Sankhyā : The Indian Journal of Statistics
1987, Volume 49, Series A, Pt. 1, pp. 58–62.

ON A CONJECTURE OF KARLIN IN SAMPLING THEORY

By SUBIR KUMAR BHANDARI
Indian Statistical Institute

SUMMARY. In 1974 Karlin conjectured that expectations of functions in a certain class, under random replacement schemes are ordered according to the coordinatewise ordering of the replacement probabilities. Krafft and Schaefer (1984) proved it for sufficiently large population size (N) compared to the sample size (n). In this paper a simple proof of this result is given along with a quantification of "sufficiently large N".

1. Introduction

A random replacement sampling plan $R(p_1, p_2, \ldots, p_{n-1})$ is a scheme for drawing a sample of n units from a population of N (distinct) units such that the i-th unit is drawn at random from the remaining and the probabilities of replacing the i-th unit (sampled) into the population is p_i.

Karlin (1974) conjectured that for all $N \geqslant n$ and for all φ satisfying the following condition K

$$(\bullet) \qquad E_{R(p)}(\varphi) \leqslant E_{R(p')}(\varphi)$$

if, and only if, $p_i \leqslant p'_i$ for all i, where $p = (p_1, \ldots, p_{n-1})$ and $p' = (p'_1, \ldots, p'_{n-1})$.

Condition K : A function $\varphi : R^n \to R$ is said to satisfy Condition K, if φ is permutationally symmetric and

$$\varphi(a, a, x_3, \ldots, x_n)+\varphi(b, b, x_3, \ldots, x_n) \geqslant 2\varphi(a, b, x_3, \ldots, x_n)$$

for all $a, b, x_3, \ldots, x_n$.

Karlin (1974) has shown that ($\bullet$) holds if $p = (0, 0, \ldots, 0)$, or if $p' = (1, 1, \ldots, 1)$ and

$$[N/(N-1)]^{n-1} \leqslant n/(n-3) \qquad \ldots \quad (1)$$

Krafft and Schaefer (1984) have shown that $p \leqslant p'$ implies

($\bullet$) if $n \leqslant 7$, or if $n \geqslant 8$ provided N is sufficiently large.

Schaefer (1987) has shown that for $n \geqslant 8$, a sufficient condition is

$$N \geqslant N_0(n) = \min[N : n(N-1)^{n-1} \geqslant (n-3)N^{n-1}]$$

for $p \leqslant p'$ to imply ($\bullet$). Furthermore, Schaefer (1987) has given an example with $n = N \geqslant 13$ for which $p \leqslant p'$ does not imply ($\bullet$).

AMS (1980) *subject classification :* Primary 62D05; Secondary 60C05, 05A20.
Key words and phrases : Random replacement Schemes.

It is easy to see that $p \leqslant p'$ does not necessarily follow from (•). (See Krafft and Schaefer, 1984).

In this paper, we have shown that $p < p'$ implies (•) if

$$N \geqslant n(n-1)/3 \equiv C_1(n), \text{ say.} \qquad \ldots \quad (2)$$

Clearly (2) implies (1); under (2), Karlin's result follows from this paper. Morever, the simple proof of our result also yield the result of Krafft and Schaefer (1984) for $n \geqslant 8$. With reference to Schaefer (1987), note that

$$\left(1-\frac{1}{N_0}\right)^{n-1} \geqslant 1-\frac{3}{n} \qquad \ldots \quad (3)$$

i.e.

$$N_0 \geqslant \frac{1}{1-\left(1-\frac{3}{n}\right)^{\frac{1}{n-1}}} \equiv C_2(n), \text{ say.}$$

The values of $C_1(n)$ and $C_2(n)$ are tabulated below:

n	8	10	15	25	50	100
$C_1(n)$	18·6	30	70	200	816·7	3300
$C_2(n)$	15·4	25·7	63·2	188·2	792·4	3250·7
percent excess of $C_1(n)$ over $C_2(n)$	19·1	16·7	10·8	6·3	3·1	1·5

The above shows that our result is slightly weaker than that of Schaefer's (1987); but the proof of our result is much simpler than that of Schaefer (1987).

2. Proof of the results

Theorem 1: *For all $p' = (1, \delta_2, \ldots, \delta_{n-1})$ and $p = (0, \delta_2, \ldots, \delta_{n-1})$ with $\delta_i = 0$ or 1, and for all φ satisfying condition K,*

$$E_{R(p')}(\varphi) \geqslant E_{R(p)}(\varphi)$$

provided $N \geqslant n(n-1)/3.$

Proof: Let $\delta_{a_1}, \ldots, \delta_{a_t}$ be the only 0's in δ_i's. Then

$$E_{R(p')}(\varphi) = \frac{1}{N^{a_1}(N-1)^{a_2-a_1} \ldots (N-t)^{n-a_t}} \Sigma'\varphi(x_{i_1}, \ldots, x_{i_n}), \qquad \ldots \quad (4)$$

$$E_{R(p)}(\varphi) = \frac{1}{N(N-1)^{a_1-1}(N-2)^{a_2-a_1} \ldots (N-t-1)^{n-a_t}} \Sigma\varphi(x_{i_1}, \ldots, x_{i_n}), \quad \ldots \quad (5)$$

where Σ' and Σ are the summations over the following sets C' and C, respectively :

$$C' = \{(x_{i_1}, \ldots, x_{i_n}) \in \Omega^n : x_{i_l} \neq x_{i_j} \quad \text{for } l > j, j \in J'\},$$

$$C = \{(x_{i_1}, \ldots, x_{i_n}) \in \Omega^n : x_{i_l} \neq x_{i_j} \quad \text{for } l > j, j \in J\},$$

where $J' = \{\alpha_1, \ldots, \alpha_l\}$, $J = \{1, \alpha_1, \ldots, \alpha_l\}$, and Ω is the set of population values.

The frequency distribution in a given n-tuple $(x_{i_1}, \ldots, x_{i_n})$ may be denoted by $(e;f) = (e_1, \ldots, e_s; f_1, \ldots, f_s)$, where exactly e_i number of x_j's occur with frequency f_i in the n-tuple, and $f_1 > \ldots > f_s$. Let

$$T(e;f) = \sum_{e;f} \varphi(x_{i_1}, \ldots, x_{i_n}),$$

where $\sum_{e;f}$ is the sum over all selections of the n-tuple of the structure

$$\left\{ \begin{array}{l} \underbrace{x_1, \ldots, x_1}_{f_1 \text{ times}}, \ldots, \underbrace{x_{e_1}, \ldots, x_{e_1}}_{f_1 \text{ times}}, \underbrace{x_{e_1+1}, \ldots, x_{e_1+1}}_{f_2 \text{ times}}, \ldots, \underbrace{x_{e_1+\ldots+e_s}, \ldots, x_{e_1+\ldots+e_s}}_{f_s \text{ times}} \end{array} \right\}$$

For a given set of $e_1 + \ldots + e_s$ distinct x's, consider an n-tuple of the above structure. Let the number of possible arrangements of the elements of this n-tuple compatible with C' and C be $C'(e;f)$ and $C(e,f)$, respectively.

It is clear that $C(e,f) > 0$ implies $C'(e,f) > 0$. Let $C'_1 \subset C'$ be the subset of these n-tuples which have a singleton in their first co-ordinate; define $C'_2 = C' - C'_1$. Consider an n-tuple in C corresponding to the frequency distribution $(e;f)$ with $f_s = 1$. Each such element of C would then generate $(n-e_s)$ elements in C'_2 by interchanging the first coordinate with any of the other coordinates which are not singletons. On the other hand, each element in C'_2 obtained in this way would be repeated e_s times in this process, since any of the e_s singletons can occupy the first position

$$(n-e_s)C(e;f) \leqslant e_s C'_2(e;f).$$

Also note that $C'_1(e;f) = C(e;f)$

$$C'(e;f) \geqslant (n/e_s)C(e;f) \qquad \ldots (6)$$

We may write

$$E_{R(p')}(\varphi) = \Sigma a'(e,f)T(e,f), \qquad \ldots (7)$$

$$E_{R(p)}(\varphi) = \Sigma a(e,f)T(e,f), \qquad \ldots (8)$$

where the summation is over all frequency distributions (e, f). Now note that

$$\left(\frac{N-1}{N}\right)^{e_1-1}\left(\frac{N-2}{N-1}\right)^{e_2-e_1}\cdots\left(\frac{N-t-1}{N-t}\right)^{n-e_t} \geqslant \frac{N-n+1}{N}. \quad \ldots (9)$$

Hence, if $e_t \leqslant n-3$

$$\frac{a'(e,f)}{a(e,f)} = \left(\frac{N-1}{N}\right)^{e_1-1}\cdots\left(\frac{N-t-1}{N-t}\right)^{n-e_t}\cdot\frac{C'(e,f)}{C(e,f)} \geqslant \frac{N-n+1}{N}\cdot\frac{n}{e_t} \geqslant 1. \quad \ldots (10)$$

Now an argument similar to that given in the proof of Theorem 3.1 of Karlin (1974, 1081–1082), can be used to complete the proof of the theorem.

Note: The condition $N \geqslant n(n-1)/3$ implies that

$$(N-t') \geqslant (n-t)(n-t-1)/3$$

for $t \geqslant t'$, and $N-t' > 0$, $n-t > 0$.

Theorem 2: *For* $p = (p_1, \ldots, p_{n-1})$ *and* $p' = (p_1', \ldots, p_{n-1}')$ *with* $0 \leqslant p_i \leqslant p_i' \leqslant 1$, $i = 1, \ldots, n-1$, *and for all* φ *satisfying condition* K,

$$E_{R(p)}(\varphi) \leqslant E_{R(p')}(\varphi), \quad \ldots (11)$$

provided $N \geqslant n(n-1)/3$.

Proof: It is sufficient to show (11) for $p \leqslant p'$, where $p_t < p_t'$ and $p_i = p_i'$ for all $i \neq t$. Now

$$E_{R(p)}(\varphi) = \Sigma\, p_1(\delta_1)\ldots p_{n-1}(\delta_{n-1})E_{R(\delta_1,\ldots,\delta_{n-1})}(\varphi),$$

$$E_{R(p')}(\varphi) = \Sigma\, p_1'(\delta_1)\ldots p_{n-1}'(\delta_{n-1})E_{R(\delta_1,\ldots,\delta_{n-1})}(\varphi),$$

where $\delta_i = 0$ or 1, $p_k(1) = p_k$, $p_k(0) = 1-p_k$, $p_k'(1) = p_k'$, $p_k'(0) = 1-p_k'$, and the above summations are over all $\delta_1, \ldots, \delta_{n-1}$.

Thus (11) will follow, if

$$E_{R(\Delta)}(\varphi) \leqslant E_{R(\Delta')}(\varphi),$$

where $\Delta = (\delta_1, \ldots, \delta_{t-1}, 0, \delta_{t+1}, \ldots, \delta_{n-1})$,

and $\Delta' = (\delta_1, \ldots, \delta_{t-1}, 1, \delta_{t+1}, \ldots, \delta_{n-1})$.

Next, note that

$$E_{R(\Delta)}(\varphi) = E_{R(\Delta)}[E_{R(\Delta)}\varphi(X_1, \ldots, X_n) \mid X_1, \ldots, X_{t-1}]$$

$$E_{R(\Delta')}(\varphi) = E_{R(\Delta')}[E_{R(\Delta')}\varphi(X_1, \ldots, X_n) \mid X_1, \ldots, X_{t-1}].$$

Since the distribution of $X_1, ..., X_{t-1}$ is the same under $R(\Delta)$ and $R(\Delta')$, it is sufficient to prove

$$E_{R(\Delta)}(\varphi(X_1, ..., X_n) \mid X_1, ..., X_{t-1}) \leqslant E_{R(\Delta')}(\varphi(X_1, ..., X_n) \mid X_1, ..., X_{t-1})$$

for all values of $X_1, ..., X_{t-1}$. After drawing $X_1, ..., X_{t-1}$, the population size N is reduced to $N-t'$ (with $t' \leqslant t-1$), and the remaining sample size becomes $n-(t-1)$. The result (11) now follows from Theorem 1 and the note after that theorem.

Acknowledgement. The author is thankful to Professor Somesh Das Gupta and Dr. Rahul Mukerjee for their help.

References

Karlin, S. (1974): Inequalities for symmetric sampling plans. I. *Ann. Statist.*, 2, 1065-1094.

Krafft, O. and Schaefer, M. (1984): On Karlin's conjecture for random replacement sampling plans. *Ann. Statist.*, 12, 1528-1535.

Marshall, A. W. and Olkin, I. (1979): *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

Schaefer, M. (1987): A counter-example to Karlin's conjecture for random replacement sampling plans. *Sankhyā*, Volume 49, Series A, Pt. 1.

Paper received : August, 1985.

Revised : April, 1986.