Wojciech Gamrot
Uniwersytet Ekonomiczny w Katowicach

ON A CLASS OF ESTIMATORS FOR A RECIPROCAL OF BERNOULLI PARAMETER

Introduction

The estimation of a reciprocal for the probability of an event is an issue of broad practical interest. Such a problem arises e.g. in empirical Horvitz--Thompson estimation for complex sampling designs considered by Fattorini [2006], Thompson and Wu [2008] and Fattorini [2009]. When sampling weights, defined as reciprocals of first order inclusion probabilities are too complex to compute exactly they are instead replaced with estimates evaluated in a simulation study. For this purpose, it is desired that the estimator of inverse probability always takes finite values, so that the Horvitz-Thompson statistic is also always finite. This effect may be achieved in many ways including the use of restricted maximum likelihood principle, the truncation of the binomial distribution considered by Stephan [1946], Rempała and Szekely [1998] and Marciniak and Wesołowski [1999] or bayesian estimation considered e.g. by Berry [1989], Marchand and MacGibbon [2000] and Marchand et al. [2005]. Another approach to this problem, proposed by Fattorini [2006] utilizes the reciprocal of the sampling fraction with a certain constant adjustment that prevents the denominator from reaching zero. In this paper, the original estimator of Fattorini is generalized by admitting varying values of the adjustment constant. This leads to a broader class of estimators. Expressions for the bias for some of these estimators are provided. The problem of setting adjustment constant is then considered.

1. Estimator of Fattorini

Let X denote the number of successes in n Bernoulli trials with success probability equal to p an let q = 1-p. Hence X has a binomial distribution with expectation E(X) = np and variance V(X) = np(1-p). Fattorini [2006] considers a consistent estimator of the probability p in the form:

$$\hat{p}_1 = \frac{X+1}{n+1}.$$

This leads to the estimator of the reciprocal:

$$\hat{d}_1 = \frac{n+1}{X+1}.$$

Fattorini [2006] derived exact formula for its bias:

$$B(\hat{d}_1) = \frac{-q^{n+1}}{p} \tag{1}$$

and by developing this estimator into factorial series he also gave an upper bound for its variance:

$$V(\hat{d}_1) \le \frac{5}{(n+2)p^3}.$$

The bias of the estimator $\hat{d}_1$ may be intuitively viewed as a superposition of biases resulting from two different sources. First, introduction of adjustment constant into sampling fraction $\hat{p}_1$ to make it strictly positive makes it also positively biased. When $\hat{p}_1$ resides in a denominator of $\hat{d}_1$, this bias for the whole reciprocal becomes negative. On the other hand, from Jensen's inequality we have E(1/X) > 1/E(X) which means that taking inverse of any random variable is by itself a transformation introducing positive bias. Hence, it is reasonable that at least to some extent these two components cancel each other. Meanwhile, the formula (1) is in general negative for $p\in(0,1)$, which means that they do not reduce to zero. This also suggests, that by choosing an adjustment constant different from unity one might perhaps reduce the bias totally. In the sequel such a possibility is investigated.

2. A more general estimator

We will now consider a more general statistic in the form:

$$\hat{p}_c = \frac{X+c}{n+c},$$

where c is some non-negative constant. In general, it is not required for c to be an integer, and hence we consider a class of estimators for $c\in\langle 0,+\infty\rangle$. It is easy to show that:

$$E(\hat{p}_c) = \frac{np+c}{n+c}.$$

And hence its bias is:

$$B(\hat{p}_c) = E\left(\frac{X+c}{n+c}\right) - p = \frac{cq}{n+c}.$$

The bias is positive and tends to zero when $n\to\infty$. Meanwhile, the variance is given by:

$$V(\hat{p}_c) = \frac{npq}{(n+c)^2},$$

which means that $\hat{p}_c$ is consistent. This lets us construct a consistent estimator for d in the form:

$$\hat{d}_c = \frac{n+c}{X+c}.$$

OBVIOUSLY, the statistic of Fattorini is a special case of the above estimator for c = 1. The statistic $\hat{p}_c$ takes discrete values from the interval [c / (n + c),1] and consequently, $\hat{d}_c$ takes values from [1,(n + c) / c]. For integer c > 0 and 0 < p < 1 we have (see Appendix 1 for a proof):

$$E\left(\frac{1}{X+c}\right)=\frac{1}{p^c}\frac{n!}{(n+c)!}\left(E_{n+c}\left(\prod_{r=1}^{c-1}(X-r)\right)+(-1)^c(c-1)!q^{n+c}\right), \qquad (2)$$

where the symbol $E_{n+c}(\ )$ represents expectation computed with respect to the binomial distribution B(n+c,p), as opposed to the symbol E( ) representing expectation calculated with respect to the B(n,p) binomial distribution. This notation assumes that the multiplication of elements in the empty set for c = 1 equals 1. Hence, for c = 1,2,3,4,... the expectation (2) depends on:

$E_{n+c}(1) = 1$

$E_{n+c}(X-1) = E_{n+c}(X) - 1$

$E_{n+c}((X-1)(X-2)) = E_{n+c}(X^2) - 3E_{n+c}(X) + 2$

$E_{n+c}((X-1)(X-2)(X-3)) = E_{n+c}(X^3) - 6E_{n+c}(X^2) + 11E_{n+c}(X) - 6$

.....

and hence on raw moments for the B(n + c,p) distribution which are easy to calculate from the moment generating function for B(n,p) that takes the well-known form:

$$M(t)=(1-p+pe^t)^n.$$

Hence raw moments of the order r = 1,2,3,... for B(n,p) are computed as:

$$E(X^r)=\left.\frac{\partial^r M(t)}{\partial t^r}\right|_{t=0}.$$

This leads to:

$$E(X)=np$$

$$E(X^2)= np\,(q+np)$$

$$E(X^3) = np(1-3p+3np+2p^2-3np^2+n^2p^2)$$

and so on. Substituting (n + c) instead of n in above formulas one gets raw moments of the B(n + c,p) distribution. Hence for c = 1,2,3,... the expectations of the estimator $\hat{d}_c$ may be expressed as (see Appendix 2):

$$E(\hat{d}_1) = \frac{1-q^{n+1}}{p} \tag{3}$$

$$E(\hat{d}_2) = \frac{(n+2)p-1+q^{n+2}}{(n+1)p^2} \tag{4}$$

$$E(\hat{d}_3) = \frac{p(n+3)\,(np+2p-2)+2-2q^{n+3}}{(n+1)(n+2)p^3} \tag{5}$$

$$E(\hat{d}_4) = \frac{(n+4)p[6+3p+(n+4)p(np+p-3)+2p^2]-6+6q^{n+4}}{p^4(n+1)(n+2)(n+3)} \tag{6}$$

and so on. Obviously the formula (3) corresponding to $E(\hat{d}_1)$ is equivalent to the result of Fattorini [2006] given by (1). The other three are not. It is easy to verify that for c = 1,2,3,4 we have $\lim_{n\to\infty} E(\hat{d}_c) = d$ which confirms that for c = 1,2,3,4 the estimator $\hat{d}_c$ is asymptotically unbiased for d. Let us use these formulas to calculate the exact bias:

$$B(\hat{d}_c) = E(\hat{d}_c) - d\,,$$

for c = 1,2,3,4, for n = 1,...,2000 and for p = 0.01, 0.05, 0.1, 0.5. The results of calculations are shown on Figure 1.

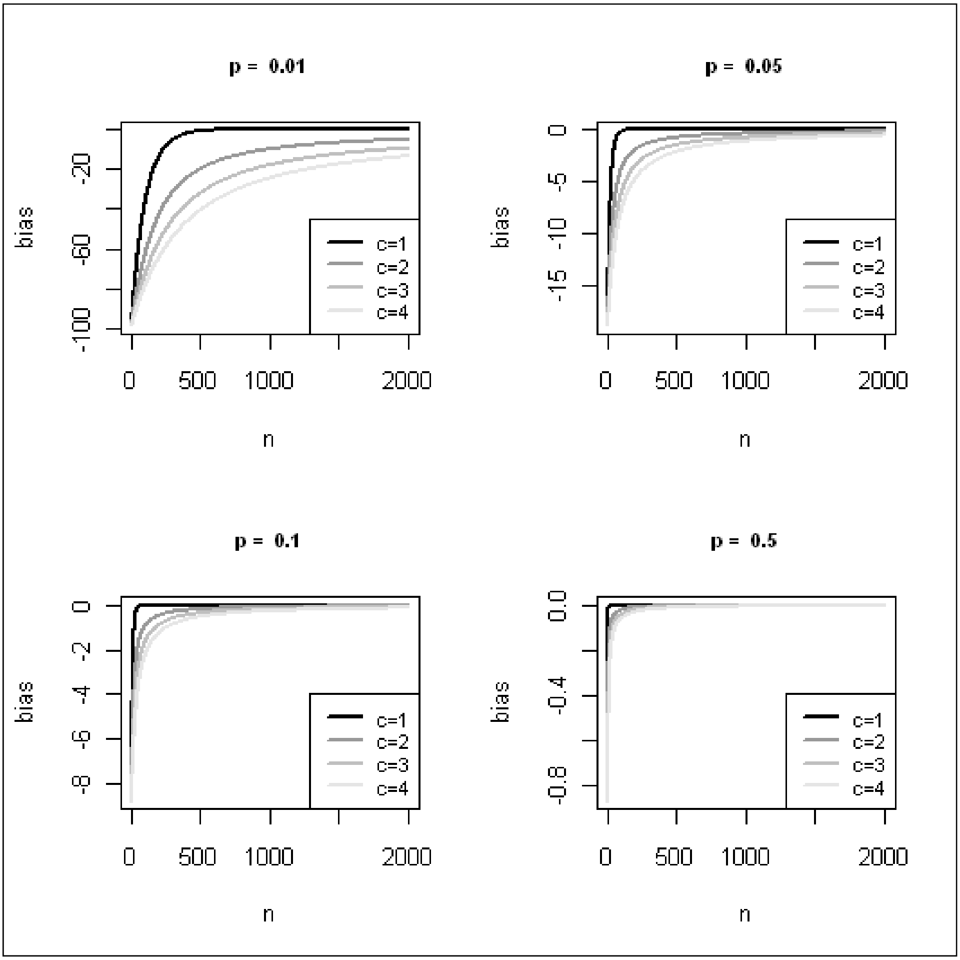

Fig. 1. Exact biases for $\hat{d}_c$, c = 1,2,3,4 and n = 1,...,2000.

For all presented values of c, n and p the bias of $\hat{d}_c$ is negative. Not surprisingly, it tends to zero when sample size grows. The lower the true probability p, the higher absolute values of bias (see the scale on the y-axis). The same tendency was also observed for $p > 0.5$ although is not illustrated with the graph. The absolute value of bias also grows with c. It is lowest for c = 1 and highest for c = 4. Results shown on Figure 1 suggest, that increasing the constant c to take integer values higher than one significantly increases the bias. However, this still does not preclude finding some non-integer values for c (perhaps in the vicinity of c = 1) that would provide lower bias.

3. Properties when c is not an integer

When n is small enough, the behavior of the proposed general estimator may be examined by computing the probabilities of all possible sample counts via probability function of binomial distribution. Stochastic properties of the estimator may then be computed explicitly using their definitions. Let us now present such a 'small-sample' study, carried out for n = 500 and varying values of c and p. The bias of the estimator as a function of c and p is shown on Figure 2. Its MSE as a function of c and p is shown on Figure 3. The Figure 4 presents the share of bias in the MSE as a function of c and p.

The bias of the estimator turns out to not always be negative, as the analysis above would suggest. Indeed, it increases when c decreases and for small values of c it grows above zero. Moreover, the bias strongly depends on p. Its absolute value seems to remain rather stable and close to zero for large p but it tends to increase dramatically when p takes values close to zero. Similar tendency is observed for the mean square error of the estimator. It seems to remain quite stable for large p, but increases very quickly when p closes to zero. The MSE also depends on c and for a constant p there is apparently a value of c that minimizes the MSE.

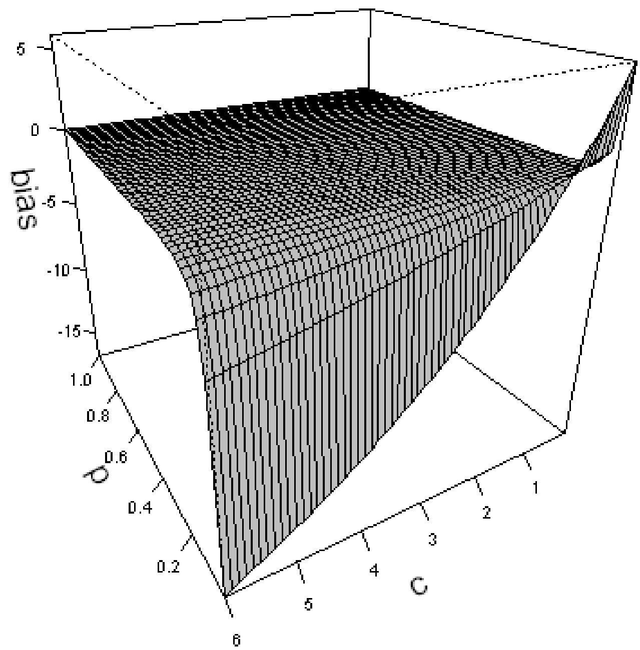

Fig. 2. The bias of the estimator $\hat{d}_c$ as a function of p and c for n = 500

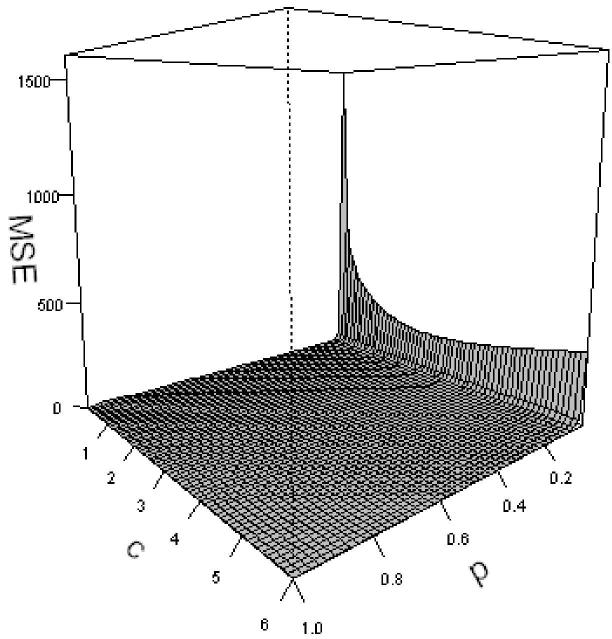

Fig. 3. The MSE of the estimator $\hat{d}_c$ as a function of p and c for n = 500

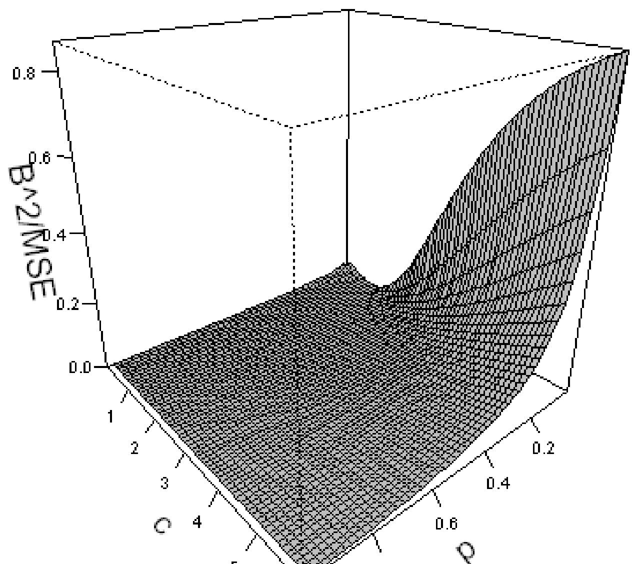

Fig. 4. The share of bias in the MSE as a function of p and c for n = 500

For large p the share of bias in the MSE is very modest, but for smaller p it grows dramatically. If c is also large then the bias dominates the MSE. The dependency of bias share on c resembles that of MSE itself: it seems that for a constant p there exists a value of c for which the share of bias in the MSE is minimized.

For a fixed n, one may be interested in choosing the optimal value of c that nullifies the bias or minimizes the MSE. However the probability p is not controlled by the sampler. Hence it is important to verify if such an optimal value of c depends on p or if it is perhaps constant. In order to do this the dependence of bias and MSE on c for certain values of p is plotted on Figures 5 and 6.

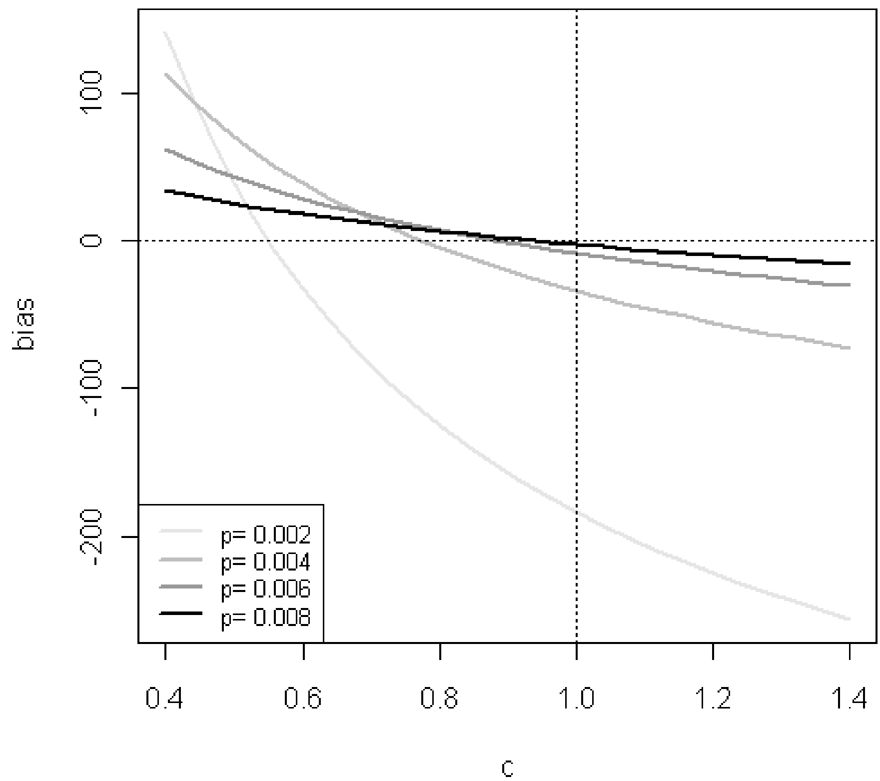

Fig. 5. The bias as a function of c for certain values of p and n = 500

The Figure 5 clearly demonstrates that unfortunately there is no universal value of c that would nullify the bias for any possible p. When p grows, the value of c for which the bias is equal to zero is shifting towards higher values (see

Figure 7), although it remains below one (even for much larger p's such as p = 0.99, since the bias curve very quickly looses steepness when p grows). It seems to be coherent with earlier observation of bias growing with c when c > 1 and sample size is large. This renders any value c above one unjustifiable. These results are also coherent with the observation that for the statistic of Fattorini [2006] based on c = 1 the bias is negative for any possible $p\in(0,1)$. On the other hand, while optimal c-values nullifying the bias are known to lie below one, setting c to some value which is too close to zero may result in strong positive bias.

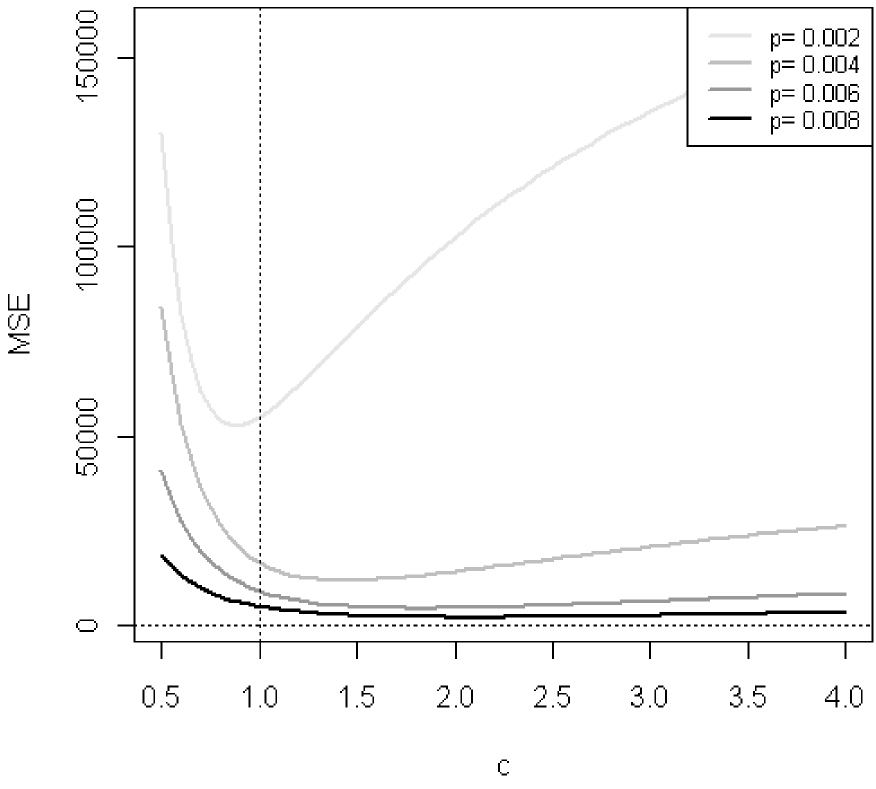

Fig. 6. The bias as a function of c for certain values of p and n = 500

The results for the MSE shown on Figure 6 indicate that unfortunately optimum values of c that minimize the MSE also change with p. They move upwards with growing p but do not seem to stabilize. The curve reflecting dependence of optimal c-values on p shown on Figure 7 is non-decreasing but its shape is somewhat complicated, with the second derivative changing sign at least two times (for $p \approx 0.02$ and $p \approx 0.96$). The value of c = 1 chosen by Fattorini [2006] apparently minimizes MSE for some value of p inside the [0.002, 0.004] interval. Hence, if there is e.g. external knowledge available that implies p > 0.004 then it is reasonable to expect that MSE is minimized for some c > 1, and any c-value lower than one would not be advisable. However again, choosing a too large c-value might also increase the MSE instead of reducing it. On the other hand, one may also notice that although optimal values swing wildly with changing p, the MSE seems to be much more dependent on p itself, than on the choice of constant c.

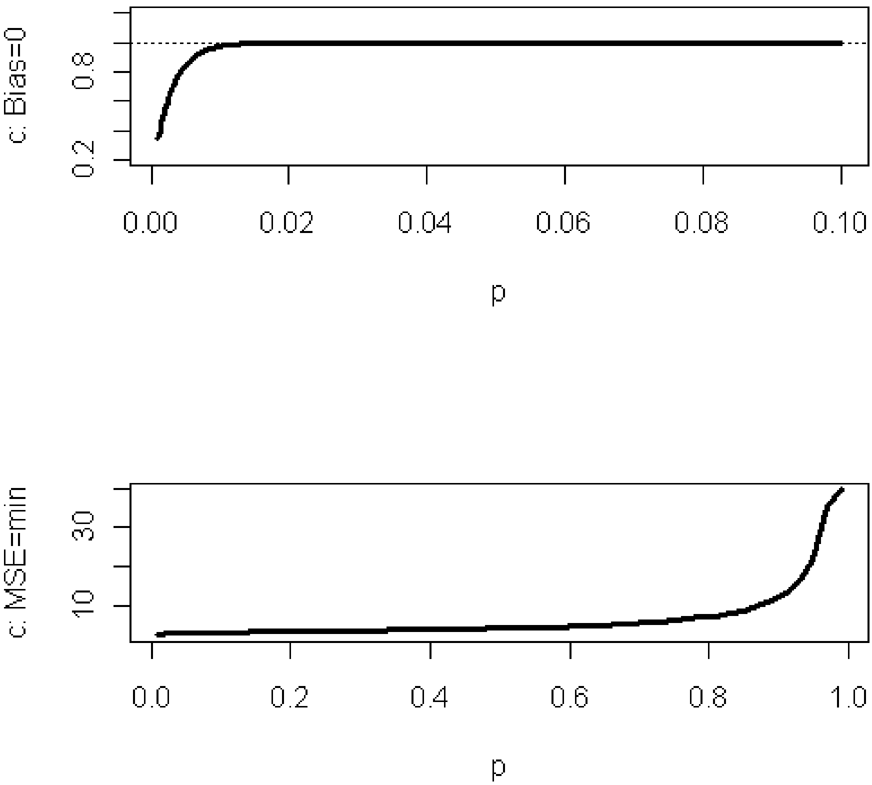

Fig. 7. The c-values nullifying the bias and minimizing the MSE for p = 0.01,0.02,...,0.99 and for n = 500

Conclusions

In this paper a class of estimators for inverse probability indexed by a parameter $c\in(0,+\infty)$ was considered. All estimators in the class are consistent.

They always take finite values, as opposed to the simple reciprocal of the sampling fraction. The class incorporates the well-known Fattorini's [2006] statistic for c = 1. The formulas for bias of these estimators were derived for c = 1,...,4 and a method for computing the bias for c = 5,6,... was suggested.

The bias of estimators depends on sample size n, parameter c and unknown probability p. It turns out that there is no single value of c that would nullify the bias or minimize the mean square error for all possible values of p. In other words, no estimator in the class dominates others in terms of accuracy for a fixed n and all values of p. This also applies to the Fattorini's statistic. However, it seems that values of c greater than one do not nullify bias for any possible p so they should be avoided.

If the exact formula (or upper bound) for MSE or absolute bias when c is not integer and the sample is large were known, then some partial knowledge on p taking e.g. form of inequality constraints might be explored to set the value of c in such a way that MSE or bias for most pessimistic (unfavorable) p is minimized. This justifies further efforts aimed at finding such a formula.

Literature

Berry C.J. (1989): *Bayes Minimax Estimation of a Bernoulli p in a Restricted Parameter Space*. "Communications in Statistics – Theory and Methods", No. 18(12).

Fattorini L. (2006): *Applying the Horvitz-Thompson Criterion in Complex Designs: A Computer-Intensive Perspective for Estimating Inclusion Probabilities*. "Biometrika", No. 93(2).

Fattorini L. (2009): *An Adaptive Algorithm for Estimating Inclusion Probabilities and Performing the Horvitz-Thompson Criterion in Complex Designs*. "Computational Statistics", No. 24.

Marciniak E., Wesołowski J. (1999): *Asymptotic Eulerian Expansions for Binomial and Negative Binomial Reciprocals*. "Proceedings of the American Mathematical Society", No. 127(11).

Marchand E., Perron F., Gueye R. (2005): *Minimax Estimation of a Constrained Binomial Proportion p When |p-1/2| is Small*. "Sankhya", No. 67(3).

Marchand E., MacGibbon B. (2000): *Minimax Estimation of a Constrained Binomial Proportion*. "Statistics & Decisions", No. 18.

Rempała G., Szekely G.J. (1998): *On Estimation with Elementary Symmetric Polynomials*. "Random Operations and Stochastic Equations", No. 6.

Stephan F.F. (1946): *The Expected Value and Variance of the Reciprocal and Other Negative Powers of a Positive Bernoullian Variate*. "Annals of Mathematical Statistics", No. 16.

Thompson M.E., Wu C. (2008): *Simulation-based Randomized Systematic PPS Sampling under Substitution of Units*. "Survey Methodology", No. 34 (1).

APPENDIX 1

Derivation of the formula (2):

$$E\left(\frac{1}{X+c}\right)=\sum_{k=0}^{n}\frac{1}{k+c}\binom{n}{k}p^k q^{n-k}=\sum_{k=0}^{n}\frac{1}{k+c}\frac{n!}{k!(n-k)!}p^k q^{n-k}=$$

$$=\frac{1}{p^c}\frac{n!}{(n+c)!}\sum_{k=0}^{n}\frac{(k+c-1)!}{k!}\binom{n+c}{k+c}p^{k+c}q^{n+c-(k+c)}=$$

$$=\frac{1}{p^c}\frac{n!}{(n+c)!}\sum_{k=c}^{n+c}\frac{(k-1)!}{(k-c)!}\binom{n+c}{k}p^k q^{n+c-k}=$$

$$=\frac{1}{p^c}\frac{n!}{(n+c)!}\sum_{k=c}^{n+c}\left(\prod_{r=1}^{c-1}(k-r)\right)\binom{n+c}{k}p^k q^{n+c-k}=$$

$$=\frac{1}{p^c}\frac{n!}{(n+c)!}\left(E_{n+c}\left(\prod_{r=1}^{c-1}(X-r)\right)-\sum_{k=0}^{c-1}\left(\prod_{r=1}^{c-1}(k-r)\right)\binom{n+c}{k}p^k q^{n+c-k}\right)=$$

$$=\frac{1}{p^c}\frac{n!}{(n+c)!}\left(E_{n+c}\left(\prod_{r=1}^{c-1}(X-r)\right)-(0-1)(0-2)(0-(c-1))\binom{n+c}{0}p^0 q^{n+c}-\right.$$

$$\left.-\sum_{k=1}^{c-1}(k-1)(k-2)(k-(c-1))\binom{n+c}{k}p^k q^{n+c-k}\right)=$$

$$=\frac{1}{p^c}\frac{n!}{(n+c)!}\left(E_{n+c}\left(\prod_{r=1}^{c-1}(X-r)\right)+(-1)^c(c-1)!q^{n+c}\right)$$

APPENDIX 2

Derivation of formulas (3)-(6) representing expectations for $\hat{d}_c$ when $c = 1,2,3,4$.

$$E(\hat{d}_1)=E\left(\frac{n+1}{X+1}\right)=(n+1)\frac{1}{p^1}\frac{n!}{(n+1)!}\left(E_{n+1}(1)+(-1)^1(1-1)!q^{n+1}\right)=\frac{1-q^{n+1}}{p}$$

$$E(\hat{d}_2) = E\left(\frac{n+2}{X+2}\right) = \frac{n+2}{p^2}\frac{n!}{(n+2)!}\left(E_{n+2}(X) - 1 + (-1)^2(2-1)!q^{n+2}\right) =$$
$$= \frac{1}{n+1}\frac{(n+2)p - 1 + q^{n+2}}{p^2}$$

$$E(\hat{d}_3) = E\left(\frac{n+3}{X+3}\right) =$$
$$= (n+3)\frac{1}{p^3}\frac{n!}{(n+3)!}\left(E_{n+3}(X^2) - 3E_{n+3}(X) + 2 + (-1)^3(3-1)!q^{n+3}\right) =$$
$$= \frac{(n+3)p\,(q + (n+3)p) - 3(n+3)p + 2 - 2q^{n+3}}{(n+1)(n+2)p^3} =$$
$$= \frac{p(n+3)\,(np + 2p - 2) + 2 - 2q^{n+3}}{(n+1)(n+2)p^3}$$

For c = 4 let us note that:

$E_{n+4}(X^3)$-6 $E_{n+4}(X^2)$+ 11E_{n+4}(X)-6 =(n+4)p(1-3p+3(n+4)p+2p^2-

-3(n+4)p^2+$(n+4)^2p^2$) -6 (n+4)p (1-p+(n+4)p)+11(n+4) p -6 =

= (n+4)p [1 -3p + (n+4)p [3-3p+(n+4)p-6] +2p^2 -6(1-p) +11] -6 =

= (n+4)p [6 +3p + (n+4)p [np+p-3] +2p^2] -6

and consequently:

$$E\left(\frac{n+4}{X+4}\right) = (n+4)\frac{1}{p^4}\frac{n!}{(n+4)!}\left(E_{n+4}\left(\prod_{r=1}^{4-1}(X-r)\right) + (-1)^4(4-1)!q^{n+4}\right) =$$
$$= \frac{1}{p^4}\frac{E_{n+4}(X^3) - 6E_{n+4}(X^2) + 11E_{n+4}(X) - 6 + 6q^{n+4}}{(n+1)(n+2)(n+3)} =$$
$$= \frac{(n+4)p[6 + 3p + (n+4)p(np + p - 3) + 2p^2] - 6 + 6q^{n+4}}{p^4(n+1)(n+2)(n+3)}$$

O KLASIE ESTYMATORÓW ODWROTNOŚCI PRAWDOPODOBIEŃSTWA

Streszczenie

Powszechnie znany estymator odwrotności prawdopodobieństwa zaproponowany przez Fattoriniego uogólniono poprzez dopuszczenie zmian wartości jednostkowej stałej występującej we wzorze definiującym go. Prowadzi to do konstrukcji klasy estymatorów odwrotności prawdopodobieństwa. Własności tych estymatorów zbadano analitycznie. Zaproponowano metodę wyznaczania asymptotycznego obciążenia dla całkowitych wartości zmodyfikowanej stałej. Własności estymatorów dla małych prób wyznaczono numerycznie, korzystając z własności rozkładu dwumianowego.