementary Principles in Statistical Mechanics*. Yale Univ. Press, New Haven, Conn.

JAYNES, E. T. (1967). Foundations of probability theory and statistical mechanics. In *Delaware Seminar in Foundation of Physics*. (Ed., M. Bunge) Springer, Berlin.

JAYNES, E. T. (1968). Prior probabilities. *IEEE Trans. System Science and Cybernetics*. **4** 227–241.

MANAGEMENT SCIENCE DEPARTMENT
WELLS FARGO BANK
475 SANSOME ST., 16TH FLOOR
SAN FRANCISCO, CA 94111