The Annals of Statistics
1989, Vol. 17, No. 4, 1886–1905

THE CONSTRUCTION OF ΠPS SAMPLING DESIGNS THROUGH A METHOD OF EMPTYING BOXES[1]

BY A. HEDAYAT, BING-YING LIN AND J. STUFKEN

University of Illinois at Chicago, National Cheng-Chi University and Iowa State University

We present a simple but universal technique for the construction of ΠPS sampling designs. A tool that is used in the construction consists of playing a game in which objects are removed from N boxes, n at a time, and at most one from each box at a time. Necessary and sufficient conditions on N, n and the contents of the boxes are established such that all boxes can be emptied by this process.

It is shown that every ΠPS design can be derived from such a game. Sampling designs with additional properties are obtained through additional restrictions on emptying the boxes. Various rigorous methods are presented, complemented by numerous suggestions. The emphasis is on controlling sample selection probabilities and inequalities for the first- and second-order inclusion probabilities. The method is very adaptive to computer use.

1. Introduction. Consider a finite population consisting of N identifiable units, say $1, 2, \dots, N$. Associated with the ith unit is an unknown quantity y_i, which, if desired, can be observed. The objective is to estimate $T = \sum_{i=1}^N y_i$, the population total, based on a sample of size n. We further assume that a known positive quantity x_i is associated with the ith unit and that there is reason to believe that the y_i's are approximately proportional to the x_i's. Various sampling and estimation techniques have been suggested in the literature, to utilize the information contained in the x_i's for obtaining estimates of the population total with a higher precision. Many of these are discussed and compared in Brewer and Hanif (1983). One of the more popular strategies is the use of so-called ΠPS sampling designs in conjunction with the Horvitz–Thompson (1952) estimator for T. To introduce this strategy explicitly we need some notation. Instead of using the quantities x_i, it is more convenient to use

$$z_i = x_i \Big/ \left(\sum_{j=1}^N x_j \right),$$

also called the (normed) size measure of unit i. A fixed size n ΠPS sampling design based on the units $1, 2, \dots, N$ with size measures $z_1, z_2, \dots, z_N$ is a pair

$$d = (S_d, P_d),$$

Received September 1987; revised November 1988.

[1]Research supported by grants AFOSR 85-0320 and 89-0221.

AMS 1980 *subject classifications*. Primary 62D05; secondary 62K10.

Key words and phrases. Inclusion probabilities to size sampling, unequal probability sampling, auxiliary size measures, controlled probability sampling, BIB designs.

with the following properties:

1. S_d is a collection of subsets of $\{1, 2, \ldots, N\}$, all of which have cardinality n.
2. P_d is a positive function on S_d.
3. $\sum_{s \in S_d} P_d(s) = 1$.
4. $\sum_{s \in S_d, s \ni i} P_d(s) = nz_i$ for $i = 1, 2, \ldots, N$.

The elements of S_d are the possible samples, together forming the support of d. Without loss of generality we will restrict our attention to what is known as reduced sampling designs. This allows us to present a sample as a subset of the units instead of an ordered n-tuple based on the units. The cardinality of S_d is the support size of d, while $P_d(s)$, $s \in S_d$, is the probability that s is the selected sample. The quantity $\sum_{s \in S_d, s \ni i} P_d(s)$ is usually denoted by π_i and is called the first-order inclusion probability for unit i under d. The second-order inclusion probability for i and j, $i \neq j$, under d is defined as $\sum_{s \in S_d, s \ni i, j} P_d(s)$ and denoted by π_{ij}. With such a sampling design, the Horvitz–Thompson estimator for T is defined as

$$\hat{T} = \sum_{i \in s} y_i/\pi_i,$$

where s is the sample selected through d. It can easily be shown that $\hat{T}$ is an unbiased estimator of T. The variance of $\hat{T}$, say V, may be expressed as

$$V = \sum_{i<j}\sum (\pi_i\pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2. \tag{1.1}$$

This expression was first derived by Sen (1953) and Yates and Grundy (1953). Unbiased estimators for V exist if and only if $\pi_{ij} > 0$, $i \neq j$. Under that assumption, the following expression provides us with such an estimator, also called the Sen–Yates–Grundy variance estimator:

$$\hat{V} = \sum_{\substack{i<j \\ i, j \in s}}\sum ((\pi_i\pi_j - \pi_{ij})/\pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2. \tag{1.2}$$

There are a few things we can learn from looking at the expressions in (1.1) and (1.2). First of all, assuming that we use the Horvitz–Thompson estimator, we would like to use a design d with the π_i's proportional to the y_i's. This would minimize the expression for V. Since the y_i's are unknown, we have to settle for the next best thing and use a design with π_i's proportional to the known x_i's. This is exactly condition 4 in the definition of a Π*PS* sampling design and is the main motivation for the use of these designs. But, if possible, we would like some other conditions to be satisfied. Obviously we like $\pi_{ij} > 0$ for all $i \neq j$, due to (1.2). Since not everyone is comfortable with a negative variance estimate, we would also like that $\pi_{ij} \leq \pi_i\pi_j$, since this guarantees that $\hat{V}$ is nonnegative. Although other selection procedures with different estimators are available, there is neither theoretical nor empirical evidence that any of these are superior to the strategy described here.

In addition to the desired properties of a ΠPS sampling design, as just pointed out, it is not inconceivable that some other properties of the design are desired. For example, certain samples may be considerably more expensive or less convenient to implement than others. We might want to manipulate the support of the design to exclude such samples from it, without violating the other properties. In a general context, the problem of manipulating the support or sample selection probabilities was already recognized by Goodman and Kish (1950). For most of the existing construction techniques for ΠPS sampling designs, the support consists of all $\binom{N}{n}$ possible samples, with no room to manipulate the function $P_d(s)$. The most noteworthy exceptions to this are formed by Hedayat and Lin (1980a, b, c), a foundation for the current work, by Nigam, Kumar and Gupta (1984) and by Gabler (1987). However, it was already shown by Wynn (1977) that for any sampling design there exists a sampling design with the same first- and second-order inclusion probabilities and a support size of at most $N(N-1)/2$. Unfortunately, the proof is not constructive.

The approach used in Gabler (1987) is to construct ΠPS sampling designs with their support imbedded in an a priori selected set. Although this is a nice idea, the method does not always work. It could be that there simply is no ΠPS design with the desired support or that capturing such a design is not possible through Gabler's theorem. Even in cases where the technique works, it is not immediately clear how the initial choice of a sampling design influences the resulting ΠPS design. It is in particular not clear how to guarantee that some of the inequalities for the inclusion probabilities are satisfied.

Some of the techniques in Nigam, Kumar and Gupta (1984) resemble those in Hedayat and Lin (1980c). Their methods heavily emphasize the use of balanced incomplete block designs and can be readily adapted to the technique described in this paper. The same authors, Gupta, Nigam and Kumar (1982) and Kumar, Gupta and Nigam (1985), also present methods based on balanced incomplete block designs to manipulate the support of ΠPS designs. These methods can be useful for some situations, but do not have the same versatility as the current method.

The technique introduced by Hedayat and Lin (1980c) is especially well-suited for manipulating the sample selection probabilities and reducing the support size of ΠPS designs. This paper describes the method and discusses various new results and ideas related to it. The method is extremely flexible and gives a lot of freedom to choose the support of the desired ΠPS sampling design. Though it is this generality that makes the method so useful, it is also the cause of some complications. If there exist designs with some specified properties, they can always be obtained through this method; however, it may occasionally be hard to uncover them.

2. Emptying of boxes. The connection between the problem to be described in this section and that of the previous section may not immediately be clear. However, an explanation is postponed to Section 3.

In this section, we assume there are N boxes, labeled from 1 to N. Each box contains a certain number of objects, denoted by k_i for the ith box. A round of size n is defined as the action of removing n objects from these boxes, with the restriction that at most one object can be removed from any box in one round. We will represent such a round by the n labels of the boxes from which an object was removed, for example $\{i_1, i_2, \dots, i_n\}$. A game is a sequence of rounds, which, if used in succession, would result in emptying all of the boxes. For example, if $N = 4$, $n = 3$, $k_1 = k_2 = 3$, $k_3 = 2$ and $k_4 = 1$, then the rounds $\{1,2,3\}$, $\{1,2,4\}$ and $\{1,2,3\}$ form a game. The class of all possible games for fixed N, n, $k_1, \dots, k_N$ will be denoted by $\mathscr{G}(N, n; k_1, \dots, k_N)$ or simply by $\mathscr{G}$. The following theorem gives necessary and sufficient conditions for $\mathscr{G}$ to be nonempty.

THEOREM 2.1. *For given $N, n, k_1, \dots, k_N$, a corresponding game exists if and only if*:

(i) $n \le N$.
(ii) $\sum_{i=1}^N k_i \equiv 0 \pmod n$.
(iii) $\sum_{i=1}^N k_i \ge n \max_i k_i$.

PROOF. For the necessity, by the definition of a round it follows that (i) holds. Also, since a game consists of $(\sum_{i=1}^N k_i)/n$ rounds, (ii) follows. Finally, if $k_j > (\sum_{i=1}^N k_i)/n$, a game does not consist of enough rounds to empty box j; thus (iii) must hold. To prove the sufficiency, it suffices to construct a game under the assumption that (i), (ii) and (iii) are satisfied. For the first round take $\{i_1, i_2, \dots, i_n\}$ such that

$$\min\{k_i\colon i \in \{i_1, \dots, i_n\}\} \ge \max\{k_i\colon i \notin \{i_1, \dots, i_n\}\}. \tag{2.1}$$

Thus we remove one object from the n boxes with the largest content. In case there are ties, an arbitrary choice is made. The next round is now selected using the same procedure, but with the k_i's replaced by k_i''s where k_i' denotes the number of objects in the ith box after the first round. Subsequent rounds are selected in the same way, each time considering the contents of the boxes after all previous rounds have been played. To see that this procedure indeed gives us a game, it suffices to show that if (i), (ii) and (iii) are satisfied and $\{i_1, i_2, \dots, i_n\}$ satisfying (2.1) is used as a round, then (i), (ii) and (iii) are again satisfied after this round is implemented. This is obvious for (i). With k_i''s as above, the new condition under (ii) becomes

$$\sum_{i=1}^N k_i' \equiv 0 \pmod n.$$

This is clearly satisfied, since $\sum_{i=1}^N k_i' = \sum_{i=1}^N k_i - n \equiv 0 \pmod n$. For (iii), if $\max_i k_i' = \max_i k_i - 1$ this is obvious, since

$$\sum_{i=1}^N k_i' = \sum_{i=1}^N k_i - n \ge n \max_i k_i - n = n \max_i k_i'.$$

However, it could be that $\max_i k_i' = \max_i k_i$. This can only happen if there are, before playing this round, at least $n+1$ boxes which are tied for the largest contents and thus a strict inequality holds in (iii). Using (ii) we obtain that

$$n \max_i k_i' = n \max_i k_i \le n\left(\left(\sum_{i=1}^{N} k_i\right)\Big/ n - 1\right)$$
$$= \sum_{i=1}^{N} k_i - n = \sum_{i=1}^{N} k_i'.$$

This concludes the proof. □

For a game $g \in \mathscr{G}$ we define S_g as the set of all subsets of $\{1,2,\ldots,N\}$ corresponding to rounds in g. We also define C_g as the function on S_g that gives the frequency with which an element of S_g is used as a round in g. We will identify g with the pair (S_g, C_g). Two games can only correspond to the same pair if the rounds in one can be obtained by a permutation of those in the other. Thus with $N=4$, $n=3$, $k_1=k_2=3$, $k_3=2$ and $k_4=1$, the game $\{1,2,3\}$, $\{1,2,4\}$ and $\{1,2,3\}$ is identified with (S_g, C_g), where $S_g = \{\{1,2,3\},\{1,2,4\}\}$, $C_g(\{1,2,3\}) = 2$, $C_g(\{1,2,4\}) = 1$.

3. The construction of ΠPS sampling designs by emptying boxes. In this section we will show how the two problems of the previous sections are related to each other. More precisely, we will show how, for given rational size, measures $0 < z_i < 1/n$, $\sum_{i=1}^{N} z_i = 1$ and $n < N$, ΠPS sampling designs of fixed size n can be constructed by playing a game as described in Section 2. The requirement that the size measures have to be rational is not a restriction, since it is satisfied in any practical problem. The condition $z_i < 1/n$ is not restrictive either, since for any ΠPS design we have that $\pi_i = nz_i \le 1$. If $z_i = 1/n$, unit i has to be included in every sample of the support, so that the problem reduces to one with $N-1$ units and fixed sample size $n-1$.

So let N, n and the z_i's be given. Find any positive integer q such that qnz_i is an integer for all i. Now label N boxes and put $k_i = qnz_i$ objects in the ith box.

THEOREM 3.1. *With the above k_i's a game consisting of rounds of size n can be played to empty the N boxes.*

PROOF. We verify the three conditions in Theorem 2.1. Condition (i) is satisfied by assumption. Further,

$$\sum_{i=1}^{N} k_i = qn \sum_{i=1}^{N} z_i = qn \equiv 0 \pmod{n}.$$

Finally

$$\sum_{i=1}^{N} k_i = qn \ge n \max_i qnz_i = n \max_i k_i.$$

Hence the result. □

Thus let $g = (S_g, C_g)$ be a game in $\mathscr{G}(N, n; k_1, \ldots, k_N)$. Now define a sampling design $d = (S_d, P_d)$ as

$$S_d = S_g,$$
$$P_d(s) = C_g(s)/q \quad \text{for } s \in S_d.$$

To see that d is a ΠPS sampling design with the desired first-order inclusion probabilities, we note that the first two conditions in the definition of such a design are trivially satisfied. Further,

$$\sum_{s \in S_d} P_d(s) = \sum_{s \in S_g} C_g(s)/q = \sum_{i=1}^{N} k_i/nq = \sum_{i=1}^{N} z_i = 1,$$

while

$$\sum_{\substack{s \in S_d \\ s \ni i}} P_d(s) = \sum_{\substack{s \in S_d \\ s \ni i}} C_g(s)/q = k_i/q = nz_i.$$

Thus conditions 3 and 4 are also satisfied and the link between the first two sections has been established. Obviously the given construction can generally lead to many different ΠPS designs. Not only can there be many games in $\mathscr{G}(N, n; k_1, \ldots, k_N)$ with different supports and/or counting functions, but we also have an infinite number of integers q that can be selected. Generally, the larger the selected value of q is, the more elements are contained in $\mathscr{G}(N, n; k_1, \ldots, k_N)$.

The generality of the technique is most accentuated by the following result.

THEOREM 3.2. *For any ΠPS sampling design with rational selection probabilities there exists a game g that induces the design through the preceding identification.*

PROOF. Let $d = (S_d, P_d)$ be a ΠPS sampling design of fixed size n based on N units with size measures $z_1, \ldots, z_N$ and with $P_d(s)$ rational for $s \in S_d$. Take a positive integer q such that $qP_d(s)$ is integral for all $s \in S_d$. Let $k_i = qnz_i$, which is now also integral, and define $g \in \mathscr{G}(N, n; k_1, \ldots, k_N)$ to be any game with

$$S_g = S_d$$

and

$$C_g(s) = qP_d(s) \quad \text{for } s \in S_g.$$

It is easy to verify that $g = (S_g, C_g)$ is indeed an element of $\mathscr{G}(N, n; k_1, \ldots, k_N)$ and that the corresponding sampling design is the design d as above. □

Obviously the above result is not useful in selecting an appropriate value of q for given size measures z_i nor is it useful to select a particular game in $\mathscr{G}(N, n; k_1, \ldots, k_N)$. The only value of the theorem is that it states that we can restrict our attention to designs derived from the games as in Section 2, without the fear of missing any sampling designs with desirable properties.

4. The maximum sample selection probability and the minimum support size. Given values for N, n and the size measures, two questions of interest arise. First of all, for a specified sample s, what is the largest selection probability it can have under a ΠPS design with the desired first-order inclusion probabilities? Second, what is the minimum support size for ΠPS sampling designs with these first-order inclusion probabilities? We will study these questions by considering the classes of games, depending on the choice of q, corresponding to these values for $N, n, z_1, \dots, z_N$.

For the moment we will assume that a particular value of q has been selected, so that nqz_i is integral for all i. The maximum frequency with which a round s can be played in a game $g \in \mathscr{G}(N, n; k_1, \dots, k_N)$ is given in the following result.

THEOREM 4.1. *Let $\mathscr{G}(N, n; k_1, \dots, k_N)$ be a nonempty collection of games. If $s \in S_g$ for a game $g \in \mathscr{G}(N, n; k_1, \dots, k_N)$, then*

$$C_g(s) \le \min\left(\min_{i \in s} k_i, \sum_{i=1}^{N} k_i/n - \max_{i \notin s} k_i\right).$$

Moreover, there exists a game for which this upper bound is achieved.

PROOF. Let $m_s = \min(\min_{i \in s} k_i, \sum_{i=1}^{N} k_i/n - \max_{i \notin s} k_i)$. It suffices to show that if s is played in the first m_s rounds, the boxes can still be emptied, while if s is played in the first $m_s + 1$ rounds this is not possible. After s is played in the first m_s rounds, we have, denoting by k_i' the contents of the ith box,

$$\sum_{i=1}^{N} k_i' = \sum_{i=1}^{N} k_i - m_s n \equiv 0 \pmod{n}$$

and

$$\begin{aligned} n \max_i k_i' &= n \max\left(\max_{i \in s} k_i - m_s, \max_{i \notin s} k_i\right) \\ &\le n \max\left(\max_{i \in s} k_i - m_s, \sum_{i=1}^{N} k_i/n - m_s\right) \\ &= n\left(\sum_{i=1}^{N} k_i/n - m_s\right) = \sum_{i=1}^{N} k_i - nm_s = \sum_{i=1}^{N} k_i'. \end{aligned}$$

Hence, by Theorem 2.1, the boxes can still be emptied. On the other hand, if s is played in $m_s + 1$ rounds, then $m_s = \sum_{i=1}^{N} k_i/n - \max_{i \notin s} k_i$, since $\min_{i \in s} k_i \ge m_s + 1$. Therefore

$$\begin{aligned} n \max_i k_i' \ge n \max_{i \notin s} k_i' = n \max_{i \notin s} k_i &= \sum_{i=1}^{N} k_i - nm_s \\ &> \sum_{i=1}^{N} k_i'. \end{aligned}$$

Thus condition (iii) in Theorem 2.1 is now violated. This proves the result. $\square$

With the minimum support size of a class $\mathscr{G}(N, n; k_1, \dots, k_N)$ we will mean

$$\min\{|S_g|: g \in \mathscr{G}(N, n; k_1, \dots, k_N)\}.$$

It should be observed that the result of Theorem 4.1 can be used to show the following.

THEOREM 4.2. *An upper bound for the minimum support size of a nonempty* $\mathscr{G}(N, n; k_1, \dots, k_N)$ *is given by* N.

PROOF. We will construct a game $g \in \mathscr{G}(N, n; k_1, \dots, k_N)$ with $|S_g| \le N$. Take any round s for which m_s, as defined in the proof of Theorem 4.1, is positive. Since $\mathscr{G}(N, n; k_1, \dots, k_N)$ is nonempty, it is obvious that such an s exists. In fact, one possible choice is $s = \{i_1, \dots, i_n\}$ for which $\max_{i \notin s} k_i \le \min_{i \in s} k_i$. Play m_s rounds using s. Repeat this procedure for the new situation, using the new contents of the boxes. Continue like this. Each time there are two situations that can occur. It could be that $m_s = \min_{i \in s} k_i$, for the k_i's after the previous rounds or $m_s = \sum_{i=1}^N k_i/n - \max_{i \notin s} k_i$. If the first happens we say that s is of type 1; otherwise, we say it is of type 2. If s is of type 1, by playing it m_s times we empty at least one box. If s is of type 2, by playing it m_s times we create a box that has to be included in any round from here on. Since the last round of the game will empty n boxes simultaneously, there are under our procedure at most $N - n + 1$ rounds of type 1. Clearly there are at most n rounds of type 2. Some rounds may actually be of both types and if we denote the distinct rounds by $s_1, s_2, \dots, s_t$, used in this order, then it is not hard to see that s_{t-1} is of both types. After s_{t-1} has been played $m_{s_{t-1}}$ times, there are only n nonempty boxes left. If s_{t-1} is not of type 1, it must be the set of labels corresponding to the n nonempty boxes; but then s_{t-1} can be played in additional rounds, a contradiction with Theorem 4.1. Hence s_{t-1} must be of type 1. Since not all of the n boxes were used in s_{t-1}, but all of them have to be used from now on, we see that it is also of type 2. Therefore, our procedure gives a sampling design with a support size of at most

$$(N - n + 1) + n - 1 = N.$$

Hence the result. □

The counterparts of Theorems 4.1 and 4.2 in the language of sampling designs are formulated in the following two corollaries. The first of these follows from Theorems 4.1 and 3.2; the second is a consequence of Theorems 4.2 and 3.2. Notice that the previously selected value of q is irrelevant for these results.

COROLLARY 4.1. *The maximum selection probability for a sample* s *in a* ΠPS *sampling design of fixed size* n *with size measures* $z_1, \dots, z_N$ *is given by* $\min(\min_{i \in s} nz_i, 1 - \max_{i \notin s} nz_i)$.

COROLLARY 4.2. *The minimum support size over all* ΠPS *sampling designs for given* N, n *and* $z_1, \dots, z_N$ *does not exceed* N.

TABLE 1

Boxes and their contents						
1	2	3	4	5	s	m_s
21	18	15	9	6	$\{3,4,5\}$	2
21	18	13	7	4	$\{1,2,5\}$	4
17	14	13	7	0	$\{1,3,4\}$	3
14	14	10	4	0	$\{1,2,3\}$	10
4	4	0	4	0	$\{1,2,4\}$	4
0	0	0	0	0		

Designs d with $P_d(s)$ as in Corollary 4.1 and designs with a support size not exceeding N can be obtained by the methods as described in the proofs of Theorems 4.1 and 4.2.

The upper bound in Theorem 4.2 is a sharp bound, at least for a bound that does not take the values of $n, k_1, \dots, k_N$ into account. It is easy to give examples for which the minimum support size equals N.

EXAMPLE 4.1. Let $N = 5$, $n = 3$, $k_1 = 21$, $k_2 = 18$, $k_3 = 15$, $k_4 = 9$ and $k_5 = 6$. It can be shown that the minimum support size for $\mathscr{G}(5, 3; 21, 18, 15, 9, 6)$ equals 5. In Table 1 we present a scheme of emptying the boxes that leads to a game with support size 5.

Obviously, many other choices of s are possible at most of the above stages. Observe that indeed three rounds were of type 1 and three rounds of type 2, only the next to last being of both types:

$$\text{Type 1:}\quad \{1,2,5\},\ \{1,2,3\},\ \{1,2,4\},$$
$$\text{Type 2:}\quad \{3,4,5\},\ \{1,3,4\},\ \{1,2,3\}.$$

Also observe that any pair (i, j) appears in at least one of the rounds simultaneously. For the corresponding sampling design this means that all second-order inclusion probabilities are positive. This however may not be true for any game obtained by such a scheme. In Table 2 we give an alternative scheme in which neither of the pairs (3, 5) and (4, 5) appear simultaneously.

TABLE 2

Boxes and their contents						
1	2	3	4	5	s	m_s
21	18	15	9	6	$\{1,2,5\}$	6
15	12	15	9	0	$\{1,2,3\}$	8
7	4	7	9	0	$\{1,3,4\}$	5
2	4	2	4	0	$\{1,2,4\}$	2
0	2	2	2	0	$\{2,3,4\}$	2
0	0	0	0	0		

Obviously there is a need to find methods for emptying the boxes that lead always or almost always to sampling designs with desirable properties. If a game is constructed by a scheme as above, it definitely leads to a game with a small (though not necessarily the minimum) support size, but may give a game without the desired properties on the π_{ij}'s. In fact, these latter properties induce certain constraints on the support size of games that possess them, as will also be seen in later sections.

5. The multiplier technique and the construction of ΠPS sampling designs. In this section we present an idea, called the multiplier technique, that, as will become clear in later sections, is a useful tool for emptying boxes when certain properties for the derived ΠPS sampling design are desired. As an illustrative example for its use, suppose we are interested in a case with $N = 5$, $n = 2$, $z_1 = 0.3$, $z_2 = z_3 = 0.2$, $z_4 = z_5 = 0.15$. We have to select a value for q and it is clear that this should be a multiple of 10. The smaller the chosen value for q, the fewer rounds we will have to play to empty the boxes. Thus we may want to choose $q = 10$. We will have to empty the boxes with contents $k_1 = 6$, $k_2 = k_3 = 4$ and $k_4 = k_5 = 3$. It may, however, be that a certain additional property, such as $\pi_{ij} > 0$ or $\pi_{ij} \le \pi_i \pi_j$ or both, is desired for the ΠPS design. It is not always easy to recognize whether this is possible with a particular choice of q. Any ΠPS design that can be realized with $q = 10$ can also be realized with $q = 20$, but not vice versa. So perhaps to meet our additional demand(s) we should have selected $q = 20$, or even more. The multiplier technique is a tool that allows us to postpone this decision and, in many cases, still achieve the desired properties for the sampling design.

To be more explicit, let N, n and the z_i's be given. Further let $k_i = nqz_i$ be the contents of the ith box, where q is a selected positive integer that makes all the k_i's integral. After playing a certain number of rounds, say q', $q' < q$, the remaining content of box i is denoted by k_i'. We realize at this stage that some desired property for the sampling design is not or no longer feasible with the rounds played so far and the remaining contents. Often it would be feasible if we had more objects in the boxes remaining. To achieve this, we multiply the residuals k_i' by a constant, say h, a positive integer independent of i. We now continue to empty the boxes in the $h(q - q')$ rounds that are needed for this. Observe that the boxes with content hk_i' can be emptied in rounds of size n if those with content k_i' can be emptied in rounds of size n. The latter is under our control. To define the corresponding sampling design, we only have to realize that the described procedure is equivalent to the following. Start with $nhqz_i$ objects in the ith box. Use each of the previously played q' rounds now in h rounds. That will give the residuals hk_i' and we empty the boxes in the same way as in the $h(q - q')$ rounds above. Hence, S_d consists of all those rounds used in some round of the original procedure. If $s \in S_d$ is used in f_1 of the first q' rounds and in f_2 of the last $h(q - q')$ rounds, then we define $P_d(s) = (hf_1 + f_2)/hq$. That we obtain in this way a sampling design with $\pi_i = nz_i$ can easily be verified.

TABLE 3

Round	Frequency	Boxes and their contents 1	2	3	4	5
		6	4	4	3	3
{1, 2}	3	3	1	4	3	3
{1, 3}	2	1	1	2	3	3
{4, 5}	2	1	1	2	1	1

TABLE 4

Round	Frequency	Boxes and their contents 1	2	3	4	5
		4	4	8	4	4
{1, 4}	1	3	4	8	3	4
{1, 5}	1	2	4	8	3	3
{2, 4}	1	2	3	8	2	3
{2, 5}	1	2	2	8	2	2
{1, 3}	2	0	2	6	2	2
{2, 3}	2	0	0	4	2	2
{3, 4}	2	0	0	2	0	2
{3, 5}	2	0	0	0	0	0

EXAMPLE 5.1. Let $N = 5$, $n = 2$, $z_1 = 0.3$, $z_2 = z_3 = 0.2$ and $z_4 = z_5 = 0.15$. With $q = 10$ we start the game by playing the rounds as shown in Table 3.

Deliberately we have not emptied any of the boxes yet. Our strategy so far may not have been very smart, but we will assume that one of our objectives was to obtain $\pi_{ij} > 0$, $i \neq j$. Indeed, it is easy to see in this simple example that with the initial contents of the boxes this could not be achieved. However, it can be achieved with the use of a multiplier. The smallest multiplier that does the job for the remaining contents is $h = 4$. Using this we could reach our objective by playing the rounds exhibited in Table 4. This yields the ΠPS sampling design as shown in Table 5.

TABLE 5

Sample	Probability	Sample	Probability
{1, 2}	12/40	{2, 4}	1/40
{1, 3}	10/40	{2, 5}	1/40
{1, 4}	1/40	{3, 4}	2/40
{1, 5}	1/40	{3, 5}	2/40
{2, 3}	2/40	{4, 5}	8/40

Clearly, in this example we could have captured more pairs $\{i, j\}$ in the first seven rounds than the three pairs we did. However, in more complicated problems this may be less clear. It is in such cases that this method may be useful. In this example, by a result in Section 6, we could have recognized the smallest value for q that allows us to pursue the objective successfully. This, too, is not generally the case.

6. Positive second-order inclusion probabilities. One of the desired properties for a ΠPS sampling design is $\pi_{ij} > 0$, $i \neq j$. In this section we will discuss methods for emptying the boxes in such a way that this property will hold for the resulting ΠPS design. At the same time we would like to be able to manipulate π_{ij}'s and sample selection probabilities.

All that is needed for $\pi_{ij} > 0$ is to empty the boxes in such a way that for any pair of boxes there is at least one round in which an object is removed from both of them. Once this has been achieved, the boxes can be emptied in any fashion. One way to achieve this objective is through the use of balanced incomplete block (BIB) designs. In the following result let N, n and z_i be as in the previous sections.

THEOREM 6.1. *Take any BIB(N, b, n). Select q such that*

$$qz_i \geq b/N \quad \text{for all } i \tag{6.1}$$

and

$$q(1 - nz_i) \geq b(1 - n/N) \quad \text{for all } i. \tag{6.2}$$

With $k_i = qnz_i$ objects in the ith box, we can start the game by playing the first b rounds according to the b blocks in the BIB(N, b, n). Irrespective of how the boxes are emptied from thereon, we obtain $\pi_{ij} > 0$, for all $i \neq j$, for the corresponding ΠPS sampling design.

PROOF. The last statement of the theorem is obvious, since each pair of units appears simultaneously in at least one of the first b rounds. The proof is completed by showing that:

(i) The boxes contain enough units to play the first b rounds.
(ii) After playing the first b rounds, the boxes can still be emptied.

The first of these follows from (6.1). For the second, we have to verify (i), (ii) and (iii) in Theorem 2.1 with the k_i's denoting the contents after the first b rounds. The only nontrivial part is (iii), which follows by using (6.2). □

Notice that there is always a value of b for which the required BIB design exists, possibly $b = \binom{N}{n}$. Also notice that a value of q satisfying (6.1) and (6.2) does always exist. Finally, in many practical problems there may be ways to cover all pairs in fewer rounds than the number required for a BIB design. Such methods may then be preferred. The result, however, shows once more how concepts from design of experiments can be utilized in sample survey methods.

For a discussion of the parallels between these two areas of statistics, we refer to Fienberg and Tanur (1987).

A special case is that of sample size $n = 2$. It is one of the more important cases, since it is a frequently occurring situation in cluster sampling. The clusters are conceptually thought of as the units and they are grouped into a large number of small strata. The y_i's correspond to cluster totals for some characteristic, while the z_i's are chosen proportional to the known or estimated cluster sizes. For each stratum, a ΠPS sampling design is used to select a small number of clusters, for example, 2. Various other papers in the literature deal explicitly with the case $n = 2$, such as Brewer and Undy (1962), Brewer (1963), Rao (1965), Durbin (1967), Hanurav (1967), Rao and Bayless (1969), Jessen (1969) and Sunter (1986). A difference between $n = 2$ and $n \geq 3$ is that for $n = 2$ all possible samples must be in the support to achieve $\pi_{ij} > 0$, for all $i \neq j$.

For both $n = 2$ and $n \geq 3$ we could, once $\pi_{ij} > 0$ has been achieved, try to manipulate π_{ij}'s or the sample selection probabilities $P_d(s)$. A useful tool is that of combining boxes. An illustration is given in the following example.

EXAMPLE 6.1. Let $N = 8$, $n = 3$, $z_1 = z_2 = 4/22$, $z_3 = z_4 = 3/22$ and $z_5 = z_6 = z_7 = z_8 = 2/22$. Suppose that practical considerations lead to the desire of a low second-order inclusion probability for units 1 and 2, while the sample $\{1, 2, 3\}$ should, if possible, be avoided. However, we do require that $\pi_{ij} > 0$, $i \neq j$. How can this be done? Due to the inequality $\pi_{ij} \geq \pi_i + \pi_j - 1$, we obtain that we must allow $\pi_{12} \geq 2/22$. With $k_i = nqz_i$ we see that all k_i's are integral if and only if $q \equiv 0 \pmod{22}$. We take the smallest possible value, $q = 22$. The contents of the boxes may then be represented by the vector $(12, 12, 9, 9, 6, 6, 6, 6)$. Units 1 and 2 have to be selected simultaneously in at least two rounds. Avoiding the round $\{1, 2, 3\}$, we fulfill this requirement by playing $\{1, 2, 4\}$ and $\{1, 2, 5\}$, say. At this stage we combine the boxes 1 and 2 to form one box, labeled $\{1, 2\}$, with $10 + 10 = 20$ units in it. The new situation may be described by the vector $(20, 9, 8, 5, 6, 6, 6)$. Notice that the box $\{1, 2\}$ has to be used in every round from now on. To achieve $\pi_{ij} > 0$, $i \neq j$, we have to play $\{\{1, 2\}, i, j\}$ at least once. That, indeed, can be done in this case. If this would not have been possible, we could have made it possible through the multiplier technique. After playing $\{\{1, 2\}, i, j\}$, $i < j$, $i, j \in \{3, \ldots, 8\}$, the contents of the boxes are described by $(5, 4, 3, 0, 1, 1, 1)$. Now empty these boxes in an arbitrary way or, if other objectives are an issue, in some particular way. One possible result is given in Table 6.

Select a sample with the selection probabilities as listed in that table. If the combined unit $\{1, 2\}$ is included, perform a Bernoulli experiment with success probability $10/(10 + 10) = 1/2$ to decide whether unit 1 or 2 should be used. The selection could also be accomplished in one step, upon replacing $\{\{1, 2\}, 3, 4\}$ by $\{1, 3, 4\}$ and $\{2, 3, 4\}$, and assigning a selection probability of $2/22$ to both of them. Similar replacements should then be made for other samples that include $\{1, 2\}$. The resulting values for π_{ij} are listed in Table 7.

Clearly more than two units may be used to form a combined unit, while more than one combined unit can be formed during a game. If the boxes contain k_i units at the time that we like to combine the units corresponding to a subset H

TABLE 6

Sample	Probability	Sample	Probability
$\{1,2,4\}$	1/22	$\{\{1,2\},4,7\}$	1/22
$\{1,2,5\}$	1/22	$\{\{1,2\},4,8\}$	1/22
$\{\{1,2\},3,4\}$	4/22	$\{\{1,2\},5,6\}$	1/22
$\{\{1,2\},3,5\}$	1/22	$\{\{1,2\},5,7\}$	1/22
$\{\{1,2\},3,6\}$	2/22	$\{\{1,2\},5,8\}$	1/22
$\{\{1,2\},3,7\}$	1/22	$\{\{1,2\},6,7\}$	1/22
$\{\{1,2\},3,8\}$	1/22	$\{\{1,2\},6,8\}$	1/22
$\{\{1,2\},4,5\}$	1/22	$\{\{1,2\},7,8\}$	2/22
$\{\{1,2\},4,6\}$	1/22		

Combined units	Contribution to combination
$\{1,2\}$	(10, 10)

TABLE 7
Values for $44\pi_{ij}$

	2	3	4	5	6	7	8
1	4	9	10	7	6	6	6
2		9	10	7	6	6	6
3			8	2	4	2	2
4				2	2	2	2
5					2	2	2
6						2	2
7							4

of units, we can only form this combination if

$$\sum_{i \in H} k_i \leq \sum_i k_i/n.$$

Units that are combined at some time will no longer be used simultaneously from thereon. Thus the second-order inclusion probabilities of such units are determined by the rounds as played before forming the combined unit. In the special case of $n = 2$, it deserves preference to start with playing each possible pair in the first $\binom{N}{2}$ rounds.

Different objectives can always be formulated, but may not always be compatible. It is not clear how compatibility can be recognized. In case of noncompatibility of requirements, we should formulate the objectives so that some are weakened or even completely deleted. Such cases will demand some inventiveness from the user.

Different situations and objectives will call for different ideas to be used in conjunction with our technique. The technique is flexible enough to allow this, but at the same time this means that there is no standard recipe to empty the boxes.

7. The inequalities $\pi_{ij} \le \pi_i \pi_j$. Simple rules for choosing a value of q and for emptying the boxes in such a way that $\pi_{ij} \le \pi_i \pi_j$, in addition to $0 < \pi_{ij}$, for the corresponding ΠPS sampling design are not easy to give. This is especially true if additional objectives related to the support of the design or the second-order inclusion probabilities have to be taken into account. We will first give two methods to empty the boxes for $n = 2$ such that $0 < \pi_{ij} \le \pi_i \pi_j$ is guaranteed for the corresponding ΠPS design and such that some choice is left to play the last rounds. This can be used, to a limited extent, to influence the second-order inclusion probabilities. We will then conclude this section with a brief discussion on techniques that can help in achieving the desired properties for ΠPS designs that are constructed through the method of emptying boxes.

Thus let $n = 2$, and N and the z_i's be given. Define

$$(7.1)\quad \pi_{ij}^* = z_i z_j\Big((1-2z_i)^{-1} + (1-2z_j)^{-1}\Big)\left(1 + \sum_{k=1}^{N} z_k^2 (1-2z_k)^{-1}\right)^{-1}.$$

This is the second-order inclusion probability for i and j under the ΠPS sampling design as given by Brewer (1963), Rao (1965) and Durbin (1967). For a positive integer q, to be further restricted in the following, define $n_{ij}^* = q\pi_{ij}^*$. If n_{ij}^* is integral for all $i \ne j$, we can play a game with $q\pi_i$ objects in the ith box by using $\{i, j\}$ in n_{ij}^* rounds. This gives the Brewer–Rao–Durbin design. Now suppose all we know is that $q\pi_i$ is integral for all i. Let $n_{ij} = [|n_{ij}^*|]$, where $[|\cdot|]$ denotes the largest integer function. Now play $\{i, j\}$ in n_{ij} rounds. Generally this will not empty all boxes. The following result states that the remaining rounds can be played in any way, without violating $0 < \pi_{ij} \le \pi_i \pi_j$, provided that q is large enough.

THEOREM 7.1. *Let q be a positive integer satisfying*:

(i) *$q\pi_i$ is integral for all i.*
(ii) *$q\pi_{ij} \ge 1$ for all $i \ne j$.*
(iii) *$q \ge (N-1)/(\pi_i\pi_j - \pi_{ij}^*)$ for all $i \ne j$, where π_{ij}^* is as in* (7.1).

Then, by playing each pair $\{i, j\}$ initially $n_{ij} = [|q\pi_{ij}^*|]$ times and, if necessary, emptying the boxes in any manner after this, we obtain a game for which the corresponding ΠPS sampling design has the properties $0 < \pi_{ij} \le \pi_i\pi_j$, $i \ne j$.

PROOF. The proof consists of three steps.

STEP 1. The boxes contain enough objects to play the initial rounds. This follows from

$$\sum_{j \ne i} n_{ij} \le \sum_{j \ne i} q\pi_{ij}^* = q\pi_i = k_i.$$

STEP 2. After the initial rounds, the boxes can still be emptied in rounds of size 2. For this we have to show that, for all i_0,

$$k_{i_0} - \sum_{j \neq i_0} n_{i_0 j} \leq \sum_{i=1}^{N} k_i/2 - \sum\sum_{i \neq j} n_{ij}/2.$$

This follows from

$$\begin{aligned}\sum\sum_{i \neq j} n_{ij}/2 - \sum_{j \neq i_0} n_{i_0 j} &= \sum_{i \neq i_0} \sum_{j \neq i, i_0} n_{ij}/2 \\ &\leq \sum_{i \neq i_0} \sum_{j \neq i, i_0} q\pi_{ij}^*/2 = \sum\sum_{i \neq j} q\pi_{ij}^*/2 - \sum_{j \neq i_0} q\pi_{i_0 j}^* \\ &= \sum_i q\pi_i/2 - q\pi_{i_0} = \sum_i k_i/2 - k_{i_0}.\end{aligned}$$

STEP 3. Irrespective of how the latter rounds are played, the corresponding ΠPS sampling design will always satisfy $0 < \pi_{ij} \leq \pi_i \pi_j$, $i \neq j$. From (ii) it follows that n_{ij}, and hence π_{ij}, is positive. Since $k_i - \sum_{j \neq i} n_{ij} \leq k_i - \sum_{j \neq i}(q\pi_{ij}^* - 1) = N - 1$, it follows that there are at most $N - 1$ objects left in the ith box after the initial rounds. Consequently, a pair $\{i, j\}$ is used in at most $n_{ij} + N - 1$ rounds. Therefore,

$$\begin{aligned}\pi_{ij} &\leq (n_{ij} + N - 1)/q \leq (q\pi_{ij}^* + N - 1)/q \\ &= \pi_{ij}^* + (N - 1)/q \leq \pi_i \pi_j,\end{aligned}$$

by condition (iii). That concludes the proof. □

Notice that neither condition (ii) nor (iii) is used in the first two steps of the proof. Condition (ii) guarantees that $\pi_{ij} > 0$, $i \neq j$, while (iii) guarantees that $\pi_{ij} \leq \pi_i \pi_j$, $i \neq j$. Especially (iii) does not always have to be satisfied for the conclusions to hold; it is used to show that even under the most adverse circumstances the inequalities $\pi_{ij} \leq \pi_i \pi_j$, $i \neq j$, hold.

The following result provides an alternative to Theorem 7.1 and generally leaves more rounds to be played freely. There is, however, a restriction on its applicability.

THEOREM 7.2. *Let N and the z_i's be given such that*

$$\max_i z_i^2 \leq \sum_{i=1}^{N} z_i^2/2. \tag{7.2}$$

Let q be a positive integer such that $2qz_i$ is integral for all i. Let $k_i = 4q^2 z_i$, the content of box i. Play each pair $\{i, j\}$ in $4q^2 z_i z_j$ rounds and complete the game in any way. The corresponding ΠPS sampling design satisfies $0 < \pi_{ij} \leq \pi_i \pi_j$, $i \neq j$.

PROOF. The same three steps as in the proof of Theorem 7.1 have to be satisfied. The first follows from

$$\sum_{j \neq i} 4q^2 z_i z_j = 4q^2 z_i(1 - z_i) \leq k_i.$$

For the second step we need that

$$4q^2 z_i^2 \leq 2q^2 \sum_{i=1}^{N} z_i^2,$$

which follows from (7.2). For the last step, it is clear that $\pi_{ij} > 0$. Without loss of generality, assume that $z_i \leq z_j$. Then

$$\pi_{ij} \leq \frac{4q^2 z_i z_j + 4q^2 z_i^2}{2q^2} = 2z_i z_j + 2z_i^2 = 2z_i(z_j + z_i) \leq (2z_i)(2z_j) = \pi_i \pi_j.$$

Hence the result. □

EXAMPLE 7.1. Let $N = 6$, $z_1 = 0.3$, $z_2 = 0.2$, $z_3 = z_4 = 0.15$ and $z_5 = z_6 = 0.1$. We use Theorem 7.1 with $q = 200$ and Theorem 7.2 with $q = 10$. Although $q = 200$ does not satisfy condition (iii), the conclusion of Theorem 7.1 still holds. In both cases the contents of the boxes are given by $k_1 = 120$, $k_2 = 80$, $k_3 = k_4 = 60$ and $k_5 = k_6 = 40$. We will have to play 200 rounds to empty the boxes. The method of Theorem 7.1 prescribes 192 of these, while that of Theorem

TABLE 8

Sample	Frequencies specified by Theorem 7.1	Frequencies specified by Theorem 7.2
{1,2}	36	24
{1,3}	25	18
{1,4}	25	18
{1,5}	16	12
{1,6}	16	12
{2,3}	13	12
{2,4}	13	12
{2,5}	8	8
{2,6}	8	8
{3,4}	9	9
{3,5}	5	6
{3,6}	5	6
{4,5}	5	6
{4,6}	5	6
{5,6}	3	4

	Residuals after specified rounds					
Box no.	1	2	3	4	5	6
Theorem 7.1	2	2	3	3	3	3
Theorem 7.2	36	16	9	9	4	4

7.2 prescribes only 161 rounds. These rounds are given in Table 8. Although Theorem 7.2 leaves 39 rounds to be played, it should be realized that the objective of emptying all boxes puts some restrictions on this. For example, in 36 of the 39 rounds box 1 has to be used.

These two theorems are useful, since they do not demand any inventiveness of the user. The first rounds are completely determined; the latter can be played in an arbitrary way, as long as the boxes are emptied. This is verified in a trivial way through Theorem 2.1. This does not mean that inventiveness is not rewarding. A smart user can use other methods to empty the boxes and gain more leverage on the sample selection probabilities without violating the inequalities $0 < \pi_{ij} \leq \pi_i \pi_j$, $i \neq j$. Such methods may, however, heavily depend on N and the z_i's, and cannot be formulated under general results as given in this section.

For $n \geq 3$ general results for emptying the boxes and guaranteeing that $\pi_{ij} \leq \pi_i \pi_j$ for the corresponding ΠPS sampling design are hard to obtain, especially if additional restrictions on the support are desired. We conclude this section with the presentation of an approach to this problem that we found useful.

For given N, n and z_i's, select a positive integer q such that $k_i = nqz_i$ is integral for all i. Before playing any rounds, use the multiplier αq, for some positive integer α. Even $\alpha = 1$ will often suffice. Thus, we start with $\alpha q k_i$ objects in box i. We will have to play αq^2 rounds of size n to empty all boxes. Suppose that in λ_{ij} of those rounds an object is removed from both box i and j. The inequalities $0 < \pi_{ij} \leq \pi_i \pi_j$ can be rephrased as $1 \leq \lambda_{ij} \leq \alpha k_i k_j$. The lower bound does not have to cause any problems. An attempt should be made to play successive rounds such that no pair $\{i, j\}$ is used more than $\alpha k_i k_j$ times. This can often be achieved. Even in cases where we did not succeed, violations were mild, both in the sense that they occurred for only very few pairs $\{i, j\}$ and, where they did, that the difference $\lambda_{ij} - \alpha k_i k_j$ was very small compared to αq^2.

8. Discussion. The construction of ΠPS sampling designs through the method of emptying boxes is remarkably simple and surprisingly versatile. Numerous designs with various properties can be obtained without much difficulty. Of course, when faced with increased demands regarding the properties of a design it may be difficult to obtain such a design, if at all possible. But contrary to methods available in the literature, this method does give us an opportunity to search for designs which do meet our demands or come close to it.

An attractive feature of the method is its adaptability to computer algorithms. Very simple to very complicated algorithms may be used to empty the boxes, depending on the particular situation or on the demands for the resulting design. Ideas discussed in the previous sections should be useful in drafting such algorithms. It is clear that neither we nor anybody else will be able to provide a design for every imaginable set of demands. In that sense, the problem will never have a complete solution. Assuming this, there are two routes that can be followed. The first is to find a solution for a particular set of demands. Most results in the literature have been obtained via this route. An example is

Sampford (1967), who gave a solution for $0 < \pi_{ij} \leq \pi_i\pi_j$, $i \neq j$. The second is to present a technique, and create tools that are valuable with that technique, to explore the existence of designs, no matter what the demands are. That has been the route followed in this paper.

The technique, the method of emptying boxes, is simple to implement. The tools from Sections 4–7 are useful to deal effectively with some demands for the resulting designs as discussed in those sections. The effectiveness can be enhanced by use of advanced computational equipment; theoretical enhancement may pose a challenge for a longer time to come.

A disadvantage of the presented method is that it requires a complete listing of the support and corresponding probabilities to implement the design. It is conceivable that future research may eliminate this problem. For example, if some algorithm is used to empty the boxes and the algorithm is such that it uniquely determines the boxes to be selected at each round, all we have to do to implement the design is randomly select a number from 1 to $\sum_{i=1}^{N} k_i/n$. If i is the selected number, use the sample corresponding to the boxes that were used in the ith round. It will depend on the algorithm how hard it is to determine the second-order inclusion probabilities for such a procedure.

Acknowledgments. We owe thanks to the Associate Editor and three referees for their constructive comments which led to various improvements of this paper. Part of this paper was delivered at a Statistical Research Conference dedicated to the memory of Jack Kiefer and Jacob Wolfowitz. The conference was held at Cornell University, July 6–9, 1983, and was a special meeting of the Institute of Mathematical Statistics cosponsored by the American Statistical Association. Financial support for expenses associated with the presentation was provided by the National Science Foundation, the Office of Naval Research, the Army Research Office–Durham and units of Cornell University.

REFERENCES

BREWER, K. R. W. (1963). A model of systematic sampling with unequal probabilities. *Austral. J. Statist.* **5** 5–13.

BREWER, K. R. W. and HANIF, M. (1983). *Sampling with Unequal Probabilities. Lecture Notes in Statist.* **15**. Springer, New York.

BREWER, K. R. W. and UNDY, G. C. (1962). Samples of two units drawn with unequal probabilities without replacement. *Austral. J. Statist.* **4** 89–100.

DURBIN, J. (1967). Design of multi-stage surveys for the estimation of sampling errors. *Appl. Statist.* **16** 152–164.

FIENBERG, S. E. and TANUR, J. M. (1987). Experimental and sampling structures: Parallels diverging and meeting. *Internat. Statist. Rev.* **55** 75–96.

GABLER, S. (1987). The nearest proportional to size sampling design. *Comm. Statist. A—Theory Methods* **16** 1117–1131.

GOODMAN, R. and KISH, L. (1950). Controlled selection—A technique in probability sampling. *J. Amer. Statist. Assoc.* **45** 350–372.

GUPTA, V. K., NIGAM, A. K. and KUMAR, P. (1982). On a family of sampling schemes with inclusion probability proportional to size. *Biometrika* **69** 191–196.

HANURAV, T. V. (1967). Optimum utilization of auxiliary information: ΠPS sampling of two units from a stratum. *J. Roy. Statist. Soc. Ser. B* **29** 374–391.

HEDAYAT, A. and LIN, B.-Y. (1980a). Controlled probability proportional to size sampling designs (abstract). *Bull. Inst. Math. Statist.* **9** 282.

HEDAYAT, A. and LIN, B.-Y. (1980b). A complete class theorem for probability proportional to size sampling designs (abstract). *Bull. Inst. Math. Statist.* **9** 297.

HEDAYAT, A. and LIN, B.-Y. (1980c). Controlled probability proportional to size sampling designs. Technical Report, Univ. Illinois at Chicago.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685.

JESSEN, R. J. (1969). Some methods of probability non-replacement sampling. *J. Amer. Statist. Assoc.* **64** 175–193.

KUMAR, P., GUPTA, V. K. and NIGAM, A. K. (1985). On inclusion probability proportional to size sampling scheme. *J. Statist. Plann. Inference* **12** 127–131.

NIGAM, A. K., KUMAR, P. and GUPTA, V. K. (1984). Some methods for inclusion probability to size sampling. *J. Roy. Statist. Soc. Ser. B* **46** 564–571.

RAO, J. N. K. (1965). On two simple schemes of unequal probability sampling without replacement. *J. Indian Statist. Assoc.* **3** 173–180.

RAO, J. N. K. and BAYLESS, D. L. (1969). An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum. *J. Amer. Statist. Assoc.* **64** 540–549.

SAMPFORD, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika* **54** 499–513.

SEN. A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *J. Indian Soc. Agric. Statist.* **5** 119–127.

SUNTER, A. (1986). Solution to the problem of unequal probability sampling without replacement. *Internat. Statist. Rev.* **54** 33–55.

WYNN, H. P. (1977). Convex sets of finite population plans. *Ann. Statist.* **5** 414–418.

YATES, F. and GRUNDY, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc. Ser. B* **15** 253–261.

A. HEDAYAT
DEPARTMENT OF MATHEMATICS,
STATISTICS AND COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS
BOX 4348
CHICAGO, ILLINOIS 60680

BING-YING LIN
NATIONAL CHENG-CHI UNIVERSITY
TAIWAN, REPUBLIC OF CHINA

J. STUFKEN
DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011