COMPARISON OF SOME STATISTICAL EXPERIMENTS ASSOCIATED WITH SAMPLING PLANS

by

Erik Torgersen
Institute of Mathematics, University of Oslo,
Blindern, Oslo 3 - Norway

Some experiments occurring in sampling theory may be described as follows:

Consider a finite population I and a characteristic of interest which, with varying amount (value, degree,...) is possessed by all individuals in I. Let $\theta(i)$ be the amount of this characteristic for individual i.

It is known that θ belongs to some set Θ of functions on I.

Let α be a sampling plan, i.e. a probability distribution on the set of finite sequences of elements from I. If this sampling plan is used and if the characteristics of sampled individuals are determined without error, then the outcome

$$x = ((i_1,\theta(i_1)),(i_2,\theta(i_2)), \quad ,(i_n,\theta(i_n)))$$

is obtained with probability $\alpha(i_1,i_2,\ldots,i_n)$.

If $\mathcal{E}_\alpha$ denotes the experiment obtained by observing x then, provided θ is not too small, $\mathcal{E}_{\alpha_1}$, is at least as informative as $\mathcal{E}_{\alpha_2}$ if and only if the sampled subset under α_1 is "stochastically contained" in the sampled subset under α_2.

We shall here, utilizing the theory of comparison of statistical experiments, discuss this and other related results.

AMS 1970 Subject Classification: Primary: G2 D05. Secondary: 62 B15

Key words: Being more informative, stochastically contained in, deficiency, replacement plans.

1. Introduction

A theory of comparison of experiments based on mathematical decision theory has developed during the last thirty years or so. It has been extensively used, see Le Cam [8], in asymptotic theory. There are so far not many applications to non-asymptotic comparison of statistical models. Some fairly general results on linear normal models may be found in Swensen [12]. The purpose of this talk is to present some simple applications on experiments associated with sampling plans. We refer to Heyer [3], Le Cam [8], [9] and Torgersen [13] for expositions of the theory of comparison of experiments. The material covered in section 2 in Torgersen [14] is adequate here.

Consider a population I which is an, and may be any, enumerable set. Suppose also that there is a characteristic of interest which, with varying amount (value, degree,...) is possessed by all individuals in I. Let $\theta(i)$ be the amount of this characteristic for individual $i \in I$. The function θ on I defined this way is our parameter of interest. We shall assume that it is a priori known that θ belongs to, and may be any element of, a set Θ of functions on I .

In order to find out about θ we may take a sample from I and measure the characteristic for each of the individuals in the sample. An essential assumption is now that the sampling is carried out according to a known sampling plan α , i.e. a probability distribution on the space I_s of finite sequences of elements from I . Before proceeding let us agree that a probability measure on an enumerable set is defined for all subsets. We may, in order to retain the possibility of making no observations at all, include the "empty" sequence $\emptyset$ in I . If the sampling plan α is used and if the characteristics of the sampled individuals are measured without errors, then the outcome $(i_1,\theta(i_1)),\ldots,(i_n,\theta(i_n))$ is obtained eith probability $\alpha(i_1,\ldots,i_n)$: Thus we mwy let our sample space consist of all sequences

$(i_1,f_1),(i_2,f_2),\ldots,(i_n,f_n)$ where $(i_1,\ldots,i_n)\in \mathcal{I}_s$, $f_1,\ldots,f_n \in \bigcup_\theta \theta[\mathcal{I}]$ and where $f_\mu = f_\nu$ whenever $i_\mu = i_\nu$.

Let $P_{\theta,\alpha}$ denote the probability distribution of the outcome when θ prevails and α is used. Thus the sampling plan α determines a statistical experiment $\mathcal{E}_\alpha = (P_{\theta,\alpha} : \theta \in \Theta)$

Let $(I_1,F_1),\ldots,(I_n,F_n)$ be the random outcome and consider the statistics U and X where $U = \{I_1,\ldots,I_n\}$ and X is the function on the set U determined by F . Now $P_{\theta,\alpha}((i_1,f_1),\ldots,(i_n,f_n)) = \alpha(i_1,\ldots,i_n)$ or $= 0$ as $(f_1,\ldots,f_n) = (\theta(i_1),\ldots,\theta(i_n))$ or not. It follows, as is well known, that (U,X) is sufficient. [Just check that conditional probabilities given (U,X) may be specified independently of θ]. It is known, see [1], that (U,X) actually is minimal sufficient, - but we shall not use this fact here. The important thing is that the reduction by sufficiency leads to another, and equivalent experiment $\bar{\mathcal{E}}_{\bar\alpha} = (\bar P_{\theta,\bar\alpha} : \theta \in \Theta)$ which may be described as follows:

Let $\mathcal{U}$ be the class of all finite subsets of $\mathcal{I}$. If $u \in \mathcal{U}$ and α is a sampling plan on $\mathcal{I}$ then $\bar\alpha$ is the probability distribution on $\mathcal{U}$ induced from α by the set valued map $(i_1,\ldots,i_n) \to \{i_1,\ldots i_n\}$. Thus $\bar\alpha$ is the probability distribution of the sampled subset of $\mathcal{I}$.

We may then let the sample space, $\bar{\mathcal{X}}$, of $\bar{\mathcal{E}}_{\bar\alpha}$ consist of all pairs (u,x) where $u \in \mathcal{U}$ and $x = \theta|u$ for some $\theta \in \Theta$. If α is used then the probability, $\bar P_{\theta,\bar\alpha}((u,x))$ of the outcome (u,x) is $\bar\alpha(u)$ or $= 0$ as $x = \theta|u$ or not.

It follows that the structure of experiments $\mathcal{E}_\alpha$ may be identified with a structure of probability measures on the set of finite subsets of the population $\mathcal{I}$.

Note that the set of experiments $\mathcal{E}_\alpha$, and hence the set of experiments $\mathcal{E}_{\bar\alpha}$ is closed under products. More precisely $\mathcal{E}_\alpha \times \mathcal{E}_\beta \sim \mathcal{E}_\gamma$ where $\gamma(k_1,k_2,\ldots,k_r) = \alpha(\emptyset)\beta(k_1,\ldots,k_r) + \alpha(k_1)\beta(k_2,\ldots,k_r) + \ldots + \ldots \alpha(k_1,\ldots,k_{r-1})\beta(k_r) + \alpha(k_1,\ldots k_r)\beta(\emptyset) ; (k_1,\ldots,k_r) \in \mathcal{I}_s$ so that

$\bar{\gamma}(u) = \Sigma\{\bar{\alpha}(u_1)\bar{\beta}(u_2) : u_1 \cup u_2 = u\}$; $u \in \mathcal{U}$.

A few notations and other terms which will be used are listed below:

$\mathcal{I}$ = a population.

$N = \#\mathcal{I}$

$\mathcal{I}_s$ = the set of finite sequences of elements from $\mathcal{I}$.

$\mathcal{U}$ = the class of finite subsets of $\mathcal{I}$.

$\#A$ = the number of elements in A or $= \infty$ as A is finite or infinite.

$\alpha, \beta, \ldots$: probability distribution on $\mathcal{I}_s$.

$\bar{\alpha}$ = the probability measure on $\mathcal{U}$ induced from α by the set valued maps $(i_1,\ldots,i_n) \to \{i_1,\ldots,i_n\}$.

$\bar{\bar{\alpha}}$ = the probability distribution on integers induced from α by the map $(i_1,\ldots,i_n) \to \#\{i_1,\ldots,i_n\}$.

$(z_1,\ldots,z_n)$ = an ordered n-tuple .

$\{z_1,\ldots,z_n\}$ = the set consisting of all elements z such that $z = z_1$ or $z = z_2$ or ... or $z = z_n$.

$\bigwedge_i z_i = \inf_i z_i$

$\bigvee_i z_i = \sup_i z_i$

$\mu(x) = \mu(\{x\})$ if μ is a measure and $\{x\}$ is the one point set containing x .

$\|\mu\|$ = total variation of μ .

$\mathcal{E} \geq \mathcal{F}$: The experiment $\mathcal{E}$ is at least as informative as the experiment $\mathcal{F}$.

$\mathcal{E} \sim \mathcal{F}$: $\mathcal{E}$ and $\mathcal{F}$ are equally informative.

$\delta(\mathcal{E},\mathcal{F})$ = the deficiency of $\mathcal{E}$ w.r.t. $\mathcal{F}$. If $\mathcal{E} = (P_\theta : \theta \in \Theta)$ and $\mathcal{F} = (Q_\theta : \theta \in \Theta)$ then $\delta(\mathcal{E},\mathcal{F})$ is, Le Cam [7], the smallest number of the form $\sup_\theta \|P_\theta M - Q\|$ where M is a Markov operator from the band generated by the P_θ's to the band generated by the Q_θ's.

$\Delta(\mathcal{E},\mathcal{F}) = \delta(\mathcal{E},\mathcal{F}) \vee \delta(\mathcal{F},\mathcal{E})$

Isotonic = monotonically increasing :

A map φ from a partially ordered set $(X,\leq)$ to a partially ordered set is called monotonically increasing (decreasing) if $\varphi(x_1) \leq \varphi(x_2)$ whenever $x_1 \leq x_2$ $(x_1 \geq x_2)$.

2. Comparability of experiments E_α

In order to simplify the writing let us agree to write "$E \geq F$" for "E is at least as informative as F". If "$E \geq F$" and "$F \geq E$" then we shall say that E and F are equivalent and write this $E \sim F$.

Among the several natural (and fortunately equivalent) ways of introducing notations of comparison is the randomization (Markov kernel, transition,...) criterion of Le Cam, which states roughly that $E \geq F$ if and only if F may be obtained from E by a randomization.

Applying this to the discrete experiments $E_\alpha \sim \bar{E}_{\bar{\alpha}}$ and $E_\beta \sim \bar{E}_{\bar{\beta}}$ we find that $E_\alpha \geq E_\beta$ if and only if

(1) $$\bar{P}_{\theta,\bar{\beta}}((v,y)) = \sum_{(u,x)} M((v,y)|(u,x))\bar{P}_{\theta\bar{\alpha}}(u,x) \; ; (v,y) \in \bar{X}$$

for numbers $M((v,y)|(u,x)) \geq 0$; $(u,x),(v,y) \in \bar{X}$ such that

$$\sum_{(v,y)} M((v,y)|(u,x)) = 1 \; ; \; (u,x) \in \bar{X}$$

Using the definitions of the measures $\bar{P}$, (1) may be rewritten:

(2) $$\bar{\beta}(v) = \sum_u M((v,\theta| v)|(u,\theta| u))\bar{\alpha}(u) \; ; \; v \in \mathcal{U}, \; \theta \in \Theta.$$

Hence:

(3) $$1 = \sum_u \Big[\sum_v M((v,\theta| v)|(u,\theta| u))\Big]\bar{\alpha}(u) \; ; \; \theta \in \Theta$$

It follows that $\sum_v M((v,\theta| v)|(u,\theta| u)) = 1$ when $\bar{\alpha}(u) > 0$.

Say that Θ satisfy (C) if:

(C) There is a θ^0 in Θ with the property that there to each $i \in I$ corresponds at least one θ in Θ such that $\theta(j) = \theta^0(j)$ or $\neq \theta^0(j)$ as $j \neq i$ or $j = i$.

Let θ^0 be as in (C). Assume $\bar{\alpha}(u^0)>0$ and put $x^0=\theta^0|u^0$. Put $\Theta^0=\{\theta:\theta\in\Theta \& \theta|u^0=x^0\}$. Then $\theta^0\in\Theta^0$. Consider so a pair (v,θ) where $v\in\mathcal{U}$ and $\theta\in\Theta^0$. If $M((v,\theta|v)|(u^0,x^0))>0$ then, by (3), $(v,\theta|v)$ is necessarily of the form $(v,\theta^0|v)$ i.e. $\theta|v=\theta^0|v$. It follows that

$$(4) \qquad M((v,\theta|v)|(u^0,x^0))\leq M((v,\theta^0|v)|(u^0,x^0)) \; ; \; v\in\mathcal{U} \; .$$

Hence, since both sides add up to 1 in v, "=" holds in (4) for each $v\in\mathcal{U}$. Consider now a particular $v^0\in\mathcal{U}$ such that $M((v^0,\theta^0|v^0)|(u^0,x^0))>0$. Then, by (4) with "$\leqq$" replaced with "=", $M((v^0,\theta|v^0)|(u^0,x^0))>0$ for each $\theta\in\Theta^0$. It follows, using (3), that $\theta|v^0=\theta^0|v^0$; $\theta\in\Theta^0$. If $v^0\not\subseteq u^0$ then we may choose a $i\in v^0-u^0$. By assumption there is a $\theta\in\Theta^0$ such that $\theta(i)\neq\theta^0(i)$ contradicting $\theta|v^0=\theta^0|v^0$. It follows that $v\subseteq u$ whenever $M((v,\theta^0|v)|(u,\theta^0|u))\bar{\alpha}(u)>0$. Define now for each pair $(u,v)\in\mathcal{U}^2$ a number $\bar{\Gamma}(v|u)$ by:
$\bar{\Gamma}(v|u)=M((v,\theta^0|v)|(u,\theta^0|u))$ when $\bar{\alpha}(u)>0$.

$$(5) \qquad \bar{\Gamma}(v|u)=0 \quad \text{if} \quad v\neq u \quad \text{and} \quad \bar{\alpha}(u)=0$$

$$\bar{\Gamma}(u|u)=1 \quad \text{if} \quad \bar{\alpha}(u)=1$$

Then $\sum_v \bar{\Gamma}(v|u)=\sum_{v\subseteq u}\bar{\Gamma}(v|u)=1$; $u\in\mathcal{U}$..

Substituting $\theta=\theta^0$ in (2) we find:

$$(6) \qquad \bar{\beta}(v)=\sum_v \bar{\Gamma}(v|u)\bar{\alpha}(u)$$

Define finally a joint distribute $\bar{\rho}$ on $\mathcal{U}^2$ by: $\bar{\rho}(u,v)=\bar{\Gamma}(v|u)\bar{\alpha}(u)$. Then $\bar{\rho}$ has marginals $\bar{\alpha}$ and $\bar{\beta}$ and $\bar{\rho}(\{(u,v):u\supseteq v\})=1$. The last established fact may be recognized as one of several usual and equivalent ways of expressing that $\bar{\alpha}$ is stochastically larger than $\bar{\beta}$ w.r.t. the inclusion ordering $\subseteq$ on $\mathcal{U}$.

Suppose now; conversely, that we have been able to construct a joint distribution $\bar{\rho}$ with this property. Specify the conditional distribution on, $\bar{\Gamma}$, of obtaining a "last" set v given that the "first" is u

such that $\sum\{\bar{\Gamma}(v|u) : v \subseteq u\} = 1$ for all $u \in \mathcal{U}$. (If $\bar{\alpha}(u) > 0$ then this holds by definition). Define a Markov kernel M from $\bar{X}$ to $\bar{X}$ by $M((v,y)|(u,x)) = \bar{\Gamma}(v|u)$ whenever $v \subseteq u$ and $y = x|v$. (If $v \nsubseteq u$ or $y \neq x|v$ then necessarily $M((v,y)|(u,x)) = 0$). It is then easily checked that M satisfies (2) so that $\bar{\mathcal{E}}_{\bar{\beta}}$ is obtained from $\bar{\mathcal{E}}_{\bar{\alpha}}$ by the randomization M .

We collect this as well as some closely related statements in:

<u>Theorem 1</u> (Comparability criterions).

Suppose Θ satisfy condition (C) above . Then the following four conditions are all equivalent:

(i) $\mathcal{E}_\alpha \geq \mathcal{E}_\beta$

($\bar{\text{i}}$) $\bar{\mathcal{E}}_{\bar{\alpha}} \geq \bar{\mathcal{E}}_{\bar{\beta}}$

(ii) There is a joint distribution ρ on pairs $(I,J) \in \mathcal{I}_s^2$ such that I is destribuled as α , J is distributed as β and $\rho(\{I\} \supseteq \{J\}) = 1$

($\bar{\text{ii}}$) There is a joint distribution $\bar{\rho}$ on pairs $(U,V) \in \mathcal{U}^2$ such that U is distributed as $\bar{\alpha}$, V is distributed as $\bar{\beta}$ and $\bar{\rho}(U \supseteq V) = 1$

<u>Remark 1</u> . Condition (C) is only needed to prove that (i) implies (ii). The implications (i)$\Leftrightarrow$($\bar{\text{i}}$)$\Leftarrow$(ii)$\Leftrightarrow$($\bar{\text{ii}}$) hold even if Θ does not satisfy (C). This follows from the theorem as stated, by enlarging Θ or directly from an inspection of its proof.

<u>Remark 2</u> . It follows from wellknown results (see the remark after theorem 7) on orderings of probability measures on partially ordered sets that ($\bar{\text{ii}}$) , and hence (ii) , may be expressed as follows:

(ii) $E_\alpha h(I) \geq E_\beta h(J)$ for each bounded function h such that $h(i_1,\ldots,i_m) \leq h(j_1,\ldots,j_n)$ whenever $\{i_1,\ldots,i_m\} \subseteq \{j_1,\ldots,j_n\}$.

$(\overline{ii})$ $\alpha(H) \geqq \beta(H)$ for any increasing class $H \subseteq U$

Here a subclass H of U is called increasing if $u \in H$ whenever $v \in H$ for some $v \subseteq u$. Trivially H is increasing if and only if H is of the form $H = \bigcup_{\nu=1}^{\infty} \{u : u \supseteq w_\nu\}$ for some sequence $w_1, w_2, \ldots$ in U.

Completion of the proof of theorem 1. The eqivalence of (i) and $(\overline{I})$ follows fom sufficiency and we saw above that $(\overline{I}) \Leftrightarrow (\overline{ii})$. The implication $(ii) \Rightarrow (\overline{ii})$ is trivial so it remains only to show that $(\overline{ii}) \Rightarrow (ii)$. Suppose then that $(\overline{ii})$ is satisfied. Let $\alpha(\cdot|\{I\})$ and $\beta(\cdot|\{J\})$ be the conditional distributions of, respectively, I given $\{I\}$ and given $\{J\}$. Construct a joint distribution ρ for I and J such that the conditional distribution of (I,J) given (U,V) has marginals $\alpha(\cdot|U)$ and $\beta(\cdot|V)$. Then ρ satisfies (ii)

□

Associated with each sampling plan α is a "cumulative distribution" function $\Phi_{\bar\alpha}$ on U defined by: $\Phi_{\bar\alpha}(w) = \sum\{\bar\alpha(u) : u \subseteq w\}$. It is easily seen that $\Phi_{\bar\alpha}$ determines $\bar\alpha$.

Corollary 2. Suppose Θ satisfies (C). Then the following conditions are equivalent:

(i) $E_\alpha \sim E_\beta$

(ii) $\bar\alpha \doteq \bar\beta$

(iii) $\Phi_{\bar\alpha} = \Phi_{\bar\beta}$

Proof: By remark 2, $\Phi_{\bar\alpha} = \Phi_{\bar\beta}$ when $E_\alpha \sim E_\beta$

□

Ordering of sampling plans according to the "distribution functions" $\Phi_{\bar\alpha}$ corresponds to ordering by affinities, or what is equivalent in this case, to ordering by Hellinger transforms. To see this consider functions $\theta^1, \theta^2, \ldots, \theta^r$ in Θ and positive numbers $t_1, \ldots, t_r$ with sum 1.

Then $\int dP^{t_1}_{\theta^1,\alpha}\, dP^{t_2}_{\theta^2,\alpha} \ldots dP^{t_r}_{\theta^r,\alpha} = \int d\bar{P}^{t_1}_{\theta^1,\bar{\alpha}} \ldots d\bar{P}^{t_r}_{\theta^r,\bar{\alpha}} = \Phi_{\bar{\alpha}}(w)$

where $w = \{i : \theta^1(i) = \theta^2(i) = \ldots = \theta^r(i)\}$. If Θ satisfies condition (C) of theorem 1 then any class $\{u : u \subseteq w\}$ where $w \in \mathcal{U}$ is of this form. It is, however, not difficult to construct examples of noncomparable sampling plans α and β such that $\Phi_{\bar{\alpha}} \leq \Phi_{\bar{\beta}}$.

If $\mathcal{E}_\alpha \geq \mathcal{E}_\beta$ then $\mathcal{E}_\alpha$ is more informative than $\mathcal{E}_\beta$ for any decision problem, in particular for all testing problems. If Θ is not too small then it suffices to consider testing problems by:

<u>Proposition 3</u>. Suppose $\Theta \geq \eta^I$ where $\#\eta \geq 2$. Then $\mathcal{E}_\alpha \geq \mathcal{E}_\beta$ if and only if $\mathcal{E}_\alpha$ is at least as informative as $\mathcal{E}_\beta$ for testing problems.

<u>Proof</u>: Suppose $\Theta \geq \eta^I$ where $\#\eta = 2$ and that $\mathcal{E}_\alpha$ is at least as informative as $\mathcal{E}_\beta$ for testing problems. Choose a $\bar{\theta} \in \eta^I$ and sets $v^1, v^2, \ldots, v^r$ in $\mathcal{U}$. Let Θ_0 consist of all $\theta \in \Theta$ such that $\theta|v^\nu \neq \bar{\theta}|v^\nu$; $\nu = 1,\ldots,r$. Let $\bar{\mathcal{E}}_{\bar{\alpha}}$ and $\bar{\mathcal{E}}_{\bar{\beta}}$ be realized by, respectively, observing (U,X) and (V,Y) . Define the test $\tilde{\delta} = \tilde{\delta}(V,Y)$ by putting $\tilde{\delta} = 1$ when there is a $\nu \in \{1,\ldots,r\}$ such that $V \supseteq v^\nu$ and $Y|v^\nu = \bar{\theta}|v^\nu$, and by putting $\tilde{\delta} = 0$ otherwise. Then $E_\theta \tilde{\delta}(V,Y) = 0$; $\theta \in \Theta_0$. By assumption there is a test $\delta = \delta(U,X)$ so that $E_\theta \delta \equiv E_\theta \tilde{\delta}$. In particular $\sum_u \delta(u,\theta|u)\alpha(u) = 0$ when $\theta \in \Theta_0$. Suppose $u \in \mathcal{U}$ is such that $u \not\supseteq v^\nu$; $\nu = 1,\ldots,r$. Then, by assumption, there is a $\theta \in \Theta_0$ such that $\theta|u = \bar{\theta}|u$. Hence $\delta(u,\bar{\theta}|u)\alpha(u) = 0$ in this case. This yields:

$\sum\{\alpha(u) : u \supseteq v^1 \text{ or } u \supseteq v^2 \text{ or} \ldots \text{or } u \supseteq v^r\} \geq \sum \delta(u,\bar{\theta}|u)\alpha(u) = E_{\bar{\theta}}\delta = E_{\bar{\theta}}\tilde{\delta} =$
$\sum \tilde{\delta}(v,\bar{\theta}|v)\beta(v) = \sum\{\beta(v) : v \supseteq v^1 \text{ or } v \supseteq v^2 \text{ or} \ldots \text{or } v \supseteq v^r\}$. Hence $\alpha(H) \geq \beta(H)$ for any increasing class H in $(\mathcal{U}, \subseteq)$. The proposition follows now from theorem 1 and the remark after that theorem.

□

If I is finite then a sampling plan α will be called (population) symmetric if $\alpha(\rho(i_1),\dots,\rho(i_n)) = \alpha(i_1,\dots,i_n)$ for each sequence $(i_1,\dots,i_n)$ in I_s and each permutation ρ of I. It is easily seen that $\bar{\alpha}(u)$ depends on u only through $\#u$ when α is symmetric. Conversely any probability distribution π on U such that $\pi(u)$ depends on u via $\#u$ is of the form $\pi = \bar{\alpha}$ for a symmetric sampling plan α without replacement.

Let for any sampling plan α, $\bar{\bar{\alpha}}$ be the probability distribution of the number of different elements in the sample sequence (set) when the sample sequence (set) is distributed according to $\alpha(\bar{\alpha})$.
Thus

$$\bar{\bar{\alpha}}(n) = \sum\{\bar{\alpha}(u):\#u=n\} = \sum\{\alpha(i_1,\dots,i_m):\ \#\{i_1,\dots,i_m\}=n\} .$$

If α is symmetric then $\bar{\alpha}$ is determined by $\bar{\bar{\alpha}}$ by:

$$\bar{\alpha}(u) = \binom{\#N}{\#u}^{-1}\bar{\bar{\alpha}}(\#u)$$

Clearly any probability distribution on $\{0,1,2,\dots,N\}$ is of the form $\bar{\bar{\alpha}}$ for a unique symmetric α without replacement. If both E_α and E_β are symmetric then $E_\gamma = E_\alpha \times E_\beta$ is symmetric as well and:

$$\bar{\bar{\gamma}}(n)/\binom{N}{n} = \sum \frac{n(n-r_1+n-r_2)}{(n-r_1)!(n-r_2)!}\,[\bar{\bar{\alpha}}(r_1)/\binom{N}{r_1}][\bar{\bar{\beta}}(r_2)/\binom{N}{r_2}]$$

where the summation is over all ordered pairs (r_1,r_2) of integers in $\{0,1,\dots,n\}$ such that $r_1 + r_2 \geq n$.

Note also, as is wellknown, that any symmetric sampling plan α is a mixture of simple random sampling plans without replacement. More precisely:

$$E_\alpha \sim \sum_{n=0}^{N} \bar{\bar{\alpha}}(n) E_{\rho_n}$$

where $\rho_n(i_1,\dots,i_n) = N_{(n)}^{-1}$ when $i_1,\dots,i_n$ are distinct, while $\rho_n(i_1,\dots,i_m) = 0$ whenever $m \neq n$. It follows then, since

$\mathcal{E}_{\rho_0} \leq \mathcal{E}_{\rho_1} \leq \ldots \leq \mathcal{E}_{\rho_n}$ that $\mathcal{E}_\alpha \geq \mathcal{E}_\beta$ whenever α and β are symmetric sampling plans such that $\bar{\bar{\alpha}}$ is stochastically greater than $\bar{\bar{\beta}}$. Suppose conversely that $\bar{\bar{\alpha}}$ is stochastically greater than $\bar{\bar{\beta}}$. Then there is a joint distribution $\bar{\bar{\rho}}$ on $\{0,1,\ldots,N\}^2$ with marginals $\bar{\bar{\alpha}}$ and $\bar{\bar{\beta}}$ and such that $\bar{\bar{\rho}}(\{(m,n):m \geq n\}) = 1$. Put $\bar{\bar{\Gamma}}(n|m) = \frac{\bar{\bar{\rho}}(m,n)}{\bar{\bar{\alpha}}(m)}$ if $\bar{\bar{\alpha}}(m) > 0$. If $\bar{\bar{\alpha}}(m) = 0$, then we may put $\bar{\bar{\Gamma}}(n|m) = 1$ or $=0$ as $n=m$ or $n \neq m$. Define a kernel $\bar{\Gamma}$ from $\mathcal{U}$ to $\mathcal{U}$ by: $\bar{\Gamma}(v|u) = \binom{\#u}{\#v}^{-1} \bar{\bar{\Gamma}}(\#v|\#u)$ if $v \subseteq u$. Put $\bar{\Gamma}(v|u) = 0$ if $v \nsubseteq u$. Let $v \in \mathcal{F}$ and put $n = \#v$. Then

$$\sum_u \bar{\Gamma}(v|u)\bar{\alpha}(u) = \sum_{m=n}^{N} \binom{N-m}{m-n} \binom{m}{n}^{-1} \bar{\bar{\Gamma}}(n|m)\bar{\bar{\alpha}}(m)\binom{N}{m}^{-1} = \binom{N}{n}^{-1} \sum_m \bar{\bar{\Gamma}}(n|m)\bar{\bar{\alpha}}(m) =$$

$$= \binom{N}{n}^{-1} \bar{\bar{\beta}}(n) = \bar{\beta}(v)$$. This, together with theorem 1, proves:

<u>Theorem 4</u>. Let Θ satisfy condition (C) and let α and β be symmetric sampling plans. Then $\mathcal{E}_\alpha \geq \mathcal{E}_\beta$ if and only if $\bar{\bar{\alpha}}$ is stochastically greater than $\bar{\bar{\beta}}$.

<u>Remark</u>. Condition (C) is, by the proof above, not needed for the "if" part of the statement.

3. Random replacement sampling plans

Define (not necessarily symmetric) sampling plans $\alpha_{p,n,\pi} = \alpha_\pi$ where p is a probability distribution on $\mathcal{I}$ such that $p(i) > 0$ for all $i \in \mathcal{I}$, n is a positive integer and π is a probability distribution on $\{0,1\}^{n-1}$ as follows:

Choose a sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n-1}$ of 0's and 1's according to π. Then draw individuals $I_1, I_2, \ldots, I_n$ one after another such that:

(i) An individual which is drawn at the m-th draw where $m < n$ is replaced or not as $\varepsilon_m = 1$ or $\varepsilon_m = 0$.

(ii) I_1 is drawn from $\mathcal{I}$ such that $Pr(I_1 = i_1) = p(i_1)$; $i_1 \in \mathcal{I}$.

(iii) If $I_1, \ldots, I_m$ has been drawn then stop whenever $m = n$ or if $m < n$ and each element of $\mathcal{I}$ has been drawn without being replaced. If otherwise then I_{m+1} is drawn from the remaining part A of the population such that $\Pr(I_{m+1}=i_{m+1}) = p(i_{m+1})/p(A)$; $i_{m+1} \in A$.

Using theorem 1 we get the following intuitively reasonable sufficient condition for comparability:

Proposition 5. Let p and n be fixed. Then $\mathcal{E}_{\alpha_\pi} \geqq \mathcal{E}_{\alpha_{\pi'}}$ whenever π is stochastically larger (for the pointwise ordering on $\{0,1\}^{n-1}$) than π' .

Remark 1. Let $n=3$. It is then easily seen that $\bar{\bar{\alpha}}_{\delta_{0,1}}$ is stochastically larger than $\bar{\bar{\alpha}}_{\delta_{1,0}}$ when $N \geq 2$. Thus the converse of the above statement is, even if we restrict attention to independent and uniformly destributed drawings, not true.

Remark 2. Suppose $N = \# I < \infty$ and that p is the uniform distribution on I . Then, by theorem 4 and the proposition, $E_{\alpha_\pi} h(\# I) \geq E_{\alpha_{\pi'}} h(\# I)$ whenever π is stochastically larger than π' and h is monotonically increasing. If, in addition, the drawings are independent (i.e. π and π' are product measures) then this proves a very particular case of a conjecture by Karlin [5]. A discussion of the relationship of the problems and results in [5] to the theory of comparison of experiments may be found in the appendix.

Proof. Note first that $\alpha_\pi(i_1,\ldots,i_n) = E\alpha_{\delta_\varepsilon}(i_1,\ldots,i_n)$ where ε is distributed according to π and δ_ε is the one point distribution in ε . Hence $\bar{\alpha}_\pi(u) = E\bar{\alpha}_{\delta_\varepsilon}(u)$; $u \in \mathcal{U}$. Suppose now that we knew that $\bar{\alpha}_{\delta_\varepsilon}$ is "stochastically contained" in $\bar{\alpha}_{\delta_{\varepsilon'}}$ whenever $\varepsilon \geq \varepsilon'$. [The terminologi is consistent with the following convention: Let P and Q be probability distributions on χ and let R be a relation on χ . Then P is stochastically in relation R to Q if $\Pr(X_P,X_Q) \in R) = 1$ for random variables X_P and X_Q with, respectively, distributions P and Q].

Let h be an isotonic function on $(\mathcal{U},\subseteqq)$. Then $\sum_u h(u)\bar{\alpha}_{\delta_\varepsilon}(u)$ is monotonically decreasing in ε. Hence $\sum_u h(u)\alpha_\pi(u) = \sum_{\varepsilon u}\sum h(u)\alpha_{\delta_\varepsilon}(u)\pi(\varepsilon) \leqq \sum_{\varepsilon u}\sum h(u)\alpha_{\delta_\varepsilon}(u)\pi'(\varepsilon) = \sum_u h(u)\alpha_{\pi'}(u)$. It follows that $\bar{\alpha}_\pi$ is stochastically contained in $\bar{\alpha}_{\pi'}$. It suffices therefore to show that $\bar{\alpha}_{\delta_\varepsilon}$ is stochastically contained in $\bar{\alpha}_{\delta_{\varepsilon'}}$, when $\varepsilon \geqq \varepsilon'$. We shall show this by showing that the sampling plans $\alpha_{\delta_\varepsilon}$; $\varepsilon \in \{0,1\}^{n-1}$ may all be imbedded within a single stochastic framework. This framework shall consist of independent I-valued random variables $V_{\mu,\nu}$; $\mu=1,2,\ldots$, $\nu=1,2,\ldots n$ such that each $V_{\mu,\nu}$ has distribution p. Before proceeding, let us for each m-tuple $(i_1,\ldots,i_m)$ with $m<n$ and for each sequence $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_m$ of 0's and 1's put $A(i_1,\ldots,i_m;\varepsilon_1,\ldots,\varepsilon_m) = I - \{i_\nu : \nu \leqq m \;\&\; \varepsilon_\nu = 0\}$. Thus $A(i_1,\ldots,i_m,\varepsilon_1,\ldots,\varepsilon_m)$ is precisely the elements left in I after $i_1,\ldots,i_m$ has been drawn and the replacement policy $(\varepsilon_1,\ldots,\varepsilon_m)$ has been used.

Define, for given ε, recurcively random variables $R_1,R_2,\ldots R_n$ by:

(i) $R_1 = 1$

(ii) If $R_1,\ldots,R_m$ are given where $m<n$ and $R_m<\infty$ then R_{m+1} is the smallest integer $\mu \geqq 1$ such that

$V_{\mu,m+1} \in A(V_{1,R_1},\ldots,V_{m,R_m},\varepsilon_1,\ldots,\varepsilon_m)$ when

$A(V_{1,R_1},\ldots,V_{m,R_m},\varepsilon_1,\ldots,\varepsilon_m) \neq \emptyset$. Put $R_{m+1} = \infty$ otherwise.

The quantities R_m, I_m and ν depend on ε. Use the notations R'_m, I'_m and ν' when ε is replaced by ε'. Suppose now that $\varepsilon \geqq \varepsilon'$ Then we have for each $m \leqq n$.

(§) $R_{m'} \geqq R_m$

(§§) If $I'_1,\ldots,I'_m$ are defined then $I_1,\ldots,I_m$ are also defined and $A(I'_1,\ldots,I'_m,\varepsilon'_1,\ldots,\varepsilon'_m) \subseteqq A(I_1,\ldots,I_m,\varepsilon_1,\ldots,\varepsilon_m)$

(§§§) If $I'_1,\ldots,I'_m$ are defined then $I_1,\ldots,I_m$ are also defined and $\{I_1,\ldots,I_m\} \subseteqq \{I'_1,\ldots,I'_m\}$.

Proofs of (§), (§§) and (§§§). The statements are trivial if $m=1$. The general case follows by induction on m. Suppose (§), (§§) and (§§§) hold with m replaced by $m-1$ where $m \geqq 2$.

Put $A_k = A(I_1,\ldots,I_k,\varepsilon_1,\ldots,\varepsilon_k)$ and $A'_k = A(I'_1,\ldots,I'_k,\varepsilon'_1,\ldots,\varepsilon'_k)$. By the induction hypothesis: $A'_{m-1} \subseteqq A_{m-1}$ whenever $R'_m < \infty$. Suppose then that $R'_m < \infty$. Then $V_{m,\mu}$; $\mu = 1,2,\ldots$ have already reached A_{m-1} when A'_{m-1} is reached. This proves (§). Now $A'_m = A'_{m-1} \cap \{I'_m : \varepsilon'_m = 0\}^c$ and $A_m = A_{m-1} \cap \{I_m : \varepsilon_m = 0\}^c$. This shows that $A'_m \subseteqq A_m$ whenever $\varepsilon_m = 1$. If $\varepsilon_m = 0$ then, since $\varepsilon' \leqq \varepsilon$, $\varepsilon'_m = 0$. The only case which then needs particular attention is the case $A'_{m-1} \ni I_m \neq I'_m$. This, however is an impossibility since $R'_m \geqq R_m$. Hence (§§) is established.

It remains to show that $I_m \in \{I'_1,\ldots,I'_m\}$. Assuming $I_m \neq I'_m$ we see, as above, that $I_m \notin A'_{m-1}$. This, however, implies that I_m has been drawn and not replaced in the sequence $I'_1,\ldots,I'_{m-1}$. Hence $I_m \in \{I'_1,\ldots,I'_{m-1}\} \subseteqq \{I'_1,\ldots,I'_m\}$. This proves (§§§).

It is now easily seen that $(I_1,\ldots,I_\nu)$ is distributed according to $\alpha_{\delta_\varepsilon}$. (Just consider the conditional probability of obtaining the sequence $i_1,\ldots,i_m$; $m \leqq \nu$ given that $i_1,\ldots,i_{m-1}$ has been obtained). Our claims concerning the sampling plans α_π follow now from (§§§) and theorem 1.

□

4. Deficiencies and distances.

Let us proceed to the slightly more difficult problem about deficiencies between experiments $\mathcal{E}_\alpha$. Thus we shall try to find out how much do we loose (in risk say), under the least favorable conditions for comparison by basing our decisions on $\mathcal{E}_\alpha$ instead of $\mathcal{E}_\beta$. Following Le Cam [7] we shall limit ourselves to decision problems with bounded loss functions. Clearly:

$\|\bar{P}_{\theta\bar{\alpha}} - \bar{P}_{\theta\bar{\beta}}\| = \sum_u |\bar{P}_{\theta\bar{\alpha}}(u,\theta|u) - \bar{P}_{\theta\bar{\beta}}(u,\theta|u)| = \|\bar{\alpha} - \bar{\beta}\|$ where $\|\bar{\alpha} - \bar{\beta}\|$ may be replaced by $\|\bar{\bar{\alpha}} - \bar{\bar{\beta}}\|$ when α and β are symmetric. It follows that

$\delta(\mathcal{E}_\alpha,\mathcal{E}_\beta) \leq ||\bar{\alpha}-\bar{\beta}||$ in general and that $\delta(\mathcal{E}_\alpha,\mathcal{E}_\beta) \leq ||\bar{\bar{\alpha}}-\bar{\bar{\beta}}||$ in the symmetric case. We shall, however, see that these upper bounds may be very bad. If, for example $\bar{\bar{\alpha}}$ and $\bar{\bar{\beta}}$ are mutually singular, then $||\bar{\alpha}-\bar{\beta}|| = ||\bar{\bar{\alpha}}-\bar{\bar{\beta}}|| = 2$ while the deficiencies $\delta(\mathcal{E}_\alpha,\mathcal{E}_\beta)$ and $\delta(\mathcal{E}_\beta,\mathcal{E}_\alpha)$ may both be, say, less than 10^{-100}.

We shall now, in order to get lower bounds for deficiencies, consider the problem of estimating the restrictions $(\theta|w_1,\theta|w_2,\ldots,\theta|w_r)$ of θ to given nonempty subsets $w_1,\ldots,w_r$ of I. If our proposals for these restrictions are respectively, $t_1,t_2,\ldots,t_r$ then we put the loss $= 0$ or $= 1$ according to whether at least one of the restrictions have been correctly estimated or not. Let $\bar{\mathcal{E}}_{\bar{\alpha}}$ be realized by (U,X) where $U \in \mathcal{U}$ is distributed according to $\bar{\alpha}$ while $X = \theta|U$ when θ prevails. Choose a $\theta^0 \in \Theta$ and define an estimator $\rho = (\rho_1,\ldots,\rho_r)$ by putting $\rho_\nu(U,X) = X|w_\nu$ or $\theta^0|w_\nu$ as $U \supseteq w_\nu$ or $U \not\supseteq w_\nu$. The risk at $\theta \in \Theta$ is then $\sum\{\bar{\alpha}(u): u \not\supseteq w_1, u \not\supseteq w_2,\ldots,u \not\supseteq w_r\}$ or 0 as $\theta^0|w_\nu \neq \theta|w_\nu; \nu=1,2,\ldots,r$ or not.

Assuming that there is a $\theta \in \Theta$ such that $\theta(i) \neq \theta^0(i)$ for all i we see that maximum risk is $C = 1 - \sum\{\bar{\alpha}(u): u \supseteq w_1 \text{ or} \ldots \text{ or } u \supseteq w_r\}$. Suppose now that there was a decision rule with smaller maximum risk. Restrict, for the moment, θ to some finite subset $\tilde{\Theta}$ of Θ. If λ_0 is a least favorable prior distribution on $\tilde{\Theta}$ then any Bayes solution for λ_0 is minimax. Thus we may assume that there is a nonrandomized decision rule $\tilde{\rho}$ with risk $< C$ for all $\theta \in \tilde{\Theta}$. Let $\mathcal{D}_1$ consist of all sets $u \in \mathcal{U}$ which does not contain any set w_ν and put $\mathcal{D}_2 = \mathcal{U} - \mathcal{D}_1$. The risk at θ may then be decomposed as $\sum_1 + \sum_2$ where $\sum_s = \sum\{\bar{\alpha}(u) : \tilde{\rho}_\nu(u,\theta|u) \neq \theta|w_\nu ; \nu=1,\ldots,r, u \in \mathcal{D}_s\}$. Our assumption implies that $\sum_1 < C = \sum\{\bar{\alpha}(u): u \in \mathcal{D}_1\}$ for all $\theta \in \tilde{\Theta}$. Hence, for all $\theta \in \tilde{\Theta}$, there is a $u \in \mathcal{D}_1$ such that $\tilde{\rho}_\nu(u,\theta|u) = \theta|w_\nu$ for some ν. If $u \in \mathcal{D}_1$ then there are points $i_{u,1},\ldots,i_{u,r}$ such that $i_{u,\nu} \in w_\nu - u; \nu=1,\ldots,r$. Put for each pair $(u,x): \rho^*_\nu(u,x) = \tilde{\rho}_t(u,x)(i_{u,t})$. Then $\tilde{\Theta} = \cup\{\tilde{\Theta}_{u,\nu}: u \in \mathcal{D}_1, \nu \in \{1,\ldots,r\}\}$ where $\tilde{\Theta}_{u,\nu} = \{\theta: \theta \in \tilde{\Theta}, g^*_\nu(u,\theta|u) = \theta(i_{u,\nu})\}$.

It follows that there is a finite subset $\{i_1,\ldots,i_m\}$ of $(w_1\cup\ldots\cup w_r)-u$ and functions $f_{i_1},f_{i_2},\ldots,f_{i_m}$ on $\tilde{\Theta}$ such that $\tilde{\Theta}=\bigcup_{\nu=1}^{m}\tilde{\Theta}_\nu$ where $\tilde{\Theta}_\nu=\{\theta:\theta(i_\nu)=f_{i_\nu}(\theta)\}$ and each f_{i_ν} depends on $\theta\in\tilde{\Theta}$ via $\theta|w_\nu$. We may without loss of generality assume that $i_1,\ldots,i_m$ are distinct.

There are several conditions which we may impose on Θ in order to ensure the impossibility of this. Suppose, for example, that $\# I=N<\infty$, $\tilde{\Theta}=\eta^N$ where $\#\eta=k>N$. Then the constuction above implies the contradiction: $Nk^{N-1}<k^N=\#\tilde{\Theta}\leq\sum_{\nu=1}^{m}\#\tilde{\Theta}_\nu\leq mk^{N-1}\leq Nk^{N-1}$. Similarly if $\# I=\infty$ and $\Theta\supseteq\eta_1^\infty$ where $\#\eta_1=\infty$. In that case $\tilde{\Theta}$ may be chosen as follows. Choose $\theta^0\in\eta_1^\infty$ and let η be some subset of η_1 containing $k>\#\{w_1\cup\ldots\cup w_r\}$ elements. Then the above arguments lead to the following contradiction:

$$\#\{w_1\cup\ldots\cup w_r\}k^{m-1}<k^m=\#\tilde{\Theta}\leq\sum_{\nu=1}^{m}\#\tilde{\Theta}_\nu\leq mk^{-1}\leq\#\{w_1\cup\ldots\cup w_r\}k^{m-1}.$$

Altogether we have shown that C is the minimax risk whenever $\Theta\supseteq\eta^I$ where $\#\eta\geq 1+\# I$. Hence, since the loss function is nonnegative and bounded by 1 :

$\frac{1}{2}\delta(\mathcal{E}_\alpha,\mathcal{E}_\beta)=\frac{1}{2}\delta(\bar{\mathcal{E}}_{\bar{\alpha}},\bar{\mathcal{E}}_{\bar{\beta}})\geq\beta(H)-\alpha(H)$ where $H=\{u:u\in U$ and $u\supseteq w_i$ for some $i\}$. As any increasing class of sets is a limit of such families we find that :

$\delta(\mathcal{E}_\alpha,\mathcal{E}_\beta)=\delta(\bar{\mathcal{E}}_{\bar{\alpha}},\bar{\mathcal{E}}_{\bar{\beta}})\geq 2\sup[\beta(H)-\alpha(H)]$ where the sup is over all increasing classes in $(U,\subseteq)$. Using a result of Strassen [10] we find the following criterions for deficiency :

<u>Theorem 7</u>. Suppose $\Theta\supseteq\eta^I$ where $\#\eta\geq 1+\# I$. Let α and β be sampling plans and let $\varepsilon\geq 0$. Then the following conditions are all equivalent :

(i) $\delta(\mathcal{E}_\alpha,\mathcal{E}_\beta)=\delta(\bar{\mathcal{E}}_{\bar{\alpha}},\bar{\mathcal{E}}_{\bar{\beta}})\leq\varepsilon$

(ii) $\bar{\beta}(H)-\bar{\alpha}(H)\leq\frac{\varepsilon}{2}$ for any increasing class H of sets in $(U,\subseteq)$

(iii) $\int h d\bar{\beta} - \int h d\bar{\alpha} \leq \frac{\varepsilon}{2}||h||$ for any isotonic function h on $(\mathcal{U},\subseteqq)$

(iv) There is a joint distribution $\bar{\rho}$ on $\mathcal{U}^2$ with marginals $\bar{\alpha}$ and $\bar{\beta}$ such that $\bar{\rho}(\{(u,v) : u \supseteqq v\}) \geq 1 - \frac{\varepsilon}{2}$

Remark. The equivalence of conditions (ii), (iii) and (iv) and the fact that these conditions imply (i) does not require any condition on Θ. It should be apparent from Strassen [11] and the proof below that these equivalences hold if $(\mathcal{U},\subseteqq)$ is replaced by quite general partially ordered sets. This has, when $\varepsilon = 0$, been noted by several authors.

Proof: If (ii) holds, then (iii) follows by writing $\int h d(\bar{\beta}-\bar{\alpha}) = \int_0^{||h||} (\bar{\beta}-\bar{\alpha})(h \geq t)dt$ and by noting that $[h \geq t]$ is an increasing class of sets. Applying (iii) to indicator functions we recover (ii). Thus (ii)$\Longleftrightarrow$(iii). By Theorem 11 in Strassen [11] condition (iv) is equivalent to the condition that $\bar{\beta}(H) \leqq \bar{\alpha}(\{u : u \supseteqq v$ for some $v \in H\}) + \frac{\varepsilon}{2}$ for each subclass H of $\mathcal{U}$. Clearly nothing is lost by restricting attention to isotonic subclasses of $(\mathcal{U},\subseteqq)$, and then this is merely a restatement of (ii).

Suppose that $\bar{\rho}$ is as in (iv). Put $\bar{\Gamma}(v|u) = \bar{\rho}(u,v)/\bar{\alpha}(u)$ when $\bar{\alpha}(u) > 0$. Put $\bar{\Gamma}(v|u) = 1$ or 0 as $v=u$ or $v \neq u$ when $\bar{\alpha}(u) = 0$. Define a function A from $\mathcal{U}$ to $[0,1]$ by: $A(u) = \sum\{\bar{\Gamma}(v|u) : v \subseteqq u\}$. Extend $\bar{\chi} = \{(u,x) : u \in \mathcal{U},\ x = \theta|u$ for some $\theta \in \Theta\}$ to a set $\hat{\chi}$ by joining a point ζ not belonging to $\bar{\chi}$. Define finally a Markov kermel M from $\hat{\chi}$ to $\hat{\chi}$ such that $M((v,y)|(u,x)) = \bar{\Gamma}(v|u)$ when $(u,x) \in \bar{\chi}$, $v \subseteqq u$ and $y = x|v$. Then, necessarily, $M(\zeta|(u,x)) = 1 - A(u)$. We find succsesively: $||\bar{P}_{\theta,\bar{\beta}} - \bar{P}_{\theta,\bar{\alpha}}M|| = \sum_v |\bar{\beta}(v) - \sum_u M((v,\theta|v)|(u,\theta|u))\bar{\alpha}(u)| + \sum_u M(\zeta|(u,\theta|u))\bar{\alpha}(u) = \sum_v |\bar{\beta}(v) - \sum_{u \supseteqq v} \bar{\Gamma}(v|u)\bar{\alpha}(u)| + \sum_u (1-A(u))\bar{\alpha}(u) = 2\sum\{\bar{\rho}(u,v) : u \not\supseteqq v\} \leqq \varepsilon$. Thus (iv) implies, without any assumption on Θ. The proof is now completed by noting that, under the stated condition on Θ, the lower bound established immediately before the formulation of this theorem yields the implication (i) $\Rightarrow$ (ii):

$\square$

If α and β are symmetric then comparison may, as we might expect, be expressed in terms of $\bar{\bar{\alpha}}$ and $\bar{\bar{\beta}}$.

<u>Corollary 8</u>. Let α and β be symmetric sampling plans and put $N = \#I$ Then conditions (ii), (iii) and (iv) of theorem 6, are, without any assumption on Θ, equivalent to each of the folloing conditions:

(ii') $\bar{\bar{\beta}}[m,N] - \bar{\bar{\alpha}}[m,w] \le \frac{\varepsilon}{2}$; $m=0,1,\dots,N$

(iii') $\int h\,d\bar{\bar{\beta}} - \int h\,d\bar{\bar{\alpha}} \le \frac{\varepsilon}{2}||h||$ for any isotonic nonnegative function h on $\{0,1,\dots,N\}$

(iv') There is a joint distribution $\bar{\bar{\rho}}$ on $\{0,1,\dots,N\}^2$ with marginals $\bar{\bar{\alpha}}$ and $\bar{\bar{\beta}}$ such that $\bar{\bar{\rho}}(\{(m,n):m \ge n\}) \ge 1-\frac{\varepsilon}{2}$.

<u>Proof</u>: The equivalence of (ii'), (iii') and (iv') follows by the remark after theorem 7. Suppose these conditions are satisfied. Let h be a nonegative isotonic function on $(\mathcal{U},\subsetneq)$. Then $E_{\bar{\bar{\alpha}}}h(U) = E_{\bar{\bar{\alpha}}}g(\#U)$ and $E_{\bar{\bar{\beta}}}h(U) = E_{\bar{\bar{\beta}}}g(\#U)$ where $g(m) = E(h(U)\,|\,\#U=m) = \binom{N}{m}^{-1}\sum\{h(u):\#u=m\}$. Clearly $||g|| \le ||h||$ and g is isotonic since

$g(m+1) = \binom{N}{m+1}^{-1}\sum\{h(u):\#u=m+1\} \ge \binom{N}{m+1}^{-1}\sum\limits_{u:\#u=m+1}\frac{1}{m+1}\sum\{h(v):v\subsetneq u,\ \#v=m\} = \binom{N}{m+1}^{-1}\frac{1}{m+1}(N-m)\sum\{h(v):\#v=m\} = g(m)$; $m=0,1,\dots,N-1$. Hence, by (iii'), $E_{\bar{\bar{\beta}}}h(U) - E_{\bar{\bar{\alpha}}}h(U) = E_{\bar{\bar{\beta}}}g(\#U) - E_{\bar{\bar{\alpha}}}g(\#U) \le \frac{\varepsilon}{2}||g|| \le \frac{\varepsilon}{2}||h||$. Thus condition (iii) of theorem 7 is established. Conversely, suppose (iii) of theorem 7 (and hence (ii)) is satisfied. Let $m \le N$ and put $H = \{u:\#u \ge m\}$. Then H is isotonic. Hence $\bar{\bar{\beta}}[m,N] - \bar{\bar{\alpha}}[m,N] = \bar{\beta}(H) - \bar{\alpha}(H) \le \frac{\varepsilon}{2}$. Thus (ii') holds.

$\square$

Example 9. (Approximation by fixed size sampling plans).

Let α be a symmetric sampling plan and let w_k be the sampling plan consisting of drawing "randomly" without replacement k elements, i.e. $\bar{w}_k(u) = \binom{N}{k}^{-1}$ when $\#u=k$. Then $\delta(\mathcal{E}_\alpha, \mathcal{E}_{w_k}) = 2\bar{\alpha}[0,k-1]$ while $\delta(\mathcal{E}_{w_k}, \mathcal{E}_\alpha) = 2\bar{\alpha}[k+1,N]$ so that $\delta(\mathcal{E}_\alpha, \mathcal{E}_{w_k}) + \delta(\mathcal{E}_{w_k}, \mathcal{E}_\alpha) = 2\|\alpha - w_k\|$.
Thus, if $\bar{\alpha}(r) = \binom{N}{r} p^r(1-p)^{N-r}$; $r=0,1,\ldots,N$ where $p \in]0,1[$ then, as $p \to 0$, $\delta(\mathcal{E}_{w_k}, \mathcal{E}_\alpha) \to 0$ although $\|\alpha - w_k\| \to 2$

Note also that the best approximation, w.r.t. Δ , to $\mathcal{E}_\alpha$ by a fixed size sampling plan $\mathcal{E}_{w_k}$ is obtained by letting k be a median in $\bar{\alpha}$. Thus it is, in general, not expected sample size but the median samplesize which yields the approximation.

Example 10. (Inequalities for symmetric sampling plans).

Define for each finite subset u of I a vector $\zeta(u) = (\zeta_1(u), \zeta_2(u), \ldots, \zeta_N(u)) \in R^N$ by: $\zeta_i(u) = [\#u]^{-1}$ or $=0$ as $i \in u$ or not; $i=1,\ldots,N$. Then $\sum_{i=1}^{N} \zeta_i(u)\theta(i)$ is the arithmetic average of the observed θ-values after repetitions in the sample sequence have been removed. If the sampling is without replacement then $\sum_{i=1}^{N} \zeta_i(u)\theta(i)$ is just the arithmetic average $\frac{1}{n}[\theta(i_1)+\ldots+\theta(i_n)]$.

Consider now a convex function φ on $[-1,1]^N$. Suppose the random sample sequence $I = (I_1,\ldots,I_n)$ is distributed according to the symmetric sampling plan α . Let K_i ; $i \in I$ be the absolute frequency of individual i in the sequence $(I_1,\ldots,I_n)$. By symmetry the distribution of K_i given $U = \{I\}$ does not depend on i as long as i is restricted to U . In particular $E(\frac{1}{n}K_i \mid \{I\}=u) = \frac{1}{\#u} \sum_{j \in u} E(K_j|n|\{I\}=u) = (\#u)^{-1}$ when $i \in u$. Writing $K = (K_1,\ldots,K_N)$ we find

$$\zeta(U) = E[(K/n)|U]$$

Hence, by Jensen's inequality:

$$(1) \qquad E\varphi(K/n) \geq E\varphi(\zeta(U))$$

Consider another symmetric sampling plan β and let $\bar{\rho}$ be a joint distribution for the random pair (U,V) satisfying condition (iv) of theorem 7 with $\varepsilon = 2\sup_m[\bar{\bar{\beta}}[m,N] - \bar{\bar{\alpha}}[m,N]]$. Then, by convexity:

$$E_{\bar{\beta}}\varphi(\zeta(v)|U) \geq \sum_{v\subseteq U}\varphi(\zeta(v))Pr(V=v|U) - \|\varphi\|\sum_{v\not\subseteq U}Pr(V=v|U) \geq$$

$\varphi(\sum\zeta(v)Pr(V=v|U,v\subseteq U)Pr(V\subseteq U|U) - \|\varphi\|Pr(V\not\subseteq U|U)$. Now, by symmetry, $\bar{\rho}$ may, and shall, be chosen so that $\bar{\rho}(\pi(u),\pi(v)) = \bar{\rho}(u,v)$ for any permutations π of I . It follows that $Pr(V=v|U,v\subseteq U)$ only depends on the cardinalities of v and U as long as $v\subseteq U$.

Hence $\sum_v\zeta(v)Pr(V=v|U,v\subseteq U) = \zeta(U)$ so that

$$E_{\bar{\beta}}\varphi(\zeta(v))|U) \geq \varphi(\zeta(U))Pr(V\subseteq U|U) - \|\varphi\|Pr(V\not\subseteq U|U) =$$

$$\varphi(\zeta(U)) - Pr(V\not\subseteq U|U)[\varphi(\zeta(U)) + \|\varphi\|] \geq \varphi(\zeta(U)) - 2Pr(V\not\subseteq U|U)\|\varphi\|$$

It follows that:

$$(2) \qquad E_{\bar{\beta}}\varphi(\zeta(U)) \geq E_{\bar{\alpha}}\varphi(\zeta(U)) - \varepsilon\|\varphi\|$$

Combining (1) and (2) we get:

$$(3) \qquad E_{\bar{\beta}}\varphi(K/n) \geq E_{\bar{\beta}}\varphi(\zeta(U)) \geq E_{\bar{\alpha}}\varphi(\zeta(U)) - 2\max_m(\bar{\bar{\beta}} - \bar{\bar{\alpha}})([m,N])\|\varphi\|$$

In particular; for any convex function ψ on $[\min_i\theta_i, \max_i\theta_i]$

$$(4) \qquad E_\beta\psi(\frac{1}{n}\sum_{\nu=1}^{n}\theta(I_\nu)) \geq E_{\bar{\beta}}\psi(\frac{1}{\#U}\sum_U\theta_i) \geq E_{\bar{\alpha}}\psi(\frac{1}{\#U}\sum_U\theta_i) -$$

$$2\|\psi\|\max_m(\bar{\bar{\beta}} - \bar{\bar{\alpha}})([m,N]) \ .$$

The left most inequalities in (3) and (4) may trivially, be replaced by equalities when β is without replacement.

(4) generalizes various generalizations (See Lanke [6] and Marshall & Olkin [10]) of the basic inequalities for sampling with and without replacement in Hoeffding [4].

Appendix. A discussion of the relationship of the problems in Karlin [5] to the theory of comparison of experiments.

The sampling plans $\alpha_{p,n,\pi}$ where p is the uniform distibution on I and π is a product measure on $\{0,1\}^{n-1}$ was considered in Karlin [5]. We shall, in order to discuss the relationship between example 5 and some results in [5], need a few concepts.

Note first that each sequence $(i_1,\ldots,i_n)$ in I_s determines its empirical probability distribution $H(\cdot|(i_1,\ldots,i_n))$ on I where $H(i|(i_1,\ldots,i_n)) = \#\{\nu: \nu \leq n : i_\nu = i\}/n$. Identify this distribution with the "probability" vector $(H(i|(i_1,\ldots,i_n);i\in I)$.

Let H_0 be the uniform distribution on I, i.e. $H_0(i) = 1/N$; $i \in I$. The ordered pair $(H_0,H(\cdot|(i_1,\ldots,i_n))$ determines a dichotomy $D(i_1,\ldots,i_n)$. If α is a sampling plan then the α-mixture, $E_\alpha D_I = \sum\alpha(i_1,\ldots,i_n)D(i_1,\ldots,i_n)$, is well defined. Let also, for each sampling plan α, S_α be the probability distribution of the random probability vector $(H(i|I)\,;i\in I)$ when $I \in I_s$ is distributed according to α. Note that S_α determines $\bar{\alpha}$ and $\bar{\bar{\alpha}}$ by:

$$\bar{\alpha}(u) = S_\alpha(\{x:x_i>0 \Leftrightarrow i\in u\})\ ;u\in U$$

and

$$\bar{\bar{\alpha}}(n) = S_\alpha(\{x:\#\{i:x_i>0\}=n\})\ ;n=0,1,2,\ldots$$

Let us from here on, for simplicity, restrict attention to symmetric sampling plans assigning mass zero to the empty sequence. Note that if α has this property then $E_\alpha H(i|\cdot) = \frac{1}{N}$; $i\in I$. Let, for each $i\in I$, $NH(i|\cdot)S_\alpha$ be the distibution on I having density $NH(i|\cdot)$ w.r.t. S_α. Then $E_\alpha = (S_\alpha, NH(i|\cdot)S_\alpha\ ;i\in I)$ is an experiment whose parameter set contains $N+1$ points.

Let now α and β be symmetric sampling plans realized by observing, respectively, I and J in I_s. Consider the following conditions:

C_1: $\quad D_J \geqq D_I$ a.s. for some joint distribution of (I,J)

C_2: $\quad F_\beta \geqq F_\alpha$

C_3: $\quad ED_J \geqq ED_I$

Then $C_1 \Rightarrow C_2 \Rightarrow C_3$ since C_i ; $i=1,2,3$, is equivalent to respectively

$$E\,\varphi(J) \geqq E\,\varphi(I)\ ;\varphi \in \Phi_i;\ i=1,2,3, \text{ where:}$$

$$\Phi_1 = \{\varphi : \varphi(i_1,\ldots,i_n) \equiv \psi(H(i|i_1,\ldots,i_n)\ ;\ i\in I)$$
where ψ is Schur convex}

$$\Phi_2 = \{\varphi : \varphi(i_1,\ldots,i_n) \equiv \psi(H(i|i_1,\ldots,i_n)\ ;\ i\in I)$$
where ψ is symmetric and convex}

$$\Phi_3 = \{\varphi : \varphi(i_1,\ldots,i_n) \equiv \sum_i g(H(i|i_1,\ldots,i_n))$$
where g is convex}

Consider also the following classes of functions on I_s:

$$\Phi_K = \{\varphi : \varphi \text{ is symmetric \& } 2\varphi(i,j,i_3,\ldots,i_n) \leqq \varphi(i,i,i_3\ldots i_n) + \varphi(j,j,i_3\ldots i_n)$$
whenever $i,j,i_3,\ldots,i_n \in I_s\}$,

$$\Phi_Z = \{\varphi : \varphi \text{ is symmetric \&}$$
$$\sum_{ij}\varphi(i,j,i_3,\ldots,i_n)\omega_i\omega_j \geqq 0$$
whenever $i,j,i_3,\ldots i_n \in I_s$ & $\sum \omega_i = 0\}$

and

$$\Phi_\# = \{\varphi : \varphi(i_1,\ldots,i_n) \equiv g(\#\{i_1,\ldots,i_n\}) \text{ where}$$
$$g(0) \geqq g(1) \geqq g(2) \geqq \text{--}\quad\}$$

The classes Φ_K and Φ_Z was considered by, respectively, Karlin [5] and Van Zwet [15]. The ordering C_1 is considered in Marshall and Olkin [10]. According to the terminology there, α dominates β if and only if C_1 is satisfied.

By trying out N-tuples $\omega_i\,; i \in I$ such that $\sum\omega_i = 0$ and $\#\{i : \omega_i \neq 0\} \leqq 2$ we see that

$$\Phi_Z \subseteqq \Phi_K$$

It is also easy to see that

$$\Phi_{\#} \subseteqq \Phi_K$$

while $(i_1,\ldots,i_n) \rightsquigarrow g(\#\{i_1,\ldots,i_n\})$ is in Φ_Z if and only g is convex and monotonically decreasing.

Any $\varphi \in \Phi_K$ defines a function of the relative frequencies which became Schur convex after symmetrization. On the other hand symmetric and convex functions of the relative frequencies are in Φ_K . Furthermore $\Phi_3 \subseteqq \Phi_Z$ so that ordering by Φ_Z implies ordering by mixtures.

Note that ordering by $\Phi_{\#}$ is equivalent to ordering by Φ_K whenever we restrict attention to symmetric sampling plans with a fixed number $n \leqq 3$ of drawings (The number of drawings corresponding to the sample $(i_1,\ldots,i_n) \in I_s$ is n). Thus, by proposition 5, C_1 holds for $\alpha = \alpha_{p,n,\pi}$ and $\beta = \alpha_{p,n,\pi'}$ when π is stochastically greater than π' and $n \leqq 3$.

Identify now each product measure π on $\{0,1\}^{n-1}$ with the vector $\pi = (\pi_1,\pi_2,\ldots,\pi_{n-1})$ where $\pi_\nu\,; \nu=1,2,\ldots,n-1$ is the probability that the ν-th individual drawn is replaced before the $(\nu+1)$-th is drawn.

Let $\alpha_{n,\pi} = \alpha_{p,n,\pi}$ where p is the uniform distribution on I and $n \leqq N$. In [5] Karlin conjectured that $E_{\alpha_{n,\pi}}\varphi \geqq E_{\alpha_{n,\pi'}}\varphi;\ \varphi \in \Phi_K$, if and only if $\pi \geqq \pi'$. By the remark after the example the "only if" does not hold in general. The "if" however appears still to be open although important progress was made by Van Zwet who proved that $E_{n,(1,\dots,1)}\varphi \geq E_{n,\pi}\varphi$ for all $\varphi \in \Phi_Z$ and all $\pi = (\pi_1,\pi_2,\dots,\pi_{n-1})$. We do not know if Karlin's conjecture implies the more general conjecture - namely that C_1 holds for $\alpha = \alpha_{n,\pi}$ and $\beta = \alpha_{n,\pi'}$ when $\pi \geq \pi'$. It appears however that several of the statements in Karlin's paper may be phrased in terms of condition C_1 . As an example we prove the following:

Proposition 6. Suppose C_1 holds for $\alpha = \alpha_{n,1,1,\dots,1}$ and $\beta = \alpha_{n,0,1,\dots,1}$ whenever $n \leq n_0 \wedge N$. Then C_1 holds also for $\alpha = \alpha_{n,1,\dots,1}$ and $\beta = \alpha_{n,\pi_1,\dots,\pi_{n-1}}$ for all $\pi_1,\dots,\pi_{n-1}$ provided $n \leq n_0 \wedge N$.

Remark: The proposition, as well as its proof, is modelled after lemma 3.1 and its proof in Karlin [5] .

Proof: Consider for each $m \leq n_0$ the statement $S_m : C_1$ holds for $\alpha = \alpha_{n,1,\dots,1}$ and $\beta = \alpha_{n,\pi_1,\dots,\pi_{n-1}}$ for all $\pi_1,\dots,\pi_{n-1}$ whenever $n \leq m \wedge N$. Clearly S_1 is true. Suppose S_{m-1} is established. Let $n \leq m \wedge N$ and $\varphi(i_1,\dots,i_n) = \psi(H(\nu|i_1,\dots,i_n);\nu \in I)$ where ψ is Schur convex. We must, in order to establish S_m, show that

$$E_{\alpha_{n,\pi_1,\dots,\pi_{n-1}}}\varphi(I) \leq E_{\alpha_{n,1,\dots,1}}\varphi(I)$$

If $n \leq m-1$ then this follows from S_{m-1}. Thus we may, and shall, assume that $n=m$. Let us, for this proof, write E_π for $E_{\alpha_{n,\pi_1,\dots,\pi_{n-1}}}$ and use superscript (ν) on E to indicate that individual ν is removed from the population.

Conditioning w.r.t. the first drawn individual and the decision on whether this individual should be replaced or not we find:

$$E_\pi \varphi(I_1,\ldots,I_m) = (\pi_1/N)\sum_\nu E_\pi[\varphi(\nu,I_2,\ldots,I_m) \mid I_1=\nu \text{ is not replaced}]$$

$$+ ((1-\pi_1)/N)\sum_\nu E_\pi[\varphi(\nu,I_2,\ldots,I_m) \mid I_1=\nu \text{ is replaced}] =$$

$$(\pi_1/N)\sum_\nu E_{\pi_2,\ldots,\pi_{m-1}}\varphi(\nu,I_2,\ldots,I_m) +$$

$$((1-\pi_1)/N)\sum_\nu E^{(\nu)}_{\pi_2,\ldots,\pi_{m-1}}\varphi(\nu,I_2,\ldots,I_m) =$$

$$= \pi_1 E_{\pi_2,\ldots,\pi_{m-1}}\tilde{\varphi}(I_2,\ldots,I_m) + ((1-\pi_1)/N)\sum_\nu E^{(\nu)}_{\pi_2,\ldots,\pi_{m-1}}\varphi(\nu,I_2,\ldots,I_m)$$

where $\tilde{\varphi}(i_2,\ldots,i_m) = \frac{1}{N}\sum_\nu \varphi(\nu,i_2,\ldots,i_m)$.

Write $(i_1,\ldots,i_m) \prec (j_1,\ldots,j_m)$ when $D_{(i_1,\ldots,i_m)} \leqq D_{(j_1,\ldots,j_m)}$.

Then $(i,i_2,\ldots,i_m) \prec (i,j_2,\ldots,j_m)$ when $(i_2,\ldots,i_m) \prec (j_2,\ldots,j_m)$ and $i \notin \{i_2,\ldots,i_m,j_2,\ldots,j_m)$. Hence, by the induction hypothesis, the last sum is

$$\leqq ((1-\pi_1)/N)\sum_\nu E^{(\nu)}_{1,1,\ldots 1}\varphi(\nu,I_2,\ldots,I_m) =$$

$$(1-\pi_1)E_{0,1,\ldots,1}\varphi(I_1,I_2,\ldots,I_m) \leqq (1-\pi_1)E_{1,1,\ldots,1}\varphi(I_1,I_2,\ldots I_m)$$

Note next that $\tilde{\varphi}(i_2,\ldots,i_m)$ is a Schur convex function of $H(\nu \mid i_2,\ldots,i_m)$; $\nu \in \mathcal{I}$. [Let $(i_2,\ldots,i_m) \prec (j_2,\ldots,j_m)$ and let K_ν L_ν be, respectively, the absolute frequency of ν w.r.t. $(i_2,\ldots,i_m)$ and $(j_2,\ldots,j_m)$. Thus, using the terminology in [10], $K = (K_1,\ldots,K_N)$ is majorized by $L = (L_1,\ldots,L_N)$.

Let $K_{(1)} \geqq K_{(2)} \geqq \dots \geqq K_{(N)}$ and $L_{(1)} \geqq L_{(2)} \geqq \dots \geqq L_{(N)}$ be, respectively, $(K_1,\dots,K_N)$ and $(L_1,\dots,L_N)$ ordered in decreasing order. Let e_t be the m-tuple whose t-th element is 1 and having all other elements equal to zero. Then:

$$N\tilde{\varphi}(i_2,\dots,i_m) = \sum_t \varphi(K_{(1)},\dots,K_{(N)}) + e_t) \leqq$$

$$\sum_t \varphi((L_{(1)},\dots,L_{(N)})) + e_t) = N\tilde{\varphi}(j_2,\dots,j_m)$$

The last inequality follows from Fulkerson and Ryser [2]. This result may also be found as lemma D.2 in Chapter 5 in Marshall and Olkin [10].] Using the induction hypothesis once more we find that the first term is $\leqq \pi_1 E_{1,\dots,1}\tilde{\varphi}(I_2,\dots,I_m) = \pi_1 E_{1,\dots,1}\varphi(I_1,\dots,I_m)$. It follows that $E_\pi \varphi(I_1,\dots,I_m) \leqq \pi_1 E_{1,\dots,1}\varphi(I_1,\dots,I_m) +$ $(1-\pi_1)E_{1,\dots,1}\varphi(I_1,\dots,I_m) = E_{1,1,\dots,1}\varphi(I_1,\dots,I_m)$. Thus S_m holds. Hence, by induction, S_{n_0} holds.

□

Referring to Karlin [5] we may now, by substituting the proposition above for lemma 3.1 in [5] and then copying part (i) of theorem 3.1 and its proof, deduce that C_1 holds for $\alpha = \alpha_{n,1\dots,1}$ and $\beta = \alpha_{n,\pi_1,\dots,\pi_{n-1}}$ for all $\pi_1,\dots,\pi_{n-1}$ when $(N/N-1)^{n-1} \leqq n/(n-3)$.

According to a theorem of Muirhead, see Marshall and Olkin [10], condition C_1 for sampling plans α and β implies:

(§) $$E_\alpha a(I_1)a(I_2)\dots a(I_n) \geqq E_\beta a(I_1)a(I_2)\dots a(I_n)$$

for all nonnegative functions a on $\mathcal{I}$. It follows directly from

Van Zwet [15] that (§) holds for $\alpha=\alpha_{n,1,1,\dots,1}$ and $\beta=\alpha_{n,0,1,\dots,1}$ when $n \leq N$. Assuming, which we without loss of generality may, that $\sum_\nu a(\nu)=1$ we see that (§) in this case may be written:

$$(\S\S) \qquad \sum_{\nu=1}^{N} a(\nu)(1-a(\nu))^{n-1} \leq (1-\tfrac{1}{N})^{n-1} \quad .$$

Now $(\sum a(\nu)(1-a(\nu))^{n-1})^{\frac{1}{n-1}} \uparrow 1-\bigwedge_\nu a(\nu)$ as $n\to\infty$. Thus (§§) does not hold for arbitrarily large n when $a(\nu)\neq \frac{1}{N}$ for at least one ν. Let us complete these comments by showing directly that (§§) holds when $n \leq N$.

Restrict a to the set of probability distributions on I. Consider $Q=\sum a(\nu)(1-a(\nu))^{n-1}$ as a function of $a(1),a(2),\dots,a(N-1)$.. We must show that $\bar{Q}=\max_a Q(a) = Q(\frac{1}{N},\dots,\frac{1}{N})$. Suppose a^0 is a probability distribution on I such that $a^0(1)=0<a^0(2)$. Then the derivative at 0 of the function $a(1)(1-a(1))^{n-1} + (a^0(2)-a(1))(1-a^0(2)+a(1))^{n-1} + \sum_{\nu=3}^{N} a^0_\nu\,(1-a^0_\nu)^{n-1}$ w.r.t. $a(1)$ at 0 is $1-(1-na^0_2)(1-a^0_2)^{n-2}$ which is positive. It follows that the minimal support of any maximizing a^0 is I. We may then without loss of generality assume that $a^0(1)\geq a^0(2)\geq\dots\geq a^0(N)>0$.

Putting $U(x)=(1-nx)(1-x)^{n-1};\ 0\leq x\leq 1$ we find $\partial Q/\partial a(\nu) = U(a(\nu))-U(a(N))$. Hence $U(a^0(\nu))=U(a^0(N))$; $\nu\in I$. Now U obtains its maximum at $x=2/n$ and is strictly decreasing (increasing) as $x<\frac{2}{n}(x>\frac{2}{n})$. This imply that either $a^0(N)=\dots=a^0(s+1)<a^0(s)=\dots=a^0(1)$ for some $s<N$ or $a^0(1)=a^0(2)=\dots=a^0(N)=\frac{1}{N}$. In the latter case $\bar{Q}=(1-\frac{1}{N})^{n-1}$ and

we are through. The first case, however, can't occure when $n \leqq N$. To see this note that $a^0(N) < \frac{1}{N} \leqq \frac{1}{n}$, that $U(x) \gtrless 0$ as $x \gtrless \frac{1}{n}$ and that $a^0(N) < \frac{1}{n} < \frac{2}{n} < a^0(1)$. Thus we obtain the contradiction: $0 < U(a^0(N)) = U(a^0(1)) < 0$.

REFERENCES

[1] Cassel, C., Särndal, C., Wretman, J.H. 1977. Foundations of inference in survey sampling. Wiley, New York.

[2] Fulkerson, D.R., Ryser, H.J. 1962. Multiplicities and minimal widths for (0-1) matrices. Canad. J. Math., 14, 498-508.

[3] Heyer, H. 1973. Mathematische Theorie Statistischer Experimente. Springer Verlag.

[4] Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc., 58, 13-30.

[5] Karlin, S. 1974. Inequalities for symmetric sampling plans I. Ann. Statist., 2, 1065-1094.

[6] Lanke, J. 1974. On an inequality of Hoeffding and Rosén. Scand. J. Statist., 1, 84-86.

[7] Le Cam, L. 1964. Sufficiency and approximate sufficiency. Ann. Math. Statist., 35, 1419-1455.

[8] Le Cam, L. 1974. Notes on asymptotic methods in statistical decision theory. Centre de Recherches Math. Univ. de Montreal.

[9] Le Cam, L. 1975. Distances between experiments. In a survey of statistical design and linear models (ed. J.N. Srivastava), 383-395. North-Holland, Amsterdam.

[10] Marshall, A.W., Olkin, I. 1979. Inequalities: Theory of majorization and its applications. Academic Press, New York.

[11] Strassen, V. 1965. The existences of probability measures with given marginals. Ann. Math. Statist., 36, 423-439.

[12] Swensen, A.R. 1980. Deficiencies between linear normal experiments. Ann. Statist., 8, 1142-1155.

[13] Torgersen, E.N. 1976. Comparison of statistical experiments. Scand. J. Statist., 3, 186-208.

[14] Torgersen, E.N. 1981. Measures of information based on comparison with total information and with total ignorance. To appear in Ann. Statist.

[15] Van Zwet, W.R. 1980. Random replacement sampling plans. Lecture given at the meeting on mathematical statistics and probability, March 23-29, 1980 at Oberwohlfach. W. Germany.