

Bayesian nonparametric estimation of finite population quantities in absence of design information on nonsampled units

Sahar Z Zangeneh* Robert W. Keener† Roderick J.A. Little‡

Abstract

In Probability proportional to size (PPS) sampling, the sizes for nonsampled units are not required for the usual Horvitz-Thompson or Hajek estimates, and this information is rarely included in public use data files. Previous studies have shown that incorporating information on the sizes of the nonsampled units through semiparametric models can result in improved estimates. When the design variables that govern the selection mechanism, are missing, the sample design becomes informative and predictions need to be adjusted for the effect of selection. We present a general framework using Bayesian nonparametric mixture modeling with Dirichlet process priors for imputing the nonsampled size variables, when such information is not available to the statistician analyzing the data.

Key Words: Bayesian nonparametric modeling; mixture models; Dirichlet process priors; Probability proportional to size sampling

1. Introduction

We consider probability samples from a finite population of size $N < \infty$, where each member of the population has a non-zero probability of inclusion in the sample. We denote the survey (or outcome) variable of the i th unit to be $Y_i; i = 1, \dots, N$, the inclusion indicator variable of the i th unit to be $I_i, i = 1, \dots, N$, and $x = (x_1, \dots, x_N)$ the vector of design variables. The aim of analysis is typically to perform inference about a finite population quantity $Q(Y)$, from a sample of size n , where $Y = (Y_1, \dots, Y_N)$ and selection in the sample is governed by the design variables x . The design-based and model-based paradigms comprise the two main approaches to survey inference. The design-based or randomization approach (Cochran, 2009), which automatically takes features of the survey design into account, is desirable for its lack of reliance on distributional assumptions. Design-based inferences are based on the distribution of $I = (I_1, \dots, I_N)$, while the outcome variables Y_1, \dots, Y_N are treated as fixed. This traditional approach yields reliable inferences for large samples, but is generally asymptotic and can be inefficient.

Model-based inference on the other hand, considers both $I = (I_1, \dots, I_N)$ and $Y = (Y_1, \dots, Y_N)$ to be random variables and inferences are based on the joint distribution of Y and I . This approach primarily aims at predicting the survey outcome variables for the nonsampled values in the population. Interpreting nonsampled observations as missing data, Rubin (1976) establishes

*Department of Statistics, The University of Michigan, Ann Arbor, MI, U.S.A.

†Department of Statistics, The University of Michigan, Ann Arbor, MI, U.S.A.

‡Department of Biostatistics, The University of Michigan, Ann Arbor, MI, U.S.A.

conditions under which the selection method can be ignored for model-based inferences from the Bayesian, likelihood or sampling theory viewpoints. He shows that under probability sampling, inferences can be based on the distribution of Y alone, provided the design variables are included in the model, and the distribution of I given Y is independent of the distribution of Y conditional on the survey design variables. See Little and Zheng (2007) and Sugden and Smith (1984) for more detailed discussions. We consider inference from a sample of size n which partitions the population into two sets $\{S, S^c\}$, where S denotes the sample and S^c the population units not observed in the sample. In particular, we consider Bayesian modeling of surveys (Ghosh and Meeden, 1997; Little, 2004), where a prior distribution is specified for the parameters θ , along with a distribution for the population values Y conditional on θ . We focus our attention to models with non-informative priors dominated by the likelihood, whose resulting point estimates are similar to those obtained from frequentist superpopulation modeling. Moreover, the Bayesian methodology enables us to perform reliable inference in an operationally simple manner.

This study focuses on Bayesian model-based inference of finite population quantities in the absence of design information on the nonsampled units. We consider situations where the design variables x_i are only available to the analyst for the observed units in the sample. Although, complete design information $x = (x_S, x_{S^c})$ is necessary at the sampling phase, since the traditional design-based methods only require knowing the design variables for the sampled units, x_{S^c} is not recorded in many public use data sets Pfeffermann et al. (1998). (Sugden and Smith, 1984) derive conditions where the sampling design becomes informative for model-based inference.

For the remainder of this paper, we restrict our attention to probability proportional to size (PPS) sampling. In PPS sampling, the size measures $x_i, i = 1, \dots, N$ comprise the design variables, which are known for all units in the population at the time of sampling, and unit i is selected with probability π_i proportional to its size x_i . We generally aim at estimating finite population quantities $Q(Y)$ of a continuous outcome Y , from a systematic PPS sample of size n , where n is assumed fixed, in situations where the size measures x_i are only reported for the sampled cases, but the number of nonsampled cases $N - n$ and the population mean of x for those cases is known.

Zheng and Little (2003), showed that incorporating design information on nonsampled units through spline models can result in improved estimates when the size variables in the entire population are known. These models had the advantage of yielding efficient estimates, while being robust to model mis-specification. Chen (2009) and her co-authors extended this methodology to estimating finite population quantiles, again when knowing the size variables in the entire population.

Under PPS sampling, the sample design becomes informative (Sugden and Smith, 1984) when the sizes of the nonsampled units are unknown, and hence the sizes of nonsampled units need to be adjusted for the effects of selection. We therefore factor the joint distribution of the survey variables and design variables as

$$f(x, Y) = f(x)f(Y|x), \quad (1)$$

where sampling is ignorable in $f(Y|x)$ (Rubin, 1976), and we are therefore able to base inference on a model for the conditional density, based on the observed units in the sample. Exploiting this

factorization, Little and Zheng (2007) proposed estimation of the finite population total when the size variables are only known for the sampled units, but the mean of the size variables is known in the population of known size N , and Zangeneh and Little (2011) studied the inferential properties of such an estimator. A Bayesian bootstrap procedure that adjusts for PPS selection was used to predict the sizes of the nonsampled units, followed by a penalized spline model to predict the survey outcome variables of the nonsampled units. As noted by these authors, imputing the sizes of the nonsampled units plays a crucial role towards model-based inference of finite population quantities.

Nonparametric methods are desirable to predict the sizes of the nonsampled units, as they are robust to model mis-specification and impose minimal prior assumptions on the distribution of the size variables. Moreover multiple imputation based on Bayesian principles is a powerful tool for missing data problems, due to their ability to propagate imputation uncertainty (Little and Rubin, 1987).

In his seminal paper, Ferguson (1973), introduced a general framework for nonparametric Bayesian inference via Dirichlet process priors, which is a distribution over the set of probability measures. Several authors further studied theoretical properties of such models, but it was not until after advancements in Bayesian computing, and the development of simulation algorithms such as the Gibbs sampler (Geman and Geman, 1993; Gelfand and Smith, 1990), that this theory gained widespread attention, due to computational feasibility.

Bayesian nonparametric mixture models have been flourishing in many areas of statistics and is an active area of research. However, minimal attention has been paid to such methodology by survey statisticians. We present a general Bayesian nonparametric framework, which encompasses the Bayesian bootstrap as a special case, for imputing the sizes of nonsampled units under PPS sampling.

2. Methods

2.1 Dirichlet Process Mixture modeling

Since the sizes for the nonsampled units are unknown, the sampling design becomes informative Sugden and Smith (1984), and therefore the draws of the sizes of nonsampled cases need to be corrected for the effects of selection. Our general framework is based on considering two separate exchangeable models for the sampled and nonsampled units. We denote $f(x_i|i = 1, \theta_i)$ to be the distribution of the i th size variable $i = 1, \dots, n$ with underlying parameter vectors $\theta_i; i = 1, \dots, n$. Note that the data points are overparametrized, making the modeling approach nonparametric (Jordan, 2005). Formally, we have the following hierarchical model

$$X_i|\theta_i \sim f(x_i|i = 1, \theta_i)$$

$$\theta_i|G \sim G$$

$$G|\alpha, G_0 \sim DP(\alpha, G_0),$$

in which each of the (X_i, θ_i) pairs are independent given G . Modeling the predictive posterior distribution of observed size variables via a Dirichlet process mixture (DPM) model, the posterior distribution of a hypothetical future observation from the sampled units is

$$(X_{n+1}|1, \theta, \text{data}) \sim \int P(X_{n+1}|\theta_{n+1})dP(\theta_{n+1}|1, \theta) \quad (2)$$

where

$$(\theta_{n+1}|1, \theta) \sim \frac{\alpha}{\alpha + n}G_0(\theta_{n+1}) + \frac{\alpha}{\alpha + n} \sum_{i=1}^n \delta_{\theta_i}(\theta_{n+1}),$$

with δ_{θ_i} denoting a point mass at θ_i .

By assuming a discrete distribution for X_i and a Dirichlet/multinomial distribution for θ , we get the Bayesian bootstrap model (Rubin, 1981; Aitkin, 2008), while a normal distribution for X_i and a Normal/Inverse Wishart distribution for θ gives rise to a Dirichlet process mixture of normals (DPMN) (Escobar and West, 1995).

2.2 Relationship between posterior predictive distribution of sampled and nonsampled units

Let X_1, \dots, X_N be the set of all size variables in the population, $T_x = \sum_{i=1}^N X_i$ denote the population mean of the sizes, and $I_i = 1(\text{unit } i \text{ in sample})$ be the inclusion indicator variable of the i th unit in the population. We denote the parameter vector corresponding to the entire population to be θ , and define $f(x, e|\theta)$ to be the joint density of (X_i, I_i) given T_x and indexed by the parameters θ , and let $f(x|\theta) = f(x, 0|\theta) + f(x, 1|\theta)$ be the marginal density of the size variables. Then, according to the sampling design, we have $f(x, 1|\theta) = cx f(x|\theta)$ for some positive constant c . Since the (unconditional) chance of selection is n/N ,

$$\frac{n}{N} = \int f(x, 1|\theta)d\nu(x) = c \int x f(x|\theta)d\nu(x) = cE(X) = c \frac{T_x}{N}.$$

Thus $c = n/T_x$ and so $f(x, 1|\theta) = nx f(x|\theta)/T_x$ and

$$f(x, 0|\theta) = f(x|\theta) - f(x, 1|\theta) = \frac{T_x - nx}{T_x} f(x|\theta). \quad (3)$$

On the other hand

$$f(x|1, \theta) = \frac{f(x, 1|\theta)}{E(I_i|\theta)} = \frac{Nxf(x|\theta)}{T_x}, \quad (4)$$

therefore from (3) and (4), we have

$$f(x|0, \theta) = \frac{f(x, 0|\theta)}{1 - E(I_i|\theta)} = \frac{T_x - nx}{T_x} \frac{1}{1 - n/N} f(x|\theta) = \frac{1}{N - n} \left(\frac{T_x}{x} - n \right) f(x|1, \theta). \quad (5)$$

Note that the conditional density described in (5) is unstable for values of x close to zero. To overcome this issue, and due to the positivity of size variables, we model the natural logarithms of the size variables instead. Thus, (5) can be re-expressed as

$$\tilde{f}(x^*|0, \theta) = \frac{1}{N - n} \left(\frac{T_x}{e^{x^*}} - n \right) \tilde{f}(x^*|1, \theta), \quad (6)$$

where $x^* = \log(x)$, and $\tilde{f}(x^*|1, \theta)$ is modeled as a DPM as explained above.

2.3 Imputing size variables for nonsampled units using Dirichlet process mixture of normals

We follow the approach of Escobar and West (1995), which models $\tilde{f}(x^*|1, \theta)$ as a Dirichlet process mixture of normals with unknown means, unknown variances, unknown number of clusters, and unknown precision α . We model

$$\begin{aligned} X_i | \mu_i, \Sigma_i &\sim N(\mu_i, \Sigma_i), i = 1, \dots, n \\ (\mu_i, \Sigma_i) | G &\sim G \\ G | \alpha, G_0 &\sim DP(\alpha G_0), \end{aligned}$$

with the baseline distribution chosen as the conjugate normal/inverse Wishart prior for the mean and covariance matrix,

$$G_0 = N \left(\mu | m_1, \frac{1}{k_0} \Sigma \right) IW(\Sigma | \nu_1, \psi_1).$$

With calculations similar to those presented in Escobar and West (1995), we find the posterior predictive distribution of a hypothetical future observation from the sampled units to be

$$(X_{n+1}|1, \theta) \sim \frac{\alpha}{\alpha + n} T_{\nu_1+1-p}(m_1, M) + \frac{1}{\alpha + n} \sum_{j=1}^K n_j N(\mu_j^*, \Sigma_j^*), \quad (7)$$

where $T_{\nu_1+1-p}(m_1, M)$ denotes a multivariate t -distribution with $\nu_1 + 1 - p$ degrees of freedom, where p is the dimension of x (in this case $p = 1$), mode m_1 and scale factor $M = \frac{1+k_0}{(\nu_1-p+1)k_0} \varphi_1$. Also, μ_j^* and Σ_j^* , $j = 1 \dots K$ are distinct values of the cluster means and variances under the normal/inverse Wishart conjugate priors for G_0 , where $K < n$.

Combining (5) and (7), we have

$$\begin{aligned} (X_{n+1}^*|0, \theta) &\sim \frac{1}{N - n} \left(\frac{T_x}{e^{x^*}} - n \right) \frac{\alpha}{\alpha + n} T_{\nu_1+1-p}(m_1, M) \\ &+ \frac{1}{N - n} \left(\frac{T_x}{e^{x^*}} - n \right) \frac{1}{\alpha + n} \sum_{j=1}^K n_j N(\mu_j^*, \Sigma_j^*). \end{aligned} \quad (8)$$

Therefore, the probability density function of the sizes for the logarithm of the nonsampled values is

$$\begin{aligned}
 f(x^*|0, \theta) &= \frac{1}{N-n} \frac{1}{\alpha+n} \sum_{j=1}^K n_j (2\pi)^{-p/2} |\Sigma_j^*|^{-1/2} \left(\frac{T_x}{e^{x^*}} - n \right) e^{-\frac{1}{2}(x^* - \mu_j^*)^t \Sigma_j^{*-1} (x^* - \mu_j^*)} \quad (9) \\
 &+ \frac{1}{N-n} \frac{\alpha}{\alpha+n} \left(\frac{T_x}{e^{x^*}} - n \right) \frac{\Gamma(\frac{\nu_1+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{(\nu_1+1-p)^{p/2} \pi^{p/2}} |M|^{-1/2} \\
 &\times \left(1 + \frac{1}{\nu_1+1-p} + (x^* - m_1)^t M^{-1} (x^* - m_1) \right)^{-\frac{\nu_1+1}{2}}.
 \end{aligned}$$

We use importance sampling to draw random variables with density given in (9). This procedure is summarized in Algorithm 1:

Algorithm 1 DPMN model for imputing nonsampled sizes.

1. Fit a DPM mixture of normals model to the logarithms of the observed size variables X^S .
 2. For each of $\theta^{(1)}, \dots, \theta^{(D)}$ obtained from fitting the model in step 1.
 - 2.1. Compute the density given in Equation (9).
 - 2.2. Use importance sampling to draw $N - n$ random variables $x_{n+1}^{*(d)}, \dots, x_N^{*(d)}$, with density (9).
 - 2.3. Exponentiate the draws $x_{n+1}^{*(d)}, \dots, x_N^{*(d)}$, to obtain draws of the nonsampled units.
-

3. Simulation Study

We performed a simulation study to assess the performance of our proposed algorithm for different choices of ν_1 , as well as in comparison with the adjusted Bayesian bootstrap for PPS (Little and Zheng, 2007; Zangeneh and Little, 2011). We considered two variants of each model: (A) unscaled (B) predictions of size variables scaled, so that they sum to their known population total. We simulated the size variables from a gamma distribution with shape parameter 2 and rate parameters 1 and 2. We also considered populations of size 250, 500 and 1,000 and sampling fractions of 10 and 20 percent, and took systematic PPS samples of size $n = r \times N$ in each iteration. Density estimation for observed size variables was performed using the method of Escobar and West (1995) for fitting the Dirichlet process mixture of normals implemented in the R package `DPpackage` (Jara et al., 2011).

In order to assess how well each imputation method captures the shape of the distribution, we augmented the drawn nonsampled sizes to the true sampled sizes and calculated summary measures for each set of draws; a total of 100 sets of draws were recorded in each iteration. The B methods all scale the predicted sizes to average to their true mean in the population, therefore these methods all estimate the population mean without error. Moreover, since PPS sampling oversamples larger

units, the maximum of the size variables will be captured in the sample with very high probability, and hence the Bayesian bootstrap estimates the maximum with very high precision, as it assigns point masses to the observed data (Zangeneh and Little, 2011). We therefore restrict our attention to the estimates of the population quantiles and the population variance. Table 1 shows the root mean square errors for the Bayes estimates of the quantiles and variance under square error loss for the population of sizes generated from a Gamma distribution with shape parameter 2 and scale parameter 1, $\text{Gamma}(2, 1)$. Table 2 shows the same summaries for the sizes following a $\text{Gamma}(2, 2)$ distribution. Finally, Table 3 displays the coverage and average width of 95% credible intervals of the estimated population quantiles and variance for the $\text{Gamma}(2, 2)$ population of size $N = 250$, with a sampling rate of 20%.

Based on our simulations, we see that the Dirichlet process mixture of normals performs better at estimating the population quantiles when the population and sample sizes are small. However, as the sample size increases, the scaled Bayesian bootstrap begins to show better performance. We also notice that scaling significantly improves the estimates of the quantiles, but may distort the estimate of the variance. Finally, we do not see the degrees of freedom of the distribution of the underlying base measure to play a significant role in estimating the population quantiles and variance.

4. Discussion

In this paper, we presented a general framework to impute the nonsampled size variables in PPS sampling using Bayesian nonparametric mixture models. We focused on two special cases of such models; the Dirichlet process mixture of normals and the Bayesian bootstrap, and compared their performances in simulation studies. The Bayesian bootstrap makes weaker parametric assumptions on the base distribution G , and hence is superior in large samples, where enough data is available. Our simulations also confirmed that, for large samples, the Bayesian bootstrap performs as well as, and sometimes better than, the DPMN models; however for small sample scenarios, the DPMN performs better in imputing the nonsampled size variables. The Bayesian bootstrap is appealing due to its simplicity and computational efficiency, however considering the current computational advancements in Bayesian nonparametrics, other forms of DPMs, such as the DPMN model studied in this paper, could also be integrated in survey inference.

5. Acknowledgements

This research was partially supported by the National Science Foundation Grant No. 0706771. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Aitkin, M. (2008). Applications of the bayesian bootstrap in finite population inference. *Journal of Official Statistics*, 24(1):21.
- Chen, Q. (2009). *Bayesian predictive inference for three topics in survey samples*. PhD thesis, The University of Michigan.
- Cochran, W. (2009). *Sampling techniques*. Wiley-India.
- Escobar, M. and West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, 90(430).
- Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 1(2):209–230.
- Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, pages 398–409.
- Geman, S. and Geman, D. (1993). Stochastic relaxation, gibbs distributions and the bayesian restoration of images*. *Journal of Applied Statistics*, 20(5-6):25–62.
- Ghosh, M. and Meeden, G. (1997). *Bayesian methods for finite population sampling*, volume 79. Chapman & Hall/CRC.
- Jara, A., Hanson, T., Quintana, F., Müller, P., and Rosner, G. (2011). Dppackage: Bayesian semi- and nonparametric modeling in r. *Journal of Statistical Software*, 40:1–30.
- Jordan, M. (2005). Dirichlet processes, chinese restaurant processes and all that. In *Tutorial presentation at the NIPS Conference*.
- Little, R. (2004). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, 99(466):546–556.
- Little, R. and Rubin, D. (1987). Statistical analysis with missing data.
- Little, R. and Zheng, H. (2007). The Bayesian Approach to the Analysis of Finite Population Surveys. *Bayesian Statistics*, 8(1):1–20.
- Pfeffermann, D., Krieger, A., and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, 8:1087–1114.
- Rubin, D. (1976). Inference and missing data. *Biometrika*, 63(3):581.
- Rubin, D. (1981). The bayesian bootstrap. *The Annals of Statistics*, 9(1):130–134.

- Sugden, R. and Smith, T. (1984). Ignorable and informative designs in survey sampling inference. *Biometrika*, 71(3):495.
- Zangeneh, S. Z. and Little, R. J. A. (2011). Two-stage semiparametric bayesian inference for the finite population total in probability proportional to size samples. to be submitted.
- Zheng, H. and Little, R. (2003). Penalized spline model-based estimation of the finite populations total from probability-proportional-to-size samples. *Journal of Official Statistics*, 19(2):99–118.

Table 1: Root mean square errors of the posterior means of quantiles and variance from their true population value (as a percentage of the true population quantity) for the Gamma distribution with shape parameter 2 and rate parameter 1.

		BB		DPMN.5		DPMN.10		DPMN.20	
		A	B	A	B	A	B	A	B
N = 250 r = 0.1	q _{.05}	56.12	45.86	37.30	24.41	40.08	27.75	44.69	31.45
	q _{.25}	26.02	11.35	22.79	9.36	23.85	10.82	25.78	12.34
	q _{.50}	19.02	5.04	16.47	3.86	16.22	3.21	16.42	2.51
	q _{.75}	13.14	6.91	14.58	6.75	14.41	7.21	13.96	7.48
	q _{.95}	11.17	7.89	10.69	5.83	10.72	5.91	10.79	6.12
	var	25.61	25.38	19.88	23.94	20.26	21.52	21.68	21.79
N = 250 r = 0.2	q _{.05}	25.64	22.62	13.45	9.04	14.26	10.18	16.65	11.86
	q _{.25}	12.13	9.40	10.27	5.84	10.53	6.15	11.82	6.90
	q _{.50}	7.05	3.70	7.77	2.68	7.55	2.58	7.69	2.39
	q _{.75}	9.77	3.80	7.69	2.08	7.71	1.91	8.22	1.87
	q _{.95}	7.20	6.92	6.37	5.21	6.27	4.90	6.43	4.61
	var	16.95	17.01	17.31	15.66	15.95	15.85	13.96	13.89
N = 500 r = 0.1	q _{.05}	32.86	28.06	18.92	14.07	21.72	16.62	25.32	19.70
	q _{.25}	12.61	8.83	10.24	5.58	10.92	6.18	12.17	6.96
	q _{.50}	7.93	3.50	6.45	1.71	6.59	1.96	7.26	2.40
	q _{.75}	6.43	5.45	4.38	3.21	3.83	3.40	3.92	3.82
	q _{.95}	4.01	5.65	3.17	4.82	2.70	4.50	2.68	4.22
	var	10.61	19.35	11.51	18.24	10.80	16.77	10.50	16.62
N = 500 r = 0.2	q _{.05}	21.62	20.13	20.74	17.31	20.90	17.90	23.10	19.29
	q _{.25}	9.98	7.00	8.22	4.24	8.31	4.35	8.97	4.74
	q _{.50}	6.73	3.36	6.00	1.92	6.28	2.07	6.59	2.10
	q _{.75}	4.97	3.36	4.10	2.80	4.39	2.88	4.20	3.01
	q _{.95}	2.83	3.87	3.85	4.00	3.63	3.53	3.63	3.21
	var	12.22	13.88	11.04	12.87	11.38	12.56	11.99	13.54
N = 1000 r = 0.1	q _{.05}	24.47	21.03	17.85	13.03	19.09	14.25	20.11	15.11
	q _{.25}	9.99	5.75	11.02	5.87	11.37	6.29	11.83	6.64
	q _{.50}	7.96	3.12	7.19	1.70	7.30	1.79	7.27	1.65
	q _{.75}	5.31	3.53	5.52	3.54	5.44	3.60	5.13	3.67
	q _{.95}	4.58	3.76	4.22	3.66	4.16	3.68	3.77	3.41
	var	11.25	14.48	10.36	13.93	11.08	14.62	10.72	14.38
N = 1000 r = 0.2	q _{.05}	12.62	11.36	16.01	13.18	16.36	13.27	16.26	13.54
	q _{.25}	5.64	2.71	7.59	3.86	7.97	4.00	7.95	4.25
	q _{.50}	6.10	2.46	5.18	0.90	5.28	0.89	5.43	1.12
	q _{.75}	3.87	1.71	3.97	2.09	3.86	2.14	3.87	2.17
	q _{.95}	2.89	2.14	2.80	1.72	2.90	1.69	2.77	1.57
	var	4.94	5.64	5.49	6.76	5.47	6.85	6.04	7.34

Table 2: Root mean square errors of the posterior means of quantiles and variance from their true population value (as a percentage of the true population quantity) for the Gamma distribution with shape parameter 2 and rate parameter 2.

		BB		DPMN.5		DPMN.10		DPMN.20	
		A	B	A	B	A	B	A	B
N = 250 r = 0.1	q _{.05}	28.26	24.90	10.25	8.83	12.37	9.93	14.45	12.06
	q _{.25}	9.38	5.75	7.24	4.64	8.08	4.74	8.53	5.48
	q _{.50}	6.18	3.70	5.85	1.95	6.49	2.09	6.61	2.51
	q _{.75}	6.31	4.39	7.41	4.23	7.31	4.77	6.88	5.10
	q _{.95}	9.87	7.28	8.16	5.76	7.91	4.99	8.15	5.15
	var	24.38	22.70	33.12	32.63	27.34	25.68	23.89	19.77
N = 250 r = 0.2	q _{.05}	20.08	19.18	12.33	8.68	12.81	9.42	15.37	11.60
	q _{.25}	6.55	3.49	7.26	1.82	6.97	1.76	7.37	2.18
	q _{.50}	6.99	2.01	6.75	0.80	6.47	0.79	6.75	0.80
	q _{.75}	8.36	3.59	6.93	2.42	6.79	2.57	6.72	2.83
	q _{.95}	3.81	3.93	4.88	3.49	4.58	3.14	4.41	2.56
	var	12.78	10.99	60.56	26.02	10.25	8.42	11.70	10.46
N = 500 r = 0.1	q _{.05}	13.19	10.91	7.59	6.65	6.23	5.23	6.71	5.21
	q _{.25}	5.83	3.58	5.80	3.86	6.21	4.32	7.20	5.15
	q _{.50}	4.88	4.10	4.80	3.01	4.08	2.53	3.89	2.38
	q _{.75}	4.69	2.85	3.95	0.89	3.59	1.01	3.58	1.20
	q _{.95}	7.89	6.68	5.49	3.94	5.15	3.76	5.44	4.22
	var	16.93	15.03	19.03	17.10	16.17	14.31	16.64	14.76
N = 500 r = 0.2	q _{.05}	7.30	6.06	13.31	13.02	12.17	12.09	10.84	11.05
	q _{.25}	5.44	2.76	4.91	2.50	4.99	2.43	4.98	2.34
	q _{.50}	5.19	2.98	4.41	2.23	4.76	2.29	4.90	2.34
	q _{.75}	3.02	2.08	2.92	1.78	3.15	1.67	3.19	1.52
	q _{.95}	3.81	4.80	2.26	3.22	2.48	3.39	2.23	3.25
	var	8.64	9.98	9.03	10.25	8.87	10.01	8.23	9.23
N = 1000 r = 0.1	q _{.05}	6.93	4.85	9.89	8.06	9.55	7.43	8.50	6.23
	q _{.25}	7.44	3.87	6.57	2.60	6.98	2.92	7.35	3.29
	q _{.50}	5.52	1.94	5.08	0.94	5.25	1.09	5.41	1.19
	q _{.75}	4.70	2.06	4.00	0.74	4.07	0.70	4.09	0.63
	q _{.95}	4.41	4.49	4.37	4.04	4.06	3.60	3.92	3.14
	var	7.62	10.04	6.98	9.59	6.54	9.39	6.61	9.26
N = 1000 r = 0.2	q _{.05}	7.81	5.19	8.43	6.47	7.96	6.13	7.75	5.56
	q _{.25}	6.72	3.33	5.66	2.38	5.39	2.18	5.70	2.20
	q _{.50}	5.03	1.86	4.73	1.65	4.55	1.68	4.94	1.77
	q _{.75}	2.86	1.72	2.89	1.20	2.78	1.05	2.94	1.04
	q _{.95}	2.80	3.77	2.03	2.87	1.95	2.76	1.97	2.77
	var	5.98	8.17	4.99	7.61	4.73	7.07	4.79	7.29

Table 3: Coverage and average width (AW) of 95% credible intervals of estimates of the quantiles and variance of the size population for the Gamma(2,2) population with $N = 250$ and $r = .1$

		BB		DPMN.5		DPMN.10		DPMN.20	
		A	B	A	B	A	B	A	B
$q_{.05}$	coverage	0.56	0.82	1.00	0.98	1.00	1.00	1.00	0.98
	AW	0.41	0.45	0.45	0.34	0.45	0.33	0.43	0.31
$q_{.25}$	coverage	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.98
	AW	0.68	0.53	0.51	0.27	0.48	0.25	0.46	0.24
$q_{.50}$	coverage	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	AW	0.83	0.53	0.57	0.18	0.53	0.17	0.51	0.17
$q_{.75}$	coverage	1.00	1.00	1.00	0.94	0.98	0.96	0.98	0.82
	AW	1.02	0.67	0.68	0.21	0.63	0.21	0.61	0.21
$q_{.95}$	coverage	0.92	0.98	1.00	0.98	1.00	0.98	1.00	0.94
	AW	1.48	1.37	1.13	0.84	1.02	0.80	0.94	0.74
var	coverage	0.98	0.98	1.00	0.98	0.98	0.98	1.00	1.00
	AW	0.55	0.55	0.59	0.64	0.54	0.55	0.47	0.48