Statistica Sinica **6**(1996), 199-214

ON BALANCED BOOTSTRAP FOR STRATIFIED MULTISTAGE SAMPLES

A. K. Nigam and J. N. K. Rao*

Institute of Applied Statistics and Development Studies, Lucknow
**Carleton University, Ottawa*

Abstract. For constructing inferential procedures from a random sample of size n, Graham et al. (1990) proposed second order balanced bootstrap designs that reduce the variance in the usual, unbalanced bootstrap simulation. Their methods, however, do not cover the following cases: (i) n, a composite odd number; (ii) $n = 4m + 1$, a prime number. Here we first give two methods that provide second-order balanced designs for all cases. We then extend the results to stratified multistage samples, and construct balanced bootstrap designs, for the important special case of equal first-stage sample sizes within strata, yielding second-order balance.

Key words and phrases: Balanced incomplete block designs, bootstrap, Hadamard matrices, quadratic residues, second-order balance, stratified sampling.

1. Introduction

Bootstrap resampling methods are simulation methods for constructing inferential procedures from a random sample $x = (x_1, \ldots, x_n)'$ of size n. To reduce the variance in the usual, unbalanced bootstrap simulation, several efficient bootstrap simulation methods have been proposed: balanced bootstrap by Davison et al. (1986) and Graham et al. (1990); centering method by Efron (1990); explicit use of linear approximation of the statistic by Davison et al. (1986). Hall (1989) studied the properties of these methods in the case of statistics expressible as smooth functions of means. He showed that the three methods are asymptotically equivalent and superior to the usual bootstrap in the sense that the simulation variance is of smaller order. The advantage of the balanced bootstrap, however, is that it is totally automatic for any statistic as in the case of the usual bootstrap.

The method of Davison et al. (1986) yields first-order balance which mainly affects bootstrap estimation of bias, while the method of Graham et al. (1990) provides second-order balance which mainly affects bootstrap estimation of variance. Graham et al. give balanced designs yielding second-order balance, but their methods do not cover the following cases: (a) n, a composite odd number;

(b) $n = 4m+1$, a prime number. In Section 2, we give two methods that provide second-order balanced designs for all cases. The solutions obtained include those of Graham et al. as particular cases.

Balanced resampling methods have long been used in sample surveys in the context of stratified multistage samples. For example, McCarthy (1969) proposed the well-known method of balanced repeated replication (BRR) which yields second-order balance. Rao and Wu (1988) and Rao, Wu and Yue (1992) extended the usual bootstrap to stratified multistage samples with strata first-stage sample sizes n_h $(h = 1, \ldots, L)$, where n_h is small and the number of strata, L, is relatively large. In Section 3, we construct balanced bootstrap designs yielding second-order balance for the important special case of equal n_h by employing the balanced designs of Section 2 in conjunction with Hadamard matrices. An advantage of these designs is that they may require fewer replicates than the BRR and cover cases where BRR designs, based on orthogonal arrays of strength two, are not available; for example $n_h = 6$ in each stratum h. Sitter (1993) proposed an extension of the BRR using orthogonal multi-arrays that allow the number of resampled units per stratum to be greater than one, unlike the BRR. Sitter's method is particularly suited for the case of even n_h.

Results of a simulation study on the performance of the balanced bootstrap for stratified samples are reported in Section 4.

2. Simple Random Sampling

2.1. Second-order balance

Let $T = t(x)$ be an estimator of a parameter of interest, θ, computed from the random sample $x = (x_1, \ldots, x_n)'$, where $t(x)$ is symmetric. Suppose we are interested in estimating $E(T)$ and $\mathrm{Var}(T)$. The bootstrap estimators of $E(T)$ and $\mathrm{Var}(T)$ are then given by $E_*(T) = E(T^*|x)$ and $\mathrm{Var}_*(T) = E\{(T^* - E_*T)^2|x\}$, where $T^* = t(x^*)$ and $x^* = (x_1^*, \ldots, x_n^*)'$ denotes a random sample drawn with replacement from x. The bootstrap estimators $E_*(T)$ and $\mathrm{Var}_*(T)$ are generally not computable, except in special cases. For example in the linear case $T^* = \bar{x}^* = n^{-1}\Sigma f_i^* x_i$, we have

$$E_*(\bar{x}^*) = \bar{x}, \quad \mathrm{Var}_*(\bar{x}^*) = (n-1)s^2/n^2, \tag{2.1}$$

where $\bar{x} = \Sigma x_i/n$, $s^2 = \Sigma(x_i - \bar{x})^2/(n-1)$ and f_i^* is the frequency count of x_i in the bootstrap sample.

The usual bootstrap approximates $E_*(T)$ and $\mathrm{Var}_*(T)$ by drawing a large number, S, of independent samples $x_s^* = (x_{s1}^*, \ldots, x_{sn}^*)'$ from x and then using

$$\bar{T}^* = S^{-1}\sum_{s=1}^{S} T_s^*, \quad V^* = \sum_{s=1}^{S}(T_s^* - \bar{T}^*)^2/(S-1), \tag{2.2}$$

where $T_s^* = T(x_s^*)$. The simulation error is not zero even in the linear case (2.1); for example,

$$\bar{x}^* = n^{-1}\sum_{i=1}^{n} x_i\Big(S^{-1}\sum_{s=1}^{S} f_{si}^*\Big) \neq n^{-1}\sum_{i=1}^{n} x_i$$

unless

$$S^{-1}\sum_{s=1}^{S} f_{si}^* = 1 \quad (i = 1, \ldots, n), \tag{2.3}$$

where f_{si}^* is the frequency count of x_i in the s-th bootstrap sample x_s^*. Note that (2.3) is in agreement with the multinominal expectation $E(f_i^*|x) = 1$.

The method of Davison et al. (1986) ensures first-order balance as given by (2.3). As a result, it eliminates the linear term in a Taylor expansion of $\bar{T}^*$ when T is a smooth function of $\bar{x}$, but does not affect V^*. In particular, V^* is not equal to $\mathrm{Var}_*(T)$ in the linear case where $T = \bar{x}$.

Second-order balanced bootstrap designs satisfy

$$S^{-1}\sum_{s=1}^{S} f_{si}^* f_{sj}^* = \delta(i-j) + 1 - n^{-1} \tag{2.4}$$

in addition to (2.3), where $\delta(k) = 0$ or 1 according as $k \neq 0$ or $k = 0$. Conditions (2.3) and (2.4) ensure that $\bar{T}^* = E_*(T)$ and $V^* = \mathrm{Var}_*(T)$ in the linear case. Note that (2.4) is in agreement with the multinominal expectation $E(f_i^* f_j^*|x)$.

2.2. Second-order balanced bootstrap designs

We now give two general methods for constructing second-order balanced bootstrap (SBB) designs.

Case 1. $n = tm$ $(m \geq t \geq 2)$

Graham et al. (1990) considered the special case of $n = 2m$ or $t = 2$. Note that Case 1 covers composite odd numbers n; for example $m = 5$ and $t = 3$ gives $n = 15$.

Suppose a balanced incomplete block design (BIBD) with parameters $v = n = tm$, b, r, $k = m, \lambda$ exists, where each of b blocks contains k plots, each of v treatments appears in r blocks, and each pair of treatments appears in λ blocks. To construct a SBB design from this BIBD, repeat each treatment within a block t times and add $\tilde{s}$ complete blocks, where $\tilde{s} = S - b$ and $S = (r-\lambda)t^2$. Identifying the treatments with the sample units $\{1, 2, \ldots, n\}$, we get S balanced bootstrap samples satisfying the conditions (2.3) and (2.4). This follows by using the BIBD relations $b = rt$, $\lambda(n-1) = r(m-1)$.

Note that $S = bt(n-m)/(n-1)$ so that S is minimized by choosing a BIBD with the smallest possible $b(\geq n)$. Also, note that $\tilde{s} = b[n(t-2)+1]/(n-1) \geq 1$ since $t \geq 2$.

Many series of BIBD with $v = tm$ and $k = m$ exist and are catalogued in Hall (1967), Raghavarao (1971), Rao (1961), Sprott (1962) and Takeuchi (1962). Some of the well-known series which exist for every p, a prime or a prime power, are as follows:

(i) $v = (p+1)(p^2+1)$, $k = p+1$, $b = (p^2+1)(p^2+p+1)$, $r = p^2+p+1$, $\lambda = 1$;
(ii) $v = p^2$, $k = p, b = p(p+1)$, $r = p+1, \lambda = 1$;
(iii) $v = p^3$, $k = p, b = p^2(p^2+p+1), r = p^2+p+1$, $\lambda = 1$;
(iv) $v = p^3, k = p^2, b = p(p^2+p+1)$, $r = p^2+p+1$, $\lambda = p+1$.

For instance, when $t = m$ is a prime or a prime power, we can use series (ii) with $p = m$, giving $\tilde{s} = m^3 - m^2 - m$ and $S = m^3$. This series is also called Yates' orthogonal series.

Example 1 ($t = 3, m = 3$). A BIBD belonging to series (ii) with parameters $v = n = 9$, $b = 12$, $r = 4$, $k = m = 3, \lambda = 1$ exists from which a set of $S = 27$ balanced bootstrap samples is obtained; see Table 1.

Table 1. Balanced design for $n = 9$, $S = 27$; unstratified case

Sample	Frequency of sample units									
s	1	2	3	4	5	6	7	8	9	
1	3	3	3	0	0	0	0	0	0	
2	0	0	0	3	3	3	0	0	0	
3	0	0	0	0	0	0	3	3	3	
4	3	0	0	3	0	0	3	0	0	
5	0	3	0	0	3	0	0	3	0	
6	0	0	3	0	0	3	0	0	3	
7	3	0	0	0	0	3	0	3	0	
8	0	3	0	3	0	0	0	0	3	
9	0	0	3	0	3	0	3	0	0	
10	3	0	0	0	3	0	0	0	3	
11	0	3	0	0	0	3	3	0	0	
12	0	0	3	3	0	0	0	3	0	
13—27	1	1	1	1	1	1	1	1	1	(repeat 15 times)

Example 2 ($t = 3, m = 7$). A BIBD with parameters $v = n = 21$, $b = 30, r = 10, k = m = 7$, $\lambda = 3$ exists from which a set of $S = 63$ balanced bootstrap samples can be constructed. This BIBD does not belong to any of the above

series.

Case 2. $n = 4m + 1$ or $4m + 3$, a prime or a prime power

Graham et al. (1990) considered the case of $n = 4m + 3$, and we extend their argument to cover $n = 4m + 1$. A design with $S = n$ balanced bootstrap samples is obtained by first constructing an initial block of size n consisting of the quadratic residues (i.e., even or odd powers of the primitive root) of the Galois field of n elements, each repeated twice together with the element n once. The remaining blocks are then obtained by developing the initial block full cycle modulo n. It follows from Saha and Dey (1973, Theorem 2.2) that this is a balanced ternary design satisfying (2.3) and (2.4).

Example 3 ($n = 13$). The primitive root of 13 is 2, and taking its even powers we get the quadratic residues as 1, 3, 4, 9, 10 and 12. The resulting design has the initial block (1, 1, 3, 3, 4, 4, 9, 9, 10, 10, 12, 12, 13) which when developed full cycle modulo 13 gives the desired balanced design; see Table 2.

Table 2. Balanced design for $n = 13$, $S = 13$; unstratified case

Sample	Frequency of sample units												
s	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2	0	2	2	0	0	0	0	2	2	0	2	1
2	1	2	0	2	2	0	0	0	0	2	2	0	2
3	2	1	2	0	2	2	0	0	0	0	2	2	0
4	0	2	1	2	0	2	2	0	0	0	0	2	2
5	2	0	2	1	2	0	2	2	0	0	0	0	2
6	2	2	0	2	1	2	0	2	2	0	0	0	0
7	0	2	2	0	2	1	2	0	2	2	0	0	0
8	0	0	2	2	0	2	1	2	0	2	2	0	0
9	0	0	0	2	2	0	2	1	2	0	2	2	0
10	0	0	0	0	2	2	0	2	1	2	0	2	2
11	2	0	0	0	0	2	2	0	2	1	2	0	2
12	2	2	0	0	0	0	2	2	0	2	1	2	0
13	0	2	2	0	0	0	0	2	2	0	2	1	2

Quadratic residues for prime numbers up to $n = 97$ are reported in Vinogradov (1954). The number of quadratic residues is $2m$ when $n = 4m + 1$ and $2m + 1$ when $n = 4m + 3$.

The case $n = 4m + 1$ leads to $n = 5$ when $m = 1$. In this case we get a new design with only $S = 5$ samples which is obtained by developing the initial block

(20021) modulo 5 (see Table 3) whereas Graham et al. (1990, p.192) obtained two designs through trial and error requiring 10 and 15 samples. This reduction in S is particularly useful in the stratified case studied in Section 3.

Table 3. Balanced design for $n = 5,\ S = 5$; unstratified case

Sample	Freq. of sample units				
s	1	2	3	4	5
1	2	0	0	2	1
2	1	2	0	0	2
3	2	1	2	0	0
4	0	2	1	2	0
5	0	0	2	1	2

3. Stratified Multistage Sampling

3.1. Second-order balance

Large-scale surveys often employ stratified multistage designs with a large number of strata, L, and relatively few first stage units or clusters, n_h, sampled within each stratum independently. It is a common practice to sample the clusters with probabilities proportional to sizes and without replacement, but at the stage of variance estimation the calculations are greatly simplified by treating the sample as if the clusters are sampled with replacement and subsampling done independently each time a cluster is selected. This approximation leads to overestimation of variance of $T = t(x)$, but the relative bias is likely to be small if the first stage sampling fractions within strata are small.

Let w_{hik} be the survey weight attached to the k-th sample element, or ultimate unit, in the i-th sample cluster belonging to h-th stratum. The estimator $T = t(x)$ of a parameter of interest θ is computed using the survey weights w_{hik}. For example, an estimator of population total X is of the form

$$\hat{X} = \sum_{(hik)\in s} w_{hik} x_{hik}, \tag{3.1}$$

where s denotes the sample of elements and x_{hik} is the value of a characteristic of interest associated with the sample element $(hik) \in s$. Under the assumption of with-replacement sampling of clusters, an unbiased estimator of variance of $\hat{X}$

is given by

$$\operatorname{var}(\hat{X}) = \sum_{h=1}^{L} s_{rh}^2/n_h, \tag{3.2}$$

where

$$(n_h - 1)s_{rh}^2 = \sum_{i=1}^{n_h}(r_{hi} - \bar{r}_h)^2 \tag{3.3}$$

$$r_{hi} = \sum_k (n_h w_{hik}) x_{hik},\ \bar{r}_h = \sum_i r_{hi}/n_h. \tag{3.4}$$

Unbiasedness of $\operatorname{var}(\hat{X})$ follows by noting that the r_{hi}'s are independent and identically distributed random variables with the same mean and the same variance in each stratum h.

Turning to the bootstrap for stratified multistage sampling, Rao and Wu (1988) and Rao, Wu and Yue (1992) have shown that a scale adjustment should be made on the weights w_{hik} in order to have valid variance estimation in the case of small n_h. Their method works as follows: (i) Draw a simple random sample of m_h clusters with replacement from the n_h sample clusters, independently for each h. Let f_{hi}^* be the number of times the (hi)-th sample cluster is selected, $\sum_i f_{hi}^* = m_h$. Define the bootstrap weights

$$w_{hik}^* = \left[\left\{1 - (m_h/(n_h-1))^{\frac{1}{2}}\right\} + (m_h/(n_h-1))^{\frac{1}{2}}(n_h/m_h)f_{hi}^*\right]w_{hik} \tag{3.5}$$

and calculate T^*, the bootstrap estimator of θ, using the weights w_{hik}^* in the formula for T; (ii) Independently replicate step (i) a large number, S, of times and calculate the corresponding estimates $T_1^*, \ldots, T_S^*$; (iii) The bootstrap estimator, E_*T, and variance estimator, $\operatorname{Var}_*(T) = E_*(T^* - ET^*)^2$, are approximated by

$$\bar{T}^* = S^{-1}\sum_{s=1}^{S} T_s^*, \quad V^* = \sum_{s=1}^{S}(T_s^* - \bar{T}^*)^2/(S-1). \tag{3.6}$$

One could also use T in place of $\bar{T}^*$ in the formula for V^*. As before, the simulation error is not zero in the linear case where $T = \hat{X}$.

To construct second-order balanced bootstrap designs, we confine ourselves to the important special case $m_h = n_h = n$ for all h. The conditions for second-order balance are then given by

$$S^{-1}\sum_{s=1}^{S} f_{shi}^* = 1, \tag{3.7}$$

$$S^{-1}\sum_{s=1}^{S} f^*_{shi}f^*_{shj} = \delta(i-j)+1-n^{-1},\ i\neq j, \tag{3.8}$$

and

$$S^{-1}\sum_{s=1}^{S} f^*_{shi}f^*_{skj} = 1, \quad h\neq k, \tag{3.9}$$

for all $i=1,\ldots,n$; $h=1,\ldots,L$, where f^*_{shi} is the frequency count of the (hi)-th sample cluster in the s-th balanced bootstrap sample, $s=1,\ldots,S$. Conditions (3.7)–(3.9) ensure that $\bar{T}^* = E_*(T)$ and $V^* = \mathrm{Var}_*(T)$ in the linear case, noting that they are in agreement with the multinominal expectations $E(f^*_{hi}|x)$, $E(f^*_{hi}f^*_{hj}|x)$, and $E(f^*_{hi}f^*_{kj}|x) = E(f^*_{hi}|x)E(f^*_{kj}|x)$, $h\neq k$.

3.2. Second-order balanced bootstrap designs

Case 1. $n=2m$

As noted in the Introduction, second-order balanced bootstrap designs are obtained by employing the balanced designs of Section 2 in conjunction with Hadamard matrices.

For the case $n=2m$ and L strata, we construct a balanced design as follows: (i) Take L columns of a Hadamard matrix of order p, consisting of $+1$'s and -1's, where $p=0$ (modulo 4) if $L\geq 3$ and $p=2$ if $L=2$. (ii) Replace each $+1$ and -1 in the Hadamard matrix by

$$N=\begin{bmatrix} N_1 \\ E\end{bmatrix} \quad \text{and} \quad N_c=\begin{bmatrix} N_{1c} \\ E\end{bmatrix}$$

respectively, where N_1 is a $b\times n$ matrix obtained by changing 1 to 2 in the transpose of the incidence matrix of a BIBD with parameters $v=n=2m$, $b=2r$, r, $k=m$, λ, and E is a $\{4(r-\lambda)-b\}\times n$ matrix of all 1's (see Graham et al. (1990)). The matrix N_{1c} is obtained from N_1 by changing 0 to 2 and 2 to 0 in N_1. For the BIBD under consideration, we have $\lambda = r(m-1)/(2m-1)$, an integer. Therefore, $r=x(v-1)$, a multiple of $v-1$. This gives $b=2x(v-1)$. We minimize b by choosing $x=1$ in which case $r=v-1$, $b=2(v-1)$ and $4(r-\lambda)-b=2$. (iii) Identify the first n columns of the resulting matrix of order $S\times(Ln)$ with the n sample clusters in stratum 1, the next n columns with the n sample clusters in stratum 2, and so on, where $S=4p(r-\lambda)$.

We now show that the above method gives S balanced stratified bootstrap samples by verifying the conditions (3.7)–(3.9) for second-order balance. Clearly, the conditions (3.7) and (3.8) for balance within each stratum are satisfied since both N and N_c satisfy (2.3) and (2.4) with $t = 2$ and S changed to $\tilde{S} = S/p = 4(r-\lambda)$. Turning to across strata balance (3.9), we first note that each of the pairs (N, N), (N, N_c) (N_c, N), (N_c, N_c), corresponding to two different strata h and k, appears the same number $p/4$ of times, if $2 < L < p$ and the column of all $+1$'s or -1's excluded from the Hadamard matrix; if $p = L$ we would also have each of the pairs $(N, N), (N, N_c)$ or (N_c, N), (N_c, N_c) appear the same number, $p/2$, of times. Next, we note that the sums of $f^*_{shi}f^*_{ski}$ for $(N, N), (N_c, N_c)$, $(N, N_c), (N_c, N)$ are respectively $4r + \{4(r-\lambda) - b\}$, $4r + \{4(r-\lambda) - b\}$, $\{4(r-\lambda) - b\}$, $\{4(r-\lambda) - b\}$. Similarly, the sums of $f^*_{shi}f^*_{skj}$, $i \neq j$ for $(N, N), (N_c, N_c)$, $(N, N_c), (N_c, N)$ are respectively $4\lambda + \{4(r-\lambda) - b\}$, $4\lambda + \{4(r-\lambda) - b\}$, $4(r-\lambda) + \{4(r-\lambda) - b\}$, $4(r-\lambda) + \{4(r-\lambda) - b\}$. Therefore, for every strata pairs (h, k), we have

$$\sum_{s=1}^{S} f^*_{shi}f^*_{ski} = \frac{p}{4}[8r + 4\{4(r-\lambda) - b\}] = 4p(r-\lambda) = S,$$

and

$$\sum_{s=1}^{S} f^*_{shi}f^*_{skj} = \frac{p}{4}[8\lambda + 8(r-\lambda) + 4\{4(r-\lambda) - b)\} = 4p(r-\lambda) = S$$

if $2 < L < p$. In the case $p = L$ we also have strata pairs (h, k) with

$$\sum_{s=1}^{S} f^*_{shi}f^*_{ski} = \frac{p}{2}[4r + 2\{4(r-\lambda) - b\}] = 4p(r-\lambda) = S$$

and

$$\sum_{s=1}^{S} f^*_{shi}f^*_{skj} = \frac{p}{2}[4\lambda + 4(r-\lambda) + 2\{4(r-\lambda) - b\}] = 4p(r-\lambda) = S.$$

Thus the across strata balance condition (3.9) is also satisfied.

Example 4. For simplicity, consider the case $L = 2$ and $n = 6$. In this case

$$H = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix},$$

and using N_1 given in Table 6 of Graham et al. (1990) and its complement N_{1c}, we get the desired design with $S = 24$ given in Table 4. For example, the first row (220020|220020) means that the first balanced bootstrap sample consists of sample clusters 1, 2 and 5 in stratum 1 repeated twice and sample clusters 1, 2 and 5 in stratum 2 repeated twice. Similarly, the thirteenth row (220020|002202) means that the thirteenth balanced bootstrap sample consists of sample clusters 1, 2 and 5 in stratum 1 repeated twice and sample clusters 3, 4 and 6 in stratum 2 repeated twice.

Table 4. Balanced design for $L = 2, n = 6, S = 24$; stratified case

Sample	Frequency for stratum 1						Frequency for stratum 2					
s	1	2	3	4	5	6	1	2	3	4	5	6
1	2	2	0	0	2	0	2	2	0	0	2	0
2	2	2	0	0	0	2	2	2	0	0	0	2
3	2	0	2	2	0	0	2	0	2	2	0	0
4	2	0	2	0	2	0	2	0	2	0	2	0
5	2	0	0	2	0	2	2	0	0	2	0	2
6	0	2	2	2	0	0	0	2	2	2	0	0
7	0	2	0	2	2	0	0	2	0	2	2	0
8	0	2	2	0	0	2	0	2	2	0	0	2
9	0	0	2	0	2	2	0	0	2	0	2	2
10	0	0	0	2	2	2	0	0	0	2	2	2
11	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1
13	2	2	0	0	2	0	0	0	2	2	0	2
14	2	2	0	0	0	2	0	0	2	2	2	0
15	2	0	2	2	0	0	0	2	0	0	2	2
16	2	0	2	0	2	0	0	2	0	2	0	2
17	2	0	0	2	0	2	0	2	2	0	2	0
18	0	2	2	2	0	0	2	0	0	0	2	2
19	0	2	0	2	2	0	2	0	2	0	0	2
20	0	2	2	0	0	2	2	0	0	2	2	0
21	0	0	2	0	2	2	2	2	0	2	0	0
22	0	0	0	2	2	2	2	2	2	0	0	0
23	1	1	1	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1	1	1

In Section 2 we considered the general case $n = tm$ when $L = 1$. This cannot be readily extended to the stratified case using the above method since the sample size for N_{1c} will be $(mt - m)t$ which is not equal to $n = mt$ unless $t = 2$.

Case 2. $n = 4m + 1$ *or* $4m + 3$, *a prime or a prime power*

As in case 1, we take L columns of a Hadamard matrix of order p. We then replace $+1$ by N and -1 by N_c, where N is the $n \times n$ incidence matrix of the balanced design for the unstratified case based on even powers of the primitive root, and N_c is obtained from N by interchanging 0 and 2. Clearly, N_c is also balanced since it can be obtained by developing modulo n the initial block with odd powers of the primitive root occurring twice and the element n occurring once. The conditions for second order balance within each stratum are therefore satisfied. The proof of the cross strata balance is established in the Appendix by showing that $\Sigma_s f^*_{shi} f^*_{skj} = pn = S$ for all pairs of strata (h, k).

4. Simulation Study

We now present the results of a limited simulation study on the finite-sample performance of the proposed balance bootstrap for stratified random samples. For this purpose, we employed a synthetic population of $N = 14,000$ pairs (x_i, z_i) generated from a bivariate gamma distribution in which the variable z has a gamma distribution with density function $f(z) = 0.04z \exp(-z/5)$ and the conditional density of x given z is also a gamma given by

$$f(x|z) = \{b^c \Gamma(c)\}^{-1} x^{c-1} \exp(-x/b),$$

where

$$b = 1.15z^{3/2}(8+5z)^{-1},\ c = 0.04z^{-3/2}(8+5z)^2.$$

Hansen et al. (1983) used this model to study the effect of model misspecification on estimators of a population mean. After generating the synthetic population, we divided it into 32 strata defined by intervals of z, such that the aggregate values of z were approximately the same from each stratum. Then 10,000 samples, each of size $n = 160$, were drawn by stratified random sampling with equal allocation $n_h = 5$, in order to simulate the mean square error (MSE) of an estimator $T(x)$ or $T(x, z)$. The parameters considered here are the finite population ratio, X/Z, and the regression and correlation coefficients of x on z. From each sample, the corresponding estimates were computed using the survey weights $w_{hi} = N_h/n_h$, where N_h is the number of population units in stratum h (see Kovar et al. (1988) for details on computation). The true MSE of an estimator T was simulated as $\text{MSE} = \sum_\ell (T_\ell - \theta)^2/10,000$, where T_ℓ is the estimate from ℓ-th sample, $\ell = 1, \ldots, 10,000$.

To study the properties of various variance estimators and confidence intervals, we selected an independent set of 2,000 samples as above. From each sample, the following variance estimates were computed: (i) the jackknife (JACK) with $n = 160$ replicates, (ii) the BRR replicates obtained from an orthogonal array with 250 runs, (iii) the balanced bootstrap (BBOOT) replicates obtained from the design in Table 3 in conjunction with a Hadamard matrix of order 32 giving

$S = 32 \times 5 = 160$; (iv) the Rao-Wu bootstrap (BOOT) replicates using $S = 160$ for comparability (see Kovar et al. (1988) for details on (i), (ii) and (iv)). Using these variance estimates and the simulated MSE, we simulated the relative bias (RB) and coefficient of variation (CV) of the variance estimators. Table 5 reports these values for the ratio, regression and correlation coefficients (RB and CV are defined in the footnotes to Table 1). It is clear from Table 5 that the four methods behave similarly with respect to RB and CV, except that BBOOT is more stable than BOOT as expected (% CV of 16.3 compared to 19.7 for the ratio and 30.9 compared to 32.4 for the regression coefficient).

Table 5. % Relative bias (RB) and % CV of variance estimators and % error rates (L,U) and standardized lengths (SL) of confidence intervals (nominal level of 5% in each tail) for the ratio, regression and correlation coefficients under stratified random sampling with 32 strata and 5 units from each stratum

Method	RB[a]	CV[b]	L	U	L+U	SL[c]
			Ratio			
JACK	−4.0	16.3	4.1	6.3	10.4	0.98
BRR	−4.0	16.3	4.1	6.3	10.4	0.98
BBOOT	−4.0	16.3	4.1	5.7	9.8	1.00
BOOT	−4.4	19.7	4.2	6.3	10.5	0.99
			Regression			
JACK	0.3	31.0	4.7	7.6	12.3	0.99
BRR	−0.1	30.6	4.7	7.5	12.2	0.99
BBOOT	0.5	30.9	4.6	6.5	11.1	1.06
BOOT	−0.7	32.4	5.2	7.2	12.4	1.03
			Correlation			
JACK	−3.0	37.0	9.1	3.4	12.5	0.97
BRR	−8.9	32.2	9.8	3.7	13.5	0.94
BBOOT	−3.8	35.3	4.9	5.7	10.6	1.08
BOOT	−6.1	35.6	5.7	5.6	11.3	1.04

[a] $RB = \frac{1}{2000}(\sum_\ell V_\ell/\text{MSE}) - 1 \times 100$, where V_ℓ = variance estimate for ℓ-th sample.

[b] $CV^2 = \frac{1}{2000}\{\sum_\ell(V_\ell - \text{MSE})^2\}/\text{MSE}^2$.

[c] SL = (average length of interval over ℓ)/$(2z_{0.05}\text{MSE}^{\frac{1}{2}})$, where $z_{0.05}$ is the upper 5% point of $N(0,1)$.

We have also studied the properties of normal-theory jackknife and BRR confidence intervals and bootstrap-t confidence intervals; the latter intervals were obtained by approximating the distribution of $t = (T - \theta)/V_J$ by its bootstrap counterpart $t^* = (T^* - T)/V_J^*$, where V_J is the jackknife variance estimator and V_J^* is obtained from V_J by changing w_{hik} to w_{hik}^* (see Rao et al. (1992) for details). Table 5 also reports the simulated error rates in the lower and upper tails (L, U) and standardized lengths (SL) of the above confidence intervals (SL is defined in the footnote to Table 5).

It is clear from Table 5 that the two bootstrap$-t$ intervals track the error rates in both the lower and upper tails better than the normal theory jackknife and BRR intervals, especially for the correlation ($L = 9.8$, $U = 3.7$ for BRR vs. $L = 4.9$, $U = 5.7$ for BBOOT). The balanced bootstrap interval seems to perform somewhat better than the Rao-Wu bootstrap interval with respect to L and U, but its standardized length is slightly larger. The jackknife and BRR perform better than the bootstrap with respect to SL in the case of regression and correlation.

5. Concluding Remarks

We have given methods for constructing second-order balanced bootstrap designs under simple random sampling and stratified multistage sampling with equal first-stage sample sizes within strata (i.e., $n_h = n$). In the latter case, the bootstrap sample sizes within strata, m_h, are taken equal to n_h (i.e., $m_h = n_h = n$). We have also noted in Section 3.2 that our method for $n = 2m$ in the stratified case does not readily extend to the general case $n = tm$, unlike under simple random sampling. It would be useful to develop a suitable method to handle the general case $n = tm$. Also, it would be useful to extend the work to the case where $m_h \neq n$. In particular, the choice $m_h = n - 1$ is useful since it simplifies the bootstrap weights w_{hik}^* to $w_{hik}^* = [n/(n-1)]f_{hi}^* w_{hik}$. Finally, we need suitable methods to handle the general case of unequal first-stage sample sizes, n_h.

Acknowledgements

We are grateful to the Natural Sciences and Engineering Research Council of Canada for supporting this research. Our thanks are also due to Randy Sitter, C.F.J. Wu and the referee for constructive comments.

Appendix: Across Strata Balance for the Cases $n = 4m+1$ and $4m+3$

We need to prove that the sum of $f_{shi}^* f_{ski}^*$ for $(N, N), (N_c, N_c), (N, N_c)$, (N_c, N) are respectively $2n - 1, 2n - 1, 1, 1$. Similarly, we need to show that the sum of $f_{shi}^* f_{skj}^*$, $i \neq j$, for $(N, N), (N_c, N_c), (N, N_c), (N_c, N)$ are respectively

$n-1, n-1, n+1, n+1$. Then

$$\sum_{s=1}^{S} f^*_{shi} f^*_{ski} = \frac{p}{4}(4n-2+2) = np = S$$

and

$$\sum_{s=1}^{S} f^*_{shi} f^*_{skj} = \frac{p}{4}(2n-2+2n+2) = np = S,\ i \neq j,$$

so that condition (3.9) for second-order balance is satisfied. We need to consider the two cases $n = 4m+1$ and $n = 4m+3$ separately.

Case 1. $n = 4m+1$

Consider a pair of columns in N with possible entries 20, 02, 21, 12, 00, 22, 01, 10. Now if there is an entry 12 then there must also be the entry 21 since the sum of $f^*_{shi} f^*_{shj}$ for N equals $n-1 = 4m$, a multiple of 4. Also in a pair of columns with 12 and 21 occurring once, 10 and 01 do not occur since in any given column 1 appears exactly once. Next consider the entry 22. Since the sum of $f^*_{shi} f^*_{shj}$ for N equals $4m$, 22 must appear $m = (n-1)/4$ times if $21, 12$ do not appear; otherwise, 22 appears $m-1$ times. Turning to the pair 20, we first note that both 0 and 2 appear $(n-1)/2 = 2m$ times in any column. Therefore, 20 appears $2m - m = m$ times if 22 appears m times or $2m - (m-1) - 1 = m$ times if 22 appears $m-1$ times; i.e., 20 always appears m times.

We are now in a position to evaluate the relevant sums of $f^*_{shi} f^*_{ski}$ and $f^*_{shi} f^*_{shj}$. First, for (N, N) we have $f^*_{ski} = f^*_{shi}$ and $f^*_{skj} = f^*_{shj}$ so that it readily follows from (2.3) and (2.4) that the relevant sums are $2n-1$ and $n-1$; similarly for (N_c, N_c). It remains to evaluate the sums for (N, N_c); those for (N_c, N) follow by symmetry. First, we note that the sum of $f^*_{shi} f^*_{ski}$ is equal to 1 since 1 appears only once in any given column of N which remains unchanged in the corresponding column of N_c while 2 and 0 are changed to 0 and 2 respectively. Turning to the sum of $f^*_{shi} f^*_{skj}$ for (N, N_c), it is clear that the contribution comes only from the entries 20, 21 and 10 in N which become 22, 21 and 12 in (N, N_c). Therefore, this sum equals $4m+2 = n+1$ when 21 appears in N, i.e., the frequency of 10 is 0. If instead 10 appears in N, then the sum again equals $4m+2 = n+1$ since 21 does not appear in N.

Case 2. $n = 4m+3$

In this case we exploit the structure of the related BIBD with $v = b = 4m+3 = n$, $r = k = 2m+1 = (n-1)/2$, $\lambda = m = (n-3)/4$. The incidence matrix N^* of BIBD has the following structure: the pairs 11, 00, 01, 10 occur with frequencies $\lambda = m$, $b - 2r + \lambda = m+1$, $r - \lambda = m+1$ and $r - \lambda = m+1$ respectively. The initial block of N is obtained by changing 1 to 2 and fixing a 1 in the n-th place. Also, as 2 occurs at quadratic residues and $n = 4m+3$, it cannot occur at the $(n-1)$-th position of the initial block. Thus, 1 is proceeded by 0 and followed by 2 in N. It is therefore clear that 01 and 10 cannot occur together and either of the two pairs 01, 12 or 10, 21 occur, each appearing exactly once.

Now the pair 10 of the BIBD leads to two pairs in N, namely 20 and 21. Since 10 in the BIBD appears $m+1$ times, the pair 20 appears $m+1$ times in N if 21 does not appear and $m+1-1=m$ times if 21 appears once.

We therefore find the frequency structure of pairs 20, 10, 21 in N is one of the following types: (a) $m,1,1$; (b) $m+1,0,0$. Since 20, 10, 21 become 22, 12, 21 in (N,N_c), the sum of $f^*_{shi}f^*_{skj}$ for (N,N_c) equals $4m+2+2=n+1$ for structure (a) and $4(m+1)+0+0=n+1$ for structure (b). Using the previous arguments for case 1, the sum of $f^*_{shi}f^*_{ski}$ for (N,N_c) is 1. These results also hold for (N_c,N) by symmetry. Finally, using (2.3) and (2.4) again the sums of $f^*_{shi}f^*_{ski}$ and $f^*_{shi}f^*_{skj}$ for both (N,N) and (N_c,N_c) are equal to $2n-1$ and $n-1$ respectively.

References

Davison, A. C., Hinkley, D. V. and Schechtman, E. (1986). Efficient bootstrap simulation. *Biometrika* **73**, 555-566.

Efron, B. (1990). More efficient bootstrap simulations. J. Amer. Statist. Assoc. **85**, 79-89.

Graham, R. L., Hinkley, D. V., John, P. W. M. and Shi, S. (1990). Balanced design of bootstrap simulations. J. Roy. Statist. Soc. Ser.B **52**, 185-202.

Hall, P. (1989). Antithetic resampling for the bootstrap. Biometrika **76**, 713-724.

Hall, M. (1967). Combinatorial Theory. Blaisdell, New York.

Hansen, M. H., Madow, W. G. and Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. J. Amer. Statist. Assoc. **78**, 776-793.

Kovar, J. G., Rao, J. N. K and Wu, C. F. J. (1988). Bootstrap and other methods to measure errors in survey estimates. Canan. J. Statist. **16**, 25-45.

McCarthy, P. J. (1969). Pseudo-replication: Half-samples. Rev. Int. Statist. Inst. **37**, 239-264.

Raghavarao, D. (1971). Constructions and Combinatorial Problems in Design of Experiments. John Wiley, New York.

Rao, C. R. (1961). A study of BIB designs with replications 11 to 15. Sankhyā Ser.A **23**, 117-127.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data. J. Amer. Statist. Assoc. **83**, 231-241.

Rao, J. N. K., Wu, C. F. J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. Survey Methodology **18**, 209-217.

Saha, G. M. and Dey, A. (1973). On construction and uses of balanced n-ary designs. Ann. Inst. Statist. Math. **25**, 439-445.

Sitter, R. R. (1993). Balanced repeated replications based on orthogonal multi-arrays. Biometrika **80**, 211-221.

Sprott, D. A. (1962). A listing of BIB designs from $r=16$ to 20. Sankhyā Ser.A **24**, 203-204.

Takeuchi, G. (1962). A table of difference set generating balanced incomplete block designs. Rev. Int. Statist. Inst. **30**, 361-366.

Vinogradov, I. M. (1954). Elements of Number Theory. Dover, New York.

Institute of Applied Statistics and Development Studies, B-16/1 First Floor, Rajajipuram, Lucknow -226 017, India.

Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6, Canada.

(Received September 1993; accepted November 1994)