Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 35–40

RANDOM REPLACEMENT SCHEMES AND MULTIVARIATE MAJORIZATION[1]

BY SAMUEL KARLIN and YOSEF RINOTT
Stanford University and the Hebrew University of Jerusalem

In this note we obtain certain inequalities comparing random replacement schemes to sampling with replacement. Some of the results are related to multivariate majorization and Schur functions.

1. Various Stochastic Comparisons and Random Replacement Schemes. Let $\mathscr{A} = \{a_1, \ldots, a_N\}$, $a_i \in \mathscr{R}$ = the real line. We shall consider a sample of size n ($n \leq N$) drawn from A, and denote the observations by $X_1, \ldots, X_n$. In a symmetric random replacement scheme the observation X_1 is drawn with equal probabilities from A, i.e., $P(X_1 = a_i) = 1/N$, $i = 1, \ldots, N$. The element drawn for X_1 is replaced in A with probability π_1, and removed from A with probability $1-\pi_1$. Then X_2 is sampled, and the element which is drawn is replaced with probability π_2. Continuing to X_{n-1}, the vector $\pi = (\pi_1, \ldots, \pi_{n-1})$ defines the random replacement scheme $R(\pi)$. Note that for $\pi = \mathbf{0} = (0, \ldots, 0)$, $R(\pi)$ is equivalent to sampling without replacement while for $\pi = \mathbf{1} = (1, \ldots, 1)$, $R(\pi)$ corresponds to sampling with replacement and $X_1, \ldots, X_n$ are i.i.d.

It follows from Joag-Dev and Proschan (1983) that under $R(\mathbf{0})$, $X_1, \ldots, X_n$ are *negatively associated, i.e.*,

$$E\{\phi(X_i, i \in A)\psi(X_j, j \in B)\} \leq E\phi(X_i, i \in A)E\psi(X_j, j \in B) \tag{1.1}$$

for any partition A,B of $1, \ldots, n$, where ϕ and ψ are increasing functions.

In particular, (1.1) implies

$$E\{\Pi_{i=1}^n \varphi_i(X_i)\} \leq \Pi_{i=1}^n E\varphi_i(X_i) \tag{1.2a}$$

for any functions φ_i, all increasing (or all decreasing) and nonnegative. Note that (1.2a) can be written as

$$E_{R(\mathbf{0})}\{\Pi_{i=1}^n \varphi_i(X_i)\} \leq E_{R(\mathbf{1})}\{\Pi_{i=1}^n \varphi_i(X_i)\}. \tag{1.2b}$$

Inequalities for sampling schemes were obtained by various authors including Sen (1970), Rosén (1972), Serfling (1973), Kemperman (1973), Karlin (1974), Van Zwet (1983), and Krafft and Schaefer (preprint). The question of characterizing the class of functions for which

$$E_{R(\pi)}\psi(X_1, \ldots, X_n) \leq E_{R(\mathbf{1})}\psi(X_1, \ldots, X_n) \tag{1.3}$$

remains unresolved. The next result provides a class of functions for which (1.3) holds.

THEOREM 1. $\mathbf{E}_{R(\pi)}\{\Pi_{i=1}^n \varphi(X_i)\} \leq \mathbf{E}_{\mathbf{R}(\mathbf{1})}\{\Pi_{i=1}^n \varphi(X_i)\}$ *for any* $\varphi \geq 0$.

Proof. We write π instead of $R(\pi)$ as an index for the expectation. For $n = 2$

[1] *Supported in part by NIH Grant 5R01 GM10452-20 and NSF Grant MSC79-24310.*
AMS 1980 subject classifications. 62D05, 62H05.
Key words and phrases: Schur functions, negative association, birthday problem.

$$
\begin{aligned}
E_{\pi_1}\{\varphi(X_1)\varphi(X_2)\} &= \pi_1/N^2\,(\Sigma_{k=1}^N\varphi(a_k))^2 + \{(1-\pi_1)/(N(N-1))\}\,\Sigma_{k=1}^N\varphi(a_k)(\Sigma_{j\neq k}\varphi(a_j)) \\
(1.4)\qquad &= \pi_1/N^2\,(\Sigma_{k=1}^N\varphi(a_k))^2 + \{(1-\pi_1)/(N(N-1))\}\,\Sigma_{k=1}^N\varphi(a_k)(\Sigma_{j=1}^N\varphi(a_j)-\varphi(a_k)) \\
&= \pi_1/N^2\,(\Sigma_{k=1}^N\varphi(a_k))^2 + \{(1-\pi_1)/(N(N-1))\}\{(\Sigma_{k=1}^N\varphi(a_k))^2-\Sigma_{k=1}^N\phi^2(a_k)\}.
\end{aligned}
$$

Now $\Sigma\varphi^2(a_k) \geqslant (\Sigma\varphi(a_k))^2/N$, and therefore with $\Sigma\varphi(a_k) = m$ the last expression in (1.4) is bounded above by

$$\pi_1 m^2/N^2 + \{(1-\pi_1)/(N(N-1))\}\ \{m^2(1 - 1/N)\} = m^2/N^2 = N^{-2}(\Sigma_{k=1}^N \varphi(a_k))^2 = E_1(\varphi(X_1)\varphi(X_2)\}$$

and the case $n = 2$ is established. We now proceed by induction.

Let $\psi(X_1, \ldots, X_n) = \Pi_{i=1}^n\phi(X_i)$. Then (see Karlin (1974), Lemma 3.1)

$$
\begin{aligned}
E_{\{\pi_1,\ldots,\pi_{n-1}\}}\psi(X_1,\ldots,X_n) &= (1/N)\,\Sigma_{k=1}^N\pi_1E_{\{\pi_2,\ldots,\pi_{n-1}\}}\psi(a_k,X_2,\ldots,X_n) \\
&\quad + (1/N)\Sigma_{k=1}^N(1-\pi_1)E'_{\{\pi_2,\ldots,\pi_{n-1}\}}\psi(a_k,X_2,\ldots,X_n)
\end{aligned}
$$

where E' computes the expectation when a_k is removed from the sample space of $X_2, \ldots, X_n$. Invoking the induction hypothesis, i.e., Theorem 1 holding for n–1 variables this leads to

$$
\begin{aligned}
E_{\{\pi_1,\ldots,\pi_{n-1}\}}\psi(X_1,\ldots,X_n) &\leqslant \pi_1E_{\{1,\ldots,1\}}\psi(X_1,\ldots,X_n). \\
&\quad + (1-\pi_1)E_{\{0,1,\ldots,1\}}\psi(X_1,\ldots,X_n).
\end{aligned}
$$

Hence in order to complete the induction argument it suffices to prove

$$(1.5)\qquad E_{\{0,1,\ldots,1\}}\psi(X_1,\ldots,X_n) \leqslant E_{\{1,\ldots,1\}}\psi(X_1,\ldots,X_n).$$

Since $\varphi(a_i)$ is only a relabeling of a_i we assume $\varphi(a_i) = a_i \geqslant 0$ and also without loss of generality $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_N$. With this (1.5) becomes

$$(1.6)\qquad \{N(N-1)^{n-1}\}^{-1}\Sigma_{k=1}^N a_k(\Sigma_{j=1}^N a_j - a_k)^{n-1} \leqslant N^{-n}(\Sigma_{j=1}^N a_j)^n$$

and with $b_k = a_k/(\Sigma_{j=1}^N a_j)$ so that $\Sigma_{k=1}^N b_k = 1$ we obtain that (1.6) is equivalent to

$$(1.7)\qquad \Sigma_{k=1}^N b_k(1-b_k)^{n-1} \leq \{1-(1/N)\}^{n-1}.$$

We now prove (1.7) by induction on N. For $N = 1$, (1.7) is trivial. If the maximum of $\Sigma_{k=1}^N b_k(1-b_k)^{n-1}$ is obtained at a boundary point of the simplex $\{b_i \geqslant 0, \Sigma_{i=1}^N b_i = 1\}$, then some $b_i = 0$ and by the induction hypothesis at the maximum point

$$\Sigma_{k=1}^N b_k(1-b_k)^{n-1} \leqslant \{1-(1/(N-1))\}^{n-1} \leqslant \{1-(1/N)\}^{n-1}.$$

If the maximum is at an interior point, then by differentiating $\Sigma b_k(1-b_k)^{n-1} - \lambda(\Sigma b_k - 1)$ we obtain the equation $(1-b_k)^{n-1}-(n-1)b_k(1-b_k)^{n-2}-\lambda = 0$, and equivalently

$$(1.8)\qquad n(1-b_k)^{n-1}-(n-1)(1-b_k)^{n-2}-\lambda = 0.$$

Summing in (1.8) over k we have

$$(1.9)\qquad n\Sigma_{k=1}^N(1-b_k)^{n-1}-(n-1)\Sigma_{k=1}^N(1-b_k)^{n-2} = N\lambda.$$

Now, using the Tchebycheff rearrangement inequality,

$$(1.10)\qquad \Sigma_{k=1}^N(1-b_k)^{n-1} \geqslant (1/N)\Sigma_{k=1}^N(1-b_k)S_{k=1}^N(1-b_k)^{n-2} = ((N-1)/N)\Sigma_{k=1}^N(1-b_k)^{n-2}$$

and therefore from (1.9)

$$(N/n)\lambda \geqslant ((N-1)/N)\Sigma_{k=1}^N(1-b_k)^{n-2}-((n-1)/n)\Sigma_{k=1}^N(1-b_k)^{n-2} \geqslant 0.$$

Returning to the expressing in (1.8), $\lambda \geq 0$ implies that the polynomial $nx^{n-1}-(n-1)x^{n-2}-\lambda$ has only one positive root by Descartes' rule of signs. Therefore, an interior maximum of

$\Sigma_{k=1}^N b_k(1-b_k)^{n-1}$ can occur only when all values of $1-b_k$ are equal to this root and hence if an interior maximum exists it must occur at $b_1 = \dots = b_n = 1/N$. At this point (1.7) holds with equality and thus (1.7) is now established. □

Remark. The inequality

$$E_{\{0,1,\dots,1\}}\{\Pi_{i=1}^n \varphi_i(X_i)\} \leq E_{\{1,\dots,1\}}\{\Pi_{i=1}^n \varphi_i(X_i)\}$$

for *different* φ_i increasing does *not* hold in general. To see this note that for $\varphi_1 \equiv 1/N$ and $\varphi_i(a_i) = a_i$, $i = 2, \dots, n$ we would have to prove instead of (1.7)

$$\Sigma_{k=1}^N N^{-1}(1-b_k)^{n-1} \leq (1-(1/N))^{n-1}. \tag{1.11}$$

However (1.11) holds with equality when all $b_i = 1/N$ and the inequality is reversed for any other choice of b_i.

As a special case of Theorem 1 we obtain

$$P_{R(\pi)}[X_1 \leq c, \dots, X_n \leq c] \leq \{P(X_1 \leq c)\}^n$$

where on the right-hand side X_1 takes the values $\{a_1, \dots, a_n\}$ with equal probabilities, i.e., $P(X_1 = a_i) = 1/N$. Also,

$$P_{R(\pi)}[X_1 \geq c, \dots, X_n \geq c] \leq \{P(X_1 \geq c)\}^n.$$

The next result should be compared with Theorem 3.1 of Karlin (1974).

THEOREM 2. *Let*

$$\psi(x_1, \dots, n_n) = \Sigma_{1\leq i_1 < \dots < i_r \leq n} \Pi_{j=1}^r \varphi(x_{i_j}), \quad r \leq n$$

where $\varphi \geq 0$. *Then*

$$E_{R(\pi)}\psi(X_1, \dots, X_n) \leq E_{R(1)}\psi(X_1, \dots, X_n). \tag{1.12}$$

The case $r = n$ coincides with Theorem 1.

Proof. The case $r = 1$ is trivial so we take $r \geq 2$. Note that it suffices to assume $0 \leq a_1 \leq \dots \leq a_N$ and take $\psi(x_1, \dots, x_n) = \Sigma_{1\leq i_1<\dots<i_r\leq n}\Pi_{j=1}^r x_{i_j}$.

Again, it suffices to prove (1.5) which for the present ψ becomes

$$\binom{n-1}{r}(N(N-1)^r)^{-1}\Sigma_{k=1}^N(1-b_k)^r + \binom{n-1}{r-1}(N(N-1)^{r-1})^{-1}\Sigma_{k=1}^N b_k(1-b_k)^{r-1} \leq N^{-r}\binom{n}{r} \tag{1.13}$$

where $\Sigma_{k=1}^N b_k = 1$, $b_i \geq 0$, which reduces to

$$((n-r)/(N-1))\Sigma_{k=1}^N(1-b_k)^r + r\Sigma_{k=1}^N b_k(1-b_k)^{r-1} \leq (1-(1/N))^{r-1}n. \tag{1.14}$$

We prove (1.14) by induction on N. As before, on the boundary where some $b_i = 0$, (1.14) follows readily from the induction hypothesis. Differentiation with respect to b_k of the left hand side of (1.14) with the constraint $\Sigma_{k=1}^N b_k = 1$ yields

$$-((n-r)/(N-1))r(1-b_k)^{r-1} + r(1-b_k)^{r-1} - r(r-1)b_k(1-b_k)^{r-2} - \lambda = 0 \tag{1.15}$$

or

$$r(r-(n-r)/(N-1))(1-b_k)^{r-1} - r(r-1)(1-b_k)^{r-2} - \lambda = 0 \tag{1.16}$$

Summation over k produces

$$r(r-(n-r)/(N-1))\Sigma_{k=1}^N(1-b_k)^{r-1} - r(r-1)\Sigma_{k=1}^N(1-b_k)^{r-2} = \lambda N$$

and invoking the inequality

$$\Sigma_{k=1}^N(1-b_k)^{r-1} \geq ((N-1)/N)\Sigma_{k=1}^N(1-b_k)^{r-2}; \quad \text{see (1.10)}$$

we have

$$(N/r)\lambda \geq (r-(n-r)/(N-1))((N-1)/N)-(r-1)\ \Sigma_{k=1}^{N}(1-b_k)^{r-2}$$
$$= ((N-n)/N)\ \Sigma_{k=1}^{N}(1-b_k)^{r-2} \geq 0.$$

Thus, again (1.16) has a unique critical point with $b_k \equiv 1/N$. Since for $b_k \equiv 1/N$ (1.13) hold with equality, the result follows. □

Examples. Theorem 2 implies inequality (1.12) for $\psi(x_1, \ldots, x_n) = (x_1 + \ldots + x_n)^\alpha$ for any integer $\alpha > 0$. For $\Sigma_{k=1}^{N} a_i < 1$ we obtain by expansion that (1.12) also holds with $\psi(x_1, \ldots, x_n) = [1-(x_1 + \ldots + x_n)^\alpha]^{-1}$ or any positive combination $\Sigma c_k(x_1 + \ldots + x_n)^{\alpha_k}$, $\alpha_k \geq 0$ integers.

The preceding inequalities are related to multivariate majorization and Schur function as explained next.

2. Multivariate Majorization and Negative Association. A function $\varphi(\mathbf{x})$ defined on $\mathcal{R}^N$ *is said to be Schur concave* if $\varphi(\mathbf{x}) \geq \varphi(\mathbf{y})$ for any $\mathbf{x},\mathbf{y} \in \mathcal{R}^N$ such that $\mathbf{x} = \mathbf{yM}$ for some matrix $\mathbf{M} \in \mathcal{D}$ = the class of $N \times N$ doubly stochastic matrices. See Marshall and Olkin (1979) for details, references and historical remarks. Let $\mathbf{X}$ and $\mathbf{Y}$ be $n \times N$ matrices whose columns are $\mathbf{x}_1, \ldots, \mathbf{x}_N$, and $\mathbf{y}_1, \ldots, \mathbf{y}_N$, respectively. The inequality $\Sigma_{i=1}^{N} g(\mathbf{x}_i) \geq \Sigma_{i=1}^{N} g(\mathbf{y}_i)$ holds for every concave function g defined on $\mathcal{R}^n$ if and only if there exists a matrix $\mathbf{M} \in \mathcal{D}$ such that $\mathbf{X} = \mathbf{YM}$. (This result is due to Hardy, Littlewood and Pólya (1934) for $n = 1$, and to Sherman (1951), Stein and Blackwell (1953)).

In particular, the matrix function $\psi(\mathbf{X}) = \Sigma_{i=1}^{N} g(\mathbf{x}_i)$ satisfies $\psi(\mathbf{X}) \geq \psi(\mathbf{Y})$ whenever $\mathbf{X} = \mathbf{YM}$ provided g is concave. Related notions of multivariate Schur concavity and probabilistic applications were studied by Rinott (1973), Marshall and Olkin (1979), Karlin and Rinott (1981), Tong (1982) and Karlin and Rinott (1983). In some of the applications one obtains the inequality $\psi(\mathbf{X}) \geqslant \psi(\mathbf{Y})$ whenever $\mathbf{X} = \mathbf{YM}$ where $\mathbf{M}$ belongs to a subclass of $\mathcal{D}$. Of particular interest is the class $\mathcal{T}$ of matrices which can be represented as products of matrices of the form $t\mathbf{I} + (1-t)\mathbf{P}$ where $\mathbf{I}$ is the $N \times N$ identity matrix, $\mathbf{P}$ is a permutation matrix which interchanges only two coordinates, and $0 \leq t \leq 1$.

Our next theorem describes an example of Schur concavity with respect to the class $\mathcal{T}$. A probabilistic interpretation of the result in terms of a birthday-problem of coincidence probabilities will be given. We first need a lemma which extends Ostrowski's (1952) well-known criterion for Schur concavity. The proof can be found in Rinott (1973), Marshall and Olkin (1979).

LEMMA 1. *A differentiable function* $\psi: \mathcal{R}^{nN} \to R$ *is multivariate Schur concave with respect to* $\mathcal{T}$, *i.e.,* $\psi(\mathbf{X}) \leq \psi(\mathbf{XT})$ *for every* $\mathbf{T} \in \mathcal{T}$ *and* $n \times N$ *matrix* $\mathbf{X} = \|x_{ij}\|$ *if and only if*

(i) $\psi(\mathbf{X}) = \psi(\mathbf{XP})$ *for every* $N \times N$ *permutation matrix* $\mathbf{P}$*; and*

(ii) $\Sigma_{i=1}^{n}(x_{ij}-x_{ik})[\partial\psi(\mathbf{X})/\partial x_{ij} - \partial\psi(\mathbf{X})/\partial x_{ik}] \leq 0$ for all $1 \leq j \neq k \leq N$.

Let $\alpha^1, \ldots, \alpha^n \in \mathcal{R}^N$ denote the n *rows* of the $n \times N$ matrix $\mathbf{A}$, $\alpha^i = (\alpha_1^i, \ldots, \alpha_n^i)$, $i = 1, \ldots, n$. We assume that the rows are similarly ordered, that is $(\alpha_j^i - \alpha_k^i)(\alpha_j^{i'} - \alpha_k^{i'}) \geq 0$ for all $1 \leq j,\ k \leq N$, $1 \leq i,\ i' \leq n$. Note that if $\mathbf{T} = t\mathbf{I} + (1-t)\mathbf{P}$ where $\mathbf{P}$ is a permutation matrix that interchanges only two coordinates, then applied to these two coordinates, $\mathbf{T}$ operates like the matrix $\begin{pmatrix} t & 1-t \\ 1-t & t \end{pmatrix}$ which preserves the order if $t \geq ½$ and reverses the order if $t \leq ½$. If the rows of $\mathbf{A}$ are similarly ordered, then so are the rows of $\mathbf{AT}$ for any $\mathbf{T} \in \mathcal{T}$.

THEOREM 3. *Let* $\psi(\mathbf{A})$ *be defined by*

(2.1) $$\psi(\mathbf{A}) = \Sigma_{j_1 \neq \ldots \neq j_n} \Pi_{k=1}^{n} \alpha_{j_k}^{k}$$

where the sum extends over all $\binom{N}{n}$ *vectors of n different indices between* 1 *and* N $\mathbf{A} = \|\alpha_j^i\|$ *is* $n \times N$ *satisfying* $\alpha_j^i \geq 0$, $i = 1, \ldots, n$, $j = 1, \ldots, N$, $n \leq N$, *and the rows of* $\mathbf{A}$, $\alpha^1, \ldots, \alpha^n$ *are similarly ordered. Then* $\psi(\mathbf{A}) \leq \psi(\mathbf{AT})$ *for all* $\mathbf{T} \in \mathcal{T}$.

Proof. In view of Lemma 1, we compute

$$\partial\psi/\partial\alpha_1^1 - \partial\psi/\partial\alpha_2^1 = \Sigma_{1 \neq j_2 \neq \ldots \neq j_n} \alpha_{j_2}^2 \ldots \alpha_{j_n}^n - \Sigma_{2 \neq j_2 \neq \ldots \neq j_n} \alpha_{j_2}^2 \ldots \alpha_{j_n}^n .$$

Let

$$u_k = \Sigma_{3 \leq j_2 \neq \ldots \neq j_{k-1} \neq j_{k+1} \neq \ldots \neq j_n} \alpha_{j_2}^2 \ldots a_{j_{k-1}}^{k+1} \alpha_{j_{k+1}}^{k+1} \ldots \alpha_{j_n}^n, \quad k = 1, \ldots, n.$$

Then

(2.2) $$\partial\psi/\partial\alpha_1^1 - \partial\psi/\partial\alpha_2^1 = (\Sigma_{k=2}^{n} \alpha_2^k u_k + u_1) - (\Sigma_{k=2}^{n} \alpha_1^k u_k + u_1)$$
$$= \Sigma_{k=2}^{n} (\alpha_2^k - \alpha_1^k) u_k .$$

Therefore

$$\Sigma_{k=1}^{n} (\alpha_1^i - \alpha_2^i)(\partial\psi/\partial\alpha_1^1 - \partial\psi/\partial\alpha_2^1) = \Sigma_{i=1}^{n} \Sigma_{k \neq i} (\alpha_1^i - \alpha_2^i)(\alpha_2^k - \alpha_1^k) u_k \leq 0$$

since similar ordering implies $(\alpha_1^i - \alpha_2^i)(\alpha_2^k - \alpha_1^k) \leqslant 0$, and replacing 1,2 by any pair of indices the required result follows from Lemma 1. □

Note that Theorem 3 involves Schur concavity with respect to $\mathcal{T}$ on the set of nonnegative $n \times N$ matrices having similarly ordered rows.

In the proof of Theorem 3 consider the subclass $\mathcal{S}$ of $\mathcal{T}$ consisting of finite products of matrices of the form $\mathbf{T} = t\mathbf{T} + (1-t)\mathbf{P}$, $\mathbf{P}$ a permutation matrix that interchanges only two adjacent coordinates and $\frac{1}{2} \leqslant t \leqslant 1$. Such a $\mathbf{T}$ preserves the ordering of the components when applied to a vector. The calculation in (2.2) implies $(\alpha_1^1 - \alpha_2^1)(\partial\psi/\partial\alpha_1^1 - \partial\psi/\partial\alpha_2^1) = (\alpha_1^1 - \alpha_2^1)\Sigma(\alpha_2^k - \alpha_1^k)u_k \leqslant 0$ and the same holds if we replace the pair of indices 1,2 by any pair. By the well known criterion of Ostrowski (1952) it follows that $\psi(\mathbf{A}) = \psi(\alpha^1, \ldots, \alpha^n)$ is Schur convex in α^1, when $\alpha^2, \ldots, \alpha^n$ are fixed and $\alpha^1, \ldots, \alpha^n$ are all similarly ordered. This implies

THEOREM 4. *Under the conditions of Theorem 3* $\psi(\alpha^1, \ldots, \alpha^n) \leq \psi(\alpha^1\mathbf{T}_1, \ldots, \alpha^n\mathbf{T}_n)$ *for all* $\mathbf{T}_1, \ldots, \mathbf{T}_n \in \mathcal{S}$.

As a special case of Theorem 3 we obtain the inequalities (1.2a)–(1.2b). This is given by

PROPOSITION 1. *Let* $0 \leqslant \varphi_i$ *be increasing functions,* $i = 1, \ldots, n$, *then*

(2.3) $$E_{R(\mathbf{0})}\{\Pi_{i=1}^{n} \varphi_i(X_i)\} \leqslant E_{R(\mathbf{1})}\{\Pi_{i=1}^{n} \varphi_i(X_i)\}.$$

Proof. Set $\alpha_j^i = \varphi_i(a_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, N$. Then $E_{R(\mathbf{0})}\{\Pi_{i=1}^{n} \varphi_i(X_i)\} = \psi(\mathbf{A})/(N(N-1) \ldots (N-n+1))$ where $\psi(\mathbf{A})$ is defined by (2.1), while $E_{R(\mathbf{1})}\{\Pi_{i=1}^{n} \varphi_i(X_i)\} = N^{-n}\Pi_{i=1}^{n}(\Sigma_{j=1}^{N} \alpha_j^i)$. It is easy to see that inequality of (2.3) is homogeneous and we can assume $\Sigma_{j=1}^{N} \alpha_j^i = 1$, $i = 1, \ldots, n$, without loss of generality. Then (2.3) becomes

(2.4) $$\psi(\mathbf{A}) = N(N-1) \ldots (N-n+1)/N^n.$$

For $\mathbf{J} \in \mathcal{T}$ having all entries equal to N^{-1} we now have $\mathbf{AJ} = \mathbf{J}$, and a simple calculation shows that $\psi(\mathbf{J}) = N(N-1) \ldots (N-n+1)/N^n$. Schur concavity of ψ implies $\psi(\mathbf{A}) \leq \psi(\mathbf{AJ}) - \psi(\mathbf{J})$ and (2.4) follows. □

3. A Generalized Birthday Problem. We finally apply Theorem 3 to obtain an extension of the "*birthday problem*" (see Marshall and Olkin, 1979, p. 305). Consider a group of n individuals. Let α_j^i, $i = 1, \ldots, n$, $j = 1, \ldots, N = 365$ denote the probability that the ith person's birthday occurs on the jth day of the year, $1 \le j \le 365 = N$. Then for $\psi(\mathbf{A})$ defined in (2.1) we have for n independent persons

$$\psi(\mathbf{A}) = \text{Probability that the } n \text{ persons have } n \text{ distinct birthdays,}$$

i.e., no coincidences of birthdays occur.

This probability was studied in the case that the likelihood of a birthday on a particular day is the same for all persons. Here we allow different persons to have different distributions of birthdays as long as the vectors $(\alpha_1^i, \ldots, \alpha_{365}^i)$ are similarly ordered, which means that if day j has a higher probability of being one person's birthday than day k, then the same holds for all individuals. We have $\psi(\mathbf{A}) \le \psi(\mathbf{AT})$ for $\mathbf{T} \in \mathcal{T}$ and in particular $\psi(\mathbf{A}) \le \psi(\mathbf{AJ})$, which says that under the above assumptions the probability of no coincidence of birthdays is maximized if all days are equally likely birthdays for all individuals.

Added in proof: Theorem 2 can be derived from the theorem in Van Zwet (1983).

REFERENCES

BLACKWELL, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist. 24* 265–272.

HARDY, G.H., LITTLEWOOD, J.E., and PÓLYA, G. (1934, 1952). *Inequalities*. 1st ed., 2nd. ed. Cambridge University Press, London and New York.

JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables with applications. *Ann. Statist. 11* 286–295.

KARLIN, S. (1974). Inequalities for symmetric sampling plans I. *Ann. Statist. 2* 1065–1094.

KARLIN, S. and RINOTT, Y. (1981). Entropy inequalities for classes of probability distributions. II. The multivariate case. *Adv. Appl. Prob. 13* 325–351.

KARLIN, S. and RINOTT, Y. (1983). Comparison of measures, multivariate majorization, and applications to statistics. *Studies in Econometrics, Time Series, and Multivariate Statistics* S. Karlin, T. Amemiya, L. A. Goodman, eds., Academic Press, New York.

KEMPERMAN, J. H. B. (1973). Moment problems for sampling without replacement. I, II, III. *Indag. Math. 35* 149–188.

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

OSTROWSKI, A. M. (1952). Sur quelques applications des fonctions convexes et concaves au sens de I. Schur. *J. Math. Pures. Appl. 31* 253–292.

RINOTT, Y. (1973). Multivariate majorization and rearrangement inequalities with some applications to probability and statistics. *Israel J. Math. 15* 60–77.

ROSÉN, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement I. *Ann. Math. Statist. 43* 373–397.

SEN, P. K. (1970). The Hájek-Rényi inequality for sampling from a finite population. *Sankhyā 32* 181–188.

SERFLING, R. J. (1973). Refinements of some probability inequalities of Hoeffding (abstract). *Bull. Inst. Math. Statist. 2* 104.

SHERMAN, S. (1951). On a theorem of Hardy, Littlewood, Pólya and Blackwell. *Proc. Natl. Acad. Sci. USA 37* 826–831.

TONG, Y. L. (1982). Rectangular and elliptical probability inequalities for Schur-concave random variables. *Ann. Statist. 10* 637–642.

VAN ZWET, W. R. (1983). An inequality for random replacement sampling plans. *A Festschrift for Erich L. Lehmann*. P. J. Bickel, K. Doksum, J. L. Hodges, Jr., eds. Wadsworth, CA.