The Annals of Probability
1985, Vol. 13, No. 1, 314–317

ON THE MAXIMUM OF A MEASURE OF DEVIATION FROM INDEPENDENCE BETWEEN DISCRETE RANDOM VARIABLES

BY ZVI GILULA AND GIDEON SCHWARZ

Hebrew University of Jerusalem

The squared n^k-dimensional Euclidean distance f_k between a given joint distribution of k random variables with values in $1, \cdots, n$ and the joint distribution of independent variables with the same respective marginals has been suggested as a measure of dependence. The following facts are established for M_k, the maximum of f_k over all joint distributions for fixed k: (1) M_k is attained among the distributions with all k variables equal to a variable X that takes on just two values. (2) For $k \le 6$, $M_k = \frac{1}{2} - (\frac{1}{2})^k$ is attained when the distribution of X is $\{\frac{1}{2}, \frac{1}{2}\}$. (3) For $k \ge 7$, M_k is not attained at $\{\frac{1}{2}, \frac{1}{2}\}$ and strictly exceeds $\frac{1}{2} - (\frac{1}{2})^k$. (4) For $k \to \infty$, the distributions of X where M_k is attained approach $\{0, 1\}$, and $M_k \nearrow 1$.

1. Introduction. Let $X_1, \cdots, X_k$ $(k \ge 2)$ be discrete random variables ranging over the finite sets $C_1, \cdots, C_k$ respectively, where, without loss of generality, we assume

$$C_j = \{1, 2, \cdots, n_j\}, \quad j = 1, \cdots, k, \quad n_j \ge 2.$$

Let

$$P(i, j) = P(X_j = i), \quad j = 1, \cdots, k, \quad i = 1, \cdots, n_j,$$
$$P(i_1, \cdots, i_k) = P(X_1 = i_1, X_2 = i_2, \cdots, X_k = i_k),$$

and let $\mathbf{P} = (p(i_1, \cdots, i_k))$.

Consider the function

$$f_k(\mathbf{P}) = \sum_{i_1,\cdots,i_k} [P(i_1, \cdots, i_k) - \prod_{j=1}^k P(i_j, j)]^2.$$

If $\mathbf{P}$ is regarded for each joint distribution as a vector in Euclidean $(n_1 n_2 \cdot \cdots \cdot n_k$-dimensional) space, then $f_k(\mathbf{P})$ is the squared distance between $\mathbf{P}$ and the corresponding independent distribution, and as such, $f_k(\mathbf{P})$ is a measure of deviation from independence between $X_1, \cdots, X_k$. This function arises in the literature regarding the establishment of a measure of complete dependence between discrete random variables (Cramér 1924, Goodman and Kruskal 1954, Lancaster 1963, and Gilula 1981). The function $f_k(\mathbf{P})$ vanishes if and only if $X_1, \cdots, X_k$ are independent.

In order to interpret the value that f_k attains for a given k-dimensional joint distribution, one naturally would like to compare it with the maximal value that f_k can attain for that k. Ideally, a measure of complete dependence should attain its maximum if and only if the variables are completely dependent. Cramér

Received February 1984; revised May 1984
AMS 1980 *subject classifications.* Primary 62H20; secondary 62H05.
Key words and phrases: Deviations from independence, upper bounds.

(1924), who considered the bivariate case ($k = 2$), showed that Max $f_2 = 1/4$, and the maximum is attained when the two variables are two-valued, completely dependent, and take on their values with probability 1/2 each. Otherwise, f_2 is strictly less than 1/4 (even for completely dependent variables).

We study here the maximum of f_k for general k. First, we reduce the search for a maximum to the case where all k variables are equal to one variable, say x, that takes on two values only, with probabilities, say p and $1 - p$. Denote $f_k(\mathbf{P})$ for such a distribution by $f_k(p)$. Next, we show that when $k \leq 6$, f_k is unimodal, attaining the maximum $1/2 - (1/2)^k$ at $p = 1/2$; when $k > 6$, the maximum is attained elsewhere, and strictly exceeds $1/2 - (1/2)^k$. For k approaching infinity, the pairs $\{p_k, 1 - p_k\}$ where the maximum of f_k is attained approach $\{0, 1\}$, and the maximum itself approaches 1.

Summarizing, we obtain the following picture. Cramér's result, that for $k = 2$ the maximum of f_k is attained for completely dependent variables that take on two values with probability 1/2 each, is valid also for $k = 3$, 4, 5 and 6. Beyond $k = 6$, a surprising change occurs: the maximum is still attained for completely dependent two-valued variables, but the probabilities are no longer (1/2, 1/2). Another result is that for large k the maximum is attained close to 0 and to 1, while at (0, 1) f_k vanishes. This reflects the fact that degenerate variables, being both independent and completely dependent, form a singularity of the complete dependence concept (compare with Kimeldorf and Sampson, 1978). The difficulty is therefore inherent in the concept of complete dependence rather than in the particular measure chosen to quantify it.

2. Upper Bounds on $f_k(P)$. By Cramér (1924), in maximizing $f_k(\mathbf{P})$ attention can be confined to the case where all k variables considered are equal so that all sets C_j are the same, namely $C_j = \{1, \cdots, n\}$ $j = 1, \cdots, k$, $n \geq 2$ and

$$P(i_1, \cdots, i_k) = \begin{cases} p_i & \text{if } \; i_1 = i_2 = , \cdots, = i_k = i, \quad (i = 1, \cdots, n) \\ 0 & \text{otherwise,} \end{cases}$$

where also $P(X_j = i) = p_i$, $j = 1, \cdots k$, $i = 1, \cdots, n$, $\sum_{i=1}^{k} p_i = 1$.

Let $\mathbf{p} = (p_1, \cdots, p_n)$. Then

$$f_k(\mathbf{p}) = \sum_{i=1}^{n} p_i^2 - 2 \sum_{i=1}^{n} p_i^{k+1} + (\sum_{i=1}^{n} p_i^2)^k.$$

We will now show that in maximizing $f_k(\mathbf{p})$, attention can be further confined to the case where sets C_j contain only two elements, or equivalently, where the k variables considered are all Bernoulli variables.

$f_k(\mathbf{p})$ is continuous in the compact simplex $\sum_{i=1}^{n} p_i = 1$, $p_i \geq 0$ and therefore attains its maximum there. Let $\mathbf{q}$ be an interior point, of the simplex, different from its centroid $(1/n, \cdots, 1/n)$. Assuming $n \geq 3$, there exists a vector $\mathbf{u}$, orthogonal to $(1, \cdots, 1)$ and to $\mathbf{q}$, but not to $(q_1^k, \cdots, q_n^k)$.

As is easily seen, the gradient of $f_k(q)$ at q is a linear combination of $\mathbf{q}$ and $(q_1^k, \cdots, q_n^k)$ with nonzero coefficients. Therefore, the directional derivative of $f_k(\mathbf{p})$ at $\mathbf{q}$ in direction $\mathbf{u}$ will not be zero, and the maximum is not attained there.

The centroid is also ruled out as a maximum point, since $f_k(0, 1/(n-1), \cdots, 1/(n-1))$ is easily shown to exceed $f_k(1/n, \cdots, 1/n)$.

We must conclude then that for $\mathbf{p}$ to be a maximum point for f_k in the simplex $\sum_{i=1}^{n} p_i = 1$, $n \geq 3$, it must be a boundary point, namely at least one coordinate of $\mathbf{p}$ must be zero. By induction we obtain the desired reduction to the case where $n = 2$.

Attention is now focused, therefore, on the function

$$f_k(p) = p^2 + (1-p)^2 + [p^2 + (1-p)^2]^k - 2p^{k+1} - 2(1-p)^{k+1}.$$

This function is a polynomial in p, that is symmetric around $p = ½$. Therefore it is a polynomial in $X = (p - ½)^2$. As X increases from 0 to ¼, $f_k(p)$ goes through its range. Carrying out the substitution yields

$$f_k(p) = g_k((p - ½)^2),$$

$$g_k(X) = ½ + 2X + \textstyle\sum_{i=0}^{k} 2^{2i-k}\binom{k}{i}X^i - \sum_{j=0}^{(k+1)/2} 2^{2j-k+1}\binom{k+1}{2j}X^j.$$

Let g_k', g_k'', g_k''' denote the first, second, and third derivative of $g_k(X)$. While f_k is stationary at $p = ½$, this is not the case for g_k at the corresponding $X = 0$. Indeed, $g_k'(0)$, the linear coefficient, is $2(1 - (k^2/2^{k-1}))$. This is negative for $k = 2, 3, 4, 5, 6$ and positive for $k \geq 7$. So g_k is increasing at 0 if $k \geq 7$, and decreasing if $2 \leq k \leq 6$. Clearly, f_k does *not* attain its maximum at $p = ½$ when $k \geq 7$. To establish that, if $2 \leq k \leq 6$, g_k and f_k *do* attain their maxima at $X = 0$ and $p = ½$ respectively, consider g_k'''. It is a polynomial whose coefficients will all be positive if this is the case for the coefficients of order ≥ 3 of g_k.

Since the expression for g_k involves minus signs only for terms of order $\leq [(k+1)/2]$, the only coefficients of g_k of order ≥ 3 that could be negative, when $k \leq 6$, are the third order coefficients of g_5 and of g_6. Since they turn out to be positive, g_k''' is positive on $0 < X$ for $2 \leq k \leq 6$.

By $g_k''' > 0$, g_k'' can at most have a simple zero on the positive axis, and g_k' at most two, when $2 \leq k \leq 6$.

At $p = 0$, f_k and its derivative vanish. By symmetry, this holds at $p = 1$ as well. This is a regular point of $X = (p - ½)^2$, hence g_k and g_k' vanish at $X = ¼$. For $0 \leq k \leq 6$, g_k' can vanish only once more on the positive axis. But since $g_k'(0)$ is negative, g_k cannot start up again to a maximum, and then decrease to zero, without incurring two points of zero derivative. Hence there is no maximum, not even a local one, except for $X = 0$, where $g_k(X) = ½ - (½)^k$ for $k \leq 6$. We conclude that $f_k(p)$ is unimodal on [0, 1] and its maximum is $½ - (½)^k$, $2 \leq k \leq 6$.

For $k \geq 7$, no explicit forms for the location and value of Max f_k were obtained; however, a study of the limiting behavior of f_k, and some numerical results serve to complete the picture.

Note that for $0 < p < 1$, $\lim f_k(p) = p^2 + (1-p)^2$, and $\max_{0 \leq p \leq 1} p^2 + (1-p)^2 = 1$ is attained at $p = 0$ or $p = 1$. Clearly, therefore, 1 is the least upper bound of $f_k(p)$, over all p and k. In fact, we also have $\lim_{k\to\infty} \mathrm{Max}_{0 \leq p \leq 1} f_k(p) = 1$. This follows from the existence of sequences (ε_k) such that $\lim_{k\to\infty} f_k(\varepsilon_k) = 1$: choose any sequence (ε_k) such that

$$\varepsilon_k \to 0 \quad \text{but} \quad k\varepsilon_k \to \infty$$

(for example, $\varepsilon_k = k^{-1/2}$). Such a choice of ε_k implies, as $k \to \infty$, (a) $\varepsilon_k^2 \to 0$, (b) $(1 - \varepsilon_k)^2 \to 1$, (c) $\varepsilon_k^{k+1} \to 0$, (d) $(1 - \varepsilon_k)^{k+1} \to 0$, (e) $[(1 - \varepsilon_k)^2 + \varepsilon_k^2]^k \to 0$. While results (a), (b), (c) are immediate, (d) and (e) follow since

$$(1 - c\varepsilon_k)^k = [(1 - c\varepsilon_k)^{1/\varepsilon_k}]^{k\varepsilon_k},$$

but $(1 - c\varepsilon_k)^{1/\varepsilon_k} \to e^{-c}$ and $(e^{-c})^{k\varepsilon_k} \to 0$.

Numerical calculations of f_k seemed to indicate that f_k is bimodal, and its larger mode p_k is of the form $k^{-1/(ak-b)}$. When $(\log k)/(\log p_k)$ was plotted against k for $k = 7, \cdots, 25$ the line $k - 3$ gives a very close fit. Thus $p_k = k^{-1/(k-3)}$ and $q_k = 1 - k^{-1/(k-3)}$ are empirical formulas for the modes of the f_k for $k \geq 7$.

Acknowledgement. The authors wish to express their gratitude to Lawrence Brown, Shelby J. Haberman and Yosef Rinott for their helpful comments and suggestions.

REFERENCES

[1] CRAMÉR, H. (1924). Remarks on correlation. *Scand. Actuar. J.* **16** 220–240.

[2] GILULA, Z. (1981). A note on measuring the degree of complete dependence between two discrete random variables measured on a nominal scale. *Comm. Statist. Theory Methods* **21** 2047–2055.

[3] GOODMAN, L. A. and KRUSKAL, W. H. (1954). Measures of association for cross-classifications. *J. Amer. Statist. Assoc.* **49** 732–764.

[4] KIMELDORF, G. and SAMPSON, A. R. (1978). Monotone dependence. *Ann. Statist.* **6** 895–903.

[5] LANCASTER, H. O. (1963). Correlation and complete dependence of random variables. *Ann. Math. Statist.* **34** 1315–1321.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVE.
CHICAGO, ILLINOIS 60637

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY
JERUSALEM, ISRAEL