PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 3, Fasc. 1 (1982), p. 1-17

COMPARISON OF SOME STATISTICAL EXPERIMENTS ASSOCIATED WITH SAMPLING PLANS

BY

ERIK TORGERSEN (OSLO)

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

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

It is known that θ belongs to some set Θ of functions on $\mathscr{I}$.

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

$$x = \big((i_1, \theta(i_1)), \ldots, (i_n, \theta(i_n))\big)$$

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

Let $\mathscr{E}_\alpha$ denote the experiment obtained by observing x and assume that Θ is not too small. Then $\mathscr{E}_{\alpha_1}$ is at least as informative as $\mathscr{E}_{\alpha_2}$ if and only if the sampled subset under α_2 is "stochastically contained" in the sampled subset under α_1.

Using the theory of comparison of statistical experiments we shall here discuss this and other related results.

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 [7]) 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 [11]. The purpose of this paper is to present some simple applications for experiments associated with sampling plans. We refer to [2], [7], [8], and [12] for expositions of the theory of comparison of experiments. The material covered in Section 2 of [13] is adequate here.

Consider a population $\mathscr{I}$ which is an (and may be any) enumerable set. Suppose also that there is a characteristic of interest which, with varying amount (value, degree, etc.), is possessed by all individuals in $\mathscr{I}$. Let $\theta(i)$ be the amount of this characteristic for an individual $i \in \mathscr{I}$. The function θ on $\mathscr{I}$ defined in 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 $\mathscr{I}$.

In order to find out about θ we may take a sample from $\mathscr{I}$ 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 $\mathscr{I}_s$ of finite sequences of elements from $\mathscr{I}$. Before proceeding let us agree that a probability measure on an enumerable set is defined for all subsets. To retain the possibility of making no observations at all we may include the "empty" sequence $\emptyset$ in $\mathscr{I}_s$. 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 with probability $\alpha(i_1, \ldots, i_n)$. Thus we may let our sample space consist of all sequences $(i_1, f_1), \ldots, (i_n, f_n)$, where $(i_1, \ldots, i_n) \in \mathscr{I}_s$, $f_1, \ldots, f_n \in \bigcup_\theta \theta[\mathscr{I}]$ 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. Then the sampling plan α determines a statistical experiment $\mathscr{E}_\alpha = (P_{\theta,\alpha}\colon \theta \in \Theta)$.

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

$$P_{\theta,\alpha}\big((i_1, f_1), \ldots, (i_n, f_n)\big) = \begin{cases} \alpha(i_1, \ldots, i_n) & \text{if } (f_1, \ldots, f_n) = \big(\theta(i_1), \ldots, \theta(i_n)\big), \\ 0 & \text{otherwise.} \end{cases}$$

As is well known, (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 equivalent experiment $\bar{\mathscr{E}}_{\bar\alpha} = (\bar P_{\theta,\bar\alpha}\colon \theta \in \Theta)$ which may be described as follows.

Let $\mathscr{U}$ be the class of all finite subsets of $\mathscr{I}$. If $u \in \mathscr{U}$ and α is a sampling plan on $\mathscr{I}$, then $\bar\alpha$ is the probability distribution on $\mathscr{U}$ 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 $\mathscr{I}$.

We may then let the sample space χ of $\bar{\mathscr{E}}_{\bar\alpha}$ consist of all pairs (u, x), where $u \in \mathscr{U}$ 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 $x \neq \theta|u$, respectively.

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

Note that the set of experiments $\mathscr{E}_\alpha$, and hence the set of experiments $\mathscr{E}_{\bar{\alpha}}$, is closed under products. More precisely, $\mathscr{E}_\alpha \times \mathscr{E}_\beta \sim \mathscr{E}_\gamma$, where

$$\gamma(k_1, k_2, \ldots, k_r) = \alpha(\emptyset)\beta(k_1, \ldots, k_r) + \alpha(k_1)\beta(k_2, \ldots, k_r) + \ldots + \\ + \alpha(k_1, \ldots, k_{r-1})\beta(k_r) + \alpha(k_1, \ldots, k_r)\beta(\emptyset), \quad (k_1, \ldots, k_r) \in \mathscr{I}_s,$$

so that

$$\bar{\gamma}(u) = \sum \{\bar{\alpha}(u_1)\bar{\beta}(u_2) \colon u_1 \cup u_2 = u\}, \quad u \in \mathscr{U}.$$

Some notation and other terms which will be used in the sequel:

$\mathscr{I}$ – a population.

$N = \# \mathscr{I}$.

$\mathscr{I}_s$ – the set of finite sequences of elements from $\mathscr{I}$.

$\mathscr{U}$ – the class of finite subsets of $\mathscr{I}$.

$\# A$ – the number of elements in A or ∞ as A is finite or infinite.

$\alpha, \beta, \ldots$ – probability distributions on $\mathscr{I}_s$.

$\bar{\alpha}$ – the probability measure on $\mathscr{U}$ 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 $z = z_n$.

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

$\|\mu\|$ – total variation of μ.

$\mathscr{E} \geqslant \mathscr{F}$: the experiment $\mathscr{E}$ is at least as informative as the experiment $\mathscr{F}$.

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

$\delta(\mathscr{E}, \mathscr{F})$ – the deficiency of $\mathscr{E}$ with respect to $\mathscr{F}$. If $\mathscr{E} = (P_\theta \colon \theta \in \Theta)$ and $\mathscr{F} = (Q_\theta \colon \theta \in \Theta)$, then $\delta(\mathscr{E}, \mathscr{F})$ is [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.

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

Isotonic = monotonically increasing: A map φ from a partially ordered set $(\chi, \leqslant)$ to a partially ordered set is called *monotonically increasing* (*decreasing*) if $\varphi(x_1) \leqslant \varphi(x_2)$ whenever $x_1 \leqslant x_2$ $(x_1 \geqslant x_2)$.

2. Comparability of experiments $\mathscr{E}_\alpha$. In order to simplify the notation we write "$\mathscr{E} \geqslant \mathscr{F}$" instead of "$\mathscr{E}$ is at least as informative as $\mathscr{F}$". If $\mathscr{E} \geqslant \mathscr{F}$ and $\mathscr{F} \geqslant \mathscr{E}$, then we say that $\mathscr{E}$ and $\mathscr{F}$ are *equivalent* and write $\mathscr{E} \sim \mathscr{F}$.

Among several natural (and fortunately equivalent) ways of introducing the notation of comparison we can use the *randomization* (Markov kernel, transition, etc.) *criterion of Le Cam*, which states roughly that $\mathscr{E} \geqslant \mathscr{F}$ if and only if $\mathscr{F}$ may be obtained from $\mathscr{E}$ by a randomization.

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

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

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

$$\sum_{(v,y)} M((v, y)|(u, x)) = 1, \quad (u, x) \in \bar{\chi}.$$

Using the definitions of the measures $\bar{P}$, we may rewrite (1) as

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

Hence

(3) $$1 = \sum_u \left[\sum_v M((v, \theta|v)|(u, \theta|u))\right]\bar{\alpha}(u), \quad \theta \in \Theta.$$

It follows that

$$\sum_v M((v, \theta|v)|(u, \theta|u)) = 1 \quad \text{for } \bar{\alpha}(u) > 0.$$

The following condition will be useful:

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

Let θ^0 be as in (C). Assume $\bar{\alpha}(u^0) > 0$ and put $x^0 = \theta^0|u^0$. Put $\Theta^0 = \{\theta\colon \theta \in \Theta \text{ and } \theta|u^0 = x^0\}$. Then $\theta^0 \in \Theta^0$. Consider so a pair (v, θ), where $v \in \mathscr{U}$ 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) $$M((v, \theta|v)|(u^0, x^0)) \leqslant M((v, \theta^0|v)|(u^0, x^0)), \quad v \in \mathscr{U}.$$

Hence, since both sides add up to 1 in v, the equality holds in (4) for each $v \in \mathscr{U}$. Consider now a particular $v^0 \in \mathscr{U}$ such that $M((v^0, \theta^0|v^0)|(u^0, x^0)) > 0$. Then, by (4) with $\leqslant$ replaced by $=$,

$$M((v^0, \theta|v^0)|(u^0, x^0)) > 0 \quad \text{for each } \theta \in \Theta^0.$$

It follows from (3) that $\theta|v^0 = \theta^0|v^0$, $\theta \in \Theta^0$. If $v^0 \nsubseteq u^0$, then we may choose an $i \in v^0 - u^0$. 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 $(u, v) \in \mathscr{U}^2$ a number $\bar{\Gamma}(v|u)$ by

$$\bar{\Gamma}(v|u) = \begin{cases} M((v, \theta^0|v)|(u, \theta^0|u))\bar{\alpha}(u) & \text{if } \bar{\alpha}(u) > 0, \\ 0 & \text{if } v \neq u \text{ and } \bar{\alpha}(u) = 0, \\ 1 & \text{if } v = u \text{ and } \bar{\alpha}(u) = 0. \end{cases}$$

Then

$$\sum_v \bar{\Gamma}(v|u) = \sum_{v \subseteq u} \bar{\Gamma}(v|u) = 1, \quad u \in \mathscr{U}.$$

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

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

Define finally a joint distribution $\bar{\varrho}$ on $\mathscr{U}^2$ by

$$\bar{\varrho}(u, v) = \bar{\Gamma}(v|u)\bar{\alpha}(u).$$

Then $\bar{\varrho}$ has marginals $\bar{\alpha}$ and $\bar{\beta}$ and $\bar{\varrho}(\{(u, v): u \supseteq v\}) = 1$.

The last established result may be recognized as one of several usual and equivalent ways of expressing the fact that $\bar{\alpha}$ is stochastically larger than $\bar{\beta}$ with respect to the inclusion ordering $\subseteq$ on $\mathscr{U}$.

Suppose now, conversely, that we have been able to construct a joint distribution $\bar{\varrho}$ with this property. Specify the conditional distribution on $\bar{\Gamma}$ of obtaining a "last" set v under the assumption that the "first" is u such that $\sum\{\bar{\Gamma}(v|u): v \subseteq u\} = 1$ for all $u \in \mathscr{U}$. (If $\bar{\alpha}(u) > 0$, then this holds by definition.) Define a Markov kernel M from $\bar{\chi}$ to $\bar{\chi}$ 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 $\bar{\mathscr{E}}_{\bar{\beta}}$ is obtained from $\bar{\mathscr{E}}_{\bar{\alpha}}$ by the randomization M.

We collect this as well as some closely related statements in

THEOREM 1 (comparability criterions). *Suppose Θ satisfies condition* (C). *Then the following four conditions are equivalent*:

(i) $\mathscr{E}_\alpha \geqslant \mathscr{E}_\beta$.

($\bar{\text{i}}$) $\bar{\mathscr{E}}_{\bar{\alpha}} \geqslant \bar{\mathscr{E}}_{\bar{\beta}}$.

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

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

Remark 1. 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.

Remark 2. From well-known results (see Remark 6) on orderings of probability measures on partially ordered sets it follows that ($\bar{\text{ii}}$), and hence (ii), may be expressed as follows:

(ii′) $\mathrm{E}_\alpha h(I) \geqslant \mathrm{E}_\beta h(J)$ *for each bounded function h such that*

$$h(i_1, \ldots, i_m) \leqslant h(j_1, \ldots, j_n) \quad \textit{whenever } \{i_1, \ldots, i_m\} \subseteq \{j_1, \ldots, j_n\}.$$

($\overline{\text{ii}'}$) $\alpha(\mathscr{H}) \geqslant \beta(\mathscr{H})$ *for any increasing class $\mathscr{H} \subseteq \mathscr{U}$.*

Here a subclass $\mathscr{H}$ of $\mathscr{U}$ is called *increasing* if $u \in \mathscr{H}$ whenever $v \in \mathscr{H}$ for some $v \subseteq u$. Trivially, $\mathscr{H}$ is increasing if and only if $\mathscr{H}$ is of the form

$$\mathscr{H} = \bigcup_{\nu=1}^{\infty} \{u\colon u \supseteq w_\nu\}$$

for some sequence $w_1, w_2, \ldots$ in $\mathscr{U}$.

Completion of the proof of Theorem 1. The equivalence (i) $\Leftrightarrow$ $(\bar{\text{i}})$ follows from the sufficiency and we have seen above that $(\bar{\text{i}}) \Leftrightarrow (\bar{\text{ii}})$. The implication (ii) $\Rightarrow$ $(\bar{\text{ii}})$ is trivial, so it remains only to show that $(\bar{\text{ii}}) \Rightarrow$ (ii). Suppose then that $(\bar{\text{ii}})$ is satisfied. Let $\alpha(\cdot|\{I\})$ and $\beta(\cdot|\{J\})$ be the conditional distributions of I given $\{I\}$ and given $\{J\}$, respectively. 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).

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

Corollary 1. *Suppose Θ satisfies* (C). *Then the following three conditions are equivalent*:

(i) $\mathscr{E}_\alpha \sim \mathscr{E}_\beta$. (ii) $\bar{\alpha} = \bar{\beta}$. (iii) $\Phi_{\bar{\alpha}} = \Phi_{\bar{\beta}}$.

Proof. By Remark 2, $\Phi_{\bar{\alpha}} = \Phi_{\bar{\beta}}$ when $\mathscr{E}_\alpha \sim \mathscr{E}_\beta$.

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

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

where $w = \{i\colon \theta^1(i) = \ldots = \theta^r(i)\}$. If Θ satisfies condition (C), then any class $\{u\colon u \subseteq w\}$, where $w \in \mathscr{U}$, is of this form. However, it is not difficult to construct examples of non-comparable sampling plans α and β such that $\Phi_{\bar{\alpha}} \leqslant \Phi_{\bar{\beta}}$.

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

Proposition 1. *Suppose $\Theta \geqslant \eta^{\mathscr{I}}$, where $\# \eta \geqslant 2$. Then $\mathscr{E}_\alpha \geqslant \mathscr{E}_\beta$ if and only if $\mathscr{E}_\alpha$ is at least as informative as $\mathscr{E}_\beta$ for testing problems.*

Proof. Suppose that $\Theta \geqslant \eta^{\mathscr{I}}$, where $\# \eta = 2$, and that $\mathscr{E}_\alpha$ is at least as informative as $\mathscr{E}_\beta$ for testing problems. Choose a $\bar{\theta} \in \eta^{\mathscr{I}}$ and sets $v^1, \ldots, v^r$ in $\mathscr{U}$. Let Θ_0 consist of all $\theta \in \Theta$ such that $\theta|v^\nu \neq \bar{\theta}|v^\nu$, $\nu = 1, \ldots, r$. Let $\bar{\mathscr{E}}_{\bar{\alpha}}$ and $\bar{\mathscr{E}}_{\bar{\beta}}$ be realized by observing (U, X) and (V, Y), respectively. Define the test $\tilde{\varphi} = \tilde{\varphi}(V, Y)$ by putting $\tilde{\varphi} = 1$ if 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 $\tilde{\varphi} = 0$ otherwise. Then $\mathrm{E}_\theta \tilde{\varphi}(V, Y) = 0$, $\theta \in \Theta_0$.

By assumption there is a test $\varphi = \varphi(U, X)$ such that $\mathrm{E}_\theta \varphi \equiv \mathrm{E}_\theta \tilde{\varphi}$. In particular,

$$\sum_u \varphi(u, \theta|u)\alpha(u) = 0 \quad \text{if } \theta \in \Theta_0.$$

Suppose $u \in \mathscr{U}$ 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 $\varphi(u, \bar{\theta}|u)\alpha(u) = 0$ in this case. Consequently,

$$\sum\{\alpha(u)\colon u \supseteq v^1 \text{ or } \ldots \text{ or } u \supseteq v^r\} \geqslant \sum \varphi(u, \bar{\theta}|u)\alpha(u) = \mathrm{E}_{\bar{\theta}}\varphi = \mathrm{E}_{\bar{\theta}}\tilde{\varphi}$$
$$= \sum \tilde{\varphi}(v, \bar{\theta}|v)\beta(v) = \sum\{\beta(v)\colon v \supseteq v^1 \text{ or } \ldots \text{ or } v \supseteq v^r\}.$$

Hence $\alpha(\mathscr{H}) \geqslant \beta(\mathscr{H})$ for any increasing class $\mathscr{H}$ in $(\mathscr{U}, \subseteq)$. The proposition follows now from Theorem 1 and Remark 1.

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

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

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

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

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

Clearly, any probability distribution on $\{0, 1, \ldots, N\}$ is of the form $\bar{\bar{\alpha}}$ for a unique symmetric plan α without replacement. If both α and β are symmetric sampling plans, then the product experiment $\mathscr{E}_\alpha \times \mathscr{E}_\beta$ is equivalent to $\mathscr{E}_\gamma$, where the symmetric sampling plan γ satisfies

$$\bar{\bar{\gamma}}(n)\binom{N}{n}^{-1} = \sum \frac{n(n-r_1+n-r_2)}{(n-r_1)!\,(n-r_2)!}\bar{\bar{\alpha}}(r_1)\binom{N}{r_1}^{-1} \bar{\bar{\beta}}(r_2)\binom{N}{r_2}^{-1},$$

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

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

$$\mathscr{E}_\alpha \sim \sum_{n=0}^{N} \bar{\bar{\alpha}}(n)\mathscr{E}_{\varrho_n},$$

where $\varrho_n(i_1, \ldots, i_n) = [N(N-1)\ldots(N-n+1)]^{-1}$ when $i_1, \ldots, i_n$ are distinct, while $\varrho_n(i_1, \ldots, i_m) = 0$ whenever $m \neq n$. It follows then, since

$$\mathscr{E}_{\varrho_0} \leqslant \mathscr{E}_{\varrho_1} \leqslant \ldots \leqslant \mathscr{E}_{\varrho_n},$$

that $\mathscr{E}_\alpha \geqslant \mathscr{E}_\beta$ 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 $\bar{\bar{\varrho}}$ on $\{0, 1, \ldots, N\}^2$ with marginals $\bar{\bar{\alpha}}$ and $\bar{\bar{\beta}}$ and such that $\bar{\bar{\varrho}}(\{(m, n): m \geqslant n\}) = 1$. Let us put

$$\bar{\bar{\Gamma}}(n|m) = \frac{\bar{\bar{\varrho}}(m, n)}{\bar{\bar{\alpha}}(m)} \quad \text{if } \bar{\bar{\alpha}}(m) > 0.$$

If $\bar{\bar{\alpha}}(m) = 0$, then we may put $\bar{\bar{\Gamma}}(n|m) = 1$ or $\bar{\bar{\Gamma}}(n|m) = 0$ as $n = m$ or $n \neq m$, respectively.

Define a kernel Γ from $\mathscr{U}$ to $\mathscr{U}$ by

$$\Gamma(v|u) = \binom{\# u}{\# v}^{-1} \bar{\bar{\Gamma}}(\# v|\# u) \quad \text{if } v \subseteq u.$$

Put $\Gamma(v|u) = 0$ if $v \nsubseteq u$. Let $v \in \mathscr{F}$ and put $n = \# v$. Then

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

This, together with Theorem 1, proves

THEOREM 2. *Let Θ satisfy condition* (C) *and let α and β be symmetric sampling plans. Then $\mathscr{E}_\alpha \geqslant \mathscr{E}_\beta$ if and only if $\bar{\bar{\alpha}}$ is stochastically greater than $\bar{\bar{\beta}}$.*

Remark 3. 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 $\mathscr{I}$ such that $p(i) > 0$ for all $i \in \mathscr{I}$, n is a positive integer, and π is a probability distribution on $\{0, 1\}^{n-1}$ defined as follows:

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

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

(ii) I_1 is drawn from $\mathscr{I}$ so that $\Pr(I_1 = i_1) = p(i_1)$, $i_1 \in \mathscr{I}$;

(iii) if $I_1, \ldots, I_m$ have been drawn, then stop whenever $m = n$ or if $m < n$ and each element of $\mathscr{I}$ has been drawn without being replaced; otherwise, I_{m+1} is drawn from the remaining part A of the population so 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 2. *Let p and n be fixed. Then $\mathscr{E}_{\alpha_\pi} \leqslant \mathscr{E}_{\alpha_{\pi'}}$ whenever π is stochastically greater (for the pointwise ordering on $\{0, 1\}^{n-1}$) than π'.*

Remark 4. Let $n = 3$. It is then easily seen that $\bar{\alpha}_{\delta_{0,1}}$ is stochastically greater than $\bar{\alpha}_{\delta_{1,0}}$ when $N \geqslant 2$. Thus the converse of the above statement is not true even if we restrict attention to independent and uniformly distributed drawings.

Remark 5. Suppose that $N = \# \mathscr{I} < \infty$ and that p is the uniform distribution on $\mathscr{I}$. Then, by Theorem 2 and Proposition 2,

$$\mathrm{E}_{\alpha_\pi} h(\# \{I_1, \ldots, I_n\}) \geqslant \mathrm{E}_{\alpha_{\pi'}} h(\# \{I_1, \ldots, I_n\})$$

whenever π is stochastically greater 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 [4]. A discussion of the relationship of the problems and results in [4] to the theory of comparison of experiments may be found in [14].

Proof. Note first that $\alpha_\pi(i_1, \ldots, i_n) = \mathrm{E}\alpha_{\delta_\varepsilon}(i_1, \ldots, i_n)$, where ε is distributed according to π and δ_ε is the one-point distribution in ε. Hence $\bar{\alpha}_\pi(u) = \mathrm{E}\bar{\alpha}_{\delta_\varepsilon}(u)$, $u \in \mathscr{U}$. Suppose now that we know that $\bar{\alpha}_{\delta_\varepsilon}$ is "stochastically contained" in $\bar{\alpha}_{\delta_{\varepsilon'}}$ whenever $\varepsilon \geqslant \varepsilon'$. (The terminology 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 distributions P and Q, respectively.) Let h be an isotonic function on $(\mathscr{U}, \subseteq)$. Then $\sum_u h(u)\bar{\alpha}_{\delta_\varepsilon}(u)$ is monotonically decreasing in ε. Hence

$$\sum_u h(u)\alpha_\pi(u) = \sum_\varepsilon \sum_u h(u)\alpha_{\delta_\varepsilon}(u)\pi(\varepsilon) \leqslant \sum_\varepsilon \sum_u h(u)\alpha_{\delta_\varepsilon}(u)\pi'(\varepsilon) = \sum_u h(u)\alpha_{\pi'}(u).$$

It follows that $\bar{\alpha}_\pi$ is stochastically contained in $\bar{\alpha}_{\pi'}$. Therefore, it suffices to prove that $\bar{\alpha}_{\delta_\varepsilon}$ is stochastically contained in $\bar{\alpha}_{\delta_{\varepsilon'}}$ when $\varepsilon \geqslant \varepsilon'$. We shall show this by proving that the sampling plans $\alpha_{\delta_\varepsilon}$, $\varepsilon \in \{0, 1\}^{n-1}$, may all be imbedded within a single stochastic framework. This framework will consist of independent $\mathscr{I}$-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, for each m-tuple $(i_1, \ldots, i_m)$ with $m < n$ and for each sequence $\varepsilon_1, \ldots, \varepsilon_m$ of 0's and 1's we put

$$A(i_1, \ldots, i_m, \varepsilon_1, \ldots, \varepsilon_m) = \mathscr{I} - \{i_\nu \colon \nu \leqslant m \text{ and } \varepsilon_\nu = 0\}.$$

Thus $A(i_1, \ldots, i_m, \varepsilon_1, \ldots, \varepsilon_m)$ are precisely the elements left in $\mathscr{I}$ after $i_1, \ldots, i_m$ have been drawn and the replacement policy $(\varepsilon_1, \ldots, \varepsilon_m)$ has been used.

For given ε we define recursively random variables $R_1, \ldots, R_n$ as follows:

(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 $\mu \geq 1$ such that

$$V_{\mu,m+1} \in A(V_{1,R_1,\ldots}, V_{m,R_m,\varepsilon_1,\ldots,\varepsilon_m}) \quad \text{as } A(V_{1,R_1,\ldots}, V_{m,R_m,\varepsilon_1,\ldots,\varepsilon_m}) \neq \emptyset.$$

Put $R_{m+1} = \infty$ otherwise.

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

(a) $R_{m'} \geq R_m$.

(b) *If $I'_1, \ldots, I'_m$ are defined, then $I_1, \ldots, I_m$ are also defined and*

$$A(I'_1, \ldots, I'_m, \varepsilon'_1, \ldots, \varepsilon'_m) \subseteq A(I_1, \ldots, I_m, \varepsilon_1, \ldots, \varepsilon_m).$$

(c) *If $I'_1, \ldots, I'_m$ are defined, then $I_1, \ldots, I_m$ are also defined and*

$$\{I_1, \ldots, I_m\} \subseteq \{I'_1, \ldots, I'_m\}.$$

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

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, $A'_{m-1} \subseteq A_{m-1}$ 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 (a).

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 $A'_m \subseteq A_m$ whenever $\varepsilon_m = 1$. If $\varepsilon_m = 0$, then $\varepsilon'_m = 0$ since $\varepsilon' \leq \varepsilon$. The only case which then needs particular attention is $A'_{m-1} \ni I_m \neq I'_m$. This, however, is impossible since $R'_m \geq R_m$. Hence (b) 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

$$I_m \in \{I'_1, \ldots, I'_{m-1}\} \subseteq \{I'_1, \ldots, I'_m\}.$$

This proves (c).

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 $i_1, \ldots, i_m$ ($m \leq \nu$) under the assumption that the sequence $i_1, \ldots, i_{m-1}$ has been obtained.) Our claims concerning the sampling plans α_π follow now from (c) and Theorem 1.

4. Deficiencies and distances. Let us proceed to the slightly more difficult problem on deficiencies between experiments $\mathscr{E}_\alpha$. Thus we shall try to find out how much do we lose (in risk say) under the least favourable conditions for comparison by basing our decisions on $\mathscr{E}_\alpha$ instead of on $\mathscr{E}_\beta$. Following Le Cam [6] 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 $\delta(\mathscr{E}_\alpha, \mathscr{E}_\beta) \leqslant \|\bar{\alpha} - \bar{\beta}\|$ in general and $\delta(\mathscr{E}_\alpha, \mathscr{E}_\beta) \leqslant \|\bar{\bar{\alpha}} - \bar{\bar{\beta}}\|$ in the symmetric case. However, we shall see that these upper bounds may be very bad. If, for example, $\bar{\alpha}$ and $\bar{\beta}$ are mutually singular, then $\|\bar{\alpha} - \bar{\beta}\| = \|\bar{\bar{\alpha}} - \bar{\bar{\beta}}\| = 2$, while the deficiencies $\delta(\mathscr{E}_\alpha, \mathscr{E}_\beta)$ and $\delta(\mathscr{E}_\beta, \mathscr{E}_\alpha)$ may both be, say, less than 10^{-100}.

In order to get lower bounds for deficiencies we now consider the problem of estimating the restrictions $(\theta|w_1, \ldots, \theta|w_r)$ of θ to given non-empty subsets $w_1, \ldots, w_r$ of $\mathscr{I}$. If our proposals for these restrictions are $t_1, \ldots, t_r$, respectively, then we put the loss equal to 0 or 1 according to whether at least one of the restrictions has been correctly estimated or not. Let $\bar{\mathscr{E}}_{\bar{\alpha}}$ be realized by (U, X), where $U \in \mathscr{U}$ is distributed according to $\bar{\alpha}$ while $X = \theta|U$ when θ prevails. Choose a $\theta^0 \in \Theta$ and define an estimator $\varrho = (\varrho_1, \ldots, \varrho_r)$ by $\varrho_v(U, X) = X|w_v$ or $\varrho_v(U, X) = \theta^0|w_v$ according as $U \supseteq w_v$ or $U \nsupseteq w_v$. The risk at $\theta \in \Theta$ is then $\sum\{\bar{\alpha}(u)\colon u \nsupseteq w_1, \ldots, u \nsupseteq w_r\}$ or 0 for $\theta^0|w_v \neq \theta|w_v$ or $\theta^0|w_v = \theta|w_v$ $(v = 1, \ldots, r)$, respectively.

Assuming that there is a $\theta \in \Theta$ such that $\theta(i) \neq \theta^0(i)$ for all i, we see that the maximum risk is

$$C = 1 - \sum\{\bar{\alpha}(u)\colon u \supseteq w_1 \text{ or } \ldots \text{ or } u \supseteq w_r\}.$$

Suppose now that there is a decision rule with smaller maximum risk. Restrict, for the moment, θ to some finite subset $\tilde{\Theta}$ of Θ. If λ_0 is the least favourable prior distribution on $\tilde{\Theta}$, then any Bayes solution for λ_0 is minimax. Thus we may assume that there is a non-randomized decision rule $\tilde{\varrho}$ with risk less than C for all $\theta \in \tilde{\Theta}$. Let $\mathscr{D}_1$ consist of all sets $u \in \mathscr{U}$ which do not contain any set w_v and put $\mathscr{D}_2 = \mathscr{U} - \mathscr{D}_1$. The risk at θ may then be decomposed as $\sum_1 + \sum_2$, where

$$\textstyle\sum_s = \sum\{\bar{\alpha}(u)\colon \tilde{\varrho}_v(u, \theta|u) \neq \theta|w_v;\ v = 1, \ldots, r,\ u \in \mathscr{D}_s\}.$$

Our assumption implies that $\sum_1 < C = \sum\{\bar{\alpha}(u)\colon u \in \mathscr{D}_1\}$ for all $\theta \in \tilde{\Theta}$. Hence, for all $\theta \in \tilde{\Theta}$ there is a $u \in \mathscr{D}_1$ such that $\tilde{\varrho}_v(u, \theta|u) = \theta|w_v$ for some v. If $u \in \mathscr{D}_1$, then there are points $i_{u,1}, \ldots, i_{u,r}$ such that $i_{u,v} \in w_v - u$, $v = 1, \ldots, r$. For each pair (u, x) we put $\varrho_v^*(u, x) = \tilde{\varrho}_t(u, x)(i_{u,t})$. Then

$$\tilde{\Theta} = \bigcup\{\tilde{\Theta}_{u,v}\colon u \in \mathscr{D}_1, v \in \{1, \ldots, r\}\},$$

$$\text{where } \tilde{\Theta}_{u,v} = \{\theta\colon \theta \in \tilde{\Theta}, g_v^*(u, \theta|u) = \theta(i_{u,v})\}.$$

It follows that there are a finite subset $\{i_1, \ldots, i_m\}$ of $(w_1 \cup \ldots \cup w_r) - u$ and functions $f_{i_1}, \ldots, f_{i_m}$ 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 $\theta | w_\nu$. Without loss of generality we may 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 $\# \mathscr{I} = N < \infty$, $\tilde{\Theta} \in \eta^N$, where $\# \eta = k > N$. Then the construction above implies the contradiction:

$$Nk^{N-1} < k^N = \# \tilde{\Theta} \leqslant \sum_{\nu=1}^{m} \# \tilde{\Theta}_\nu \leqslant mk^{N-1} \leqslant Nk^{N-1}.$$

Similarly for $\# \mathscr{I} = \infty$ 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:

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

We have shown altogether that C is the minimax risk whenever $\Theta \supseteq \eta^{\mathscr{I}}$, where $\# \eta \geqslant 1 + \# \mathscr{I}$. Hence, since the loss function is non-negative and bounded by 1, we have

$$\tfrac{1}{2}\delta(\mathscr{E}_\alpha, \mathscr{E}_\beta) = \tfrac{1}{2}\delta(\bar{\mathscr{E}}_{\bar{\alpha}}, \bar{\mathscr{E}}_{\bar{\beta}}) \geqslant \beta(\mathscr{H}) - \alpha(\mathscr{H}),$$

where $\mathscr{H} = \{u: u \in \mathscr{U} \text{ and } u \supseteq w_i \text{ for some } i\}$. As any increasing class of sets is a limit of such families, we infer that

$$\delta(\mathscr{E}_\alpha, \mathscr{E}_\beta) = \delta(\bar{\mathscr{E}}_{\bar{\alpha}}, \bar{\mathscr{E}}_{\bar{\beta}}) \geqslant 2 \sup [\beta(\mathscr{H}) - \alpha(\mathscr{H})],$$

where the supremum is over all increasing classes in $(\mathscr{U}, \subseteq)$. Using a result of Strassen [10] we find the following criterions for deficiency:

THEOREM 3. *Suppose* $\Theta \supseteq \eta^{\mathscr{I}}$, *where* $\# \eta \geqslant 1 + \# \mathscr{I}$. *Let* α *and* β *be sampling plans and let* $\varepsilon \geqslant 0$. *Then the following conditions are all equivalent*:

(i) $\delta(\mathscr{E}_\alpha, \mathscr{E}_\beta) = \delta(\bar{\mathscr{E}}_{\bar{\alpha}}, \bar{\mathscr{E}}_{\bar{\beta}}) \leqslant \varepsilon$.

(ii) $\bar{\beta}(\mathscr{H}) - \bar{\alpha}(\mathscr{H}) \leqslant \varepsilon/2$ *for any increasing class* $\mathscr{H}$ *of sets in* $(\mathscr{U}, \subseteq)$.

(iii) $\int h d\bar{\beta} - \int h d\bar{\alpha} \leqslant 2^{-1}\varepsilon \|h\|$ *for any isotonic function* h *on* $(\mathscr{U}, \subseteq)$.

(iv) *There is a joint distribution* $\bar{\varrho}$ *on* $\mathscr{U}^2$ *with marginals* $\bar{\alpha}$ *and* $\bar{\beta}$ *such that* $\bar{\varrho}(\{(u, v): u \supseteq v\}) \geqslant 1 - \varepsilon/2$.

Remark 6. The equivalence of conditions (ii), (iii) and (iv), and the fact that these conditions imply (i) do not require any condition on Θ. It should be apparent from [10] and the proof below that these equivalences hold if $(\mathscr{U}, \subseteq)$

is replaced by quite general partially ordered sets. For $\varepsilon = 0$ this has been noted by several authors.

Proof. If (ii) holds, then (iii) follows from

$$\int h d(\bar{\beta}-\bar{\alpha}) = \int_0^{\|h\|} (\bar{\beta}-\bar{\alpha})(h \geqslant t)\, dt$$

and from noting that $[h \geqslant t]$ is an increasing class of sets. Applying (iii) to indicator functions we obtain (ii). Thus (ii) $\Leftrightarrow$ (iii).

By Theorem 11 in [10], (iv) is equivalent to the condition

$$\bar{\beta}(\mathscr{H}) \leqslant \bar{\alpha}(\{u\colon u \supseteq v \text{ for some } v \in \mathscr{H}\})+\varepsilon/2$$

for each subclass $\mathscr{H}$ of $\mathscr{U}$. Clearly, nothing is lost by restricting attention to isotonic subclasses of $(\mathscr{U}, \subseteq)$, and then this is merely a restatement of (ii).

Suppose that $\bar{\varrho}$ is as in (iv). Put $\bar{\Gamma}(v|u) = \bar{\varrho}(u, v)/\bar{\alpha}(u)$ when $\bar{\alpha}(u) > 0$. Put $\bar{\Gamma}(v|u) = 1$ and $\bar{\Gamma}(v|u) = 0$ as $v = u$ and $v \neq u$, respectively, when $\bar{\alpha}(u) = 0$. Define a function A from $\mathscr{U}$ to $[0, 1]$ by

$$A(u) = \sum\{\bar{\Gamma}(v|u)\colon v \subseteq u\}.$$

Extend $\bar{\chi} = \{(u, x)\colon u \in \mathscr{U}, x = \theta|u \text{ for some } \theta \in \Theta\}$ to a set $\hat{\chi}$ by joining a point ζ not belonging to $\bar{\chi}$. Finally, define a Markov kernel M from $\hat{\chi}$ to $\hat{\chi}$ by $M((v, y)|(u, x)) = \bar{\Gamma}(v|u)$ when $(u, x) \in \bar{\chi}$, $v \subseteq u$, and $y = x|v$. Then, necessarily, $M(\zeta|(u, x)) = 1-A(u)$. We find successively

$$\begin{aligned}\|\bar{P}_{\theta,\bar{\beta}}-\bar{P}_{\theta,\bar{\alpha}} M\| &= \sum_v |\bar{\beta}(v)-\sum_u M((v, \theta|v)|(u, \theta|u))\bar{\alpha}(u)| +\sum_u M(\zeta|(u, \theta|u))\bar{\alpha}(u)\\ &= \sum_v |\bar{\beta}(v)-\sum_{u\supseteq v} \bar{\Gamma}(v|u)\bar{\alpha}(u)| +\sum_u (1-A(u))\bar{\alpha}(u)\\ &= 2\sum\{\bar{\varrho}(u, v)\colon u \not\supseteq v\} \leqslant \varepsilon.\end{aligned}$$

Thus (iv) implies (i) 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):

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

COROLLARY 2. *Let α and β be symmetric sampling plans and put $N = \# \mathscr{I}$. Then conditions* (ii), (iii), *and* (iv) *of Theorem* 3 *are, without any assumption on Θ, equivalent to each of the following conditions*:

(ii′) $\bar{\bar{\beta}}[m, N]-\bar{\bar{\alpha}}[m, w] \leqslant \varepsilon/2$, $m = 0, 1, \ldots, N$.

(iii′) $\int hd\bar{\bar{\beta}}-\int hd\bar{\bar{\alpha}} \leqslant 2^{-1}\varepsilon\|h\|$ *for any isotonic non-negative function h on* $\{0, 1, \ldots, N\}$.

(iv′) *There is a joint distribution $\bar{\bar{\varrho}}$ on $\{0, 1, \ldots, N\}^2$ with marginals $\bar{\bar{\alpha}}$ and $\bar{\bar{\beta}}$ such that* $\bar{\bar{\varrho}}(\{(m, n)\colon m \geqslant n\}) \geqslant 1-\varepsilon/2$.

Proof. The equivalence of (ii′), (iii′) and (iv′) follows by Remark 6. Suppose these conditions are satisfied. Let h be a non-negative isotonic function on $(\mathscr{U}, \subseteq)$. Then

$$\mathrm{E}_{\bar{\alpha}} h(U) = \mathrm{E}_{\bar{\alpha}} g(\# U) \quad \text{and} \quad \mathrm{E}_{\bar{\beta}} h(U) = \mathrm{E}_{\bar{\beta}} g(\# U),$$

where

$$g(m) = \mathrm{E}\big(h(U) \mid \# U = m\big) = \binom{N}{m}^{-1} \sum \{h(u) \colon \# u = m\}.$$

Clearly, $\|g\| \leqslant \|h\|$ and g is isotonic since

$$\begin{aligned} g(m+1) &= \binom{N}{m+1}^{-1} \sum \{h(u) \colon \# u = m+1\} \\ &\geqslant \binom{N}{m+1}^{-1} \sum_{u: \# u = m+1} \frac{1}{m+1} \sum \{h(v) \colon v \subseteq u,\ \# v = m\} \\ &= \binom{N}{m+1}^{-1} \frac{1}{m+1} (N-m) \sum \{h(v) \colon \# v = m\} = g(m), \end{aligned}$$

$$m = 0, 1, \ldots, N-1.$$

Hence, by (iii′),

$$\mathrm{E}_{\bar{\beta}} h(U) - \mathrm{E}_{\bar{\alpha}} h(U) = \mathrm{E}_{\bar{\beta}} g(\# U) - \mathrm{E}_{\bar{\alpha}} g(\# U) \leqslant \frac{\varepsilon}{2}\|g\| \leqslant \frac{\varepsilon}{2}\|h\|.$$

Thus condition (iii) of Theorem 3 is established. Conversely, suppose (iii) of Theorem 3 (and hence (ii)) holds. Let $m \leqslant N$ and put $\mathscr{H} = \{u \colon \# u \geqslant m\}$. Then $\mathscr{H}$ is isotonic. Hence $\bar{\bar{\beta}}[m, N] - \bar{\bar{\alpha}}[m, N] = \bar{\beta}(\mathscr{H}) - \bar{\alpha}(\mathscr{H}) \leqslant \varepsilon/2$. Thus (ii′) holds.

Example 1 (approximation by fixed size sampling plans). Let α be a symmetric sampling plan and let w_k be the sampling plan consisting of k elements drawn "randomly" without replacement, i.e.

$$\bar{w}_k(u) = \binom{N}{k}^{-1} \quad \text{if } \# u = k.$$

Then $\delta(\mathscr{E}_\alpha, \mathscr{E}_{w_k}) = 2\bar{\bar{\alpha}}[0, k-1]$ while $\delta(\mathscr{E}_{w_k}, \mathscr{E}_\alpha) = 2\bar{\bar{\alpha}}[k+1, N]$, so that $\delta(\mathscr{E}_\alpha, \mathscr{E}_{w_k}) + \delta(\mathscr{E}_{w_k}, \mathscr{E}_\alpha) = 2\|\alpha - w_k\|$. Thus, if

$$\bar{\bar{\alpha}}(r) = \binom{N}{r} p^r (1-p)^{N-r}, \qquad r = 0, 1, \ldots, N,$$

where $p \in]0, 1[$, then $\delta(\mathscr{E}_{w_k}, \mathscr{E}_\alpha) \to 0$ as $p \to 0$ although $\|\alpha - w_k\| \to 2$.

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

Example 2 (inequalities for symmetric sampling plans). For each finite subset u of $\mathscr{I}$ define a vector $\zeta(u) = (\zeta_1(u), \ldots, \zeta_N(u)) \in R^N$ by $\zeta_i(u) = [\# u]^{-1}$ as $i \in u$ and $\zeta_i(u) = 0$ as $i \notin u$, $i = 1, \ldots, N$. Then $\sum_{i=1}^{N} \zeta_i(u)\theta(i)$ is the arithmetic mean 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 mean $n^{-1}[\theta(i_1)+ \ldots +\theta(i_n)]$.

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 $(i \in \mathscr{I})$ be the absolute frequency of an 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,

$$\mathrm{E}\left(\frac{1}{n}K_i \middle| \{I\} = u\right) = \frac{1}{\# u} \sum_{j \in u} \mathrm{E}(K_j|n|\{I\} = u) = (\# u)^{-1} \quad \text{as } i \in u.$$

Writing $K = (K_1, \ldots, K_N)$ we get $\zeta(U) = \mathrm{E}[(K/n)|U]$. Hence, by Jensen's inequality,

$$\mathrm{E}\varphi(K/n) \geqslant \mathrm{E}\varphi(\zeta(U)). \tag{5}$$

Consider another symmetric sampling plan β and let $\bar{\varrho}$ be a joint distribution for the random pair (U, V) satisfying condition (iv) of Theorem 3 with

$$\varepsilon = 2 \sup_m [\bar{\beta}[m, N] - \bar{\alpha}[m, N]].$$

Then, by convexity,

$$\begin{aligned} \mathrm{E}_{\bar{\beta}}\varphi(\zeta(v)|U) &\geqslant \sum_{v \subseteq U} \varphi(\zeta(v))\Pr(V = v|U) - \|\varphi\| \sum_{v \nsubseteq U} \Pr(V = v|U) \\ &\geqslant \varphi(\textstyle\sum \zeta(v)\Pr(V = v|U, v \subseteq U)\Pr(V \subseteq U|U)) - \|\varphi\| \Pr(V \nsubseteq U|U). \end{aligned}$$

Now, by symmetry, $\bar{\varrho}$ may (and shall) be chosen so that $\bar{\varrho}(\pi(u), \pi(v)) = \bar{\varrho}(u, v)$ for any permutations π of $\mathscr{I}$. It follows that $\Pr(V = v|U, v \subseteq U)$ depends only 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

$$\begin{aligned} &\mathrm{E}_{\bar{\beta}}\varphi(\zeta(v)|U) \geqslant \varphi(\zeta(U))\Pr(V \subseteq U|U) - \|\varphi\| \Pr(V \nsubseteq U|U) \\ &= \varphi(\zeta(U)) - \Pr(V \nsubseteq U|U)[\varphi(\zeta(U)) + \|\varphi\|] \geqslant \varphi(\zeta(U)) - 2\Pr(V \nsubseteq U|U)\|\varphi\|. \end{aligned}$$

It follows that

$$E_{\bar{\beta}}\varphi(\zeta(U)) \geqslant E_{\bar{\alpha}}\varphi(\zeta(U)) - \varepsilon\|\varphi\|. \tag{6}$$

Combining (5) and (6) we get

$$E_{\bar{\beta}}\varphi(K/n) \geqslant E_{\bar{\beta}}\varphi(\zeta(U)) \geqslant E_{\bar{\alpha}}\varphi(\zeta(U)) - 2\max_m(\bar{\beta}-\bar{\alpha})([m, N])\|\varphi\|. \tag{7}$$

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

$$\begin{aligned} E_{\beta}\psi\left(\frac{1}{n}\sum_{\nu=1}^{n}\theta(I_\nu)\right) &\geqslant E_{\bar{\beta}}\psi\left(\frac{1}{\# U}\sum_U \theta_i\right) \\ &\geqslant E_{\bar{\alpha}}\psi\left(\frac{1}{\# U}\sum_U \theta_i\right) - 2\|\psi\|\max_m(\bar{\beta}-\bar{\alpha})([m, N]). \end{aligned} \tag{8}$$

The most left inequalities in (7) and (8) may trivially be replaced by equalities when β is without replacement.

Formula (8) generalizes various extended versions (see [5] and [9]) of the basic inequalities for sampling with and without replacement in [3].

REFERENCES

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

[2] H. Heyer, *Mathematische Theorie statistischer Experimente*, Springer Verlag, 1973.

[3] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. 58 (1963), p. 13-30.

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

[5] J. Lanke, *On an inequality of Hoeffding and Rosén*, Scand. J. Statist. 1 (1974), p. 84-86.

[6] L. Le Cam, *Sufficiency and approximate sufficiency*, Ann. Math. Statist. 35 (1964), p. 1419-1455.

[7] – *Notes on asymptotic methods in statistical decision theory*, Centre de Recherches Math. Univ. de Montréal, 1974.

[8] – *Distances between experiments*, p. 383-395 in: *A survey of statistical design and linear models* (ed. J. N. Srivastava), North-Holland, Amsterdam 1975.

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

[10] V. Strassen, *The existences of probability measures with given marginals*, Ann. Math. Statist. 36 (1965), p. 423-439.

[11] A. R. Swensen, *Deficiencies between linear normal experiments*, Ann. Statist. 8 (1980), p. 1142-1155.

[12] E. N. Torgersen, *Comparison of statistical experiments*, Scand. J. Statist. 3 (1976), p. 186-208.

[13] – *Measures of information based on comparison with total information and with total ignorance*, Ann. Statist. 9 (1981), p. 638-657.

[14] – *Comparison of some statistical experiments associated with sampling plans*, Stat. Res. Report, Univ. of Oslo, 1981.

Institute of Mathematics
University of Oslo
Blindern, Oslo 3, Norway

Received on 20. 5. 1981
