Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/020209

An approximation method to calculate complicated expectations

Mohammad Moradi[1] *and Jennifer Brown*[2]

[1]Department of Statistics, College of Sciences, Razi University, Kermanshah, Iran
[2]Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

Received: 10 Mar. 2013, Revised: 18 Mar. 2013, Accepted: 24 Mar. 2013
Published online: 1 Jul. 2013

Abstract: In some adaptive sampling designs, expectations need to be calculated as sampling progresses. For situations when these calculations need to be done in the field where there is no access to computers, or for very complicated and computer intensive calculations, we introduce an approximation method. The method to approximate the expectation is based on conditioning on small set of points. An application of the approximation method is described for a sample design with a complicated estimator.

Keywords: Conditional expectation, Continues variable, Discrete variable, Permutation, Rao-Blackwell estimator

1 Introduction

In most statistical applications the true underlying population is assumed unknown. A sample is taken and inference used to estimate the population model. In situations where the population's model is known the parameters of the population are considered known and further statistical methods are not required. For example, when the probability function of a discrete variable is known the expectation of any function can be calculated easily. However, sometimes, in practice, when there are a large number of data points and the expectation form cannot be simplified or calculated easily, the exact value of the expectation cannot be found. In such cases, Monte Carlo simulation methods can be used to estimate the expectations. These methods typically require considerable computational work.

In this paper we introduce an approximation method that does not require a lot of computational work for complicated expectations. This method can be useful in field surveys where the researcher needs to calculate an expectation, for example, for decisions to be made to terminate sampling. In adaptive sampling [7] the process of sampling is dependent on the sequence of observed sample values. In sequential sampling, for example, selection of sample units continue until the variance of the estimator is smaller than a predetermined value. In general, in adaptive sampling, expectations are calculated when a new unit is selected. Calculation methods for these expectations should be fast and easy for field-based sampling e.g., in ecological studies, environmental studies, geographic studies, and biological studies.

The approximation method is introduced for discrete variables, but it can also be used for a complicated expectation of a continuous variable. With continuous variables an expectation is calculated from an integral which can be approximated by numerical integration techniques like Riemann Integration. Usually in numerical integration algorithms, the integral value is approximated by calculating a summation over a function in a finite set of points. While summation over a finite set of points may be simpler than calculating a complicated integral, the summation itself may need to be simplified by other approximation methods.

In this paper an approximation method is used to calculate expectations by conditioning on a small number of points. We use the example of a complicated sampling design with unequal selection probabilities without replacement and a Rao-Blackwell estimator. Rao-Blackwellization is a powerful method to improve the efficiency of trivial estimators. This estimator is given by calculating expectation of a trivial estimator and it is used frequently in different fields ([2], [3], and [5]).

* Corresponding author e-mail: moradi m@razi.ac.ir.

2 Select a subset from all values of a discrete variable

Let X be a discrete variable with values $x_1, x_2, ..., x_N$ and probability function $p(x)$. The expectation of each function on X as $f(X)$ is defined as following,

$$E(f(X)) = \sum_{i=1}^{N} f(x_i)p(x_i) \tag{1}$$

For a large N and a complicated $p(x)$, calculating $E(f(X))$ is difficult or computationally expensive.

A Monte Carlo simulation method could be used. Firstly, the random set of numbers $x_1, x_2, ..., x_n$ is generated from the distribution $p(x)$. Then the sample mean of generated $f(x)$ values defined as following is calculated.

$$\bar{f}(x) = \sum_{i=1}^{n} f(x_i)/n, \tag{2}$$

Using the law of large numbers, if n is large, then $\bar{f}(x)$ converges to $E(f(X))$.
Although the Monte Carlo simulation method has some desirable properties and calculating the estimator generally is not difficult, generating the random numbers from the distribution $p(x)$ is not straightforward. Generating such random numbers, especially for a complicated form of $p(x)$, requires sophisticated methods, e.g., as Markov Chain Monte Carlo (MCMC), and access to computers for numerical processing.

In our approximation method we do not need to generate random numbers, and instead, only some numbers are selected by the researcher from the range of x values. We should note that the numbers can be selected randomly or non-randomly. The method is described as follows.

In order to approximate $E(f(X))$, firstly a few points are selected from all N points such that the distribution shape of selected points is as similar as possible to the distribution shape of all N points as judged by symmetry, skewness and kurtosis of both histograms, or bar plots. For example, when $X \sim B(N,p)$ the representative set can be a systematic set selected from $0,1,...,N$, where the first value can be selected non-randomly. In order to select a systematic set of size n from $0,1,...,N$, firstly the range of points are divided into n equal intervals. Next, the first point is selected from the first interval $0,1,...,(N+1)/n$. Each of remainder $n-1$ points can be selected systematically by adding n to the previous selected point, sequentially. Selecting the mid interval point as the first selected point helps to select a set with most similar distribution shape to the all N points one.

When sorting X values is more difficult than sorting $p(x)$ values, we can select the representative set from the list of X values ordered corresponding to the ascending, or descending, $p(x)$ values. This can be used for example, in sorting multivariate X values where sorting the related single $p(X)$ numbers will be easier.

Let $x_1, x_2, ..., x_n$ be selected from $x_1, x_2, ..., x_N$. The approximated expectation of $f(X)$ shown by $AE(f(X))$ is obtained by calculating the expectation of $f(X)$ conditioned on the selected points, as following:

$$AE(f(X)) = \frac{\sum_{i=1}^{n} f(x_i)p(x_i)}{\sum_{i=1}^{n} p(x_i)} \tag{3}$$

Example: Assume $X \sim B(20,p)$, it is clear $E(X) = np = 20p$. To examine the precision of the approximation method, we calculate the difference of the approximated expectations with the exact expectations $E(X)$ and $E(Log(X+1))$, respectively, for $p = .1,\ .3,\ .5$ and set points $\{3,10,17\}$ and $\{1,4,7,10,13,16,19\}$. The difference and relative difference of approximated expectations and exact expectations of X and $Log(X+1)$ are given as following:

$$dif(X) = AE(X) - E(X) = \sum_{i=1}^{ns} x_i P(x_i|x_1,...,x_{ns}) - np = \frac{\sum_{i=1}^{ns} x_i P(x_i)}{\sum_{i=1}^{ns} P(x_i)} - np, \tag{4}$$

$$dif(Log(X+1)) = AE(Log(X+1)) - E(Log(X+1)) = \sum_{i=1}^{ns} Log(x_i+1)P(x_i|x_1,...,x_{ns}) - \sum_{x=0}^{n} Log(x+1)P(x)$$

$$= \frac{\sum_{i=1}^{ns} Log(x_i+1)P(x_i)}{\sum_{i=1}^{ns} P(x_i)} - \sum_{x=0}^{n} Log(x+1)P(x). \tag{5}$$

Relative difference is defined as $rdif(f(X)) = dif(f(X))/E(f(X))$. The results of relative difference for all combinations of $ns = 3,\ 7$ and $p = .1,\ .3,\ .5$ are summarized in Table (1).

In this small example, the approximated expectations for sample set size, 7, were very close to the exact approximations, and relative differences in expectations were very small with larger p values. The relative differences in the skewed distribution were larger than non-skewed distribution.

3 Applying approximated expectation in a complicated survey sampling

Here we illustrate the approximation method for a more complicated sampling design where an auxiliary variable has been used. The auxiliary variable is correlated with the response variable, an approach used to increase the efficiency of an estimator for a given sample size. Auxiliary variables can be used to improve both the sampling design and the estimator. Stratified sampling is an example of the use of an auxiliary variable in the sample design when it is used to partition the population. Sampling with unequal selection probabilities which are proportional to an auxiliary variable, is another example of the use of an auxiliary variables to improve the sampling design so that the efficiency of the estimator is increased. Ratio estimators, regression estimators, the Hansen-Hurwitz estimator and the Horvitz-Thompson estimator are examples of estimators which use auxiliary variables in their formula [6].
In general, if a correlated variable with the response variable is available, sampling with unequal selection probabilities is preferable to equal selection probabilities. Further, sampling without replacements generally ensures a more precise estimator than sampling with replacement.
With unequal probability sampling, one drawback is that calculating an efficient estimator is computationally difficult and time consuming. The Rao-Blackwell estimator has a complicated calculation, for instance. As an alternative method, we illustrate an approximate expectation method for the the Rao-Blackwell estimator, for sampling with unequal selection probabilities.

Assume a sample of size m is selected from a finite population $U = \{1,2,..,N\}$ corresponding to unequal selection probabilities $p_1, p_2, ..., p_N$ and without replacement. Let $S_o = (i_1, ..., i_m)$ be the selected units which the order units in S_o is corresponding to the order of selection. In the selected units set $S = \{i_1, ..., i_m\}$, the order is not important.

In finite populations the ordered sample set is sufficient and the set of distinct sample units is minimally sufficient for each sampling design [1].

Some estimators like Raj's estimator [4] are functions defined on the ordered sample. The efficiency of such estimators can be increased by calculating the Rao-Blackwell estimator. Let the estimator $\hat{\theta}$ be a function of the ordered set $S_o = (i_1, ..., i_m)$, and the set of unordered units be $S = \{i_1, ..., i_m\}$. For simplicity, we will eliminate subscript i from S_o and S. If $p_1, p_2, ..., p_m$ are the probabilities of selecting units $1, 2, ..., m$, respectively, the probabilities of obtaining $p(S_o)$ and $p(S)$ are given as:

$$p(S_o) = p_1 \tfrac{p_2}{1-p_1} \cdots \tfrac{p_m}{1-p_1-\ldots-p_{m-1}},$$

$$p(S) = \sum_{g=1}^{m!} p_{1_g} \frac{p_{2_g}}{1-p_{1_g}} \cdots \frac{p_{m_g}}{1-p_{1_g}-\ldots-p_{m-1_g}}. \tag{6}$$

Where subscript g determines the number of each permutation. The Rao-Blackwell estimator of $\hat{\theta}$ is shown by $\hat{\theta}_R$ and is calculated as:

$$\hat{\theta}_R = E(\hat{\theta}|S) = \frac{1}{p(S)} \sum_{g=1}^{m!} \hat{\theta}_g p_{1_g} \frac{p_{2_g}}{1-p_{1_g}} \cdots \frac{p_{m_g}}{1-p_{1_g}-\ldots-p_{m-1_g}} \tag{7}$$

Variance and estimate of variance for $\hat{\theta}$ are given as:
$V(\hat{\theta}_R) = V(\hat{\theta}) - E(V(\hat{\theta}|S))$
$\hat{V}(\hat{\theta}_R) = \hat{V}(\hat{\theta}) - V(\hat{\theta}|S)$

$$= \hat{V}(\hat{\theta}) - \frac{1}{p(S)} \sum_{g=1}^{m!} (\hat{\theta}_g - \hat{\theta}_R)^2 p_{1_g} \frac{p_{2_g}}{1-p_{1_g}} \cdots \frac{p_{m_g}}{1-p_{1_g}-\ldots-p_{m-1_g}} \tag{8}$$

Table 1: Relative difference

ns	p	$rdif(X)$	$rdif(Log(X+1))$
3	0.1	0.5	0.4
3	0.3	-0.15	-0.11
3	0.5	0	0.01
7	0.1	-0.11	-0.06
7	0.3	-0.00	0.00
7	0.5	0	0.00

It is clear that calculating $E(\hat{\theta}|S)$ for $m > 4$ is too difficult without a computer on hand, and for $m > 180$ it is near impossible.
In order to calculate $AE(\hat{\theta}|S)$, firstly, a small number of all $m!$ possible permutations are selected. Then $AE(\hat{\theta}|S)$ and $\hat{AV}(\hat{\theta}_R)$ are given by conditioning the expectations on the selected permutations. Let a subset of size $ns < m!$ be selected from all $m!$ possible permutations. Then $AE(\hat{\theta}|S)$ and $\hat{AV}(\hat{\theta}_R)$ are given as:

$$AE(\hat{\theta}|S) = \frac{1}{p(S^\star)} \sum_{g=1}^{ns} \hat{\theta}_g p_{1_g} \frac{p_{2_g}}{1-p_{1_g}} \dots \frac{p_{m_g}}{1-p_{1_g}-\dots-p_{m-1_g}}, \tag{9}$$

and

$$\hat{AV}(\hat{\theta}_R) = \hat{V}(\hat{\theta}) - AV(\hat{\theta}|S)$$

$$= \hat{V}(\hat{\theta}) - \frac{1}{p(S^\star)} \sum_{g=1}^{ns} (\hat{\theta}_g - \hat{\theta}_R)^2 p_{1_g} \frac{p_{2_g}}{1-p_{1_g}} \dots \frac{p_{m_g}}{1-p_{1_g}-\dots-p_{m-1_g}} \tag{10}$$

where $p(S^\star) = \sum_{g=1}^{ns} p(S_{o_g})$.

The permutations are vectors with m elements and sorting all of them will be more difficult than sorting the x values in a binomial distribution. As an alternative method a set of permutations can be selected such that their probability values cover the range of possible $p(S_o)$ values. Although sorting all $p(S_o)$ values and selecting a systematic set from them is not straightforward we can arrange some of the permutations such that their probability values are distinct and cover the range of $p(S_o)$ values homogenously. Before introducing this method, the following Theorem is proved.

Theorem 1 *If in a given permutation, say*
$S_{og} = (1_g, \dots, k-1_g, k_g, k+1_g, \dots, l-1_g, l_g, l+1_g, \dots, m_g)$, *two elements like* k_g *and* l_g *be replaced together and the new permutation*
$S_{og'} = (1_g, \dots, k-1_g, l_g, k+1_g, \dots, l-1_g, k_g, l+1_g, \dots, m_g)$ *be formed, the following results are given:*

$$p_{k_g} < p_{l_g} \Leftrightarrow P(S_{og}) < P(S_{og'})$$

$$p_{k_g} \geq p_{l_g} \Leftrightarrow P(S_{og}) \geq P(S_{og'}) \tag{11}$$

Proof. The first inequality is proved, and the second one can be proved similarly. Let $p_k < p_l$, the subscript g is deleted for simplicity, probability $P(S_{og})$ can be written as:

$$P(S_{og}) = p_1 \frac{p_2}{1-p_1} \dots \frac{p_m}{1-p_1-\dots-p_{m-1}} = \frac{\prod_{i=1}^{m} p_i}{\prod_{a=1}^{m-1}(1-\sum_{j=1}^{a} p_j)}$$
$$= \frac{\prod_{i=1}^{m} p_i}{\prod_{a=1}^{k-1}(1-\sum_{j=1}^{a} p_j)\prod_{a=k}^{l-1}(1-\sum_{j=1}^{a} p_j)\prod_{a=l}^{m-1}(1-\sum_{j=1}^{a} p_j)}$$

$$= \frac{A}{B\prod_{a=k}^{l-1}(1-\sum_{j=1}^{a} p_j)C} \tag{12}$$

Also, $P(S_{og'})$ can be written as:

$$P(S_{og'}) = \frac{\prod_{i=1}^{m} p_i}{\prod_{a=1}^{k-1}(1-\sum_{j=1}^{a} p_j)\prod_{a=l,k+1,\dots,l-1}(1-\sum_{j=1}^{a} p_j)\prod_{a=k,l+1,\dots,m-1}(1-\sum_{j=1}^{a} p_j)}$$

$$= \frac{A}{B\prod_{a=l,k+1,\dots,l-1}(1-\sum_{j=1}^{a} p_j)C} \tag{13}$$

It is clear that A, B and C in $P(S_{og})$ are equal to their corresponding terms in $P(S_{og'})$. Now, in order to compare $P(S_{og})$ with $P(S_{og'})$, we just need to compare $\prod_{a=k}^{l-1}(1-\sum_{j=1}^{a} p_j)$ and $\prod_{a=l,k+1,\dots,l-1}(1-\sum_{j=1}^{a} p_j)$. For a given a in the first production, we have:

$$1-\sum_{j=1}^{a} p_j = 1-\sum_{k\neq j=1}^{a} p_j - p_k \tag{14}$$

And for the same a in the second one:

$$1-\sum_{j=1}^{a} p_j = 1-\sum_{l\neq j=1}^{a} p_j - p_l \tag{15}$$

We know $p_k < p_l$ then for a given a the value of $1-\sum_{j=1}^{a} p_j$ in the $P(S_{og})$ is greater than similar one in $P(S_{og'})$, therefore $\prod_{a=k}^{l-1}(1-\sum_{j=1}^{a} p_j) > \prod_{a=l,k+1,\ldots,l-1}(1-\sum_{j=1}^{a} p_j)$ and finally we conclude $P(S_{og}) < P(S_{og'})$.

Let the ordered p values in the sample be as follows:

$$p_{(1)} < p_{(2)} < \ldots < p_{(m)} \tag{16}$$

Using Theorem (1), the smallest and largest values of $p(S_o)$ are given as:

$$p_{Min}(S_o) = p_{(1)}\frac{p_{(2)}}{1-p_{(1)}}\cdots\frac{p_{(m)}}{1-p_{(1)}-\cdots-,p_{(m-1)}},$$

$$p_{Max}(S_o) = p_{(m)}\frac{p_{(m-1)}}{1-p_{(m)}}\cdots\frac{p_{(1)}}{1-p_{(m)}-\ldots-,p_{(2)}}, \tag{17}$$

which are correspond to the ascending arrangement $(1),(2),\ldots(m)$ and the descending arrangement $(m),(m-1),\ldots,(1)$, respectively. An arrangement of some permutations which covers the interval $(p_{Min}(S_o), p_{Max}(S_o))$ homogenously, can be introduced as follows:

$$\begin{array}{c} Ar_1 = (1),(2),(3),\ldots,(m-2),(m-1),(m) \\ Ar_2 = (m),(2),(3),\ldots,(m-2),(m-1),(1) \\ Ar_3 = (m),(m-1),(3),\ldots,(m-2),(2),(1) \\ Ar_4 = (m),(m-1),(m-2),\ldots,(3),(2),(1) \\ \vdots \\ Ar_{ns} = (m),(m-1),(m-2),\ldots,(3),(2),(1) \end{array}$$

where, in this method, $ns = [m/2]+1$, and $P_{Min}(S_o) = P(Ar_1) < P(Ar_2) < \ldots < P(Ar_{ns}) = P_{Max}(S_o)$.
Another method yielding a larger ns is carried out as follows, such that in each arrangement two neighborhood elements of the previous arrangement are replaced together.

$$\begin{array}{c} Ar_1 = (1),(2),(3),\ldots,(m-2),(m-1),(m) \\ Ar_2 = (2),(1),(3),\ldots,(m-2),(m-1),(m) \\ Ar_3 = (2),(3),(1),\ldots,(m-2),(m-1),(m) \\ \vdots \\ Ar_m = (2),(3),(4),\ldots,(m-1),(m),(1) \\ Ar_{m+1} = (3),(2),(4),\ldots,(m-1),(m),(1) \\ Ar_{m+2} = (3),(4),(2),\ldots,(m-1),(m),(1) \\ \vdots \\ Ar_{2m-2} = (3),(4),(5),\ldots,(m),(2),(1) \\ Ar_{2m-1} = (4),(3),(5),\ldots,(m),(2),(1) \\ \vdots \\ Ar_{ns} = (m),(m-1),(m-2),\ldots,(3),(2),(1) \end{array}$$

where $ns = m(m-1)/2+1$, and $P_{Min}(S_o) = P(Ar_1) < P(Ar_2) < \ldots < P(Ar_{ns}) = P_{Max}(S_o)$.
Other arrangements can be constructed similarly to get different values of ns. The smallest arrangement set consisting of two or three permutations can be constructed as follows:

$$Ar_1 = (1),(2),(3),\ldots,(m-2),(m-1),(m) \tag{18}$$

$$Ar_2 = (m),(m-1),(m-2),\ldots,(3),(2),(1) \tag{19}$$

If the observed S_o is different from the two first arrangements, then the AE is calculated based on three points. Such an arrangement set can be useful in adaptive sampling designs which the value of Rao-Blackwell estimator is used to decide on when to terminate sequential sample selection. The final Rao-Blackwell estimator can be calculated based on a larger arrangement set.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] D. Basu, Role of the sufficiency and likelihood principles in sample survey theory, *Sankhyā* **A 31**, 441-451 (1969)
[2] C. T. Chao, A. L. Dryver, and T. C. Chiang, Leveraging the Rao-Blackwell theorem to improve ratio estimators in adaptive cluster sampling, *Environ Econ Stat* **18**,543-568 (2011).
[3] K. H. Low, G. J. Gordon, J. M. Dolan, and P. Khosla, Adaptive sampling for multi-robot wide area exploration, *Proc. ICRA* 755-760 (2007).
[4] D. Raj, On the method of the overlapping maps in a sample surveys, *Sankhyā* **17**, 89-98 (1957).
[5] M. M. Salehi, Rao-Blackwell version of the Hurvitz-Thompson and Hansen-Hurwitz in adaptive cluster sampling, *Environmental and Ecological Statistics*, **6**, 185-195 (1999).
[6] C. E. Särndal, B. Swensson, and J. Wretman, , *Model assisted survey sampling*, New York; Springer-Verlag 1992.
[7] S. K. Thompson, G. A. F. and Seber, . *Adaptive Sampling*, New York, Wiley, 1996.

Mohammad Moradi received the PhD degree in Statistics at Isfahan University of Technology. His research interests are in the areas of applied statistics like Sampling, Multivariate Analysis and Simulation. He has published research articles in some international journals of statistics and environmental sciences. He is referee of statistical journals.

Jennifer Brown is Professor of Statistics at University of Canterbury. Her research interests are in environmental statistics, survey sampling and data modelling. She is currently Head of the Department of Mathematics and Statistics and immediate past president of the New Zealand Statistical Association.