Applications of quasi-Monte Carlo methods in inference for complex survey data

Stanislav Kolenikov*

Abstract

This paper proposes a new method for estimating variances of complex survey estimators based on the recent developments in quasi-Monte Carlo methods. The method can be effectively used to create replication schemes in complex surveys with designs more complex than 2 PSU/stratum, while other methods such as the survey bootstrap carry with them a substantial computational burden, as well as somewhat larger instability. The new method is based on quasi-Monte Carlo methods, such as multidimensional Halton sequences. The motivation and main advantages of QMC are briefly discussed. Several possible implementations of variance estimators for complex surveys are offered. A simulation shows that the QMC methods with additional balancing can achieve performance comparable to that of the balanced bootstrap.

KEY WORDS: quasi-Monte Carlo, Halton sequence, BRR, bootstrap, complex survey data

1. Replication variance estimation in complex surveys

This paper considers design variance estimation for complex surveys by using replication, or resampling, methods. In general, those methods create a series of subsets, or replicates, of the observed data, and estimate the variance of interest through variability in the estimates of the same parameter computed from those replicates. Those methods are strictly applicable when the first stage units are taken with replacement, and are approximate in WOR designs with small sampling fractions (Rust & Rao 1996, Shao 1996, Brick, Morganstein & Valliant 2000, Shao 2003, Phillips 2004).

*University of Missouri, Columbia. Mailing address: 146 Middlebush Hall, Department of Statistics, University of Missouri, Columbia, MO 65211. Email: kolenikovs@missouri.edu

The three major replication, or resampling, methods used in complex survey inference include: balanced repeated replication (BRR), the jackknife and the bootstrap. In each of those methods, the PSUs are carefully reshuffled, so that in r-th replication, some PSUs are omitted, and some are included (may be multiple times, as in the bootstrap). The statistic of interest $\hat{\theta}^{(r)}$ for this subsample of data is computed, and the process is repeated R times. The resulting estimator of variance is defined by a standard formula

$$\hat{\mathbb{V}}_R[\hat{\theta}] = \frac{A}{R}\sum_{r=1}^{R}(\hat{\theta}^{(r)} - \tilde{\theta})^2 \qquad (1)$$

where A is a scaling constant equal to 1 for the bootstrap and jackknife, 1/4 for the basic BRR. Some possible variations of (1) include (i) taking $R-1$ in the denominator, and (ii) using either the estimate based on the original sample $\hat{\theta}$ for the estimate of location $\tilde{\theta}$, leading to MSE-type estimator, or the average of the resampled values $\tilde{\theta} = 1/R\sum_r \hat{\theta}^{(r)}$, leading to variance-type estimator. The replication methods of variance estimation then differ in the resampling designs, i.e., the patterns of included and excluded PSUs.

In balanced repeated replication (strictly applicable only for designs with $n_h = 2$ for all strata), only 1 PSU per stratum is selected into the subsample (McCarthy 1969). In jackknife method, (Krewski & Rao 1981), each of $n = \sum_h n_h$ PSUs is omitted, one at a time. In survey bootstrap (Rao & Wu 1988), $m_h \le n_h$ units are resampled with replacement from each stratum.

In designs more complex than 2 PSU/strata, only the latter two methods are strictly applicable. Certain extensions of BRR relaxing the 2 PSU/stratum requirement have been proposed (Gurney & Jewett 1975, Gupta & Nigam 1987, Wu 1991, Sitter 1993), but they still suffer the limitations of availability of the appropriate orthogonal or mixed orthogonal arrays. The computational burden of the bootstrap can be somewhat meliorated by the use of balanced bootstrap schemes (Davison, Hinkley & Schechtman 1986, Graham, Hinkley, John & Shi 1990, Nigam & Rao 1996), but exact second order balancing is only feasible for a limited range of designs.

This paper proposes to use highly uniform quasi-Monte Carlo sequences (Niederreiter 1992, Morokoff & Caflisch 1994) that are routinely employed in multivariate integration problems in physics, applied mathematics, as well as choice modeling. Those methods achieve asymptotic, in the number of points of the sequence, balance, and that has good promise in construction of approximately balanced designs. Application of QMC sequences in bootstrap problems have been considered by Do & Hall (1991).

2. Quasi-Monte Carlo methods

The term "quasi-Monte Carlo" refers to the set of methods of generating deterministic *point sets* and *nets* that achieve highly uniform coverage of the unit cube $[0,1]^s$. Those methods find most applications in numerical integration (Niederreiter 1992, Morokoff & Caflisch 1994) and they also proliferate in econometrics of discrete choice modeling (Train 2001, Train 2003, Bhat 2001).

In this work, we shall concentrate on Halton sequence obtained as follows. For an integer $b > 2$, the *radical inverse function in base* b is

$$\phi_b(n) = \sum_{j=0}^{\infty} a_j(n) b^{-j-1} \qquad (2)$$

where $a_j(n)$ are coefficients of the digit expansion of n in base b,

$$n = \sum_{j=0}^{\infty} a_j(n) b^j \qquad (3)$$

Note that $\phi_b(n) \in [0, 1)$. The *Halton sequence* for a dimension s is obtained as

$$x_n = (\phi_{b_1}(n), \dots, \phi_{b_s}(n)) \qquad (4)$$

where $b_1, \dots, b_s$ are pairwise relatively prime integers greater than 1 (typically the first s primes). It has been shown (Niederreiter 1992) that the discrepancy (the supremum of the difference between the fraction of hits by the sequence of length $\mathcal{N}$ and the Lebesgue measure over s-dimensional rectangles in $[0,1)^s$) is of the order $A\mathcal{N}^{-1}\ln^s\mathcal{N}$. That is a major asymptotic improvement over the (almost sure) $O\big((\ln\ln\mathcal{N}/\mathcal{N})^{1/2}\big)$ rate achievable by the random Monte Carlo methods. The numbers b are usually taken to be the first s primes, as that delivers the minimum to the leading term A. The bound is conservative (Morokoff & Caflisch 1994), and the constants are known to be too generous, but the rates of convergence cannot be improved. There is also an associated curse of dimensionality, as $\ln A \sim s \ln s$, i.e., A grows exponentially fast with s, and the discrepancy may still be large for small to moderate $\mathcal{N}$ and large s.

Various modification of the Halton sequences aimed at reducing "autocorrelations" between components of the sequence with high b's; combating the dimensionality curse; and providing for estimation of the (randomization, or design) standard errors have been proposed. Owen (1998) reviews the work on randomized quasi-Monte Carlo, such as random rotations and permutations of the Halton sequences. Bhat (2003) applies similar ideas to the empirical research in individual transportation choices.

3. The proposed estimator

We are proposing to use quasi-Monte Carlo methods to generate resampling designs in replication variance estimation in complex surveys. The highly uniformly distributed elements of quasi-Monte Carlo sequences will be mapped to the sampling units, to achieve asymptotically, in number of replicates R, balanced resampling designs. The method produces approximately balanced designs, and can be viewed as extension of the balanced repeated replication for arbitrary designs, including those that do not necessarily have $n_h = 2$. It relaxes the limitations on the number of replicates imposed by the availability of mixed orthogonal arrays, although to achieve exact balance, other conditions may need to be imposed. We believe computations are somewhat more straightforward with the current software than constructions of the mixed arrays.

Alternatively, the quasi-Monte Carlo based resampling designs can be viewed as a way to balance the bootstrap replication schemes, in a way more flexible than the existing methods (Nigam & Rao 1996) that, similarly to BRR, utilize the traditional balanced incomplete block design approaches such as those based on Hadamard matrices, and are currently described in the case of equal number of PSUs per strata.

In what follows we shall assume for simplicity a stratified single stage sampling design with L strata and n_h units taken in stratum h, possibly with different probabilities of selection.

A successful replication design must satisfy certain balance conditions. Wu (1991) showed how those balance conditions are related to estimation of bias and variance, and that violations of those balancing conditions generally lead to biases of those estimates. *The first order balance* condition states that all units in a given stratum appear equal number of times in the resampling plan, i.e., the number of times a unit is included in the replication design should be constant for all units within strata, although may differ between strata. The number of replications R would then have to be proportional to the least common multiple of n_h's:

$$R \propto n_h, \quad \forall\, h = 1, \ldots, L \tag{5}$$

If there is substantial variability in the number of units per stratum, this condition may be hard to satisfy. *The second order balance* condition states that all pairs of units need to appear an equal number of times in the replication scheme. This condition is even harder to satisfy. R would need to be proportional to all products $n_h(n_h - 1)$ and cross-products $n_h n_{h'}$, $h, h' = 1, \ldots, L$, for units within and between strata, respectively.

Those balance conditions are ensured by the use of orthogonal arrays in case of BRR, and hold by the law of large numbers in bootstrap. Nigam & Rao (1996) identified certain classes of designs for which second-order balanced bootstrap schemes can be constructed, but in general the problem is difficult.

We consider several possible uses of quasi-Monte Carlo sequences to generate replication designs.

The first method that we will refer as "stratified QMC" generates an L-dimensional Halton sequence, with each component mapped to a stratum in the original sampling design. If $x_{hk} = \phi_{b_h}(k)$ is the h-th component of the k-th element of this Halton sequence, then the unit with the number $[n_h x_{hk} + 1]$ will be included into k-th replicate, where $[\cdot]$ is the integer part operator. If $m_h > 1$ units need to be included in the resampling scheme, the series can be wrapped back after the first R units have been selected: $r = \big((k-1) \mod R\big) + 1$. This scheme fixes m_h to be constant across strata, and thus the sequence length and the number of replications are related as $\mathcal{N} = Rm_h, \quad m_h =$ const w.r.t. h

The aforementioned balance conditions have certain geometric interpretations. To achieve the first order balance, the projections of the Halton sequence within L-dimensional unit cube on the interval $[0, 1]$ corresponding to h-th stratum should throw the same number of points into each interval corresponding to a PSU. The second order balance condition between strata means that all projections of the Halton sequence within L-dimensional unit cube on the unit square h-th and h'-th strata should throw the same number of points into each rectangle corresponding to two PSUs of those strata. The consequence of those conditions is that the length of the Halton sequence $\mathcal{N}$ should be proportional to the least common multiple of $n_1, \ldots, n_L$ to ensure the first order balance. If the design is such that the number of PSU per stratum varies between 2 and k, then $R \propto (k!)^2$ to satisfy the second order condition. Note that those are the necessary conditions only, but the balance will hold asymptotically in $\mathcal{N}$ (or R) from the general theory of Halton sequences.

To overcome the monotonic "autocorrelation" patterns for which Halton sequences with large b's are notorious, the components of the Halton sequence can be "scrambled", by permuting the digits in bases b_k, or "shuffled", by permuting the resulting series. Owen (1995, 1998) and Hess, Polak & Daly (2003) present results showing good performance of the resulting sequences without deterioration of discrepancy.

Stratified QMC method will be prone to high dimensionality curse as mentioned above. First, the upper bound for discrepancy increases almost exponentially with L. Second, the length of the sequence to achieve balance in all dimensions must be proportional to the product of the first L primes, and that grows very fast with L: for values up to $L = 5$, the minimal balancing lengths are 2, 6, 30, 210, 2310. Unfortunately, that's exactly the typical survey situation, with high number of strata L and small number of PSU per stratum.

To arrive at other types of replication designs based on QMC ideas, we would want to restrict the dimensionality of the Halton sequence to ensure fast convergence onto tight discrepancy bounds. One way to do that is to represent the data as two-dimensional array, with PSUs being the rows of it, and individual replications, the columns. In this 2D situation, one can generate the Halton sequence and map the pairs into the two-dimensional plot representing the data. Thus if the PSUs are numbered 1 through $n = n_1 + \ldots + n_L$ the k-th element of Halton sequence, $x_k = (x_{1k}, x_{2k})$, associates the PSU with number $[x_{2k}n + 1]$ with the replicate number $[x_{1k}R + 1]$. This method will be called "data matrix QMC".

The balancing condition on the length of the sequence to achieve uniform coverage over the rectangular array is that $\mathcal{N}$ is proportional to $2 \cdot 3 = 6$, which is a much milder condition that does not depend on the design parameters. On the other hand, to approximately achieve the first order balance, the length of the sequence should be proportional to the number of replicates, and also proportional to the total number of replicates taken: $\mathcal{N} = R(m_1 + \ldots + m_L)$. Finally, the necessary condition imposed by the design is (5).

The variations of this scheme can be as follows. If there is substantial variability in the number of units m_h taken from each stratum, one can represent the h-th stratum with a horizontal strip in the data array of the relative height proportional to m_h rather than n_h. Just like in the stratified QMC case, the resulting Halton sequence can be shuffled (apparently it suffices to only shuffle one of the two dimensions) before being mapped to the data matrix. In the context of survey sampling, it might also be argued that shuffling is beneficial for privacy protection purposes, as it disguises the unit IDs with associated random component better than a fully deterministic method does. Finally, the first order balance can be forced on the design by ordering the Halton sequence by its unit-coordinate and allocating the first m_1R/n_1 units to the first PSU, the next m_1R/n_1 units to the next PSU, etc. Shuffling and balancing can be combined together.

In all of those replication designs, an internal scaling (Wu 1986, Rao & Wu 1988) is implemented through the sets of replicate weights (Rao, Wu & Yue 1992). When applied in practical data collection situations, the typical reservation holds that the post-stratification and non-response adjustments would need to be performed separately for each replicate.

Implementation of the proposed procedures is available with the use of existing software supporting Halton sequences and replication weights. Stata software (Stata Corp. 2005) is an obvious choice, as it has the widest range of design-based estimators for survey data, including the replication estimators, as well as a set of tools for generating Halton sequences (Drukker & Gates 2006). Another package that has both the complex survey package and the modules for Halton sequences is R (Maechler 2006, Lumley 2007).

4. Simulation study

A simulation study was performed on a small artificial population. Five strata of $N_h = 1000$ units each were created with the following (model) distributions:

$$\begin{aligned}
&X_{1i} \sim \Gamma(2,1), && E_{1i} \sim N(0,1),\\
&X_{2i} \sim \Gamma(2,1), && E_{2i} \sim N(0,X_{hi}),\\
&X_{3i} \sim \Gamma(2,1), && E_{3i} \sim t(4+3/(X_{hi}+1)),\\
&X_{4i} \sim \Gamma(2,3), && E_{4i} \sim N(0,X_{hi}),\\
&X_{5i} \sim \Gamma(2,9), && E_{5i} \sim N(0,X_{hi}),\\
&Y_{hi} = X_{hi} + E_{hi}, && h = 1,\ldots,5
\end{aligned} \tag{6}$$

where $h = 1,\ldots,5$ enumerates the strata, $i = 1,\ldots,N_h$, the units within strata, and X_i and E_i are independent of one another and across observations. The population parameters of interest are the mean of X_{hi} (with model expectation of 6) and the regression coefficients (with model expectations of $\beta_0 = 0$ and $\beta_1 = 1$ for intercept and slope, respectively). The coefficient of variation of X_i is 1.45.

SRS of size $n_h = 18$ were taken from each strata, and means $\bar{X}$ and regression coefficients b_0, b_1 were estimated. The following variance estimators were considered: linearization estimator; jackknife; the bootstrap estimator through the method of weights as described in Rao et al. (1992); the first order balanced bootstrap where each unit is sampled the same number of times (Davison et al. 1986, Gleason 1988,

Table 1: Performance of variance estimators.

Estimator		IQR	Stability	Coverage, %		Satterthwaite
				d.f.= $n-L$	Effective	d.f.
S.e. $[\bar{X}] = 0.6500$						
Linearized		(0.5256, 0.7143)	0.0482	92.5	94.4	9.43
Bootstrap		(0.5253, 0.7126)	0.0509	92.8	94.7	8.84
Balanced bootstrap		(0.5214, 0.7173)	0.0504	92.1	94.4	9.02
Stratified QMC:	Method 1	(0.3895, 0.5866)	0.1074	83.9	91.8	4.70
	Method 2	(0.5194, 0.7154)	0.0510	92.3	94.4	8.89
Data matrix QMC:	Method 3	(0.4389, 0.6143)	0.0719	87.2	91.3	7.73
	Method 4	(0.7934, 1.0281)	0.2501	97.9	98.8	12.67
	Method 5	(0.4092, 0.5746)	0.0906	84.7	89.9	7.22
	Method 6	(0.5212, 0.7155)	0.0522	92.3	94.4	8.56
S.e. $[b_X] = 0.0682$						
Linearized		(0.0430, 0.0690)	0.1170	84.6	96.8	3.27
Jackknife		(0.0487, 0.0810)	0.1520	90.0	99.5	2.50
Bootstrap		(0.0466, 0.0716)	0.0929	89.0	97.5	4.33
Balanced bootstrap		(0.0473, 0.0717)	0.0932	88.8	97.5	4.19
Stratified QMC:	Method 1	(0.0340, 0.0602)	0.1676	78.6	95.1	2.50
	Method 2	(0.0471, 0.0716)	0.0949	89.4	96.7	4.15
Data matrix QMC:	Method 3	(0.0379, 0.0596)	0.1350	80.7	94.3	3.23
	Method 4	(0.0480, 0.0727)	0.0935	89.9	97.4	4.14
	Method 5	(0.0375, 0.0590)	0.1379	80.8	94.1	3.29
	Method 6	(0.0469, 0.0716)	0.0937	89.6	96.7	4.25

Method 1: stratified QMC. Method 2: stratified shuffled QMC. Method 3: data matrix 2D QMC. Method 4: shuffled data matrix 2D QMC. Method 5: balanced data matrix 2D QMC. Method 6: balanced shuffled data matrix 2D QMC.

Nigam & Rao 1996, Rao & Shao 1999); and all of the six QMC-based methods described in previous section (stratified QMC with and without shuffling; data array 2D QMC with and without either of shuffling and balancing). The number of replications for the bootstrap and for 2D QMC methods was $R = 156$, while that for stratified and stratified-shuffled QMC methods was $R = 154$, thus ensuring that strata 1, 4 and 5 are balanced, while strata 2 and 3 remain unbalanced. The number of sampled units for each of the resampling methods was $m_h = n_h - 3 = 15$, according to recommendations of Rao & Wu (1988). 1000 samples were taken.

The simulations were performed in Stata software. The code for both simulations and QMC resampling designs is available from author upon request.

Table 1 gives the simulation results comparing the different variance estimators in terms of the reported standard errors, stability, and coverage of the nominal 95% CI. For the simulation size of 10000, the margin of simulation error is $1.96 \times \sqrt{0.95 \cdot 0.05/1000} = 1.35\%$, so the results between 93.65% and 96.35% can be considered acceptable. Column 4 gives the results for coverage with nominal degrees of freedom, $n - L = 90 - 5 = 85$, which is essentially the normal distribution, while column 5 gives coverage with the effective Satterthwaite degrees of freedom (Rust & Rao 1996). The latter are reported in the last column, and seem to be quite low in this design. Fig. 1 represents those results graphically with kernel density estimates of the reported standard errors distributions, along with the Monte Carlo standard deviations.

For both estimation of the variances of $\bar{X}$ and b_1, the balanced bootstrap and the shuffled+balanced 2D QMC come closest to the true variability. The linearization estimator *over*estimates the variability of the sample mean, and *under*estimates the variability of the regression slope. The basic versions of QMC without shuffling tend to underestimate the true variance. The explanation might be that there is not enough variability in the resampling weights generated from the basic QMC, or that the Halton sequences are not balanced enough at those lengths thus leading to biases of the variance estimates, in accordance with results of Wu (1991). With the exception of the shuffled 2D QMC in the sample mean case, the variation of stability is usually within a factor of 1.5, so some gains in inference precision would be expected from the more stable estimators.

In the sample mean case, the most stable estimators were found to be the linearization (and the jackknife equivalent to it), the bootstraps, and the "advanced" QMC methods (shuffled stratifed, Method 2, and shuffled balanced 2D, Method 6). The downward bias of the basic QMC estimators (stratified, Method 1; 2D, Method 3) led to substantial undercoverage of the confidence intervals. The balanced 2D QMC (Method 4), even though also biased downward, overcovers.

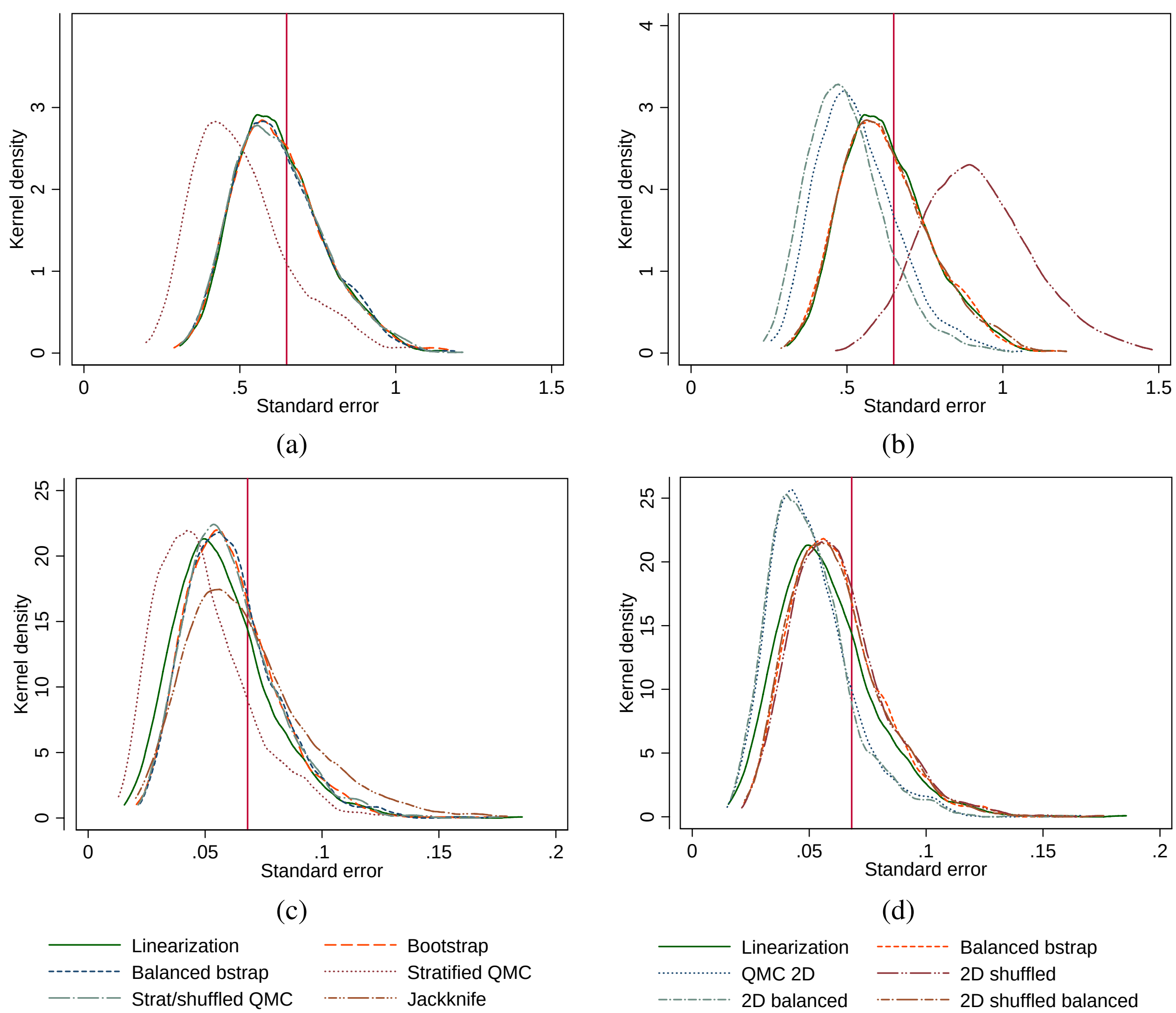

Figure 1: Estimates of variance for an artificial population. Top row: $v^{\frac{1}{2}}[\bar{X}]$; bottom row: $v^{\frac{1}{2}}[b_1]$. Left column: linearization, jackknife, the bootstraps, and stratified QMC. Right column: linearization, the balanced bootstrap, and data-matrix QMC. The vertical lines are Monte Carlo standard deviations.

None of the estimators, however, achieve the desired nominal level, with the linearization, both versions of the bootstrap, and "advanced" QMC coming closest. A correction for Satterthwaite degrees of freedom of the design (Rust & Rao 1996) brings the linearization, the bootstraps, and "advanced" QMC estimators within the simulation error limits of 95%.

In the regression slope case, the stablest estimators were found to be both versions of the bootstrap and both original and shuffled versions of the 2D QMC, all of them beating the linearization and the jackknife by some 20% stability margin. They are also demonstrating the best performance in terms of coverage probabilities, with the shuffled+balanced 2D QMC taking somewhat of a lead, although all estimators are very far from the nominal 95%. The explanation might be in high kurtosis of the explanatory variable which is known to be responsible for finite sample biases of the sandwich estimator, see Kauermann & Carroll (2001). A correction for Satterthwaite degrees of freedom of the design again appears to be overcorrecting for undercoverage.

It is also worth noting that all of the balancing procedures (balanced bootstrap, balanced data-matrix-based QMC with or without shuffling) achieve first order balance: all the units are included equal number of times.

5. Conclusion

A new procedure for obtaining approximately balanced replication designs based on quasi-Monte Carlo methods has been proposed. The performance of certain flavors of the method were found to be superior of the regular bootstrap, and comparable to those of the balanced bootstrap. It also has an appeal of being computationally feasible through modern software.

Further work will include a formal establishment of asymptotic properties of various implementation of the QMC-based variance estimators, as well as larger simulation studies with wider ranges of the QMC settings.

Acknowledgements

The author would like to thank Jon Rao, Milorad Kovacevic and John Eltinge for helpful discussions and comments. The errors, omissions, inconsistencies and typos remain the author's responsibility.

References

Bhat, C. R. (2001), 'Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model', *Transportation Research Part B: Methodological* **35**(7), 677–693.

Bhat, C. R. (2003), 'Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences', *Transportation Research Part B: Methodological* **37**(9), 837–855.

Brick, J. M., Morganstein, D. & Valliant, R. (2000), Analysis of complex sample data using replication, technical report, Westat. http://www.westat.com/wesvar/techpapers/ACS-Replication.pdf.

Davison, A. C., Hinkley, D. V. & Schechtman, E. (1986), 'Efficient bootstrap simulation', *Biometrika* **73**(3), 555–566.

Do, K.-A. & Hall, P. (1991), 'Quasi-random resampling for the bootstrap', *Statistics and Computing* **1**(1), 13–22.

Drukker, D. M. & Gates, R. (2006), 'Generating Halton sequences using Mata', *Stata Journal* **6**(2), 214–228.

Gleason, J. R. (1988), 'Algorithms for balanced bootstrap simulations', *The American Statistician* **42**(4), 263–266.

Graham, R. L., Hinkley, D. V., John, P. W. M. & Shi, S. (1990), 'Balanced design of bootstrap simulations', *Journal of the Royal Statistical Society* **52**(1), 185–202.

Gupta, V. K. & Nigam, A. K. (1987), 'Mixed orthogonal arrays for variance estimation with unequal numbers of primary selections per stratum', *Biometrika* **74**(4), 735–742.

Gurney, M. & Jewett, R. S. (1975), 'Constructing orthogonal replications for variance estimation', *Journal of the American Statistical Association* **70**(352), 819–821.

Hess, S., Polak, J. & Daly, A. (2003), On the performance of shuffled halton sequences in the estimation of discrete choice models, European Transport Conference, Strasbourg.

Kauermann, G. & Carroll, R. J. (2001), 'A note on the efficiency of sandwich covariance matrix estimation', *Journal of the American Statistical Association* **96**(456), 1387–1396.

Krewski, D. & Rao, J. N. K. (1981), 'Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods', *The Annals of Statistics* **9**(5), 1010–1019.

Lumley, T. (2007), *The* `survey` *package*, R Comprehensive Network Archive (CRAN). http://cran.r-project.org/doc/packages/survey.pdf.

Maechler, M. (2006), *The* `sfsmisc` *package*, R Comprehensive Network Archive (CRAN). http://cran.hu.r-project.org/doc/packages/sfsmisc.pdf.

McCarthy, P. J. (1969), 'Pseudo-replication: Half samples', *Review of the International Statistical Institute* **37**(3), 239–264.

Morokoff, W. J. & Caflisch, R. E. (1994), 'Quasi-random sequences and their discrepancies', *SIAM Journal on Scientific Computing* **15**(6), 1251–1279.

Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, Vol. 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, Society for Industrial and Applied Mathematics, Philadelphia.

Nigam, A. K. & Rao, J. N. K. (1996), 'On balanced bootstrap for stratified multistage samples', *Statistica Sinica* **6**(1), 199–214.

Owen, A. B. (1995), Randomly permuted (t, m, s)-nets and (t, s)-sequences, *in* H. Niederreiter & P. J.-S. Shiue, eds, 'Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing', Springer-Verlag, New York, pp. 299–317.

Owen, A. B. (1998), Monte Carlo extension of quasi-Monte Carlo, *in* 'WSC '98: Proceedings of the 30th conference on Winter simulation', IEEE Computer Society Press, Los Alamitos, CA, USA, pp. 571–578.

Phillips, O. (2004), Using bootstrap weights with WesVar and SUDAAN, Technical Report 2, Statistics Canada.

Rao, J. N. K. & Wu, C. F. J. (1988), 'Resampling inference with complex survey data', *Journal of the American Statistical Association* **83**(401), 231–241.

Rao, J. N. K., Wu, C. F. J. & Yue, K. (1992), 'Some recent work on resampling methods for complex surveys', *Survey Methodology* **18**(2), 209–217.

Rao, J. & Shao, J. (1999), 'Modified balanced repeated replication for complex survey data', *Biometrika* **86**(2), 403–415.

Rust, K. F. & Rao, J. N. (1996), 'Variance estimation for complex surveys using replication techniques', *Statistical Methods in Medical Research* **5**(3), 283–310.

Shao, J. (1996), 'Resampling methods in sample surveys', *Statistics* **27**, 203–254. with discussion.

Shao, J. (2003), 'Impact of the bootstrap on sample surveys', *Statistical Science* **18**, 191–198.

Sitter, R. R. (1993), 'Balanced repeated replications based on orthogonal multi-arrays', *Biometrika* **80**(1), 211–221.

Stata Corp. (2005), *Stata Statistical Software: Release 9*, College Station, TX, USA.

Train, K. (2001), Halton sequences for mixed logit, Econometrics Working Paper 0012002, EconWPA. available at http://ideas.repec.org/p/wpa/wuwpem/0012002.html.

Train, K. E. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press.

Wu, C. F. J. (1986), 'Jackknife, bootstrap and other resampling methods in regression analysis', *The Annals of Statistics* **14**(4), 1261–1295.

Wu, C. F. J. (1991), 'Balanced repeated replications based on mixed orthogonal arrays', *Biometrika* **78**(1), 181–188.