The Annals of Mathematical Statistics
1972, Vol. 43, No. 2, 541-552

THE MAXIMIZATION OF ENTROPY OF DISCRETE DENUMERABLY-VALUED RANDOM VARIABLES WITH KNOWN MEAN[1]

BY ROMAN A. MUREIKA

University of Alberta

1. Introduction. Let $S = \{u_k\}$ be a set of real numbers. A discrete random variable will be said to be S-valued if S represents the totality of all possible values of the random variable. If $P[X = u_k] = p_k$, then the entropy of X is defined to be $H_x = -\sum_k p_k \log p_k$ where the logarithm is taken to the base 2. It is of some interest to find the maximum value of H_x over the set of all S-valued random variables.

When S is a finite set consisting of n elements, $\max H_x = \log n$ [1]. However, if the set S is countably infinite, then an S-valued random variable may have infinite entropy. Since this is the case, it is natural to place some restrictions on the set of random variables and to then determine the maximum entropy. Again, if S is a finite set, this has been done and the result will be stated in Theorem 1.

Let $S = \{u_1, \cdots, u_n\}$ be a set of real numbers. Let $f_1, f_2, \cdots, f_m$ $(m < n)$ be m linearly independent real-valued functions. We define

$$Z(x_1, \cdots, x_m) = \sum_{k=1}^{n} \exp\{-\sum_{j=1}^{m} x_j f_j(u_k)\}$$

and $H = \max H_x$ where the maximum is taken over the set of S-valued random variables under the condition that $Ef_j = f_j^{(0)}$, $j = 1, 2, \cdots, m$, for a fixed collection of numbers $\{f_j^{(0)}\}$.

THEOREM 1. (i) $H = \sum_{j=1}^{m} \hat{x}_j f_j^{(0)} + \log Z(\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_m)$ *where* $(\hat{x}_1, \cdots, \hat{x}_m)$ *is the unique solution of the set of equations*

$$\frac{\partial}{\partial x_j}[\log Z(x_1, \cdots, x_m)] = -f_j^{(0)}, \qquad j = 1, 2, \cdots, m$$

(ii) $H = -\sum_{k=1}^{n} p_k \log p_k$ *and* $\sum_{k=1}^{m} p_k f_j(u_k) = f_j^{(0)}$, $j = 1, 2, \cdots, m$ *if and only if*

$$p_k = \exp\{-\lambda - \sum_{j=1}^{m} \hat{x}_j f_j(u_k)\}, \qquad k = 1, 2, \cdots, n$$

where $\lambda = \log Z(\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_m)$.

A proof of this result may be found in [2] or [3]. We will use it in the case $m = 1$ and $f(x) = x$.

Received January 30, 1970.

[1] This forms a part of the author's dissertation directed by Professor Eugene Lukacs and submitted in partial satisfaction of the requirements of a graduate program in Mathematics for the degree of Doctor of Philosophy at the Catholic University of America, Washington, D.C. This research was supported in part by NSF Grant GP-6175.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org