Taylor & Francis
Taylor & Francis Group

IISE Transactions

ISSN: 2472-5854 (Print) 2472-5862 (Online) Journal homepage: http://www.tandfonline.com/loi/uiie21

A Simulation Based Estimation Method for Bias Reduction

Jin Fang & L. Jeff Hong

A Simulation Based Estimation Method for Bias Reduction

Jin Fang[a,*], L. Jeff Hong[b]

[a]*Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong*
[b]*Department of Management Sciences and Department of Economics and Finance, College of Business, City University of Hong Kong, Kowloon Tong, Hong Kong*

Abstract

Models are often built to evaluate system performance measures or to make quantitative decisions. These models sometimes involve unknown input parameters that need to be estimated statistically using data. In these situations, a statistical method is typically used to estimate these input parameters and the estimates are then plugged into the models to evaluate system output performances. The output performance estimators obtained from this approach usually have large bias when the model is nonlinear and the sample size of the data is finite.

Phillips and Yu (2009) proposed a simulation based estimation method to reduce the bias of performance estimators for models that have a closed-form expression. In this paper, we extend the method to more general situations where the models have no closed-form expression and can only be evaluated through simulation. A stochastic root-finding problem is formulated to obtain the simulation based estimators and several algorithms are designed correspondingly. Furthermore, we give a thorough asymptotic analysis of the properties of the simulation based estimators, including the consistency, the order of the bias, the asymptotic variance and so on. Our numerical experiments show that the experimental results are consistent with the theoretical analysis.

Keywords: simulation based estimation, bias reduction, stochastic root-finding, asymptotic analysis

1. Introduction

Models that describe system behaviors are often used to evaluate system performance measures or to make quantitative decisions. The inputs of these models are often assumed to follow certain known distribution families but with unknown parameters that need to be estimated using data. Even though this approach may be subject to input uncertainty (i.e., the mis-specification of the input distribution families), it is widely used and generally simple and effective. In this approach,

*Corresponding author.
Email addresses: jfang@connect.ust.hk (Jin Fang), jeffhong@cityu.edu.hk (L. Jeff Hong)

the key is to estimate the unknown parameters, as long as the model and the input distribution families are determined. Once the parameters are estimated, their estimates are then plugged into the models to evaluate performance measures and to select decisions. For instance, in call centers, staffing decisions are often made according to estimates of the arrival rates; in manufacturing companies, procurement decisions and production plannings are often made based on estimates of the future demands.

Parameter estimation is a widely studied problem in the area of statistics. Many methods have been proposed to solve the problem and the parameter estimators often possess some desirable statistical properties, e.g., consistency, unbiasedness and minimum variance. When the parameter estimators are plugged into the models, however, the resulted estimators of the system performance measures, which we call *performance estimators* throughout this paper, may not inherit the same properties. For instance, we consider a M/M/1/J queueing model where the service rate and the buffer size are given, but the arrival rate is estimated through observed inter-arrival times. Suppose the true value of the arrival rate is 0.9, then the average steady-state queue length can be obtained from a function (Allen, 1990) and is calculated to be 8.098. We may obtain the maximum likelihood estimator (MLE) of the arrival rate from the inter-arrival times. The expectation of the MLE of the arrival rate approximated from numerical experiments is 0.905 when the sample size of the inter-arrival times is 200, indicating that the relative bias of the parameter estimator is only about 0.6%. However, we find that the expectation of the MLE of the average queue length is approximately 15.39 (i.e., with a relative bias of 90%), showing that the performance estimator is heavily biased even though the parameter estimator is not.

This phenomenon is quite ubiquitous when a nonlinear model is used to evaluate a performance measure. In the classical estimation approach, the input parameter is estimated solely from the observed data and is separated from the output performance evaluation. Throughout the paper, we call the classical estimation approach the *two-step approach* and its corresponding performance estimator the *two-step estimator*. In the two-step approach, the input parameter is estimated higher or lower than the true value, however, has different effects on the estimation of the output performance measures. For instance, in the M/M/1/J example, as the average queue length is an increasing and convex function of the arrival rate, an overestimated arrival rate may return us a queue length estimate which is much greater than the true queue length, especially when the traffic intensity is high, while an equally underestimated arrival rate may return us a queue length estimate which does not deviate too much away from the true queue length. In this sense, we should try to avoid overestimating the arrival rate. Similarly, it is also possible that we need to avoid underestimating the input parameters as the model changes. Therefore, we should take a holistic view to link the parameter estimation and the performance evaluation together.

Phillips and Yu (2009) proposed a simulation based estimation method (SBE method) to estimate option prices through the Black-Scholes pricing formula, where the input parameter is the volatility of the underlying asset that is estimated using MLE method. The basic idea of the SBE method can be described as follows. For an arbitrary input parameter, simulate data based on it and obtain the simulated two-step estimator, i.e., the performance estimator estimated from simulated data. Repeat such a process many times and calculate the average of the simulated two-step estimators. Minimize the distance between the average of the simulated two-step estimator and the two-step estimator estimated from the observed data by choosing the appropriate input parameter. The chosen input parameter is then plugged into the model to get a new performance estimator, which is called the simulation based estimator (SBE). Phillips and Yu (2009) showed, numerically, that the bias of the SBE is reduced compared to that of the two-step estimator.

For the SBE method, the performance measures of models are required to be evaluated many times to get the average of the two-step estimators. In practice, the performance functions of many models do not have a closed-form expression, but have to be evaluated through simulation. In this sense, the method may introduce high simulation cost. Moreover, the existing SBE method does not provide an efficient algorithm to find the SBE. Therefore, a noticeable issue is the implementation of the SBE method. In this paper, we formulate a stochastic root-finding problem to help find the SBE. We suggest to solve the root of the stochastic root-finding problem with a sample average approximation (SAA) method when the closed-form expression of the model is available and is

ACCEPTED MANUSCRIPT

smooth, and with a stochastic approximation (SA) method when the closed-form expression of the model is not available but can be evaluated through simulation and is monotone. When the model meets these structure requirements, both algorithms can significantly reduce the simulation cost, compared with the Phillips and Yu's method.

In addition, we construct a mathematical formulation for the SBE method, based on which we provide a theoretical explanation of the method. Particularly, the SBE method builds a functional relationship between the input parameter and the bias of its corresponding two-step estimator through simulation, which sheds light on how to adjust the input parameter estimators. We can also interpret the SBE method as a special case of the method of moments in terms of our analysis. Moreover, we provide a thorough asymptotic analysis of the statistical properties of the estimators. Under some smoothness and monotonicity conditions, we can show that the SBE of the performance measure retains some desirable properties such as consistency, asymptotic normality and unchanged variance when the sample size of the observed data goes to infinity. We also show that the SBE has smaller bias than the two-step estimator does when the sample size of the observed data is sufficiently large.

Literature Review

The study of parameter uncertainty in decision models is an important research area of operations research. In the stochastic simulation literature, how to account for parameter estimation errors has always been a very important research problem. The problem is often formulated as the construction of a valid confidence interval for an output performance measure; see, for instance, Henderson (2003) for a review. Cheng and Holland (1997) proposed to use a first-order approximation to propagate the confidence intervals of input parameters to the confidence interval of the output performance measure. Barton et al. (2014) provided a confidence interval for the output performance that accounts for uncertainty in the input parameters via metamodel-assisted bootstrapping. Chick (2001) applied a Bayesian model averaging approach to this problem, and Biller and Corlu (2011) further extended the Bayesian model to handle correlated inputs.

When it comes to reduce the bias of estimators, a lot of methods have been proposed. Most of these methods directly estimate the bias and then do the bias correction. For instance, the Jackknife method (Wu, 1986) omits one data point from the original data each time and calculates the mean jackknife estimator to help estimate the Jackknife bias. Asmussen and Glynn (2007) proposed a method that applies Taylor expansion to the performance function and estimates the bias of the performance estimators based on the bias and variance of the input parameters. Parametric bootstrap method resamples data from the estimated parametric distribution and then uses the simulated sample to estimate the bias (see, for instance, Efron and Tibshirani (1994) for details). The bias is then corrected by subtracting the estimated bias from the previous estimator.

The use of simulation methods in estimation is in general called simulation based estimation in econometrics literature; see, for instance, Chapter 12 of Cameron and Trivedi (2005). Its basic idea is to use a Monte Carlo method to approximate likelihood functions or expectations. For instance, McFadden (1989) proposed the method of simulated moments for estimation of discrete response models where moments are approximated through a Monte Carlo method. Another SBE method is called indirect inference, introduced by Smith (1993) and Gouriéroux et al. (1993). It uses simulation experiments performed under the initial model to correct for the asymptotic bias of input parameters estimated from an approximated model, which is used to replace the complicated and intractable initial model. Unlike the above SBE methods that focus on estimating the parameters, the SBE method in this paper focuses on the estimation of model performance measures.

The rest of the paper is organized as follows. We introduce the basic ideas and algorithms of the SBE method in Section 2 and analyze the statistical properties of the SBE in Section 3. We then illustrate the performance of SBEs through numerical examples in Section 4. The conclusions and a discussion on future studies are made in Section 5. All mathematical proofs are included in the Appendix.

ACCEPTED MANUSCRIPT

2. Basic ideas and algorithms of SBE method

Let θ denote the input parameter (or parameter vector) of a model and θ lies in a parameter space $\Theta \subset R^N$, where N is the dimension of θ. Denote the model by $p(\cdot)$, which is a function that maps θ to the output performance measure $p(\theta)$. The true value of the input parameter θ_0 is unknown and is estimated through a sample of observed data $\mathbf{X}_0 = \{X_1, \ldots, X_n\}$. Denote $\hat{\theta}_n(\cdot)$ an estimator that maps the data to an estimate. An estimate of θ_0 obtained from data $\mathbf{X}_0$ is actually $\hat{\theta}_n(\mathbf{X}_0)$. For simplicity, we use $\hat{\theta}_n$ to denote $\hat{\theta}_n(\mathbf{X}_0)$. Then, a widely used estimator is $p(\hat{\theta}_n)$, which is a two-step estimator of $p(\theta_0)$.

In statistical analysis, the observed data $\mathbf{X}_0$ is typically assumed to be a sample from a known distribution family with an unknown parameter θ_0. In this paper, we relax this assumption by assuming that we can simulate a sample $\mathbf{X}(\theta)$ for any given $\theta \in \Theta$, and the simulated sample $\mathbf{X}(\theta)$ has an identical joint distribution as the observed data $\mathbf{X}_0$ when $\theta = \theta_0$. In this sense, a known input distribution family is not mandatory for the data simulation process. For instance, $X(\theta) = f(\theta, Y)$, where $f(\theta, Y)$ is a mapping that maps the input parameter θ and other random variable Y into a random variable $X(\theta)$.

Focusing on estimating option prices, Phillips and Yu (2009) proposed the following SBE method to estimate $p(\theta_0)$.

Algorithm 1 Simulation based estimation method

Step 1: Obtain the two-step estimator $p(\hat{\theta}_n)$ from the real data $\mathbf{X}_0$.
Step 2: For a given $\theta \in \Theta$, simulate data $\mathbf{X}(\theta) = \{X_1(\theta), \ldots, X_n(\theta)\}$, use the same two-step approach to obtain $p(\hat{\theta}_n(\mathbf{X}(\theta)))$
Step 3: Repeat **Step 2** for K times, and get the average of the performance estimators $\frac{1}{K}\sum_{k=1}^{K} p(\hat{\theta}_n(\mathbf{X}_k(\theta)))$.
Step 4: Try different θ and choose one that minimizes the distance between $\frac{1}{K}\sum_{k=1}^{K} p(\hat{\theta}_n(\mathbf{X}_k(\theta)))$ and $p(\hat{\theta}_n)$. Denote the chosen θ by θ^* and let $p(\theta^*)$ be the new estimator of $p(\theta_0)$.

The new estimators θ^* and $p(\theta^*)$ are the SBEs of the unknown quantities θ_0 and $p(\theta_0)$. A drawback of this algorithm is that finding the θ^* is nontrivial. In this algorithm, data need to be simulated and the function $p(\cdot)$ needs to be evaluated K times to get the average of the performance estimators for any given θ. Things get worse if the closed-form expression of $p(\cdot)$ is unknown and can

ACCEPTED MANUSCRIPT

only be obtained through simulation in reality. If we run M simulations to approximate $p(\cdot)$, then, to get the average of the performance estimators for a given θ, KM simulation runs are needed. This is quite computational intensive, especially when the simulation of the model is expensive. Besides, the algorithm does not provide an efficient method to search for θ^*. In this paper, we figure out that if the model has some appealing structures, the computational cost can be significantly reduced.

We begin with formulating a stochastic root-finding problem to find the SBEs. Define

$$b_n(\theta) = \mathrm{E}_{\mathbf{X}}\left[p\left(\hat{\theta}_n(\mathbf{X}(\theta))\right)\right]$$

for all $\theta \in \Theta$. We use $\mathrm{E}_{\mathbf{X}}$ to denote that the expectation is taken with respect to the simulated data $\mathbf{X}(\theta)$. Let $\tilde{\theta}_n$ denote a solution of the following root-finding problem (i.e., finding θ that satisfies the following equation):

$$b_n(\theta) = p(\hat{\theta}_n). \tag{1}$$

Assume that $\tilde{\theta}_n$ exists and $\tilde{\theta}_n \in \Theta$. Then, $\tilde{\theta}_n$ and $p(\tilde{\theta}_n)$ are the SBEs. Notice that $\tilde{\theta}_n$ is an intuitive estimator as $p(\hat{\theta}_n)$ is not an unbiased estimator of $p(\theta_0)$, but an unbiased estimator of $b_n(\theta_0)$. This is because $\mathrm{E}[p(\hat{\theta}_n)] = \mathrm{E}_{\mathbf{X}_0}[p(\hat{\theta}_n(\mathbf{X}_0))] = b_n(\theta_0)$. The second equality holds because $\mathbf{X}_0$ and $\mathbf{X}(\theta_0)$ are identically distributed. Then, the SBE method may be viewed as a special case of the method of moments with one observation point $p(\hat{\theta}_n)$.

To better understand the SBE method, we plot $p(\cdot)$ and $b_n(\cdot)$ in Figure 1 based on the example discussed in the introduction. The vertical distance between $p(\cdot)$ and $b_n(\cdot)$ represents the bias of the two-step estimator and the horizontal distance between the $p(\cdot)$ and $b_n(\cdot)$ represents the correction of the input parameter estimator (i.e., $|\hat{\theta}_n - \tilde{\theta}_n|$). When the sample size n is small, the function $b_n(\cdot)$ lies away from $p(\cdot)$, the bias is large and the correction is large as well; as n increases, the function $b_n(\cdot)$ gets closer to $p(\cdot)$, the bias decreases and so does the correction.

We propose solving the root-finding problem (1) with a sample-average approximation (SAA) method or a stochastic approximation method in terms of different models.

ACCEPTED MANUSCRIPT

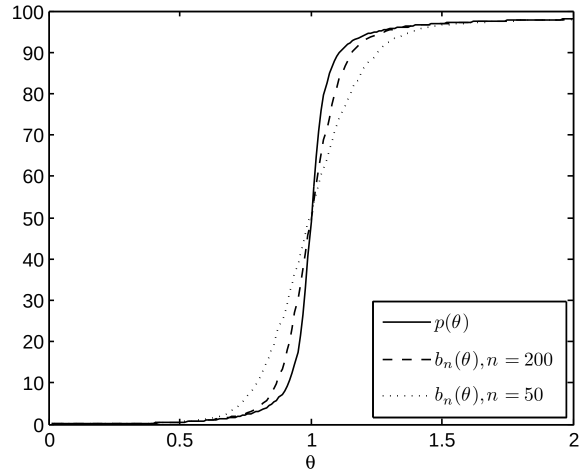

Figure 1: Change of $b_n(\cdot)$ with respect to n

2.1. Sample Average Approximation Method

In many problems we may write $\mathbf{X}(\theta) = \mathbf{X}(\theta, \omega)$, where ω incorporates all the randomness in $\mathbf{X}(\theta)$ and does not depend on θ. For instance, when $\mathbf{X}(\theta)$ denotes an i.i.d. sample of an exponential random variable with mean θ, we may write $X_i(\theta) = -\theta \log(\omega_i)$, where ω_i is a uniform$(0, 1)$ random variable; when $\mathbf{X}(\theta)$ denotes an i.i.d. sample of a normal random variable with mean θ and variance 1, we may write $X_i(\theta) = \theta + \omega_i$, where ω_i follows a standard normal distribution. While simulating data, we can first simulate an i.i.d. sample of ω, which is denoted by $\{\omega_1, \ldots, \omega_K\}$. And then, for any value θ, the simulated data $\mathbf{X}(\theta, \omega)$ can be simulated by $\{X(\theta, \omega_1), \ldots, X(\theta, \omega_K)\}$.

We may approximate $b_n(\theta)$ by a sample-average function

$$\bar{b}_n(\theta) = \frac{1}{K} \sum_{k=1}^{K} p\big(\hat{\theta}_n(\mathbf{X}(\theta, \omega_k))\big).$$

The SAA method proposes solving the root-finding problem

$$\bar{b}_n(\theta) - p(\hat{\theta}_n) = 0, \tag{2}$$

with a large value of K. Shapiro et al. (2009) showed that, under some mild conditions, the solution of the root-finding problem (2) converges to the solution of the root-finding problem (1) as $K \to \infty$. Notice that, once K and $\{\omega_1, \ldots, \omega_K\}$ are fixed, the root-finding problem (2) becomes a deterministic problem and we may use Newton's method to solve it. Let $\bar{b}'_n(\cdot)$ denote the first-order derivative of $\bar{b}_n(\cdot)$. We have the following Algorithm 2.

Algorithm 2 SBE method with SAA

Step 1: Obtain the two-step estimator $p(\hat{\theta}_n)$ from real data $\mathbf{X}_0$. Set $j = 1$ and $\theta_j = \hat{\theta}_n$.
Step 2: Simulate data $\{\omega_1, \ldots, \omega_K\}$, and obtain the mappings $\bar{b}_n(\cdot)$ and $\bar{b}'_n(\cdot)$ by $\frac{1}{K}\sum_{k=1}^{K} p\big(\hat{\theta}_n(\mathbf{X}(\cdot, \omega_k))\big)$ and $\frac{1}{K}\sum_{k=1}^{K} \big[p\big(\hat{\theta}_n(\mathbf{X}(\cdot, \omega_k))\big)\big]'$.
Step 3: Set $\theta_{j+1} = \theta_j - \frac{1}{\bar{b}'_n(\theta_j)}\Big[\bar{b}_n(\theta_j) - p(\hat{\theta}_n)\Big]$.
Step 4: Repeat **Step 3** until θ_{j+1} converges and let $p(\theta_{j+1})$ be the new estimator of $p(\theta_0)$.

The notation $\big[p\big(\hat{\theta}_n(\mathbf{X}(\cdot, \omega_k))\big)\big]'$ denotes the first-order derivative of $p\big(\hat{\theta}_n(\mathbf{X}(\cdot, \omega_k))\big)$.

In general, the algorithm is very efficient when it is applicable because we only need to simulate data $\{\omega_1, \ldots, \omega_K\}$ for one time and Newton's method typically converges fast. However, this method may be difficult to apply if calculating $\bar{b}'_n(\theta_j)$ is not easy (e.g., when the closed-form expression of $p(\cdot)$ is not available).

2.2. Stochastic Approximation Method

Sometimes, $p(\cdot)$ has no closed-form expression but can only be expressed as follows:

$$p(\theta) = \mathrm{E}_G[G(\theta)], \tag{3}$$

where $G(\theta)$ is a random variable depending on θ. This case is suitable for simulation modeling problems where $G(\theta)$ is an observation from running a stochastic simulation experiment at θ. It is important to distinguish the expectation E_G with the expectation $\mathrm{E}_{\mathbf{X}}$, where E_G is taken with respect to the simulated observation $G(\theta)$.

Then, the root-finding problem (1) may be written as

$$\mathrm{E}_{\mathbf{X}}\left\{\mathrm{E}_G\left[G(\hat{\theta}_n(\mathbf{X}(\theta)))\right]\right\} - \mathrm{E}_G[G(\hat{\theta}_n)] = 0. \tag{4}$$

We combine $\mathrm{E}_{\mathbf{X}}$ and E_G together to rewrite Problem (4) as

$$\mathrm{E}_{\mathbf{X},G}\left[G(\hat{\theta}_n(\mathbf{X}(\theta))) - G(\hat{\theta}_n)\right] = 0. \tag{5}$$

We propose to solve Problem (5) using the Robbins-Monro algorithm, which is a well known stochastic approximation algorithm. Define $f_n(\theta) = \mathrm{E}_{\mathbf{X},G}\left[G(\hat{\theta}_n(\mathbf{X}(\theta))) - G(\hat{\theta}_n)\right]$. When $f_n(\cdot)$ is non-decreasing, we have the following Algorithm 3.

In the algorithm, $\{a_j : j = 1, 2, \ldots\}$ is a sequence of positive step-sizes satisfying that $\sum_{j=1}^{\infty} a_j =$

Algorithm 3 SBE method with SA

Step 1: Obtain the two-step estimator $p(\hat{\theta}_n)$ from real data $\mathbf{X}_0$. Set $j = 1$ and $\theta_j = \hat{\theta}_n$.
Step 2: Simulate data $\mathbf{X}(\theta_j)$, and then simulate $G\big(\hat{\theta}_n(\mathbf{X}(\theta_j))\big)$ and $G(\hat{\theta}_n)$. Set
$\theta_{j+1} = \theta_j - a_j\left[G\big(\hat{\theta}_n(\mathbf{X}(\theta_j))\big) - G(\hat{\theta}_n)\right]$.
Step 3: Repeat **Step 2** until θ_{j+1} converges and let $p(\theta_{j+1})$ be the new estimator of $p(\theta_0)$.

∞ and $\sum_{j=1}^{\infty} a_j^2 < \infty$. When $f_n(\cdot)$ is non-increasing, we just need to set

$$\theta_{j+1} = \theta_j + a_j\left[G\big(\hat{\theta}_n(\mathbf{X}(\theta_j))\big) - G(\hat{\theta}_n)\right]$$

in step 2 of the algorithm. Robbins and Monro (1951) proved that θ_j converges to the root in L^2 as $j \to \infty$, if $G\big(\hat{\theta}_n(\mathbf{X}(\theta_j)\big)$ is uniformly bounded on Θ and $f_n(\cdot)$ is differentiable and monotone.

This algorithm is typically computationally efficient, because there is no need to evaluate the expected performance estimator at every point θ. In each iteration, only one simulation observation of $\mathbf{X}(\theta)$ and two evaluations of $G(\cdot)$ are enough. Moreover, the convergence rate is of polynomial order (Nemirovsky and Yudin, 1983). However, it is also possible that the algorithm may perform poorly in some problems, because it may be sensitive to the choice of the gain sequence $\{a_j : j = 1, 2, \ldots\}$ and the starting point θ_1. If this is the case, we suggest using common random numbers to introduce a positive correlation between $G\big(\hat{\theta}_n(\mathbf{X}(\theta_j))\big)$ and $G(\hat{\theta}_n)$, so as to reduce the variance of their difference.

3. Properties of SBE

In this section we analyze the consistency, bias and variance of SBEs. To simplify the explanation and analysis, we assume that the closed-form expression of $p(\cdot)$ is available throughout Section 3. If $p(\cdot)$ is obtained from simulation, the theoretical results hold as long as the simulation is replicated enough times such that the simulation uncertainty can vanish. To concentrate on the main idea we only consider the one-dimensional case, i.e., θ is a scalar, throughout the paper.

3.1. Consistency

To consider the consistency of SBEs, we make the following assumptions.

ACCEPTED MANUSCRIPT

Assumption 3.1. *There exists an open interval* $\Theta_0 = (a, b)$ *with* $a, b \in \mathbb{R}$ *and* $a < b$ *such that* $\theta_0 \in \Theta_0$.

Assumption 3.1 is generally a weak assumption, because it only requires that we have a prior knowledge on the range of θ_0, where the range may be very wide as long as it is not unbounded. In practice, system analysts often have some knowledge about the parameter of the system. For instance, the demand of a product or the arrival rate of customers cannot be infinity.

Assumption 3.2. *The performance function* $p(\cdot)$ *is Lipschitz continuous and strictly monotone on* Θ_0.

When $p(\cdot)$ is available, Assumption 3.2 is easy to verify. Even when $p(\cdot)$ is not available, Assumption 3.2 can sometimes be verified based on system knowledge. For instance, average queue length is typically continuous and strictly monotone with respect to the average arrival rate.

Assumption 3.3. *The function* $b_n(\cdot)$ *is Lipschitz continuous and strictly monotone.*

When $p(\cdot)$ is Lipschitz continuous and strictly monotone, it may be possible to verify the continuity and monotonicity of $b_n(\cdot)$. Here we show one typical case that the assumption holds. Suppose $\mathbf{X}(\theta)$ can be expressed as $\mathbf{X}(\theta, \omega)$. Then, for a fixed value of ω, if $\hat{\theta}_n(\mathbf{X}(\theta, \omega))$ is Lipschitz continuous and monotone with respect to θ (e.g., $\mathbf{X}(\theta, \omega)$ follows exponential distribution or normal distribution), by the continuity and strict monotonicity of $p(\cdot)$, $b_n(\cdot)$ is also Lipschitz continuous and strictly monotone.

Let $b_n(\Theta_0) = \{b_n(\theta) : \theta \in \Theta_0\}$ be the range of $b_n(\cdot)$. Then, we have the following lemma.

Lemma 3.1. *Suppose that Assumptions 3.1 to 3.3 hold. If* $p(\hat{\theta}_n) \in b_n(\Theta_0)$, *then* $\tilde{\theta}_n$ *exists and is unique, and* $\tilde{\theta}_n \in \Theta_0$.

The proof of Lemma 3.1 is quite straightforward. By Assumption 3.3, $b_n(\cdot)$ is strictly monotone and, thus, invertible. Therefore, the solution to the root-finding problem (1) is $\tilde{\theta}_n = b_n^{-1}(p(\hat{\theta}_n))$, and is existing and unique.

By Newey (1991), a sequence of random functions $\{Y_1(\theta), \ldots, Y_n(\theta)\}$ is said to converge to a function $y(\theta)$ in probability uniformly in $\theta \in \Theta$ if $\sup_{\theta\in\Theta} \mathbb{P}(|Y_n(\theta) - y(\theta)| > \epsilon) \to 0$ as $n \to \infty$ for any $\epsilon > 0$. In the next assumption, we assume that the estimator $\hat{\theta}_n(\mathbf{X}(\theta))$ is uniformly convergent.

ACCEPTED MANUSCRIPT

Assumption 3.4. *The estimator $\hat{\theta}_n(\mathbf{X}(\theta))$ converges to θ in probability uniformly in $\theta \in \Theta_0$ as $n \to \infty$.*

The uniform convergence in probability can be easily verified for some frequently used distributions. For instance, let $\hat{\theta}_n(\mathbf{X}(\theta))$ be the MLE of the mean of an exponential distribution with mean equal to θ. We have

$$\begin{aligned}\sup_{\theta \in (a,b)} \mathbb{P}\left\{\left|\hat{\theta}_n(\mathbf{X}(\theta)) - \theta\right| > \epsilon\right\} &= \sup_{\theta \in (a,b)} \mathbb{P}\left\{\left|\left(-\frac{1}{n}\sum_{i=1}^{n}\log(\omega_i) - 1\right)\theta\right| > \epsilon\right\} \\ &= \mathbb{P}\left\{\left|\left(-\frac{1}{n}\sum_{i=1}^{n}\log(\omega_i) - 1\right)\right| > \frac{\epsilon}{b}\right\} \to 0,\end{aligned}$$

because $-\frac{1}{n}\sum_{i=1}^{n}\log(\omega_i)$ converges to 1 in probability by the weak law of large numbers (Feller, 1968). Therefore, $\hat{\theta}_n(\mathbf{X}(\theta))$ converges to θ in probability uniformly. Generally, it can also be verified that under some conditions, the MLE of parameters of an exponential family of distributions converge in probability uniformly. The details are included in the Appendix.

Assumption 3.5. *For some $r > 0$,*

$$\lim_{n\to\infty} \sup_{\theta \in \Theta_0} \mathrm{E}_{\mathbf{X}}\left[\left|\hat{\theta}_n(\mathbf{X}(\theta))\right|^{1+r}\right] < \infty.$$

Assumption 3.5 guarantees the uniform integrability of $\hat{\theta}_n(\mathbf{X}(\theta))$ over all n. Then, we have the following lemma on the uniform convergence of $b_n(\theta)$.

Lemma 3.2. *Suppose Assumptions 3.2, 3.4 and 3.5 hold. Then, $b_n(\theta) \to p(\theta)$ uniformly on Θ_0 as $n \to \infty$.*

Notice that $b_n(\theta) - p(\theta)$ is the bias of $p\big(\hat{\theta}_n(\mathbf{X}(\theta))\big)$ when the input parameter is θ. Lemma 3.2 shows that $b_n(\theta)$ converges to $p(\theta)$ and the bias vanishes as n gets large. Then, we have the following theorem on the consistency of the SBEs $\tilde{\theta}_n$ and $p(\tilde{\theta}_n)$.

Theorem 3.1. *Suppose that Assumptions 3.1 to 3.5 hold and $p(\hat{\theta}_n) \in b_n(\Theta_0)$. Then, $\tilde{\theta}_n \to \theta_0$ and $p(\tilde{\theta}_n) \to p(\theta_0)$ in probability as $n \to \infty$.*

Theorem 3.1 illustrates that, if the two-step estimators themselves are consistent, the SBEs keep the consistency.

3.2. Bias and Variance

Throughout this paper, we follow the definitions of Lehmann (1999) when using the notation of $o(\cdot)$ and $O(\cdot)$. In specific, for two sequences of numbers a_n and b_n, we write $a_n = o(b_n)$ if $|a_n/b_n| \to 0$, which states that for large n, a_n is of order smaller than that of b_n. And we write $a_n = O(b_n)$ if $|a_n/b_n|$ is bounded, which states that a_n is of order smaller than or equal to that of b_n.

We make the following assumption on the existence of an asymptotic expansion of $\hat{\theta}_n(\mathbf{X}(\theta))$.

Assumption 3.6. *The estimator $\hat{\theta}_n(\mathbf{X}(\theta))$ admits the following asymptotic expansion:*

$$\hat{\theta}_n(\mathbf{X}(\theta)) = \theta + \frac{A(\theta)}{n^\alpha} + \frac{B(\theta)}{n^{2\alpha}} + o(n^{-2\alpha}), \tag{6}$$

where $\alpha \in (0, +\infty)$, $A(\theta)$ and $B(\theta)$ are random terms with the parameter θ, $A(\theta)$ is differentiable and $B(\theta)$ is continuous with respect to θ, and $A(\theta)$ has a finite second moment at θ_0 and $B(\theta)$ has a finite first moment at θ_0.

Many estimators have asymptotic expansions in the form of Equation (6) under some regularity conditions, see Hall (1992), Gouriéroux and Monfort (1995), Kolassa (1997) and Gouriéroux et al. (2000). For instance, if $\hat{\theta}_n(\mathbf{X}(\theta))$ is an MLE, which is consistent and asymptotically normal under some conditions (Newey and McFadden, 1994), $\hat{\theta}_n(\mathbf{X}(\theta))$ can often be expanded as

$$\hat{\theta}_n(\mathbf{X}(\theta)) = \theta + \frac{\sigma Z}{\sqrt{n}} + o(n^{-1/2}), \tag{7}$$

where σ is the standard deviation of $\sqrt{n}\left[\hat{\theta}_n(\mathbf{X}(\theta)) - \theta\right]$ and Z is a standard normal random variable. Let $p'(\cdot)$ denote the first-order derivative of $p(\cdot)$. If $p(\cdot)$ is differentiable, by Assumption 3.6 and Taylor's Theorem, we have

$$p\left(\hat{\theta}_n(\mathbf{X}(\theta))\right) = p(\theta) + p'(\theta)\frac{A(\theta)}{n^\alpha} + \left[p'(\theta)\frac{B(\theta)}{n^{2\alpha}} + \frac{1}{2}p''(\theta)\frac{A(\theta)^2}{n^{2\alpha}}\right] + o(n^{-2\alpha}).$$

We can apply the above expansions to estimators $\hat{\theta}_n$ and $p(\hat{\theta}_n)$ as well. Then, we have

$$\hat{\theta}_n = \theta_0 + \frac{A_0(\theta_0)}{n^\alpha} + \frac{B_0(\theta_0)}{n^{2\alpha}} + o(n^{-2\alpha}), \tag{8}$$

$$p(\hat{\theta}_n) = p(\theta_0) + p'(\theta_0)\frac{A_0(\theta_0)}{n^\alpha} + \left[p'(\theta_0)\frac{B_0(\theta_0)}{n^{2\alpha}} + \frac{1}{2}p''(\theta_0)\frac{A_0(\theta_0)^2}{n^{2\alpha}}\right] + o(n^{-2\alpha}), \tag{9}$$

where $A_0(\theta_0)$ and $B_0(\theta_0)$ are identically distributed as $A(\theta_0)$ and $B(\theta_0)$. We add the subscript 0 to A and B to denote that the randomness of $A_0(\theta_0)$ and $B_0(\theta_0)$ comes from the observed data

$\mathbf{X}_0$. The randomness of $A(\theta)$ and $B(\theta)$ for a given θ, on the other hand, comes from the simulated data $\mathbf{X}(\theta)$. According to the Equation (9), the bias[1] of $p(\hat{\theta}_n)$ is

$$\begin{aligned}\mathrm{E}_0\left[p(\hat{\theta}_n)-p(\theta_0)\right] &= p'(\theta_0)\mathrm{E}_0\left[A_0(\theta_0)\right]\frac{1}{n^{\alpha}} \\ &+\left\{p'(\theta_0)\mathrm{E}_0\left[B_0(\theta_0)\right]+\frac{1}{2}p''(\theta_0)\mathrm{E}_0\left[A_0(\theta_0)^2\right]\right\}\frac{1}{n^{2\alpha}}+o(n^{-2\alpha}),\end{aligned} \tag{10}$$

where E_0 denotes that the expectation is taken with respect to the observed data $\mathbf{X}_0$. Notice that E_0 is different from $\mathrm{E}_{\mathbf{X}}$. By Equation (10), if $\mathrm{E}_0\left[A_0(\theta_0)\right]\neq 0$, the bias of $p(\hat{\theta}_n)$ is $O(n^{-\alpha})$. If $\mathrm{E}_0\left[A_0(\theta_0)\right]=0$, e.g., when $\hat{\theta}_n$ is a consistent MLE and $p'(\theta_0)\mathrm{E}_0\left[B_0(\theta_0)\right]+\frac{1}{2}p''(\theta_0)\mathrm{E}_0\left[A_0(\theta_0)^2\right]\neq 0$, the bias of $p(\hat{\theta}_n)$ is $O(n^{-2\alpha})$.

We summarize the properties of the SBE $p(\tilde{\theta}_n)$ in the following theorem.

Theorem 3.2. *Suppose that Assumptions 3.1 to 3.6 hold, $p(\cdot)$ has continuous third derivative and $p'(\theta_0)\neq 0$. Then, $p(\tilde{\theta}_n)$ admits the following asymptotic expansion:*

$$p(\tilde{\theta}_n)=p(\theta_0)+p'(\theta_0)\frac{A_0(\theta_0)-\mathrm{E}_{\mathbf{X}}[A(\theta_0)]}{n^{\alpha}}+\frac{S_0(\theta_0)}{n^{2\alpha}}+o(n^{-2\alpha}), \tag{11}$$

where $S_0(\theta_0)$ is a random term depending on θ_0. Furthermore, $p(\tilde{\theta}_n)$ has the following properties:

(a) $\mathrm{E}_0\left[p(\tilde{\theta}_n)\right]-p(\theta_0)=o(n^{-2\alpha})$;

(b) $\lim_{n\to\infty}\mathrm{Var}\left[p(\tilde{\theta}_n)\right]/\mathrm{Var}\left[p(\hat{\theta}_n)\right]=1$;

(c) If $A_0(\theta_0)$ follows a normal distribution, then

$$n^{\alpha}\left[p(\tilde{\theta}_n)-p(\theta_0)\right]\Rightarrow p'(\theta_0)\sqrt{\mathrm{Var}\left[A_0(\theta_0)\right]}\cdot Z$$

as $n\to\infty$, where "$\Rightarrow$" denotes convergence in distribution and Z is a standard normal random variable.

Theorem 3.2 shows that the bias of $p(\tilde{\theta}_n)$ is $o(n^{-2\alpha})$. Therefore, $p(\tilde{\theta}_n)$ is asymptotically less biased than the two-step estimator $p(\hat{\theta}_n)$. Furthermore, Theorem 3.2 also shows that the asymptotic variance of the SBE $p(\tilde{\theta}_n)$ is the same as that of the two-step estimator $p(\hat{\theta}_n)$. In other words, the bias reduction achieved by the SBE method is not accompanied by an increase in variance, which makes the bias reduction more meaningful.

Theorem 3.2 also provides the asymptotic distribution of the SBE $p(\tilde{\theta}_n)$ when $\hat{\theta}_n$ is asymptotically normally distributed. Notice that a variance estimator of $\hat{\theta}_n$ is often available. Let $\hat{\sigma}_n^2$ denote

[1]When analyzing the bias, we need the uniform integrability of the remainder. We follow the convention of some statistical literature such as Gouriéroux and Monfort (1995) and choose to ignore this issue for simplicity.

ACCEPTED MANUSCRIPT

an estimator of $\mathrm{Var}(\hat{\theta}_n)$. By the Delta method (Casella and Berger, 2002), the variance of $p(\hat{\theta}_n)$ can be estimated by $[p'(\hat{\theta}_n)]^2\hat{\sigma}_n^2$. Therefore, by Theorem 3.2, $\mathrm{Var}(\tilde{\theta}_n)$ may be estimated by $[p'(\hat{\theta}_n)]^2\hat{\sigma}_n^2$ as well. Then, an asymptotically valid $(1-\alpha)\times 100\%$ confidence interval of $p(\theta_0)$ is

$$\left(p(\tilde{\theta}_n) - z_{\alpha/2}p'(\hat{\theta}_n)\hat{\sigma}_n,\ p(\tilde{\theta}_n) + z_{\alpha/2}p'(\hat{\theta}_n)\hat{\sigma}_n\right).$$

Remark: The bias of the SBE can be further reduced by keeping applying the SBE method to the SBE. The basic idea is to treat $\tilde{\theta}_n$ as $\hat{\theta}_n$ and get a step further SBE by solving the following equation $\mathrm{E}_{\mathbf{X}}[p(\tilde{\theta}_n(\theta))] = p(\tilde{\theta}_n)$, where $\tilde{\theta}_n(\theta)$ is the simulated SBE which is estimated from the simulated data $\mathbf{X}(\theta)$. If we apply the SBE method recursively, we can theoretically prove that the bias of the estimators will vanish eventually. However, the recursive implementation of the SBE method typically requires a prohibitively large amount of computation, it is in general difficult to implement the recursive SBE method in practice.

4. Numerical Results

In many service systems, such as restaurants, banks and hospitals, the average queue length is often used as a measure of service quality and thus used to determine the quantity of resources. Many of these systems may be described by a queueing model. In this section, we consider different queueing models based on whether their closed-form expressions are available or not.

4.1. A closed-form expression of $p(\theta)$ is available

Consider a M/M/1/J queue, where J denotes the buffer size. Let the service rate $\mu = 1$ and buffer size $J = 100$. The arrival rate θ is estimated from observed inter-arrival times. The long run average queue length is denoted by $p(\theta)$, which has a closed-form expression by Allen (1990).

Suppose the true arrival rate θ_0 is 0.9, then, the true value $p(\theta_0)$ is 8.098. Let the sample size $n = 500$. We replicate the experiment 10000 times, and plot in Figure 2 the histograms of both the MLEs $\hat{\theta}_n$ and $p(\hat{\theta}_n)$ and the SBEs $\tilde{\theta}_n$ and $p(\tilde{\theta}_n)$. From the histograms, we see that the MLE $\hat{\theta}_n$ is almost unbiased and the estimates are distributed symmetrically around the true value $\theta_0 = 0.9$. However, due to the nonlinearity of the performance function $p(\cdot)$, the performance estimator $p(\hat{\theta}_n)$

is right skewed. This causes $p(\hat{\theta}_n)$ to be heavily biased. After applying the SBE method, the estimator $\tilde{\theta}_n$ moves to the left and is biased low, but the performance estimator $p(\tilde{\theta}_n)$ is less skewed and less biased.

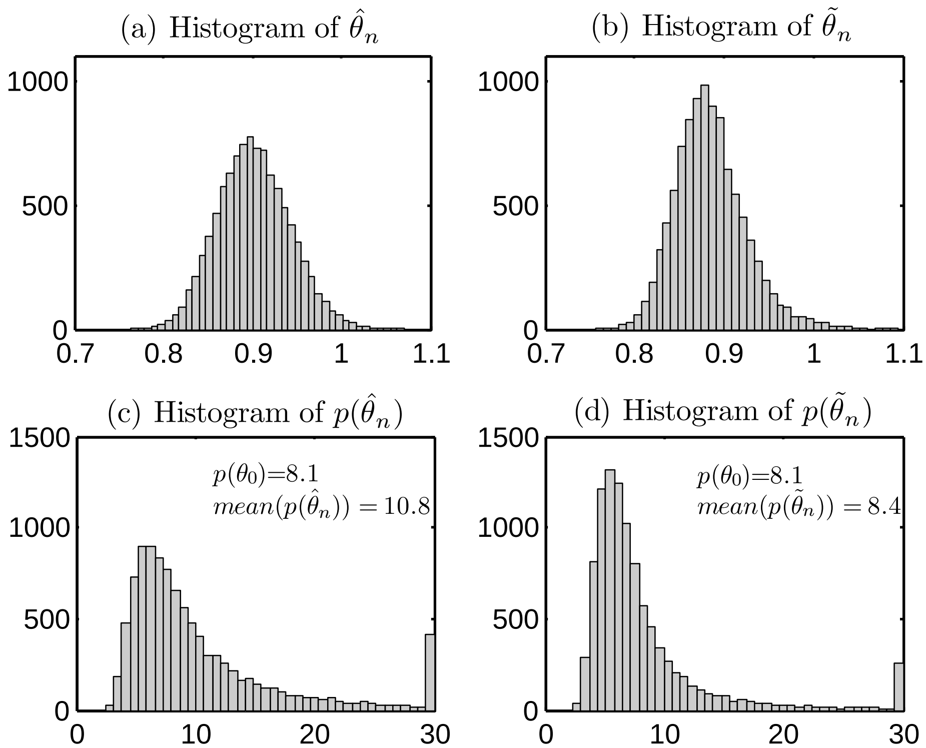

Figure 2: Histograms of the MLEs and SBEs

We then compare the SBE method with other bias reduction methods. The jackknife method estimates $p(\theta_0)$ by p_n^J, where

$$p_n^J = np(\hat{\theta}_n) - \frac{n-1}{n}\sum_{i=1}^{n} p(\hat{\theta}_n^i), \tag{12}$$

and $\hat{\theta}_n^i$ is estimated from the data with the ith observation removed. The parametric bootstrap method estimates the bias by

$$\text{Bias}^* = \frac{1}{B}\sum_{b=1}^{B} p(\hat{\theta}_n(\mathbf{X}_b(\hat{\theta}_n))) - p(\hat{\theta}_n),$$

where $\mathbf{X}_b(\hat{\theta}_n)$ denotes the data sampled (or simulated) from the distribution with the parameter $\hat{\theta}_n$ and b denotes the bth sample. Notice that $\frac{1}{B}\sum_{b=1}^{B} p(\hat{\theta}_n(\mathbf{X}_b(\hat{\theta}_n)))$ is actually a sample average

Table 1: Comparison of MLEs, JK estimators, Bootstrap estimators and SBEs

Sample size		500	650	800	1000	1250	1500	2000	2500	3000
mean(θ)	MLE	0.9020	0.9011	0.9016	0.9008	0.9008	0.9009	0.9004	0.9005	0.9005
	SBE	0.8843	0.8871	0.8900	0.8914	0.8932	0.8946	0.8957	0.8968	0.8973
std(θ)	MLE	0.040	0.036	0.032	0.029	0.026	0.023	0.021	0.018	0.016
	SBE	0.039	0.034	0.030	0.027	0.024	0.022	0.020	0.018	0.016
mean(p)	MLE	10.796	10.076	9.708	9.214	8.977	8.835	8.627	8.513	8.435
	JK	7.450	7.698	7.716	7.842	7.906	7.949	8.018	8.056	8.054
	Boots	7.467	7.524	7.484	7.698	7.817	7.892	7.950	8.006	8.046
	AG	7.478	7.621	7.755	7.839	7.927	7.986	8.036	8.058	8.067
	SBE	8.442	8.238	8.179	8.029	8.047	8.074	8.081	8.091	8.092
Relative Bias (%)	MLE	33.32	24.43	19.88	13.79	10.86	9.10	6.53	5.13	4.17
	JK	8.00	4.94	4.72	3.16	2.37	1.84	0.99	0.52	0.54
	Boots	7.79	7.09	7.58	4.94	3.47	2.54	1.83	1.14	0.64
	AG	7.65	5.89	4.23	3.19	2.10	1.37	0.76	0.48	0.38
	SBE	4.26	1.73	1.01	0.84	0.62	0.30	0.20	0.08	0.06
std(p)	MLE	8.58	6.98	5.64	4.35	3.60	3.06	2.58	2.11	1.86
	JK	7.77	5.75	4.18	3.14	2.78	2.38	2.02	1.81	1.67
	Boots	8.15	5.85	4.18	3.38	2.79	2.43	2.04	1.81	1.67
	AG	7.55	5.05	3.93	3.19	2.69	2.39	2.04	1.84	1.67
	SBE	7.44	5.90	4.55	3.46	2.89	2.50	2.21	1.86	1.67
MSE(p)	MLE	80.896	52.633	34.402	20.168	13.733	9.907	6.936	4.624	3.573
	JK	60.793	33.223	17.618	9.925	7.765	5.687	4.087	3.278	2.791
	Boots	66.821	34.552	17.849	11.584	7.863	5.947	4.184	3.285	2.792
	AG	57.387	25.730	15.563	10.243	7.265	5.725	4.165	3.387	2.790
	SBE	55.472	34.830	20.709	11.976	8.355	6.251	4.884	3.460	2.789
Coverage Probability (%)	MLE	90.03	90.55	91.46	91.70	92.44	92.79	93.14	93.44	93.85
	JK	87.21	88.19	89.10	89.72	90.43	91.10	91.68	92.36	92.64
	Boots	94.41	93.96	93.92	94.05	93.56	93.83	93.86	93.77	94.38
	AG	86.79	88.07	88.99	89.42	90.33	90.99	91.81	92.18	92.72
	SBE	87.50	87.67	88.93	90.66	90.43	91.09	91.64	92.09	92.65
Computation Cost (seconds)	MLE	- -	- -	- -	- -	- -	- -	- -	- -	- -
	JK	0.03	0.03	0.05	0.07	0.10	0.14	0.22	0.34	0.46
	Boots	0.35	0.39	0.43	0.49	0.55	0.63	0.77	0.91	1.05
	AG	- -	- -	- -	- -	- -	- -	- -	- -	- -
	SBE	1.20	1.20	1.23	1.22	1.28	1.31	1.35	1.44	1.60

The experiment is run on a computer with 3.40 GHz Intel i7-2600 Processor. The computation cost is measured by the average computation time of each run. The “- -” represents that the computation time is smaller than 0.001 second.

ACCEPTED MANUSCRIPT

approximation of $b_n(\hat{\theta}_n)$. The corrected performance estimator can be written as:

$$p_n^* = p(\hat{\theta}_n) - \text{Bias}^* = 2p(\hat{\theta}_n) - b_n(\hat{\theta}_n). \tag{13}$$

We call the estimator of the bias correction method of Asmussen and Glynn (2007) the AG estimator. If the input parameter estimator is unbiased, then the AG estimator p_n^{AG} is

$$p_n^{AG} = p(\hat{\theta}_n) - \frac{1}{2}p''(\hat{\theta}_n)\hat{\sigma}_n,$$

where $\hat{\sigma}_n^2$ is an estimator of $\text{Var}(\hat{\theta}_n)$.

We repeat the experiments 10000 times to calculate the means and standard deviations of different estimators and report them in Table 1. The SBEs are obtained by applying the SAA method. The results show that the SBEs have largest bias reduction compared with other methods. To analyze the rate of convergence of the biases of each estimators, we apply linear regression to log(|bias|) with respect to log(n), the results are plotted in Figure 3 and listed as follows:

$$\log\left(|\text{bias}(p(\hat{\theta}_n))|\right) = -1.16\log(n) + 8.20,$$

$$\log\left(|\text{bias}(p_n^J)|\right) = -1.59\log(n) + 9.52,$$

$$\log\left(|\text{bias}(p_n^*)|\right) = -1.41\log(n) + 8.66,$$

$$\log\left(|\text{bias}(p_n^{AG})|\right) = -1.77\log(n) + 10.69,$$

$$\log\left(|\text{bias}(p(\tilde{\theta}_n))|\right) = -2.23\log(n) + 12.71.$$

The regression results indicate that the bias of MLEs is roughly $O(n^{-1})$ and the bias of SBEs is roughly $O(n^{-2})$. This is consistent with the asymptotical results reported in Theorem 3.2, where $\alpha = 1/2$. Moreover, the biases of jackknife estimators, bootstrap estimators and AG estimators are roughly $O(n^{-1.5})$, which outperform that of MLEs, but are not as good as that of SBEs.

In addition to the significant bias reduction, the results also show that the variance of the SBEs does not increase compared with that of the MLEs. The MSE of SBEs is smaller than that of the MLEs but is slightly greater than those of other estimators. The coverage probability of the 95% confidence interval built based on the SBEs appears a little smaller than that of the MLEs. Considering the fact that the confidence interval is constructed based on the asymptotic analysis,

ACCEPTED MANUSCRIPT

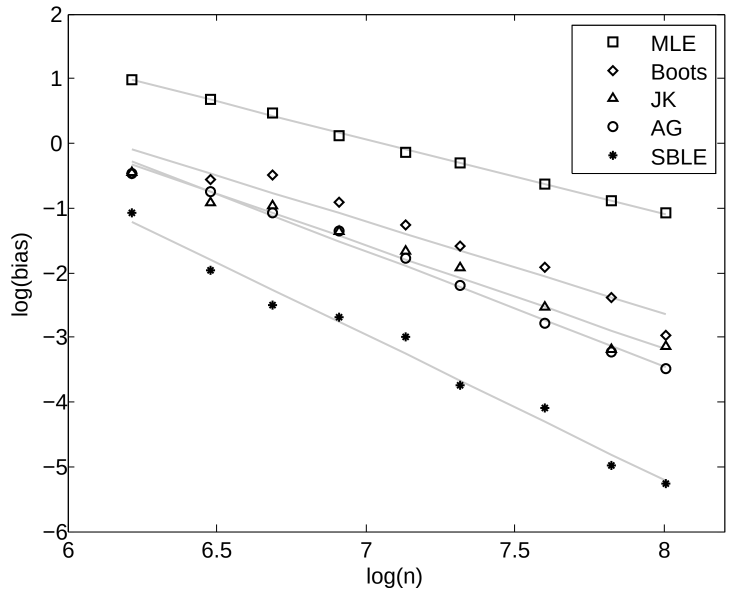

Figure 3: Bias Reduction Rate of multiple methods

it is possible that the insufficient coverage probability is caused by the small sample size. In the table we do observe that the coverage probability increases as the sample size increases.

4.2. Lack of a Closed-Form Expression of $p(\cdot)$

We consider a M/M/1 queueing model, which starts empty and the performance measure $p(\cdot)$ that we are interested in is the expected average waiting time for the first m customers, which can be expressed as follows:

$$p(\theta) = \mathrm{E}\left[\frac{1}{m}\sum_{i=1}^{m} W_i(\theta)\right],$$

where $W_i(\theta)$ is the waiting time of the i_{th} customer when the arrival rate is θ. The $W_i(\theta)$ can be simulated according to $W_i(\theta) = \max(0, W_{i-1}(\theta) + S_{i-1} - X_i(\theta))$, where S_{i-1} is the service time of the $(i-1)_{th}$ customer and $X_i(\theta)$ is the inter-arrival time of the i_{th} customer.

Let the service rate be 1 and $m = 1000$. The true value of the arrival rate is 0.9 and we repeat the simulation for 10^9 times to evaluate the true value $p(\theta_0)$, which is approximately 8.1106 with 95% confidence interval being $(8.1103, 8.1109)$. Once the estimate of θ_0, say $\hat{\theta}_n$ or $\tilde{\theta}_n$, is obtained, we repeat simulating the queueing system for 1000 times to get the performance estimator $p(\hat{\theta}_n)$ or $p(\tilde{\theta}_n)$. We repeat the numerical experiment for 10000 times to obtain the bias and standard

ACCEPTED MANUSCRIPT

Table 2: Comparison of the MLEs, the SBEs and other estimators

Sample size		500	750	1000	1250	1500
mean(θ)	MLE	0.9017	0.9004	0.9009	0.9008	0.9004
	SAA	0.8901	0.8925	0.8949	0.8960	0.8963
	SA	0.8904	0.8929	0.8953	0.8958	0.8963
std(θ)	MLE	0.040	0.033	0.029	0.025	0.0232
	SAA	0.040	0.033	0.029	0.025	0.0233
	SA	0.041	0.034	0.030	0.0266	0.0242
mean(p)	MLE	9.0861	8.7096	8.5919	8.5013	8.4120
	JK	9.5220	9.9936	7.8608	15.741	7.3440
	Boots	8.0101	8.0115	8.0757	8.0964	8.0996
	SAA	8.1686	8.0644	8.1330	8.1244	8.1040
	SA	8.2692	8.1715	8.1691	8.1538	8.1279
Relative Bias (%)	MLE	12.028	7.3854	5.9342	4.8172	3.7161
	JK	17.401	23.216	3.0799	94.080	9.4800
	Boots	1.2391	1.2219	0.4303	0.1751	0.1356
	SAA	0.7151	0.5696	0.2762	0.1701	0.0814
	SA	1.9555	0.7509	0.7213	0.5326	0.2133
std(p)	MLE	3.823	2.902	2.419	2.109	1.886
	JK	82.42	175.4	225.9	269.2	320.8
	Boots	3.424	2.646	2.255	1.995	1.802
	SAA	3.395	2.648	2.259	1.987	1.792
	SA	3.673	2.842	2.402	2.127	1.870
MSE(p)	MLE	15.567	8.780	6.083	4.601	3.648
	JK	- -	- -	- -	- -	- -
	Boots	11.734	7.011	5.086	3.980	3.247
	SAA	11.529	7.014	5.104	3.948	3.211
	SA	13.516	8.081	5.773	4.526	3.497
Coverage Probability (%)	MLE	91.91	91.95	92.98	93.63	93.74
	JK	78.35	63.28	51.13	40.41	35.16
	Boots	93.39	92.66	93.08	93.79	93.61
	SAA	89.53	89.76	91.33	92.08	92.26
	SA	89.53	89.75	91.15	91.51	91.62

The "- -" represents that the MSEs are greater than 5000.

deviations of the estimators. The numerical results are reported in Table 2.

Table 2 shows that the SBE method and the parametric bootstrap method can both reduce the bias of the performance estimators. The SAA method can outperform the parametric bootstrap method. The SA method can also reduce the bias, but is not as good as the SAA method and the bootstrap method. This is mainly because the Robbins-Monro algorithm brings errors while finding the root. The Jackknife method is not applicable in this example because the noise involved from the evaluation of the function $p(\cdot)$ is enlarged, according to the Equation (12).

ACCEPTED MANUSCRIPT

5. Conclusions and Future Research

In this paper, we extend the SBE method to general problems for bias reduction of the performance estimators. The estimators are obtained by solving a stochastic root-finding problem. We prove that after applying the SBE method, the estimators are consistent and the bias of the performance estimators may be reduced to smaller orders of the sample size. Furthermore, the variance of the SBEs may not necessarily increase and can often be well approximated. Numerical studies verify the theoretical results and show that the SBE method works well for practical problems.

The SBE method also has some weaknesses. The SBE method introduces extra simulation cost, which may be a severe problem when the simulation is expensive. Therefore, in terms of bias reduction, there is a tradeoff between the cost of using simulation and the cost of collecting more data. The results on consistency and bias reduction developed in this paper require the existence and uniqueness of the root in the stochastic root-finding problem which, unfortunately, may not always hold for practical problems. Furthermore, in this paper, we only study the case where the parameter is of one dimension and the performance function is of one dimension as well. Even though we do not study the multi-dimensional cases in this paper, we believe that the SBE method may be extended and we leave them to future studies.

Acknowledgement

The authors would like to thank the associate editor and two anonymous referees for their insightful and helpful comments that have improved this work a lot. A preliminary version of this paper (Fang and Hong, 2013) was published in the *Proceedings of the* 2013 *Winter Simulation Conference*. This research was supported in part by the Hong Kong Research Grants Council [GRF 613213, GRF 11270116 and T32-102/14N].

ACCEPTED MANUSCRIPT

Appendix A.

Appendix A.1. Proof of Lemma 3.2

To prove $b_n(\theta)$ converges to $p(\theta)$ uniformly on Θ_0 is to show that

$$\lim_{n\to\infty}\sup_{\theta\in\Theta_0}\left|\mathrm{E}_{\mathbf{X}}\left[p\big(\hat\theta_n(\mathbf{X}(\theta))\big)\right]-p(\theta)\right|=0.$$

Since $p(\cdot)$ is Lipschitz continuous, there exists a constant Γ such that

$$\sup_{\theta\in\Theta_0}\left|\mathrm{E}_{\mathbf{X}}\left[p\big(\hat\theta_n(\mathbf{X}(\theta))\big)\right]-p(\theta)\right|\le\sup_{\theta\in\Theta_0}\Gamma\cdot\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))-\theta\right|\right].$$

Therefore, we only need to show that $\lim\limits_{n\to\infty}\sup\limits_{\theta\in\Theta_0}\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))-\theta\right|\right]=0.$

For any $M>0$, we define an auxiliary function $\varphi_M(x)$ to be M if $x>M$, to be x if $|x|\le M$ and to be $-M$ if $x<-M$. Then, for all $\theta\in\Theta_0$,

$$\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))-\theta\right|\right]\le\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))-\varphi_M\big(\hat\theta_n(\mathbf{X}(\theta))\big)\right|\right]+\mathrm{E}_{\mathbf{X}}\left[\left|\varphi_M\big(\hat\theta_n(\mathbf{X}(\theta))\big)-\varphi_M(\theta)\right|\right]+\mathrm{E}_{\mathbf{X}}\left[\,|\varphi_M(\theta)-\theta|\,\right]. \tag{A.1}$$

The first term of the right hand side of (A.1) satisfies

$$\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))-\varphi_M\big(\hat\theta_n(\mathbf{X}(\theta))\big)\right|\right]\le\frac{1}{M^r}\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))\right|^{1+r}\mathbf{1}_{\{|\hat\theta_n(\mathbf{X}(\theta))|>M\}}\right]\ \le\ \frac{1}{M^r}\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))\right|^{1+r}\right].$$

By Assumption 3.5, there exists a constant $B>0$ such that $\lim\limits_{n\to\infty}\sup\limits_{\theta\in\Theta_0}\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))\right|^{1+r}\right]\le B$. Then,

$$\lim_{n\to\infty}\sup_{\theta\in\Theta_0}\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))-\varphi_M\big(\hat\theta_n(\mathbf{X}(\theta))\big)\right|\right]\le\frac{B}{M^r}. \tag{A.2}$$

For any given $\epsilon>0$, the second term satisfies

$$\begin{aligned}\mathrm{E}_{\mathbf{X}}\left[\left|\varphi_M\big(\hat\theta_n(\mathbf{X}(\theta))\big)-\varphi_M(\theta)\right|\right]&\le2M\mathbb{P}\left\{\left|\varphi_M\big(\hat\theta_n(\mathbf{X}(\theta))\big)-\varphi_M(\theta)\right|>\epsilon\right\}+\epsilon\mathbb{P}\left\{\left|\varphi_M\big(\hat\theta_n(\mathbf{X}(\theta))\big)-\varphi_M(\theta)\right|\le\epsilon\right\}\\&\le2M\mathbb{P}\left\{|\hat\theta_n(\mathbf{X}(\theta))-\theta|>\frac{\epsilon}{\Gamma_1}\right\}+\epsilon,\end{aligned} \tag{A.3}$$

where the inequality (A.3) holds because $\varphi_M(\cdot)$ is Lipschitz continuous. Then,

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\mathrm{E}_{\mathbf{X}}\left[\left|\varphi_M\big(\hat\theta_n(\mathbf{X}(\theta))\big)-\varphi_M(\theta)\right|\right]\le\epsilon, \tag{A.4}$$

where (A.4) holds because of the uniform convergence of $\hat\theta_n(\mathbf{X}(\theta))$. Then, by (A.1), (A.2) and (A.4), we have

$$\lim_{n\to\infty}\sup_{\theta\in\Theta_0}\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))-\theta\right|\right]\le\frac{B}{M^r}+\epsilon+\lim_{n\to\infty}\sup_{\theta\in\Theta_0}\mathrm{E}_{\mathbf{X}}\left[\,|\varphi_M(\theta)-\theta|\,\right]. \tag{A.5}$$

Since M can be arbitrarily large and ϵ can be arbitrarily small, by (A.5), we have

$$\lim_{n\to\infty}\sup_{\theta\in\Theta_0}\mathrm{E}_{\mathbf{X}}\left[\left|\hat\theta_n(\mathbf{X}(\theta))-\theta\right|\right]=0.$$

ACCEPTED MANUSCRIPT

Appendix A.2. Proof of Theorem 3.1

We first prove the uniform convergence of $b_n^{-1}(\cdot)$ to $p^{-1}(\cdot)$. Notice that $p(\cdot)$ is invertible because $p(\theta)$ is continuous and strictly monotone. By Assumption 3.3, $b_n(\cdot)$ is also invertible. Theorem 1 of Barvínek et al. (1991) states that:

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of real injection functions defined on $(a, b) \subseteq \bigcap_{n=1}^{\infty} Dom f_n$. If the sequence converges uniformly to a function f_0 on this interval, and if f_0 is a continuous injection on (a, b) and $(\alpha, \beta) \subseteq \bigcap_{n=0}^{\infty} f_n((a, b))$, then f_n^{-1} converges uniformly to f_0^{-1} on (α, β).

Denote the range of $p(\theta)$ on the domain Θ_0 by $p(\Theta_0)$. When Θ_0 is an open interval, $p(\Theta_0)$ is also an open interval. We denote $p(\Theta_0)$ by (α, β). Lemma 3.2 shows that $b_n(\cdot)$ converges to $p(\cdot)$ uniformly on Θ_0, then, for any $\epsilon_1 > 0$ there exists N_1 such that for all $n > N_1$, we have $\sup_{\theta\in\Theta_0} |b_n(\theta) - p(\theta)| < \epsilon_1$. Then, combining with the continuity of $b_n(\cdot)$ assumed in Assumption 3.3, we have $(\alpha + \epsilon_1, \beta - \epsilon_1) \subseteq \bigcap_{n=N_1}^{\infty} b_n(\Theta_0)$.

Notice that we can choose ϵ_1 small enough to guarantee that $\alpha+\epsilon_1 < \beta-\epsilon_1$. Then, by applying Theorem 1 of Barvínek et al. (1991) directly, we can show that $b_n^{-1}(\cdot)$ converges uniformly to $p^{-1}(\cdot)$ on $(\alpha + \epsilon_1, \beta - \epsilon_1)$, i.e., for any $\epsilon_2 > 0$, there exists N_2 such that for all $n > N_2$,

$$\sup_{x\in(\alpha+\epsilon_1,\beta-\epsilon_1)} |b_n^{-1}(x) - p^{-1}(x)| < \epsilon_2. \tag{A.6}$$

By Lemma 3.1, we have $\tilde{\theta}_n = b_n^{-1}(p(\hat{\theta}_n))$. Notice that

$$\left|\tilde{\theta}_n - \theta_0\right| = \left|b_n^{-1}(p(\hat{\theta}_n)) - \theta_0\right| \leq \left|b_n^{-1}(p(\hat{\theta}_n)) - p^{-1}(p(\hat{\theta}_n))\right| + \left|\hat{\theta}_n - \theta_0\right|. \tag{A.7}$$

Then, to prove that $\tilde{\theta}_n$ converges to θ_0 in probability, we only need to show that the two terms at the right hand side of (A.7) converge to 0 in probability respectively. First, consider the first term of the right hand side of (A.7). By (A.6), if $p(\hat{\theta}_n) \in (\alpha+\epsilon_1, \beta-\epsilon_1)$, then, $|b_n^{-1}(p(\hat{\theta}_n))-p^{-1}(p(\hat{\theta}_n))| < \epsilon_2$. Notice that $p(\theta_0) \in (\alpha+\epsilon_1, \beta-\epsilon_1)$ if ϵ_1 is small and $p(\hat{\theta}_n)$ converges to $p(\theta_0)$ in probability. Therefore, there exists $\epsilon_3 > 0$ such that $(p(\theta_0) - \epsilon_3, p(\theta_0) + \epsilon_3) \subseteq (\alpha + \epsilon_1, \beta - \epsilon_1)$ and for any $\epsilon > 0$, there exists

N_3 such that for any $n > N_3$, $\mathbb{P}\left\{|p(\hat{\theta}_n) - p(\theta_0)| > \epsilon_3\right\} < \epsilon$. Choose $N = \max(N_1, N_2, N_3)$, then for all $n > N$, we have

$$\begin{aligned}\mathbb{P}\left\{|b_n^{-1}(p(\hat{\theta}_n)) - p^{-1}(p(\hat{\theta}_n))| > \epsilon_2\right\} &\leq \mathbb{P}\left\{p(\hat{\theta}_n) \notin (\alpha + \epsilon_1, \beta - \epsilon_1)\right\} \\ &\leq \mathbb{P}\left\{p(\hat{\theta}_n) \notin (p(\theta_0) - \epsilon_3, p(\theta_0) + \epsilon_3)\right\} < \epsilon. \qquad (A.8)\end{aligned}$$

Therefore, $|b_n^{-1}(p(\hat{\theta}_n)) - p^{-1}(p(\hat{\theta}_n))|$ converges to 0 in probability.

The second term of the right hand side of (A.7) converges to 0 in probability by Assumption 3.4. Therefore, $\tilde{\theta}_n$ converges to θ_0 in probability. Then, $p(\tilde{\theta}_n)$ converges to $p(\theta_0)$ in probability as $n \to \infty$ as well.

Appendix A.3. Proof of Theorem 3.2

By Assumption 3.6, we have the following asymptotic expansion for $\hat{\theta}_n(\mathbf{X}(\tilde{\theta}_n))$:

$$\hat{\theta}_n(\mathbf{X}(\tilde{\theta}_n)) = \tilde{\theta}_n + \frac{A(\tilde{\theta}_n)}{n^\alpha} + \frac{B(\tilde{\theta}_n)}{n^{2\alpha}} + o(n^{-2\alpha}).$$

Applying Taylor's theorem to Equation (1) and replacing $\hat{\theta}_n$ and $\hat{\theta}_n(\mathbf{X}(\tilde{\theta}_n))$ by their expansions gives

$$\begin{aligned}&\mathrm{E}_{\mathbf{X}}\left[p(\tilde{\theta}_n) + \left(\frac{A(\tilde{\theta}_n)}{n^\alpha} + \frac{B(\tilde{\theta}_n)}{n^{2\alpha}}\right)p'(\tilde{\theta}_n) + \frac{1}{2}\left(\frac{A(\tilde{\theta}_n)}{n^\alpha} + \frac{B(\tilde{\theta}_n)}{n^{2\alpha}}\right)^2 p''(\tilde{\theta}_n) + o(n^{-2\alpha})\right] \\ &= p(\theta_0) + \left(\frac{A_0(\theta_0)}{n^\alpha} + \frac{B_0(\theta_0)}{n^{2\alpha}}\right)p'(\theta_0) + \frac{1}{2}\left(\frac{A_0(\theta_0)}{n^\alpha} + \frac{B_0(\theta_0)}{n^{2\alpha}}\right)^2 p''(\theta_0) + o(n^{-2\alpha}). \qquad (A.9)\end{aligned}$$

Because $\tilde{\theta}_n$ is consistent, we can expand it as

$$\tilde{\theta}_n = \theta_0 + A_n + o(A_n), \qquad (A.10)$$

where A_n converges to 0 in probability as $n \to \infty$.

Now we apply the Taylor Expansion to $\tilde{\theta}_n$ at θ_0 to the order of $n^{-2\alpha}$ for the left hand side of Equation (A.9) and substitute $\tilde{\theta}_n$ by its expansion (A.10), the Equation (A.9) can be written as

$$\begin{aligned}&p(\theta_0) + A_n p'(\theta_0) + \frac{1}{2}A_n^2 p''(\theta_0) + \mathrm{E}_{\mathbf{X}}\left[\left(\frac{A(\theta_0)}{n^\alpha} + \frac{1}{n^\alpha}A'(\theta_0)A_n\right)\left(p'(\theta_0) + p''(\theta_0)A_n\right)\right] \\ &+\mathrm{E}_{\mathbf{X}}\left[\frac{B(\theta_0)}{n^{2\alpha}}p'(\theta_0)\right] + \frac{1}{2}\mathrm{E}_{\mathbf{X}}\left[\frac{A(\theta_0)^2}{n^{2\alpha}}p''(\theta_0)\right] + o(n^{-2\alpha}) + o(n^{-2\alpha}) + o(A_n) \\ &= p(\theta_0) + \left[\frac{A_0(\theta_0)}{n^\alpha} + \frac{B_0(\theta_0)}{n^{2\alpha}}\right]p'(\theta_0) + \frac{1}{2}\left[\frac{A_0(\theta_0)}{n^\alpha} + \frac{B_0(\theta_0)}{n^{2\alpha}}\right]^2 p''(\theta_0) + o(n^{-2\alpha}).\end{aligned}$$

Based on the above equation, we have

$$A_n p'(\theta_0) = \frac{A_0(\theta_0)}{n^\alpha}p'(\theta_0) - \frac{\mathrm{E}_{\mathbf{X}}[A(\theta_0)]}{n^\alpha}p'(\theta_0) + o_{p_0}(n^{-\alpha}) + o(A_n)$$

Hence, when $p'(\theta_0) \neq 0$, we conclude that A_n is of order $n^{-\alpha}$. More specifically, we have $A_n = (A_0(\theta_0) - \mathrm{E}_{\mathbf{X}}[A(\theta_0)])/n^{\alpha}$. Then, Equation (A.10) can be written as $\tilde{\theta}_n = \theta_0 + A^*/n^{\alpha} + o(n^{-\alpha})$, where

$$A^* = A_0(\theta_0) - \mathrm{E}_{\mathbf{X}}[A(\theta_0)]. \tag{A.11}$$

Now, we consider higher orders. We expand $\tilde{\theta}_n$ as

$$\tilde{\theta}_n = \theta_0 + \frac{A^*}{n^{\alpha}} + B_n + o(B_n), \tag{A.12}$$

where B_n converges to 0 in probability as $n \to \infty$. Moreover, B_n is of higher order than $n^{-\alpha}$, i.e., $B_n = o(n^{-\alpha})$. Then, substitute $\tilde{\theta}_n$ of Equation (A.9) by its expansion (A.12) and collect all the terms that are of orders higher than $n^{-\alpha}$, we have

$$B_n p'(\theta_0) + \frac{1}{2}\frac{(A^*)^2}{n^{2\alpha}} p''(\theta_0) + \mathrm{E}_{\mathbf{X}}\left[A'(\theta_0)\frac{A^*}{n^{2\alpha}}p'(\theta_0)\right] + \mathrm{E}_{\mathbf{X}}\left[\frac{A(\theta_0)}{n^{2\alpha}}p''(\theta_0)A^*\right] + \mathrm{E}_{\mathbf{X}}\left[\frac{B(\theta_0)}{n^{2\alpha}}p'(\theta_0)\right]$$
$$+\frac{1}{2}\mathrm{E}_{\mathbf{X}}\left[\frac{A(\theta_0)^2}{n^{2\alpha}}p''(\theta_0)\right] + o(n^{-2\alpha}) + o(B_n) = \frac{B_0(\theta_0)}{n^{2\alpha}}p'(\theta_0) + \frac{1}{2}\frac{A_0(\theta_0)^2}{n^{2\alpha}}p''(\theta_0) + o(n^{-2\alpha}).$$

By the above equation, we conclude that B_n is of order $n^{-2\alpha}$. Then, $\tilde{\theta}_n$ admits the following expansion:

$$\tilde{\theta}_n = \theta_0 + \frac{A^*}{n^{\alpha}} + \frac{B^*}{n^{2\alpha}} + o(n^{-2\alpha}). \tag{A.13}$$

Substitute $\tilde{\theta}_n$ by its expansion (A.13) to the Equation (A.9). Collecting the terms of order $O(n^{-2\alpha})$ gives:

$$B^* p'(\theta_0) + \frac{1}{2}(A^*)^2 p''(\theta_0) = B_0(\theta_0)p'(\theta_0) + \frac{1}{2}A_0(\theta_0)^2 p''(\theta_0) - \mathrm{E}_{\mathbf{X}}\left[B(\theta_0)p'(\theta_0)\right] - \frac{1}{2}\mathrm{E}_{\mathbf{X}}\left[A(\theta_0)^2 p''(\theta_0)\right]$$
$$- \mathrm{E}_{\mathbf{X}}\left[A'(\theta_0)A^* p'(\theta_0)\right] + \mathrm{E}_{\mathbf{X}}\left[A(\theta_0)p''(\theta_0)A^*\right]. \tag{A.14}$$

Now, we apply the Taylor expansion to $p(\tilde{\theta}_n)$ at θ_0, and we can get the expansion (11),

$$p(\tilde{\theta}) = p(\theta_0) + \frac{A^*}{n^{\alpha}}p'(\theta_0) + \frac{S_0(\theta_0)}{n^{2\alpha}} + o(n^{-2\alpha}),$$

where $S_0(\theta_0) = B^* p'(\theta_0) + \frac{1}{2}(A^*)^2 p''(\theta_0)$.

We then analyze $\mathrm{E}_0[p(\tilde{\theta}_n) - p(\theta_0)]$ to analyze the bias of $p(\tilde{\theta}_n)$. By Equation (A.11), we get

$$\mathrm{E}_0\left[A^*\right] = \mathrm{E}_0\big[A_0(\theta_0) - \mathrm{E}_{\mathbf{X}}[A(\theta_0)]\big] = 0.$$

ACCEPTED MANUSCRIPT

Then, we consider the term of order $n^{-2\alpha}$. Based on Equation (A.14), we have

$$\mathrm{E}_0\left[\frac{B^*}{n^{2\alpha}}p'(\theta_0)+\frac{1}{2}\frac{(A^*)^2}{n^{2\alpha}}p''(\theta_0)\right]=0.$$

Therefore, we show that $\mathrm{E}_0[p(\tilde{\theta}_n)-p(\theta_0)]=o(n^{-2\alpha})$.

Now, we prove statement (b). By Equations (9) and (11), we have

$$n^{\alpha}\big(p(\tilde{\theta}_n)-p(\theta_0)\big)=p'(\theta_0)\big(A_0(\theta_0)-\mathrm{E}_{\mathbf{X}}[A(\theta_0)]\big)+o(1),$$

$$n^{\alpha}\big(p(\hat{\theta}_n)-p(\theta_0)\big)=p'(\theta_0)A_0(\theta_0)+o(1).$$

If $\lim_{n\to\infty}\mathrm{Var}(o_p(1))=0$, then $\lim_{n\to\infty}\mathrm{Var}[n^{\alpha}p(\tilde{\theta}_n)]=p'(\theta_0)^2\mathrm{Var}[A_0(\theta_0)]$ and $\lim_{n\to\infty}\mathrm{Var}[n^{\alpha}p(\hat{\theta}_n)]=p'(\theta_0)^2\mathrm{Var}[A_0(\theta_0)]$. Therefore, we have $\lim_{n\to\infty}\mathrm{Var}[p(\tilde{\theta}_n)]/\mathrm{Var}[p(\hat{\theta}_n)]=1$.

For statement (c). Based on the assumption that $A_0(\theta_0)$ follows a normal distribution, we have

$$n^{\alpha}\left[p(\tilde{\theta}_n)-p(\theta_0)\right]\Rightarrow p'(\theta_0)\sqrt{\mathrm{Var}\left[A_0(\theta_0)\right]}\cdot Z$$

as $n\to\infty$, where Z is a standard normal random variable.

Appendix A.4. Verification of the uniform convergence in probability of MLE for exponential family of distributions

The probability density function of an exponential family has the form:

$$f(x)=h(x)exp\{\theta T(x)-A(\theta)\},\theta\in\Theta$$

where $h(x)$ is the base density, $T(x)$ is the sufficient statistic vector, $A(\theta)$ is the cumulative generating function. We assume that Θ is open and bounded. If the data is $\mathbf{X}(\theta)=\{X_1(\theta),X_2(\theta),\ldots,X_n(\theta)\}$, then, the MLE of θ is the solution of

$$\frac{\partial}{\partial\theta}A(\theta)=\frac{\sum_{i=1}^{n}T(X_i(\theta))}{n}$$

By Lehmann and Casella (2006), we have $\frac{\partial}{\partial\theta}A(\theta)=\mathrm{E}\big[T\big(X(\theta)\big)\big]$ and $\frac{\partial}{\partial\theta}A(\theta)$ is a one-to-one mapping. We denote $\frac{\partial}{\partial\theta}A(\theta)$ by function $g(\theta)$. Then, the MLE of θ is

$$\hat{\theta}_n(\mathbf{X}(\theta))=g^{-1}\left(\frac{\sum_{i=1}^{n}T(X_i(\theta))}{n}\right).$$

ACCEPTED MANUSCRIPT

If $g(\theta)$ is bi-Lipschitz continuous, which can easily be checked, then we have

$$\begin{aligned}\lim_{n\to\infty}\sup_{\theta\in\Theta}\mathbb{P}\left\{\left|\hat{\theta}_n(\mathbf{X}(\theta))-\theta\right|>\epsilon\right\} &= \lim_{n\to\infty}\sup_{\theta\in\Theta}\mathbb{P}\left\{\left|g^{-1}\left(\frac{\sum_{i=1}^n T(X_i(\theta))}{n}\right)-\theta\right|>\epsilon\right\}\\ &\leq \lim_{n\to\infty}\sup_{\theta\in\Theta}\mathbb{P}\left\{K\left|\frac{\sum_{i=1}^n T(X_i(\theta))}{n}-g(\theta)\right|>\epsilon\right\}\\ &= \sup_{\theta\in\Theta}\lim_{n\to\infty}\mathbb{P}\left\{\left|\frac{\sum_{i=1}^n T(X_i(\theta))}{n}-g(\theta)\right|>\frac{\epsilon}{K}\right\}\\ &= 0\end{aligned}$$

The inequality holds because $g^{-1}(\cdot)$ is Lipschitz continuous on Θ. We can interchange the supreme and the limit because $g(\cdot)$ is Lipschitz continuous on a bounded set Θ, therefore $g(\cdot)$ is bounded on Θ. The last equality holds because of the weak law of large numbers. Therefore, under the condition that $\frac{\partial}{\partial\theta}A(\theta)$ is bi-Lipschitz continuous, we verified the uniform convergence in probability for MLE of the exponential family of distributions.

References

Allen, A. O. (1990). *Probability, Statistics, and Queueing Theory: With Computer Science Applications*. Gulf Professional Publishing.

Asmussen, S. and P. W. Glynn (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer.

Barton, R. R., B. L. Nelson, and W. Xie (2014). Quantifying input uncertainty via simulation confidence intervals. *INFORMS Journal on Computing 26*(1), 74–87.

Barvínek, E., I. Daler, and J. Francu (1991). Convergence of sequences of inverse functions. *Archivum Mathematicum 27*(2), 201–204.

Biller, B. and C. G. Corlu (2011). Accounting for parameter uncertainty in large-scale stochastic simulations with correlated inputs. *Operations Research 59*(3), 661–673.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Casella, G. and R. L. Berger (2002). *Statistical Inference*. Belmont, Califonia: Duxbury Press.

Cheng, R. C. H. and W. Holland (1997). Sensitivity of computer simulation experiments to errors in input data. *Journal of Statististical Computation and Simulation 57*, 219–241.

Chick, S. (2001). Input distribution selection for simulation experiments: accounting for input uncertainty. *Operations Research 49*(5), 744–758.

Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap*, Volume 57. CRC press.

Fang, J. and L. J. Hong (2013). Linking statistical estimation and decision making through simulation. *Proceedings of the 2013 Winter Simulation Conference* (IEEE, Washington, DC), 766–777.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*. New York: John Wiley and Sons.

Gouriéroux, C. and A. Monfort (1995). *Statistics and Econometric Models*. Cambridge University Press.

Gouriéroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, 85–118.

Gouriéroux, C., E. Renault, and Z. Touzi (2000). Calibration by simulation for small sample bias correction. In *Simulation-Based Inference in Econometrics: Methods and Applications*, pp. 328–358. Cambridge University Press.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Henderson, S. G. (2003). Input model uncertainty: Why do we care and what should we do about it? *Proceedings of the 2003 Winter Simulation Conference* (IEEE, Piscataway, NJ), 90–100.

Kolassa, J. E. (1997). *Series Approximation Methods in Statistics*. New York: Springer-Verlag.

Lehmann, E. L. (1999). *Elements of Large-sample Theory*. New York: Springer-Verlag.

Lehmann, E. L. and G. Casella (2006). *Theory of point estimation*. Springer Science & Business media.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*, 995–1026.

Nemirovsky, A. S. and D. B. Yudin (1983). *Problem complexity and method efficiency in optimization*. New York: Wiley.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica 59*, 1161–1167.

ACCEPTED MANUSCRIPT

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Phillips, P. C. B. and J. Yu (2009). Simulation-based estimation of contingent-claims price. *Review of Financial Studies 22*(9), 3669–3705.

Robbins, M. and S. Monro (1951). A stochastic approximation method. *The Annals of Mathematical Statistics 22*, 400–407.

Shapiro, A., D. Dentcheva, and A. Ruszczyński (2009). *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia: SIAM.

Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*, 63–84.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics 14*(4), 1261–1295.