111

by 1111 111

The Bootstrap Method for Monte Carlo Integration Inference

Asamh Saleh M. Al Luhayb

Department of Mathematics, College of Science, Qassim University, P.O. Box 6644, Buraydah 51452, Saudi Arabia..

Corresponding author(s). E-mail(s): a.alluhayb@qu.edu.sa;

Abstract

In this paper, the use of bootstrap method with Monte Carlo integration is introduced for one dimension. This approach is based on generating observations from a known distribution for the bootstrap samples, then apply the Monte Carlo method on each bootstrap sample to estimate the integral of interest. The empirical distribution, or the bootstrap distribution, of the estimation results can be used as a good proxy for the distribution of the integral of interest. Based on the bootstrap distribution, the standard error of the estimate of the integral of interest can be derived. Also, the percentile and Normal confidence intervals with confidence level $(1 - \alpha)\%$ can be derived as well. The bootstrap method with Monte Carlo integration is easy to implement and straightforward to provide well results. Moreover, it provides small variance for the estimate of the integral of interest. Four examples with different functions and different domains are used to present the performance of the proposed method. From the study, we find that the method provides nearly identical results for the standard errors, regardless of the distributions used for generating observations for the bootstrap samples.

Keywords: Parametric bootstrap method, Monte Carlo integration, percentile confidence interval, Normal confidence interval, standard error.

1 Introduction

In the real world applications, the standard or classical mathematical methods have been widely used to compute the exact integral of a function g defined

on an interval $[a, b]$. This can be achieved if the situation is simple; however, if the situation is complicated, the computation is quite hard or impossible in some cases. Therefore, the approximation methods can be good choices to compute the integrals for the complicated functions with minor errors. One of these approximation methods is the Monte Carlo integration method, which is described in many references, see e.g. [15, 21, 26] for more details.

This approximation method for integrals is built with respect to the sampling distribution. It is crucial to choose a suitable distribution, which should be close to the function g, to generate random observations. The suitable distribution can lead to have a better approximation for the integral with small variance. Many references in the literature discussed the importance sampling or variance reduction, see e.g. [12, 20, 21, 22, 24, 25] for detailed presentations.

Surely, the importance sampling is crucial for the Monte Carlo integration to derive a well approximation with small variance, but the importance sampling requires a long computational time and some information about the shape of g. These requirements motivate to use the bootstrap method with the Monte Carlo integration. Based on the bootstrap distribution, it is possible to derive well approximations for integrals with small standard errors. Moreover, the percentile and Normal confidence intervals of any integral can be derived with high accuracy.

This paper is organized as follows: Section 2 presents the Monte Carlo integration for one dimension along with some descriptions. In Section 3, the bootstrap method is described with explanations of computing the standard error and deriving the percentile and Normal confidence intervals for the integral of interest θ. Section 4 presents the performance of bootstrap with Monte Carlo integration. The last section presents some concluding remarks.

2 Monte Carlo integration

The Monte Carlo method is a well-known concept used to derive the approximate integrals for complicated functions defined on certain domains. This method is built based on the probability theorem [3, 14, 16], where the integral of a function g can be computed by taking the expectation of function g, $E\{g\}$. For univariate data, suppose that X is a random variable following a probability density function f. The expectation of $g(X)$ can be written mathematically as follows

$$E\{g(X)\} = \int_{D_f} g(x) f(x) dx \tag{1}$$

where D_f is the support of probability density function f.

Let $\theta = \int_a^b g(x) \, dx$ and the random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed from the Uniform distribution with parameters a and b. Further suppose that $x_1, x_2, \ldots, x_n$ are the corresponding observations to the random variables $X_1, X_2, \ldots, X_n$. The estimated result of θ can

be found by

$$\hat{\theta} = \frac{b-a}{n} \sum_{i=1}^{n} g(x_i) \qquad (2)$$

It should be noted that the estimated value $\hat{\theta}$ converges to $E\{g(X)\} = \frac{\theta}{b-a}$ as $n \to \infty$ with probability 1 based on the strong law of large numbers theorem, which is presented in [10, 11]. To make the Monte Carlo integration easy to follow and apply, we list its algorithm in three steps as follows:

1. Generate n observations from the Uniform distribution with parameters a and b.
2. Find $g(x_i)$ for all $i = 1, 2, \ldots, n$.
3. Compute $\hat{\theta} = \frac{b-a}{n} \sum_{i=1}^{n} g(x_i)$.

The Uniform distribution can be replaced by any known distribution to generate the observations in the first step, but we choose the Uniform distribution here for simplicity in application.

3 The bootstrap method

The bootstrap method is a resampling concept proposed to measure the accuracy of a statistical estimate and to make inferences about unknown population parameters, e.g. mean, median, variance and confidence interval. In the literature, it has been widely used due to its simplicity to apply and efficiency to provide well estimates. With high accuracy, the bootstrap distribution for any statistic of interest can mimic the sampling distribution, where the sampling distribution is not always easy to obtain in real applications.

To implement the bootstrap method, parametric and nonparametric models are used to create multiple bootstrap samples, then the statistic of interest is computed based on each bootstrap sample. The empirical distribution of the results can be used as a proxy distribution for the sampling distribution and this allows making inferences about the statistic of interest. For more detailed presentation, it is beneficial to see the book of "An Introduction to The Bootstrap" by Efron and Tibshirani [9] and the book of "Bootstrap Methods and Their Application" by Davison and Hinkley [5].

To use the bootstrap method with Monte Carlo integration, we first need to introduce some notations. Suppose that the random quantities $X_1, X_2, \ldots, X_n$ are independent and identically distributed following the probability distribution f and supported on $[a, b]$. Let $x_1, x_2, \ldots, x_n$ be the observations corresponding to these random quantities. Furthermore, let the integral of interest be $\theta = \int_a^b g(x)dx$. Now, it is easy to present the algorithm of the bootstrap method with Monte Carlo integration through the following steps.

1. Generate n observations from the probability distribution f.
2. Find $g(x_i)$ for all $i = 1, 2, \ldots, n$.
3. Compute $\hat{\theta}^* = \frac{b-a}{n} \sum_{i=1}^{n} g(x_i)$.
4. Perform steps (1),(2) and (3) B times; this leads to $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$.

For better performance, the value of B is suggested to be large; e.g. $B = 1000$ [1, 2, 6, 7]. To derive a bootstrap estimate of θ, we compute the average of $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$, and to provide a bootstrap standard error estimate for $\overline{\hat{\theta}^*_{boot}} = \frac{1}{B} \sum_{j=1}^{B} \hat{\theta}^{*j}$, we compute the standard deviation of $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$ by [8]

$$SE(\overline{\hat{\theta}^*_{boot}}) = \sqrt{\frac{\sum_{j=1}^{B} (\hat{\theta}^{*j})^2 - \frac{(\sum_{j=1}^{B} \hat{\theta}^{*j})^2}{B}}{B - 1}} \tag{3}$$

For the $(1 - \alpha)\%$ percentile confidence interval, we order the values $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$ from least to largest, then take the $(\frac{\alpha}{2})$th and $(1 - \frac{\alpha}{2})$th ordered values, where the former is the lower bound and the latest is the upper bound. This can be written as follows

$$\theta \in \left(\hat{\theta}^*_{(\frac{\alpha}{2})}, \hat{\theta}^*_{(1 - \frac{\alpha}{2})} \right) \tag{4}$$

For the $(1 - \alpha)\%$ Normal confidence interval, it is needed to compute $\overline{\hat{\theta}^*_{boot}}$ and $SE(\overline{\hat{\theta}^*_{boot}})$, then we use the following equation [13]

$$\theta \in \left(\overline{\hat{\theta}^*_{boot}} - Z_{(1-\alpha)} SE(\overline{\hat{\theta}^*_{boot}}), \overline{\hat{\theta}^*_{boot}} + Z_{(1-\alpha)} SE(\overline{\hat{\theta}^*_{boot}}) \right) \tag{5}$$

where $Z_{(1-\alpha)}$ is the $(1 - \alpha)$ percentile of the standard Normal distribution.

4 Simulation studies

In this section, we present different bounded integrals needed to be computed analytically based on the bootstrap method with Monte Carlo integration. We choose different domains and functions as shown in Figure 1. Also, we determine the Uniform and Normal distributions, restricted to the domain of the integral of interest, to generate observations for the bootstrap samples. To compute the bootstrap estimate of integral θ with the standard error and the percentile and Normal confidence intervals, we set B equal to 1000 and the sample size of each bootstrap sample n equal to 1000. This means that the iteration of bootstrap is 1000 as $B = 1000$. Table 1 presents different bounded integrals with their analytical results, and these integrals will be estimated based on the bootstrap method with Monte Carlo integration to make comparisons with the true results. By this strategy, we can make investigations on the performance of our method.
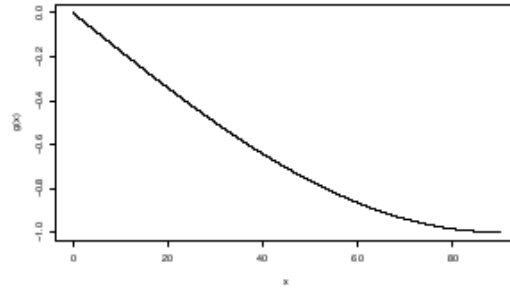
Tables 2 to 5 present the bootstrap estimates $\hat{\theta}$ along with the standard errors of $\hat{\theta}$ and the 90% percentile and Normal confidence intervals of θ for all examples presented in Table 1. It is obvious that the bootstrap estimates are nearly identical to the exact results. This is the power of using the bootstrap method with Monte Carlo integration, which is more needed for complicated integrals that is impossible to be computed theoretically. Through
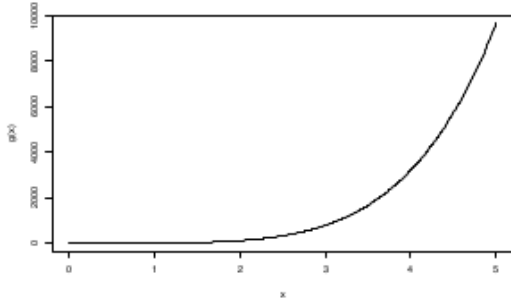
Table 1 The exact values of different integrals.

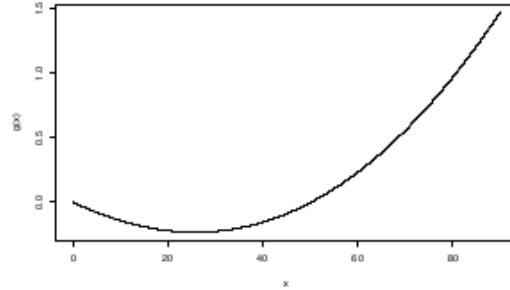| Example	$\theta =$
1	$\int_{-2}^{2} x^2 \ dx = 5.333$
2	$\int_{0}^{\frac{\pi}{2}} -sin(x) \ dx = -1$
3	$\int_{0}^{5} 3x^5 + 2x^3 + 1 \ dx = 8130$
4	$\int_{0}^{\frac{\pi}{2}} x^2 - sin(x) \ dx = 0.292$
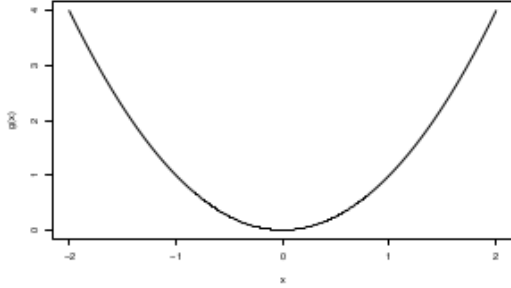
(a) $g(x) = x^2$

(b) $g(x) = -sin(x)$

(c) $g(x) = 3x^5 + 2x^3 + 1$

(d) $g(x) = x^2 - sin(x)$

Fig. 1 The shape of each function $g(x)$ presented in Table 1.

Table 2 The bootstrap estimate of θ in Example 1 along with the standard error and the percentile and Normal confidence intervals with confidence level 90%.

Bootstrap estimates	Uniform(-2, 2)	Normal(0, 5) (restricted on [-2, 2])
$\hat{\theta}$	5.339	5.228
Standard Error	0.158	0.151
Percentile confidence interval	(5.093, 5.596)	(4.982, 5.463)
Normal confidence interval	(5.136, 5.541)	(5.036, 5.421)

Table 3 The bootstrap estimate of θ in Example 2 along with the standard error and the percentile and Normal confidence intervals with confidence level 90%.

Bootstrap estimates	Uniform(0, $\frac{\pi}{2}$)	Normal(0,5) (restricted on [0, $\frac{\pi}{2}$])
$\hat{\theta}$	-1.000	-0.993
Standard Error	0.015	0.015
Percentile confidence interval	(-1.026, -0.976)	(-1.020, -0.968)
Normal confidence interval	(-1.020, -0.981)	(-1.013, -0.973)

Table 4 The bootstrap estimate of θ in Example 3 along with the standard error and the percentile and Normal confidence intervals with confidence level 90%.

Bootstrap estimates	Uniform(0, 5)	Normal(3,5) (restricted on [0, 5])
$\hat{\theta}$	8146.495	8155.398
Standard Error	385.488	387.729
Percentile confidence interval	(7514.391, 8783.074)	(7550.591, 8808.027)
Normal confidence interval	(7652.471, 8640.518)	(7658.503, 8652.292)

Table 5 The bootstrap estimate of θ in Example 4 along with the standard error and the percentile and Normal confidence intervals with confidence level 90%.

Bootstrap estimates	Uniform(0, $\frac{\pi}{2}$)	Normal(3,5) (restricted on [0, $\frac{\pi}{2}$])
$\hat{\theta}$	0.293	0.317
Standard Error	0.024	0.024
Percentile confidence interval	(0.254, 0.332)	(0.280, 0.355)
Normal confidence interval	(0.262, 0.323)	(0.286, 0.347)

the bootstrap procedure, we use Uniform and Normal distributions with different parameters, but restricted to the integral's bounds, and this leads to have nearly identical standard errors. From this observation, we can assure that the bootstrap distribution can be a good proxy distribution for the integral θ; there is no need to estimate the shape of $g(x)$ for importance sampling, or variance reduction. This helps to conserve time to running codes and it is possible to do analysis with less information. Also, based on the bootstrap method, the 90% percentile and Normal confidence intervals can be easily derived. From Tables 2 to 5, we note that the 90% percentile confidence intervals are all wider than the 90% Normal confidence intervals.

5 Concluding remarks

In this paper, we illustrated the bootstrap method with Monte Carlo integration for one dimension, and the method was used through multiple examples. From the examples, it can be concluded that the bootstrap method with Monte Carlo integration is a good approach to compute well approximate integrals for different functions in different domains with small variances, regardless to the distribution being used to generate observations for the bootstrap samples. Based on the bootstrap estimates, the $(1 - \alpha)\%$ percentile and Normal confidence intervals for θ can be derived with high accuracy. Our method can be beneficial for integrals that may be difficult or impossible to compute. Also, the method is easy to implement and straightforward, which is only relying on sampling from a known distribution, then taking the evaluation. This is repeated multiple times and this should be large, e.g. $B = 1000$. To put the method into a practical use, we included the R codes in the appendix. To run the codes, it requires about ten seconds, which is nothing in real applications and this is one of the advantages.

As a future research, the method will be generalized for multiple dimensions with more complicated functions and domains. To achieve this generalization,

we may use the copula concept, which is able to take the dependence structure between the variables into account, to generate observations for the bootstrap samples. For more detailed presentations about the copula concept, it is advised to see [4, 17, 18, 19, 23].

Conflict of interest

The author states that there is no conflict of interest.

Acknowledgement

References

[1] Al Luhayb ASM (2021) Smoothed bootstrap methods for right-censored data and bivariate data. PhD thesis, Durham University

[2] Al Luhayb ASM, Coolen FPA, Coolen-Maturi T (2023) Smoothed bootstrap for right-censored data. Communications in Statistics - Theory and Methods 0(0):1–25. https://doi.org/10.1080/03610926.2023.2171708

[3] Chung KL, Zhong K (2001) A course in probability theory. Academic press

[4] Coolen-Maturi T, Coolen FPA, Muhammad N (2016) Predictive inference for bivariate data: Combining nonparametric predictive inference for marginals with an estimated copula. Journal of Statistical Theory and Practice 10:515–538

[5] Davison AC, Hinkley DV (1997) Bootstrap Methods and Their Application. New York, NY: Cambridge University Press

[6] Efron B (1979) Bootstrap methods: Another look at the jackknife. The Annals of Statistics 7:1–26

[7] Efron B (1981) Censored data and the bootstrap. Journal of the American Statistical Association 76:312–319

[8] Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1:54–77

[9] Efron B, Tibshirani RJ (1993) An Introduction to The Bootstrap. Boca Raton, FL: Chapman & Hall

[10] Etemadi N (1981) An elementary proof of the strong law of large numbers. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 55(1):119–122

[11] Fazekas I, Klesov O (2001) A general approach to the strong law of large numbers. Theory of Probability & Its Applications 45(3):436–449

[12] Hammersley J, Morton K (1956) A new monte carlo technique: antithetic variates. In: Mathematical proceedings of the Cambridge philosophical society, Cambridge University Press, pp 449–475

[13] Hazra A (2017) Using the confidence interval confidently. Journal of thoracic disease 9(10):4125

[14] Jaynes ET (2003) Probability theory: The logic of science. Cambridge university press

[15] Kalos MH, Whitlock PA (2009) Monte carlo methods. John Wiley & Sons

[16] Loeve M (2017) Probability theory. Courier Dover Publications

[17] Muhammad N (2016) Predictive inference with copulas for bivariate data. PhD thesis, Durham University, UK, https://npi-statistics.com/pdfs/theses/NM16.pdf

[18] Muhammad N, Coolen FPA, Coolen-Maturi T (2016) Predictive inference for bivariate data with nonparametric copula. The American Institute of Physics (AIP) Conference Proceedings 1750:0600,041–0600,048. https://doi.org/10.1063/1.4954609, https://aip.scitation.org/doi/abs/10.1063/1.4954609, https://arxiv.org/abs/https://aip.scitation.org/doi/pdf/10.1063/1.4954609

[19] Muhammad N, Coolen-Maturi T, Coolen FPA (2018) Nonparametric predictive inference with parametric copulas for combining bivariate diagnostic tests. Statistics, Optimization & Information Computing 6:398–408

[20] Oh MS, Berger JO (1992) Adaptive importance sampling in monte carlo integration. Journal of Statistical Computation and Simulation 41(3-4):143–168

[21] Rizzo ML (2019) Monte carlo integration and variance reduction. In: Statistical Computing with R. Chapman and Hall/CRC, p 147–182

[22] Rubinstein RY, Marcus R (1985) Efficiency of multivariate control variates in monte carlo simulation. Operations Research 33(3):661–677

[23] Sklar AW (1959) Fonctions de re'partitiona' n-dimensions et leurs marges. Publications de l'Institut de Statistique de l'Universite' de Paris 8:229–231

[24] Tokdar ST, Kass RE (2010) Importance sampling: a review. Wiley Interdisciplinary Reviews: Computational Statistics 2(1):54–60

[25] Van Dijk HK, Kloek T (1983) Experiments with some alternatives for simple importance sampling in monte carlo integration. Tech. rep.

[26] Yang XS (2014) Introduction to computational mathematics. World Scientific Publishing Company

Appendix

```
##########################################################

# First Example

# The domain of x is (-2,2)

a=-2
b=2
```

```r
# The exact integration is equal to

theta=(1/3)*((b)^3-(a)^3)
theta

# The bootstrap method with Monte Carlo integration

h_theta=NULL

for (i in 1:1000){

# generating observations from the Uniform distribution with a=-2 and b=2

x=runif(1000,a,b)

g=x^2

h_theta[i]=(b-a)*mean(g)
}

# To provide the bootstrap estimate of theta

mean(h_theta)

# To provide the bootstrap standard error estimate

sd(h_theta)

# To provide the 90% percentile confidence interval

c(quantile(h_theta,0.05),quantile(h_theta,0.95))

# To provide the 90% Normal confidence interval

c(mean(h_theta)-qnorm(0.90,mean=0,sd=1)*sd(h_theta),
mean(h_theta)+qnorm(0.90,mean=0,sd=1)*sd(h_theta))

##########################################################
```

1	etheses.dur.ac.uk Internet Source	**6**%
2	E.F. Casas, C. Leung. "OFDM for data communication over mobile radio FM channels. I. Analysis and experimental results", IEEE Transactions on Communications, 1991 Publication	**1**%
3	www.hindawi.com Internet Source	**1**%
4	downloads.hindawi.com Internet Source	**1**%
5	wordeal.com Internet Source	**1**%
6	Submitted to 7034 Student Paper	**1**%
7	Submitted to University of Leicester Student Paper	**<1**%
8	cyberleninka.org Internet Source	**<1**%
9	repository.kaust.edu.sa Internet Source	**<1**%
10	Gabriel Escarela, Carlos Erwin Rodríguez, Gabriel Núñez-Antonio. "Copula modeling of receiver operating characteristic and predictiveness curves", Statistics in Medicine, 2020 Publication	**<1**%