Original Research Article

Comparison of the Bootstrap and Delta Method Variances of the Variance Estimator of the Bernoulli Distribution

Abstract

It is interesting to calculate the variance of the variance estimator of the Bernoulli distribution. In this paper, we compare the Bootstrap and Delta Method variances of the variance estimator of the Bernoulli distribution. First, we provide the correct Bootstrap, Delta Method, and true variances of the variance estimator of the Bernoulli distribution for three parameter values in Table 1. Then, we provide the estimates of the variance of the variance estimator of the Bernoulli distribution by the Delta Method (analytically), the true method (analytically), and the Bootstrap Method (algorithmically). Finally, three figures which compare the variance estimates, the errors, and the absolute errors for 101 parameter values in [0, 1] are given to illustrate the comparison of the Bootstrap and Delta Method.

Keywords: bootstrap, delta method, variance estimate, Bernoulli distribution
2010 MSC: 62F10, 62F12, 62F40

1. Introduction

The Bootstrap Method is a resampling technique used to obtain estimates of summary statistics. It has wide applications, e.g., Felsenstein (1985); Meyer et al. (1986); Hillis and Bull (1993); Shao and Tu (1995); Visscher et al. (1996); Briggs et al. (1997); Carpenter and Bithell (2000); Alfaro et al. (2003); Stamatakis et al. (2008); Wang and Hutson (2014); Ju (2015); Shen and Machado (2016). The Delta Method is a result concerning the approximate probability distribution for a function of an asymptotically normal statistical estimator from knowledge of the limiting variance of that estimator (Casella and Berger (2002); Shao (2003)).

Casella and Berger (2002) is a worldwidely used textbook for the courses of Statistical Inference or Advanced Mathematical Statistics for first-year graduate students majoring in statistics or in a field where a statistics concentration is desirable. In Example 10.1.21 of Casella and Berger (2002), they compared the Bootstrap and Delta Method variances of $\hat{p}\,(1-\hat{p})$, which is the variance estimator of the Bernoulli distribution. However, the variances of the Bootstrap method and the true method are wrongly calculated. The right variances are given in Table 1.

The rest of the paper is organized as follows. In the next Section 2, we provide the right variances in Table 1. We also provide the estimates of the variance of $\hat{p}\,(1-\hat{p})$ by the Delta Method (analytically), the true method (analytically), and the Bootstrap Method (algorithmically). Three figures which compare the variance estimates, the errors, and the absolute errors for 101 p values in $[0, 1]$ are given. Section 3 concludes.

2. Main Results

The correct Bootstrap and Delta Method variances of $\hat{p}\,(1-\hat{p})$ are given in Table 1. In Table 1, sample size $n = 24$, bootstrap sample size $B = 1000$. The variance of the Delta Method corresponds to $p = 2/3$ should be rounded to 0.00103, since the variance is calculated as 0.001028807. From Table 1 we see that, the estimate of the variance of $\hat{p}\,(1-\hat{p})$ by the Bootstrap Method is better than the first-order Delta Method at $p \neq 1/2$, but is worse than the second-order Delta Method at $p = 1/2$.

Table 1: Bootstrap and Delta Method variances of $\hat{p}\,(1-\hat{p})$. The second-order Delta Method is used when $p = 1/2$. The Delta Method variance is calculated numerically assuming that $\hat{p} = p$.

	$p = 1/4$	$p = 1/2$	$p = 2/3$
Bootstrap	0.00190	0.00025	0.00108
Delta Method	0.00195	0.00022	0.00103
True	0.00191	0.00021	0.00111

The original values from Table 10.1.1 in Casella and Berger (2002) are provided in Table 2 so that potential readers do not need to be referred to the book.

Table 2: (Table 10.1.1 in Casella and Berger (2002)) Bootstrap and Delta Method variances of $\hat{p}\,(1-\hat{p})$. The second-order Delta Method is used when $p = 1/2$. The Delta Method variance is calculated numerically assuming that $\hat{p} = p$.

	$p = 1/4$	$p = 1/2$	$p = 2/3$
Bootstrap	0.00508	0.00555	0.00561
Delta Method	0.00195	0.00022	0.00102
True	0.00484	0.00531	0.00519

The potential reasons why the estimates in Table 10.1.1 of Casella and Berger (2002) are not right are summarized as follows. Firstly and the most importantly, the exact expression for $\text{Var}_p\,(\hat{p}\,(1-\hat{p}))$ may be wrongly calculated by Casella and Berger. Secondly, the Bootstrap procedure are wrongly programmed by Casella and Berger.

The estimate of the variance of $\hat{p}\,(1-\hat{p})$ by the first-order Delta Method is (see Casella and Berger (2002) Example 10.1.15)

$$\widehat{\text{Var}}_p^{\text{Delta1}}\,(\hat{p}\,(1-\hat{p})) = \frac{\hat{p}\,(1-\hat{p})\,(1-2\hat{p})^2}{n} = f_1\,(\hat{p})\,.$$

Since $\widehat{\text{Var}}_p^{\text{Delta1}}\,(\hat{p}\,(1-\hat{p}))\Big|_{\hat{p}=1/2} = 0$, a clear underestimate of the variance of $\hat{p}\,(1-\hat{p})$. Therefore, when $\hat{p} = 1/2$, we need to use a second-order Delta Method. When $\hat{p} = 1/2$, the estimate of the variance of $\hat{p}\,(1-\hat{p})$ by the second-order

Delta Method is

$$\widehat{\mathrm{Var}}_p^{\mathrm{Delta2}}\left(\hat{p}\left(1-\hat{p}\right)\right) = \frac{2\hat{p}^2\left(1-\hat{p}\right)^2}{n^2} = \frac{1}{8n^2}. \tag{1}$$

The derivation of (1) can be found in the appendix. Therefore, the estimate of the variance of $\hat{p}\left(1-\hat{p}\right)$ by the Delta Method is formed by combining the first-order and the second-order Delta Method, and is given by

$$\widehat{\mathrm{Var}}_p^{\mathrm{Delta}}\left(\hat{p}\left(1-\hat{p}\right)\right) = \begin{cases} \hat{p}\left(1-\hat{p}\right)\left(1-2\hat{p}\right)^2/n, & \text{if } \frac{1}{2} \neq \hat{p} \in [0,1], \\ 2\hat{p}^2\left(1-\hat{p}\right)^2/n^2 = 1/\left(8n^2\right), & \text{if } \hat{p} = \frac{1}{2}. \end{cases}$$

The true variance of $\hat{p}\left(1-\hat{p}\right)$ is

$$\mathrm{Var}_p\left(\hat{p}\left(1-\hat{p}\right)\right) = \frac{1}{n^4}\left[\begin{array}{c} 2n\left(n-1\right)\left(3-2n\right)p^4 + 4n\left(n-1\right)\left(2n-3\right)p^3 \\ +n\left(n-1\right)\left(7-5n\right)p^2 + n\left(n-1\right)^2 p \end{array}\right] = f\left(p\right). \tag{2}$$

The derivation of (2) can be found in the appendix. We note that Exercise 10.10 in Casella and Berger (2002) asks people to calculate the exact expression for $\mathrm{Var}_p\left(\hat{p}\left(1-\hat{p}\right)\right)$. However, in their solution manual, there is no solution for this exercise.

The estimate of the variance of $\hat{\theta} = \hat{p}\left(1-\hat{p}\right)$ by the Bootstrap Method is calculated as follows (Casella and Berger (2002); Shao and Tu (1995); Shao (2003)).

Step 1. Given $p \in [0,1]$, generate an $n \times B$ matrix

$$\mathbf{X}^* = \left(x_{ki}^*\right)_{n\times B}, \quad x_{ki}^* \sim \mathrm{Bernoulli}\left(p\right).$$

Step 2. Calculate

$$\hat{p}_i^* = \bar{x}_i^* = \frac{1}{n}\sum_{k=1}^{n} x_{ki}^*, \quad i = 1, \ldots, B.$$

Step 3. Calculate

$$\hat{\theta}_i^* = \hat{p}_i^*\left(1-\hat{p}_i^*\right), \quad i = 1, \ldots, B.$$

Step 4. Calculate the bootstrap approximator

$$\mathrm{Var}_B^*\left(\hat{\theta}\right) = \frac{1}{B-1}\sum_{i=1}^{B}\left(\hat{\theta}_i^* - \overline{\hat{\theta}^*}\right)^2,$$

where

$$\overline{\hat{\theta}^*} = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i^*.$$

Table 1 just compares three p values of the three methods. What are the variance comparisons at other p values in $[0,1]$? Let us see Figure 1. In Figures 1-3, there are 101 p values $[0, 0.01, 0.02, \ldots, 0.99, 1]$, sample size $n = 24$, bootstrap sample size increases to $B = 10000$. In Figure 1, the red full line is the variance estimate by the Bootstrap Method, the blue dashed line is the variance estimate by the Delta Method, and the black dotted line is the true variance. We see that the variance estimate by the Bootstrap Method approximates the true variance very well. The second-order Delta Method at $p = \frac{1}{2}$ has a good variance estimate, while the first-order Delta Method at $p = \frac{1}{2}$ has a very bad variance estimate. The true variance curve has two peaks and one valley, and the true variances are equal to 0 at the two end points. We also see that the three curves exhibit symmetries about $p = \frac{1}{2}$. It is easy to check that the first-order Delta Method curve is symmetric about $p = \frac{1}{2}$ by checking

$$f_1\left(\frac{1}{2}+\varepsilon\right) = f_1\left(\frac{1}{2}-\varepsilon\right).$$

We can also check that the true variances curve is symmetric about $p = \frac{1}{2}$ by checking

$$f\left(\frac{1}{2}+\varepsilon\right) = f\left(\frac{1}{2}-\varepsilon\right).$$

We can exploit the Mathematica software to do this job. The Mathematica codes can be found in the supplemental file "TestSymmetric.nb". We see that the Bootstrap Method curve is numerically symmetric about $p = \frac{1}{2}$.

The error comparison of the two methods is shown in Figure 2. The error curve is calculated by the difference of the estimated curve and the true curve. Thus the error will be positive if the estimated variance is higher than the true variance, and negative otherwise. From Figure 2 we see that the Bootstrap Method has small errors on the whole interval $[0,1]$, while the Delta Method has a sine shaped error (in fact the curve is a polynomial of p of order 4) with

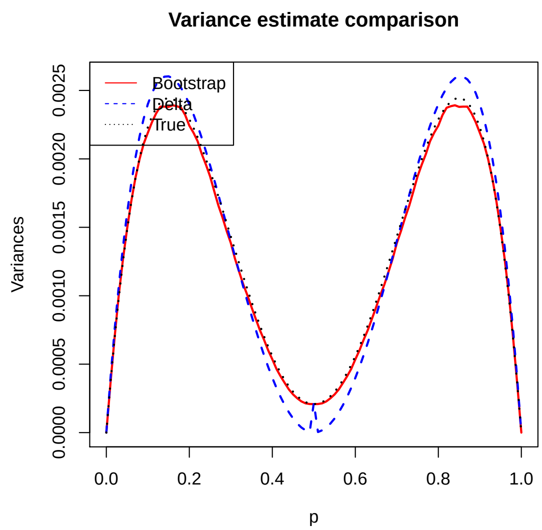

Figure 1: Variance estimate comparisons of the three methods.

an exception at $p = \frac{1}{2}$, as expected. The error curves of the two methods are symmetric about $p = \frac{1}{2}$.

Figure 3 shows the absolute error comparison of the two methods. By absolute error, we mean the absolute value of the error. We see that in most cases, the Bootstrap Method outperforms the Delta Method. The magnitude of the absolute errors of the Bootstrap Method is smaller than that of the Delta Method. The absolute error curves of the two methods are symmetric about $p = \frac{1}{2}$. The variances of the two methods agree with the true variances at the two endpoints 0 and 1. The variances of the Delta Method agree with the true variances at two intermediate values, and thus it behaves better than the Bootstrap Method near the two values. While the variance of the Bootstrap Method seems to agree with the true variance at $p = \frac{1}{2}$, and thus it is better than the Delta Method at the point. This phenomenon seems to contradict with that seen in Table 1. However, there is no contradiction because in Table 1, $B = 1000$, while in Figure 3, $B = 10000$.

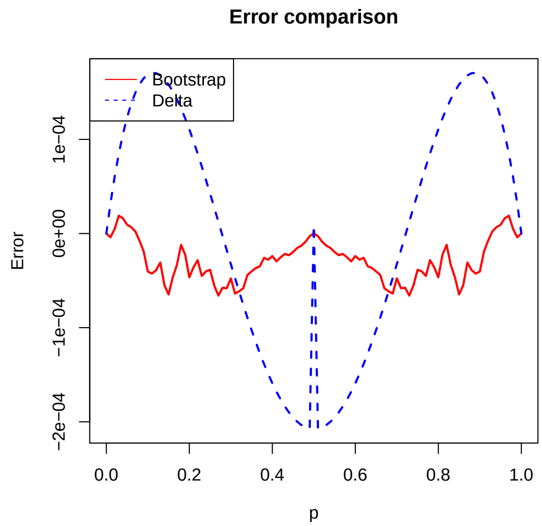

Figure 2: Error comparisons of the Bootstrap and Delta Method.

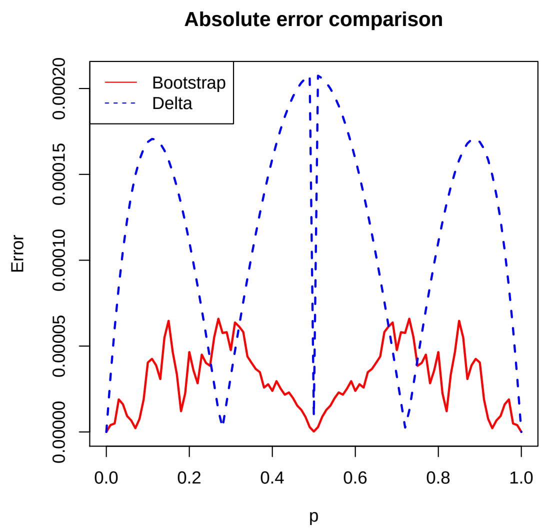

Figure 3: Absolute error comparisons of the Bootstrap and Delta Method.

3. Conclusion

We compare the Bootstrap and Delta Method variances of $\hat{p}\,(1-\hat{p})$, which is the variance estimator of the Bernoulli distribution. First, we provide the right Bootstrap, Delta Method, and true variances of $\hat{p}\,(1-\hat{p})$ in Table 1. The parameter values of Table 1 and Table 10.1.1 in Casella and Berger (2002) are the same. Then, we provide the estimates of the variance of $\hat{p}\,(1-\hat{p})$ by the Delta Method, the true method, and the Bootstrap Method. The derivations of the estimates of the variance of $\hat{p}\,(1-\hat{p})$ by the second-order Delta Method and the true method are given in the appendix. Finally, three figures which compare the variance estimates, the errors, and the absolute errors for 101 p values in $[0,1]$ are given. It is worth noting that, the three variance estimate curves exhibit symmetries about $p=\frac{1}{2}$, and in most cases, the Bootstrap Method outperforms the Delta Method.

Supporting Information: Additional information for this article is available.

R folder: R codes used in the article. The R folder will be supplied after acceptance of the article.

Mathematica folder: Mathematica codes used in the article. The Mathematica folder will be supplied after acceptance of the article.

References

Alfaro, M. E., Zoller, S., and Lutzoni, F. (2003). Bayes or bootstrap? a simulation study comparing the performance of bayesian markov chain monte carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution*, 20:255–266.

Briggs, A. H., Wonderling, D. E., and Mooney, C. Z. (1997). Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Economics*, 6:327–340.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? a practical guide for medical statisticians. *Statistics in Medicine*, 19:1141–1164.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury, USA. 2nd edition.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39:783–791.

Hillis, D. M. and Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, 42:182–192.

Ju, H. (2015). Moving block bootstrap for analyzing longitudinal data. *Communications in Statistics - Theory and Methods*, 44:1130–1142.

Meyer, J. S., Ingersoll, C. G., McDonald, L. L., and Boyce, M. S. (1986). Estimating uncertainty in population growth rates: Jackknife vs. bootstrap techniques. *Ecology*, 67:1156–1166.

Shao, J. (2003). *Mathematical Statistics*. Springer, New York. 2nd edition.

Shao, J. and Tu, D. (1995). *The Jackknife and the Bootstrap*. Springer, New York.

Shen, M. Y. and Machado, S. G. (2016). Bioequivalence evaluation of sparse sampling pharmacokinetics data using bootstrap resampling method. *Journal of Biopharmaceutical Statistics*, pages 1–8.

Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the raxml web servers. *Systematic Biology*, 57:758–771.

Visscher, P. M., Thompson, R., and Haley, C. S. (1996). Confidence intervals in qtl mapping by bootstrapping. *Genetics*, 143:1013–1020.

Wang, D. L. and Hutson, A. D. (2014). A smooth bootstrap procedure towards deriving confidence intervals for the relative risk. *Communications in Statistics - Theory and Methods*, 43:1979–1990.

Appendix

The derivations of (1) and (2) are given in the appendix.

The derivation of (1). We have, by the Central Limit Theorem,

$$\frac{\bar{X} - \mathrm{E}\bar{X}}{\sqrt{\mathrm{Var}\left(\bar{X}\right)}} \xrightarrow{d} N\left(0, 1\right), \text{ as } n \to \infty,$$

where $\bar{X}$ is the sample mean of $X_1, X_2, \ldots, X_n$, which are iid from Bernoulli (p). The Maximum Likelihood Estimator (MLE) of p is $\hat{p} = \bar{X}$. And

$$\mathrm{E}\hat{p} = \mathrm{E}\bar{X} = \mathrm{E}X = p,$$
$$\mathrm{Var}\left(\hat{p}\right) = \mathrm{Var}\left(\bar{X}\right) = \frac{\mathrm{Var}\left(X\right)}{n} = \frac{p\left(1-p\right)}{n}.$$

Therefore,

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N\left(0, 1\right), \text{ as } n \to \infty.$$

Rearranging, we obtain

$$\sqrt{n}\left(\hat{p} - p\right) \xrightarrow{d} \sqrt{p\left(1-p\right)} N\left(0, 1\right) = N\left(0, p\left(1-p\right)\right), \text{ as } n \to \infty.$$

Let $g\left(p\right) = p\left(1-p\right)$. Then by the second-order Delta Method (Theorem 5.5.26 in Casella and Berger (2002)), we have

$$n\left[g\left(\hat{p}\right) - g\left(p\right)\right] \xrightarrow{d} \frac{g^{''}\left(p\right)}{2}\sigma^2\chi_1^2, \text{ as } n \to \infty,$$

where $\sigma^2 = p(1-p)$ and χ_1^2 is the chi-square random variable with 1 degree of freedom. Therefore, for large n,

$$\mathrm{Var}_p(g(\hat{p})) \approx \mathrm{Var}_p\left(\frac{g''(p)}{2n}\sigma^2\chi_1^2\right) = \mathrm{Var}_p\left(\frac{-2}{2n}p(1-p)\chi_1^2\right) = \frac{2p^2(1-p)^2}{n^2}.$$

Replacing p by its MLE $\hat{p}$ in the above equation, we obtain the estimate of the variance of $\hat{p}(1-\hat{p})$ by the second-order Delta Method, namely,

$$\widehat{\mathrm{Var}}_p^{\mathrm{Delta2}}(\hat{p}(1-\hat{p})) = \frac{2\hat{p}^2(1-\hat{p})^2}{n^2}.$$

Therefore, (1) is established. □

The derivation of (2). We have

$$\hat{p} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}Y,$$

where $X_1, X_2, \ldots, X_n$ are iid from Bernoulli(p) and

$$Y = \sum_{i=1}^{n} X_i \sim \mathrm{Binomial}(n, p).$$

Therefore,

$$\begin{aligned}\mathrm{Var}_p(\hat{p}(1-\hat{p})) &= \mathrm{Var}_p\left(\bar{X}\left(1-\bar{X}\right)\right) = \mathrm{Var}_p\left(\frac{Y}{n}\left(1-\frac{Y}{n}\right)\right) \\ &= \mathrm{Var}_p\left(\frac{Y(n-Y)}{n^2}\right) = \frac{1}{n^4}\mathrm{Var}_p(Y(n-Y)).\end{aligned}$$

Now

$$\mathrm{Var}_p(Y(n-Y)) = \mathrm{E}\left[Y^2(n-Y)^2\right] - \{\mathrm{E}[Y(n-Y)]\}^2.$$

To calculate $\mathrm{Var}_p(Y(n-Y))$, we need to know the first four moments of Y which can be calculated by the derivative of the moment generating function evaluated at $t = 0$. The first four moments of Y are given by:

$$\begin{aligned}\mathrm{E}Y &= np, \\ \mathrm{E}Y^2 &= np + n(n-1)p^2, \\ \mathrm{E}Y^3 &= np + 3n(n-1)p^2 + n(n-1)(n-2)p^3, \\ \mathrm{E}Y^4 &= np + 7n(n-1)p^2 + 6n(n-1)(n-2)p^3 + n(n-1)(n-2)(n-3)p^4.\end{aligned}$$

Now

$$\begin{aligned}\mathrm{E}\left[Y\left(n-Y\right)\right] &= \mathrm{E}\left[nY-Y^2\right] = n\mathrm{E}Y-\mathrm{E}Y^2\\ &= n\times np-np-n\left(n-1\right)p^2 = n\left(n-1\right)p\left(1-p\right),\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}\left[Y^2\left(n-Y\right)^2\right] &= \mathrm{E}\left[Y^2\left(Y^2-2nY+n^2\right)\right] = \mathrm{E}\left[Y^4-2nY^3+n^2Y^2\right]\\ &= \mathrm{E}Y^4-2n\mathrm{E}Y^3+n^2\mathrm{E}Y^2.\end{aligned}$$

Therefore, by calculating, we obtain

$$\begin{aligned}\mathrm{Var}_p\left(Y\left(n-Y\right)\right) &= \mathrm{E}Y^4-2n\mathrm{E}Y^3+n^2\mathrm{E}Y^2-\left[n\left(n-1\right)p\left(1-p\right)\right]^2\\ &= 2n\left(n-1\right)\left(3-2n\right)p^4+4n\left(n-1\right)\left(2n-3\right)p^3\\ &\quad +n\left(n-1\right)\left(7-5n\right)p^2+n\left(n-1\right)^2p.\end{aligned}$$

Dividing $\mathrm{Var}_p\left(Y\left(n-Y\right)\right)$ by n^4, we obtain (2). □