Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

HIGHER-ORDER COVERAGE ERROR ANALYSIS FOR BATCHING AND SECTIONING

Shengyi He
Henry Lam

Department of Industrial Engineering and Operations Research
Columbia University
500 West 120th Street
New York, NY 10027, USA

ABSTRACT

While batching and sectioning have been widely used in simulation, it is open regarding their higher-order coverage behaviors and whether one is better than the other in this regard. We develop techniques to obtain higher-order coverage errors for sectioning and batching. We theoretically argue that none of batching or sectioning is uniformly better than the other in terms of coverage, but sectioning usually has a smaller coverage error when the number of batches is large. We also support our theoretical findings via numerical experiments.

1 INTRODUCTION

Sectioning and batching are widely used methods in simulation analysis. The basic idea of these methods is to divide the data into batches and quantify the variability of point estimates by suitably combining the batch estimates. They are especially useful tools to construct confidence intervals (CI) when the variance of the output is hard to compute, such as quantile (Nakayama 2014) whose variance estimation involves density estimation, and in serially dependent problems and steady-state estimation (Asmussen and Glynn 2007; Nakayama 2007).

While widely used, the detailed coverage behaviors of sectioning and batching are not well understood. To understand the statistical performances of these methods and to conduct comparisons, however, this question seems imminent. To put things in perspective, note that a good CI should have a small half width and coverage error. By construction, with the same choice of batch size, the CI half widths of sectioning and batching are equal. Therefore, their difference lies in the coverage errors. Nonetheless, under regularity conditions, both sectioning and batching lead to asymptotically exact CIs. Thus, both methods only have higher order coverage errors, and it is these errors that can differ from each other.

There are very few studies on the higher-order coverage probabilities for sectioning and batching. The challenge is that the statistics used in sectioning and batching have an asymptotic t-distribution rather than a normal distribution, so Edgeworth expansion cannot be directly applied. The most relevant result is the heuristic argument given in Nakayama (2014), which argued that since the estimator based on the whole empirical distribution has smaller bias, sectioning appears to lead to better coverage. Nakayama (2014) supported this claim with numerical results.

In this paper, we develop tools to study the higher-order expansion for the coverage probabilities of sectioning and batching. Under regularity conditions, we show that the coverage errors of sectioning and batching can be expanded as series of $n^{-1/2}$ where n is the data size in each batch. For a symmetric CI, we show that both methods have coverage errors of order $O(n^{-1})$. The coefficients in the expansion involve some integration which can not be explicitly calculated in general, but we provide examples where explicit calculation is possible and sufficient to draw some conclusions. In terms of methodology, our analysis utilizes Edgeworth expansion and Taylor expansion techniques combined with oddness and evenness arguments for functions. To the best of our knowledge, we are the first to study the higher-order coverage of sectioning and batching using this type of analysis.

From our analyses, we conclude that whether sectioning or batching has smaller higher-order coverage error depends on the problem parameters, so none of them is uniformly better. But for a fixed problem, when the number of batches is large, batching suffers from a significant bias and sectioning has better coverage.

We briefly review the literature on sectioning and batching techniques. Pope (1995) analyzes the coverage error of sectioning using Edgeworth expansion, but it focuses on the case when the number of batches goes to infinity so that the problem statistic can be approximated by normal. This is different from our analysis for the t distribution approximation which is our key novelty and faced challenge in this problem. For the CI half width, Schmeiser (1982) shows that if we assume the data size is large enough so that the non-normality of the batch estimators is negligible, then the expected half width would decrease as the number of batches increases, but the rate of decrease would become much slower when the number of batches is large. Similar observations are also made in Glynn and Lam (2018). Jackknife can be used to reduce small-sample bias within sections, but at the cost of greater computation time and uncertainty about the variance inflation (Lewis and Orav 1989). Finally, as shown in Example 3.1 of Glynn and Iglehart (1990), batching can also be seen as a special type of the more general umbrella technique of standardized time series.

The rest of this paper is organized as follows: Section 2 gives the formulation for batching, sectioning and modified sectioning. Section 3 studies the coverage error of batching, and Section 4 for sectioning and modified sectioning. Section 5 considers a specific example to reveal comparisons on the coverage errors for different methods. Section 6 provides numerical examples.

2 SECTIONING AND BATCHING

Consider the problem of constructing a CI for $\psi(P)$ where P is an unknown distribution, ψ is a known statistical functional and we have data drawn i.i.d. from P. Suppose the data size is $N = nK$. Divide the data into K batches each with size n and denote $\hat{P}_i$ as the empirical distribution for the i-th batch where $i = 1,2,\ldots,K$. Denote $\hat{P}$ as the whole empirical distribution using all of the nK data. Then the batching CI is given by $CI_B := \left(\frac{1}{K}\sum_i \psi\left(\hat{P}_i\right) \pm t_{K-1,\alpha/2}\frac{S_{\text{batch}}}{\sqrt{K}}\right)$ where $S^2_{\text{batch}} = \frac{1}{K-1}\sum_{i=1}^K\left(\psi\left(\hat{P}_i\right) - \frac{1}{K}\sum_j \psi\left(\hat{P}_j\right)\right)^2$ and $t_{K-1,\alpha/2}$ is the upper $\alpha/2$-quantile for the t_{K-1} distribution. This CI is asymptotically exact since

$$W_B := \frac{\sqrt{nK}\left(\frac{1}{K}\sum_i \psi\left(\hat{P}_i\right) - \psi\right)}{\sqrt{\frac{1}{K-1}\sum_{i=1}^K\left(\sqrt{n}\psi\left(\hat{P}_i\right) - \frac{1}{K}\sum_j \sqrt{n}\psi\left(\hat{P}_j\right)\right)^2}} \Rightarrow t_{K-1}.$$

Here $\psi = \psi(P)$ is the target value and the limit is as $n \to \infty$ with K fixed. For sectioning, the CI is given by $CI_S := \left(\psi\left(\hat{P}\right) \pm t_{K-1,\alpha/2}\frac{S_{\text{sec}}}{\sqrt{K}}\right)$ where $S^2_{\text{sec}} = \frac{1}{K-1}\sum_{i=1}^K\left(\psi\left(\hat{P}_i\right) - \psi\left(\hat{P}\right)\right)^2$. This CI is also asymptotically exact since

$$W_S := \frac{\sqrt{nK}\left(\psi\left(\hat{P}\right) - \psi\right)}{\sqrt{\frac{1}{K-1}\sum_{i=1}^K\left(\sqrt{n}\psi\left(\hat{P}_i\right) - \sqrt{n}\psi\left(\hat{P}\right)\right)^2}} = W_B + o_p(1) \Rightarrow t_{K-1}.$$

We also consider modified sectioning (Nakayama 2014) that is viewed as a middle ground between batching and sectioning. This approach uses the same variance estimator as batching, but the center estimate as sectioning. The resulting CI is $CI_{SB} := \left(\psi\left(\hat{P}\right) \pm t_{K-1,\alpha/2}\frac{S_{\text{batch}}}{\sqrt{K}}\right)$. We call this method sectioning-batching (SB). This CI is also asymptotically exact since

$$W_{SB} := \frac{\sqrt{nK}\left(\psi\left(\hat{P}\right) - \psi\right)}{\sqrt{\frac{1}{K-1}\sum_{i=1}^K\left(\sqrt{n}\psi\left(\hat{P}_i\right) - \frac{1}{K}\sum_j \sqrt{n}\psi\left(\hat{P}_j\right)\right)^2}} = W_B + o_p(1) \Rightarrow t_{K-1}.$$

Note that $-t_{K-1,\alpha/2} \leq W_B \leq t_{K-1,\alpha/2} \Leftrightarrow \psi \in CI_B$ and similar arguments hold for sectioning and SB. Therefore, to study the higher order coverage errors, it suffices to study the distributions W_B, W_S and W_{SB}, and how much they deviate from t_{K-1}.

3 ERROR ANALYSIS FOR BATCHING

The following theorem states that the coverage error for batching admits an expansion as a series of $n^{-1/2}$, and the coverage error is of order $O(n^{-1})$ for a symmetric CI. For the one-sided CIs, since we cannot cancel the first-order coverage error, the coverage error is typically of order $O(n^{-1/2})$.

Theorem 1 Suppose that $\psi\left(\hat{P}_1\right)$ has a valid Edgeworth expansion, in the sense that for some $0 < \sigma < \infty$,

$$P\left(\frac{\sqrt{n}\left(\psi\left(\hat{P}_1\right)-\psi\right)}{\sigma} \leq q\right) = \Phi(q) + \sum_{j=1}^{r} n^{-j/2} p_j(q)\phi(q) + O\left(n^{-(r+1)/2}\right)$$

holds uniformly over $q \in \mathbb{R}$, and p_j is an even polynomial when j is odd and is an odd polynomial when j is even. Here Φ and ϕ are the cdf and pdf of standard normal. Then $P(W_B \leq q)$ can be expanded as a series of $n^{-1/2}$ with residual $O\left(n^{-(r+1)/2}\right)$ and we have $P\left(-q \leq W_B \leq q\right) = P(-q \leq t_{K-1} \leq q) + O(n^{-1})$.

Proof. As long as we have a valid Edgeworth expansion for $\sqrt{n}\left(\psi\left(\hat{P}_i\right)-\psi\right)$, noting that W_B is a function of $\left(\sqrt{n}\left(\psi\left(\hat{P}_i\right)-\psi\right)\right)_{i=1}^{K}$, we can evaluate the probability $P\left(W_s \leq q\right)$ based on integration. let $f(\mathbf{z}) = \frac{\sqrt{K}\frac{1}{K}\sum_i z_i}{\sqrt{\frac{1}{K-1}\sum_{i=1}^{K}(z_i - \frac{1}{K}\sum_j z_j)^2}}$, then one can check that $W_B = f\left(\left(\sqrt{n}\left(\psi\left(\hat{P}_i\right)-\psi\right)/\sigma\right)_{i=1}^{K}\right)$. So we have

$$P\left(W_B \leq q\right) = \int_{f(\mathbf{z})\leq q} \Pi_{j=1}^{K} d\left(\Phi(z_j) + \sum_{j=1}^{r} n^{-j/2} p_j(z_j)\phi(z_j)\right) + O\left(n^{-(r+1)/2}\right).$$

For a symmetric confidence interval, we have that

$$\begin{aligned} P\left(-q \leq W_B \leq q\right) &= \int_{-q\leq f(\mathbf{z})\leq q} \Pi_{j=1}^{K} d\left(\Phi(z_j) + n^{-1/2} p_1(z_j)\phi(z_j)\right) + O\left(n^{-1}\right) \\ &= P\left(-q \leq t_{K-1} \leq q\right) + Kn^{-1/2}\int_{-q\leq f(\mathbf{z})\leq q} \phi(z_1)\phi(z_2)\ldots\phi(z_K)\left(-z_1 p_1(z_1) + p_1'(z_1)\right) d\mathbf{z} + O(n^{-1}). \end{aligned}$$

Here p_1' is the derivative of p_1. Since p_1 is an even polynomial, $-z_1 p_1(z_1) + p_1'(z_1)$ is an odd polynomial. In addition, note that the area $\{-q \leq f(\mathbf{z}) \leq q\}$ is symmetric around $\mathbf{0}$ since $f(\mathbf{z}) = -f(-\mathbf{z})$. Thus, the integration in the RHS above is 0. As a result, we have that $P\left(-q \leq W_s \leq q\right)$ can be expanded as a power of $n^{-1/2}$ and its leading term is of order n^{-1}. □

In the statement of Theorem 1, the reasonableness of the assumption on the evenness and oddness of p_i can be checked from Theorem 2.2 of Hall (1992). In general, the integrals in the expansion do not have explicit expressions and we may need numerical methods to evaluate them. Nonetheless, we provide explicit results for some simple models in Section 5 and use them to draw comparisons among different methods.

4 ERROR ANALYSIS FOR SECTIONING

The analysis for sectioning is a bit harder, since W_S cannot be expressed as merely a function of $\left(\sqrt{n}\left(\psi\left(\hat{P}_i\right)-\psi\right)\right)_{i=1}^{K}$, but also dependent on $\psi(\hat{P})$. Moreover, it is difficult to study the joint distribution of

$$\Lambda := \left(\sqrt{nK}\left(\psi\left(\hat{P}\right)-\psi\right), \left(\sqrt{n}\left(\psi\left(\hat{P}_i\right)-\psi\right)\right)_{i=1}^{K}\right)$$

via Edgeworth expansion, since its asymptotic joint distribution is degenerate. By this we mean that, under regularity conditions, $\sqrt{nK}\left(\psi\left(\hat{P}\right)-\psi\right)-\sqrt{K}\frac{1}{K}\sum_{i=1}^{K}\sqrt{n}\left(\psi\left(\hat{P}_i\right)-\psi\right)=o_p(1)$, which implies that the limiting distribution of $\Lambda\in\mathbb{R}^{K+1}$ has only K degrees of freedom, hence is degenerate.

To handle this issue, we consider the smooth function model where $\psi(P)=f(E_PX)$ for some vector X. This assumption is understandably restrictive, e.g., the quantile of $X\sim P$ does not belong to the smooth function model. However, we also note that the smooth function model has been widely assumed to allow tractable analyses in the bootstrap (Hall 1992) and empirical likelihood (DiCiccio, Hall, and Romano 1991).

For this model, we can express W_S in terms of the sectioned averages denoted by $\bar{X}_1,\ldots,\bar{X}_K$. It is also known that $\psi(\hat{P})$ admits a valid edgeworth expansion under regularity conditions (Bhattacharya and Ghosh 1978). For this model, we can show the following theorem that is similar to our batching result.

Theorem 2 Suppose that $\psi(P)=f(E_PX)$ for some vector X, the following Cramér's condition holds for the distribution of X (denoted by P_0):

$$\limsup_{|t|\to\infty}|E_{P_0}(\exp\{i\langle t,X\rangle\})|<1,$$

X has finite moments up to order $r+2$, and f is r times differentiable in a neighborhood of $E_{P_0}X$. Then $P(W_S\leq q)$ can be expanded as a series of $n^{-1/2}$ with residual $O(n^{-r/2})$. In the symmetric case, we have $P(-q\leq W_S\leq q)=P(-q\leq t_{K-1}\leq q)+O(n^{-1})$. The same result holds if W_S is replaced by W_{SB}.

Proof. We let $A_0=\sqrt{n}\sqrt{K}\frac{\bar{X}_1+\cdots+\bar{X}_K}{K}-E_{P_0}X, A_1=\sqrt{n}\sqrt{K}\left[\bar{X}_1-\frac{\bar{X}_1+\cdots+\bar{X}_K}{K}\right],\ldots,A_K=\sqrt{n}\sqrt{K}\left[\bar{X}_K-\frac{\bar{X}_1+\cdots+\bar{X}_K}{K}\right]$. By scale invariance, WLOG, suppose that $\nabla f^T\cdot Var_{P_0}X\cdot\nabla f=1$, and also suppose that $E_{P_0}X=0$. Then for the asymptotic distribution of $(A_0,A_1,\ldots,A_K)$, $\nabla f^T\cdot A_0$ is standard normal independent of $A_j, j\geq 1$. Moreover,

$$\psi\left(\hat{P}\right)=\psi\left(\frac{A_0}{\sqrt{nK}}\right),\psi\left(\hat{P}_i\right)=\psi\left(\frac{A_0+A_i}{\sqrt{nK}}\right).$$

When $r\geq 2$, with a Taylor expansion argument, we can show that

$$\sqrt{nK}(\psi(\hat{P})-\psi(P_0))=\nabla f^T\cdot A_0+\frac{1}{2\sqrt{nK}}A_0^T\left[\nabla^2 f\right]A_0+O_p\left(n^{-1}\right)\tag{1}$$

and for each $k=1,2,\ldots,K$,

$$\sqrt{n}(\psi(\hat{P}_k)-\psi(P_0))=\frac{1}{\sqrt{K}}\nabla f^T\left(A_0+A_k\right)+\frac{n^{-1/2}}{2\sqrt{K}}\left(A_0+A_k\right)^T\left[\nabla^2 f\right]\left(A_0+A_K\right)+O_p\left(n^{-1}\right).\tag{2}$$

Then W_S is given by

$$\begin{aligned}
&W_S\\
=&\frac{\nabla f^T\cdot A_0+\frac{1}{2\sqrt{nK}}A_0^T\left[\nabla^2 f\right]A_0}{\sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\left(\frac{\nabla f^T\cdot A_k}{\sqrt{K}}+\frac{n^{-1/2}}{2K}\left(A_0+A_k\right)^T\left[\nabla^2 f\right]\left(A_0+A_K\right)-\frac{n^{-1/2}}{2K}A_0^T\left[\nabla^2 f\right]A_0\right)\right)^2}}+O_p\left(n^{-1}\right)\\
=&\frac{\nabla f^T\cdot A_0+\frac{1}{2\sqrt{nK}}A_0^T\left[\nabla^2 f\right]A_0}{\sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\left(\frac{\nabla f^T\cdot A_k}{\sqrt{K}}+\frac{n^{-1/2}}{2K}\left(A_k+2A_0\right)^T\left[\nabla^2 f\right]A_k\right)\right)^2}}+O_p\left(n^{-1}\right)\\
=&\frac{\nabla f^T\cdot A_0+\frac{1}{2\sqrt{nK}}A_0^T\left[\nabla^2 f\right]A_0}{\sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\left(\frac{\nabla f^T\cdot A_k}{\sqrt{K}}\right)^2+\frac{\nabla f^T\cdot A_k}{K\sqrt{nK}}\left(A_k^T\left[\nabla^2 f\right]A_k+2A_0^T\left[\nabla^2 f\right]A_k\right)\right)}}+O_p\left(n^{-1}\right)
\end{aligned}$$

Denote $a = \frac{\sum_k \frac{\nabla f^T \cdot A_k}{K} A_k^T \left[\nabla^2 f\right] A_k}{\sum_{k=1}^K \left(\frac{\nabla f^T \cdot A_k}{\sqrt{K}}\right)^2}, b = 2\frac{\sum_k \frac{\nabla f^T \cdot A_k}{K} \left[\nabla^2 f\right] A_k}{\sum_{k=1}^K \left(\frac{\nabla f^T \cdot A_k}{\sqrt{K}}\right)^2}$ (both of them are functions of $A_1, \ldots, A_K$). We have that the RHS above

$$
\begin{aligned}
&= \frac{\nabla f^T \cdot A_0 + \frac{1}{2\sqrt{nK}} A_0^T \left[\nabla^2 f\right] A_0}{\sqrt{\frac{1}{K-1}\sum_{k=1}^K \left(\left(\frac{\nabla f^T \cdot A_k}{\sqrt{K}}\right)^2 + \frac{1}{\sqrt{nK}}\left(A_0^T b + a\right)\right)}} + O_p\left(n^{-1}\right) \\
&= \frac{\nabla f^T \cdot A_0 + \frac{1}{2\sqrt{nK}} A_0^T \left[\nabla^2 f\right] A_0}{\sqrt{\frac{1}{K-1}\sum_{k=1}^K \left(\frac{\nabla f^T \cdot A_k}{\sqrt{K}}\right)^2}} \left(1 - \frac{1}{2\sqrt{nK}}\left(A_0^T b + a\right)\right) + O_p\left(n^{-1}\right). \qquad (3)
\end{aligned}
$$

Denote the above function as $g_n(A_0, A_1, \ldots, A_{K-1})$ (note that $A_K = -(A_1 + \cdots + A_{K-1})$). Since $(A_0, \ldots, A_{K-1})$ is a linear transformation of the sectioned averages, and the sectioned averages have valid Edgeworth expansions (see e.g. Theorem 20.1 of Bhattacharya and Rao 2010), we have that the joint distribution of $A_0, \ldots, A_{K-1}$ admits a valid multivariate Edgeworth expansion:

$$
P((A_0, \ldots, A_{K-1}) \in B) = \int_B \phi_\Sigma(\mathbf{z})(1 + n^{-1/2} p(\mathbf{z})) d\mathbf{z} + O(n^{-1})
$$

for all Borel sets B. Here $\phi_\Sigma(\mathbf{z})$ denotes the density of the limit distribution of $(A_0, A_1, \ldots, A_{K-1})$, and $p(\mathbf{z})$ is an odd polynomial. For the probability that $-q \le W_B \le q$, we have that

$$
\begin{aligned}
&P(-q \le W_B \le q) \\
=&P(-q \le g_n(A_0, A_1, \ldots, A_{K-1}) \le q) + O(n^{-1}) \\
=&\int_{-q \le g_n(\mathbf{z}) \le q} \left(\phi_\Sigma(\mathbf{z}) + n^{-1/2}\phi(\mathbf{z}) p(\mathbf{z})\right) d\mathbf{z} + O(n^{-1}) \\
=&\int_{-q \le g_n(\mathbf{z}) \le q} \phi_\Sigma(\mathbf{z}) d\mathbf{z} + n^{-1/2}\int_{-q \le g_\infty(\mathbf{z}) \le q} \phi(\mathbf{z}) p(\mathbf{z}) + O(n^{-1}).
\end{aligned}
$$

Here $g_\infty(\mathbf{z}) = \lim_{n\to\infty} g_n(\mathbf{z}) = \frac{\nabla f^T \cdot z_0}{\sqrt{\frac{1}{K-1}\sum_{k=1}^K \left(\frac{\nabla f^T \cdot z_k}{\sqrt{K}}\right)^2}}$ and in the last equality above, we used that $g_\infty(\mathbf{z}) - g(\mathbf{z}) = O(n^{-1/2})$. Also note that g_∞ satisfies $g_\infty(\mathbf{z}) = g_\infty(-\mathbf{z})$ so $\int_{-q \le g_\infty(\mathbf{z}) \le q} \phi(\mathbf{z}) p(\mathbf{z}) = 0$. Hence from the above displayed equality,

$$
\begin{aligned}
&P(-q \le W_S \le q) \\
=&\int_{-q \le g_n(\mathbf{z}) \le q} \phi_\Sigma(\mathbf{z}) d\mathbf{z} + O(n^{-1}) \\
=&P(-q \le t_{K-1} \le q) + \int_{-q \le g_n(\mathbf{z}) \le q} \phi_\Sigma(\mathbf{z}) d\mathbf{z} - \int_{-q \le g_\infty(\mathbf{z}) \le q} \phi_\Sigma(\mathbf{z}) d\mathbf{z} + O(n^{-1}). \qquad (4)
\end{aligned}
$$

Here we used $P(-q \le t_{K-1} \le q) = \int_{-q \le g_\infty(\mathbf{z}) \le q} \phi_\Sigma(\mathbf{z}) d\mathbf{z}$ since the limiting distribution of W_S is t_{K-1}. Now it suffices to study the difference between $P(-q \le g_n(Z_0, Z_1, \ldots, Z_{K-1}) \le q)$ and its counterpart as $n \to \infty$, where $Z_0, Z_1, \ldots, Z_{K-1}$ follows from the limiting normal distribution of $(A_0, \ldots, A_{K-1})$. In particular, we have that $\nabla^T f \cdot Z_0 \sim N(0,1)$ and Z_0 is independent of $Z_1, \ldots, Z_{K-1}$. For this probability, we can do the following computation

$$
\begin{aligned}
&P(-q \le g_n(Z_0, Z_1, \dots, Z_{K-1}) \le q) \\
=&P\left(-q \le \frac{\nabla f^T \cdot Z_0 + \frac{1}{2\sqrt{nK}} Z_0^T \left[\nabla^2 f\right] Z_0}{\sqrt{\frac{1}{K-1}\sum_{k=1}^K \left(\frac{\nabla f^T \cdot Z_k}{\sqrt{K}}\right)^2}} \left(1 - \frac{1}{2\sqrt{nK}}\left(Z_0^T b + a\right)\right) \le q\right) \\
=&P\left(-q \le \frac{\nabla f^T \cdot Z_0 + \frac{1}{2\sqrt{nK}} Z_0^T \left[\nabla^2 f - b\nabla f^T\right] Z_0}{\sqrt{\frac{1}{K-1}\sum_{k=1}^K \left(\frac{\nabla f^T \cdot Z_k}{\sqrt{K}}\right)^2}} \left(1 - \frac{a}{2\sqrt{nK}}\right) \le q\right) \\
=&P\left(-\frac{q\sqrt{\frac{1}{K-1}\sum_{k=1}^K \left(\frac{\nabla f^T \cdot Z_k}{\sqrt{K}}\right)^2}}{1 - \frac{a}{2\sqrt{nK}}} \le \nabla f^T \cdot Z_0 + \frac{1}{2\sqrt{nK}} Z_0^T \left[\nabla^2 f - b\nabla f^T\right] Z_0 \le \frac{q\sqrt{\frac{1}{K-1}\sum_{k=1}^K \left(\frac{\nabla f^T \cdot Z_k}{\sqrt{K}}\right)^2}}{1 - \frac{a}{2\sqrt{nK}}}\right).
\end{aligned}
$$

Conditional on $Z_1, \dots, Z_{K-1}$, by normality of the distribution of Z_0 and Edgeworth expansion, we know that the conditional distribution function for $\nabla f^T \cdot Z_0 + \frac{1}{2\sqrt{nK}} Z_0^T \left[\nabla^2 f - b\nabla f^T\right] Z_0$ can be expanded as $\Phi(q) + n^{-1/2}\tilde{p}_1(q)\phi(q) + O(n^{-1})$ where $\tilde{p}_1$ is even (this evenness claim follows since $\nabla f^T \cdot Z_0 + \frac{1}{2\sqrt{nK}} Z_0^T \left[\nabla^2 f - b\nabla f^T\right] Z_0$ is a polynomial with the same form as Theorem 2.1 of Hall 1992). Thus, conditional on $Z_1, \dots, Z_{K-1}$, the above probability is given as (denote $q' = q\sqrt{\frac{1}{K-1}\sum_{k=1}^K \left(\frac{\nabla f^T \cdot Z_k}{\sqrt{K}}\right)^2}$)

$$
\Phi\left(q'\left(1 + \frac{a}{2\sqrt{nK}}\right)\right) + n^{-1/2}\tilde{p}_1(q')\phi(q') - \Phi\left(-q'\left(1 + \frac{a}{2\sqrt{nK}}\right)\right) - n^{-1/2}\tilde{p}_1(q')\phi(q') + O(n^{-1})
$$

which is (by the evenness of $\tilde{p}_1$)

$$
\Phi\left(q'\right) - \Phi\left(-q'\right) + \phi(q')q'\frac{a}{\sqrt{nK}} + O(n^{-1}).
$$

From this, we conclude that

$$
\begin{aligned}
&P(-q \le g_n(Z_0, Z_1, \dots, Z_{K-1}) \le q | Z_1, \dots Z_{K-1}) - P(-q \le g_\infty(Z_0, Z_1, \dots, Z_{K-1}) \le q | Z_1, \dots Z_{K-1}) \\
=&E_{Z_1, \dots Z_{K-1}}\left[\phi(q')q'\frac{a}{\sqrt{nK}}\right] + O(n^{-1})
\end{aligned}
$$

Noting that a is odd (and q' is even) in $Z_1, \dots, Z_{k-1}$, we have that the expectation is 0, so the above difference is indeed $O(n^{-1})$. Then from (4), we have shown that $P(-q \le W_S \le q) = P(-q \le t_{K-1} \le q) + O(n^{-1})$. To see that W_{SB} has the same property, notice that

$$
nS_{\text{sec}}^2 = nS_{\text{batch}}^2 + \frac{K}{K-1}\left(\sqrt{n}\psi(\hat{P}) - \frac{1}{K}\sum_{i=1}^K \sqrt{n}\psi(\hat{P}_i)\right)^2 = nS_{\text{batch}}^2 + O_p(1)
$$

Here the second equality holds since $\sqrt{n}\psi(\hat{P}) - \frac{1}{K}\sum_{i=1}^K \sqrt{n}\psi(\hat{P}_i) = O(n^{-1/2})$, which can be seen by plugging in expansions (1), (2) and using the equation $\sum_{i=1}^K A_i = 0$. This implies $W_{SB} = W_S + O_p(n^{-1})$. Since we have shown $P(-q \le W_S \le q) = O(n^{-1})$, the same holds for W_{SB}.

In general, similar to (3), with a Taylor expansion argument, we can then give an expansion for W_S (and similarly for W_{SB}) of the following form

$$W_S = \frac{\nabla f^T A_0}{\sqrt{\frac{1}{K-1}\sum_{i=1}^K \left(\frac{\nabla f^T A_K}{\sqrt{K}}\right)^2}} + \sum_{j=1}^r \left(n^{-j/2}\sum_{m=1}^j a_{i_1 i_2 \dots i_m} A_{0,i_1} A_{0,i_2}\dots A_{0,i_m}\right) + o_p(n^{-r/2}).$$

where $A_{0,i}$ is the i-th coordinate of A_0 and $a_{i_1,i_2,\dots,i_j}$ is a function of $(A_1,\dots,A_{K-1})$ that does not depend on n. Since $A_0, A_1, \dots A_K$ have Edgeworth expansions, we have that conditional on $A_1,\dots,A_K$, the distribution of A_0 can also be expanded as a series of $n^{-1/2}$. So we can expand $P(W_S \le q | A_1,\dots,A_{K-1})$ as a series of $n^{-1/2}$. Then by integrating over the domain of $(A_1,\dots,A_{K-1})$, we have an expansion for $P(W_S \le q)$ as a series of $n^{-1/2}$. □

5 COMPARISONS ON COVERAGE ERRORS

Whether batching or sectioning has a smaller higher-order coverage error depends on the problem specifics. In Section 5.1, we illustrate this using a $K = 2$ example. Then in Section 5.2, we discuss how our conclusions change when K increases.

5.1 Higher-Order Coverage Error When $K = 2$: An Example

Suppose that $K = 2$, $\psi(P) = f(E_P X) := E_P X + \lambda (E_P X)^2$ and P_0 is standard normal. In this case, the sectioned estimates are $\psi(\hat{P}_i) = f(E_{\hat{P}_i} X) \overset{d}{=} f(\frac{1}{\sqrt{n}} U_i)$ where $U_i \sim N(0,1)$. For this model, the higher-order coverage errors can be computed explicitly via the following lemma.

Lemma 1. *With the model introduced above, the higher-order coverage errors for batching, sectioning, and SB can be expressed as*

$$P(-q \le W_B \le q) - P(-q \le t_1 \le q) = \frac{\lambda^2}{n}\left(-q\left(q^2-1\right)^2\left(\frac{1}{q^2+1}\right)^3 \frac{4}{\pi}\right) + O(n^{-3/2}), \tag{5}$$

$$P(-q \le W_S \le q) - P(-q \le t_1 \le q) = \frac{\lambda^2}{n}\left(-q^5\left(\frac{1}{q^2+1}\right)^3 \frac{4}{\pi} + q\left(\frac{1}{q^2+1}\right)^2 \frac{1}{\pi}\right) + O(n^{-3/2}), \tag{6}$$

$$P(-q \le W_{SB} \le q) - P(-q \le t_1 \le q) = \frac{\lambda^2}{n}\left(-q^5\left(\frac{1}{q^2+1}\right)^3 \frac{4}{\pi}\right) + O(n^{-3/2}). \tag{7}$$

Proof. The test statistic for batching can be expressed as

$$W_B = \frac{\sqrt{nK}\left(\frac{f\left(\frac{1}{\sqrt{n}}U_1\right)+f\left(\frac{1}{\sqrt{n}}U_2\right)}{2}\right)}{\sqrt{\frac{n}{2}\left(f\left(\frac{1}{\sqrt{n}}U_1\right) - f\left(\frac{1}{\sqrt{n}}U_2\right)\right)^2}} = \frac{\sqrt{2}\left(\frac{U_1+U_2}{2} + \frac{\lambda}{\sqrt{n}}\frac{U_1^2+U_2^2}{2}\right)}{\sqrt{\frac{1}{2}\left(U_1 - U_2 + \frac{\lambda}{\sqrt{n}}\left(U_1^2 - U_2^2\right)\right)^2}} = \frac{\sqrt{2}\left(\frac{U_1+U_2}{2} + \frac{\lambda}{\sqrt{n}}\frac{U_1^2+U_2^2}{2}\right)}{\frac{\sqrt{2}}{2}|U_1 - U_2|\sqrt{\left(1 + \frac{\lambda}{\sqrt{n}}(U_1+U_2)\right)^2}}$$

Denote $A_0 = \sqrt{2}\frac{U_1+U_2}{2}$ and $A_1 = \sqrt{2}\frac{U_1-U_2}{2}$ (note that they are independent and this notation is indeed consistent with the notation used in the proof of Theorem 2). Then from the above,

$$W_B = \frac{A_0 + \frac{\lambda}{\sqrt{2}\sqrt{n}}\left(A_0^2 + A_1^2\right)}{|A_1|\sqrt{\left(1 + \frac{\sqrt{2}\lambda}{\sqrt{n}}A_0\right)^2}} = \frac{A_0 + \frac{\lambda}{\sqrt{2}\sqrt{n}}\left(A_0^2 + A_1^2\right)}{|A_1|\left(1 + \frac{\sqrt{2}\lambda}{\sqrt{n}}A_0\right)}$$

Here the second equality holds as long as $1+\frac{\sqrt{2}\lambda}{\sqrt{n}}A_0>0$ which happens with probability $1-O(e^{-n})$. Based on this expression for W_S, we study the event $W_B\le q$:

$$W_B\le q\Leftrightarrow A_0+\frac{\lambda}{\sqrt{2}\sqrt{n}}\left(A_0^2+A_1^2\right)\le q\left|A_1\right|\left(1+\frac{\sqrt{2}\lambda}{\sqrt{n}}A_0\right)$$

which is a quadratic function in A_0. It can be equivalently written as

$$A_0+\frac{\lambda}{\sqrt{2n}}\left(A_0^2-2q\left|A_1\right|A_0\right)\le q\left|A_1\right|-\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2 \tag{8}$$

We want to write the above as $A_0<V+O(n^{-3/2})$ for some critical value V that is independent of A_0. From the above inequality, V satisfy

$$V=q\left|A_1\right|+O_p(n^{-1/2})=:q\left|A_1\right|+V_1$$

for some $V_1=O_p(n^{-1/2})$. Plugging this in (8) and solving for V_1, we get

$$V_1=\left(q^2-1\right)\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2+O(n^{-1})=:\left(q^2-1\right)\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2+V_2.$$

Again, by plugging this in (8) and solving for V_2, we have that V_2 satisfy

$$V_2=0+O_p(n^{-3/2}).$$

So as a conclusion, with exponentially small error, $W_B\le t\Leftrightarrow A_0\le q\left|A_q\right|+\left(q^2-1\right)\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2+O_p(n^{-3/2})$. Similarly, $W_B\ge -q\Leftrightarrow A_0\ge -q\left|A_1\right|+\left(q^2-1\right)\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2+O_p(n^{-3/2})$. Based on this, for the coverage error we have

$$\begin{aligned}
&P(-q\le W_B\le q)-P(-q\le t_1\le q)\\
=&P\left(-q\left|A_1\right|+\left(q^2-1\right)\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2\le A_0\le q\left|A_1\right|+\left(q^2-1\right)\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2\right)-P\left(-q\left|A_1\right|\le A_0\le q\left|A_1\right|\right)+O(n^{-3/2})\\
=&E_{A_1}\left[\Phi\left(q\left|A_1\right|+\left(q^2-1\right)\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2\right)-\Phi\left(-q\left|A_1\right|+\left(q^2-1\right)\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2\right)-\Phi\left(q\left|A_1\right|\right)+\Phi\left(-q\left|A_1\right|\right)\right]+O(n^{-3/2})\\
=&E_{A_1}\left[-\phi\left(q\left|A_1\right|\right)\frac{q\lambda^2\left(q^2-1\right)^2A_1^5}{2n}\right]+O(n^{-3/2})\\
=&\frac{1}{\sqrt{2\pi}}\frac{\lambda^2}{n}\left(-\frac{q\left(q^2-1\right)^2}{2}\left(\frac{1}{q^2+1}\right)^3\mu_5\right)+O(n^{-3/2})\\
=&\frac{\lambda^2}{n}\left(-q\left(q^2-1\right)^2\left(\frac{1}{q^2+1}\right)^3\frac{4}{\pi}\right)+O(n^{-3/2}).
\end{aligned}$$

Here in the second equality, we condition on A_2 first and use that A_0 and A_1 are independent standard normals. Also μ_i is the i-th absolute moment of the standard normal. So we have shown (5).

For sectioning, we can do a similar computation

$$W_S=\frac{A_0+\frac{\lambda}{\sqrt{2}\sqrt{n}}A_0^2}{\left|A_1\right|\sqrt{\left(1+\frac{\sqrt{2}\lambda}{\sqrt{n}}A_0\right)^2+\frac{\lambda^2A_1^2}{2n}}}=\frac{A_0+\frac{\lambda}{\sqrt{2}\sqrt{n}}A_0^2}{\left|A_1\right|\left(1+\frac{\sqrt{2}\lambda}{\sqrt{n}}A_0\right)\left(1+\frac{\lambda^2A_1^2}{4n}\right)}+O_p(n^{-3/2}).$$

and

$$W_S \leq q \Leftrightarrow A_0 + \frac{\lambda}{\sqrt{2n}}\left(A_0^2 - 2q\,|A_1|\,A_0\right) \leq q\,|A_1| + q\,|A_1|\,\frac{\lambda^2 A_1^2}{4n}.$$

After some algebra, we get

$$W_S \leq q \Leftrightarrow A_0 \leq q\,|A_1| + \frac{\lambda}{\sqrt{2n}} q^2 A_1^2 + q\,|A_1|\,\frac{\lambda^2 A_1^2}{4n} + O_p(n^{-3/2})$$

and

$$W_S \geq -q \Leftrightarrow A_0 \geq -q\,|A_1| + \frac{\lambda}{\sqrt{2n}} q^2 A_1^2 - q\,|A_1|\,\frac{\lambda^2 A_1^2}{4n} + O_p(n^{-3/2}).$$

Then we have that

$$\begin{aligned}
&P(-q \leq W_S \leq q)\\
=&P\left(-q\,|A_1| + q^2\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2 - q\,|A_1|\,\frac{\lambda^2 A_1^2}{4n} \leq A_0 \leq q\,|A_1| + q^2\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2 + q\,|A_1|\,\frac{\lambda^2 A_1^2}{4n}\right) + O(n^{-3/2})\\
=&E_{A_1}\left[\Phi\left(q\,|A_1| + q^2\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2 + q\,|A_1|\,\frac{\lambda^2 A_1^2}{4n}\right) - \Phi\left(-q\,|A_1| + q^2\frac{\lambda}{\sqrt{2}\sqrt{n}}A_1^2 - q\,|A_1|\,\frac{\lambda^2 A_1^2}{4n}\right)\right] + O(n^{-3/2})\\
=&E_{A_1}\left[\Phi\left(q\,|A_1|\right) - \Phi\left(-q\,|A_1|\right) + \frac{\lambda^2}{n}\phi\left(q\,|A_1|\right)\left(-\frac{q^5\,|A_1|^5}{2} + \frac{1}{2}q\,|A_1|^3\right)\right] + O(n^{-3/2})
\end{aligned}$$

Here the second equality follows by conditioning on A_1. Thus, the coverage error of sectioning is given by

$$\begin{aligned}
P(-q \leq W_S \leq q) - P(-q \leq t_1 \leq q) &= E_{A_1}\left[\frac{\lambda^2}{n}\phi\left(q\,|A_1|\right)\left(-\frac{q^5\,|A_1|^5}{2} + \frac{1}{2}q\,|A_1|^3\right)\right] + O(n^{-3/2})\\
&= \frac{\lambda^2}{n}\left(-q^5\left(\frac{1}{q^2+1}\right)^3\frac{4}{\pi} + q\left(\frac{1}{q^2+1}\right)^2\frac{1}{\pi}\right) + O(n^{-3/2})
\end{aligned}$$

so (6) is proved. The algebra for SB is quite similar to sectioning. Starting from the following expression for the sectioning statistic:

$$W_{SB} = \frac{A_0 + \frac{\lambda}{\sqrt{2}\sqrt{n}}A_0^2}{|A_1|\sqrt{\left(1 + \frac{\sqrt{2}\lambda}{\sqrt{n}}A_0\right)^2}},$$

We can do similar computations as above and get (7). □

Lemma 1 indicates that these three methods have different higher-order coverage errors. More specifically, their leading term (n^{-1} order term) in the error expansion is different. When $q \geq 1$ (which is usually the case; since 1 is the 75-percentile of the t_1 distribution), the RHS of each of (5), (6) and (7) is negative, which implies that the actual coverage probability is smaller than the nominal coverage probability. With a litte algebra, we can show that RHS of (7) $<$ RHS of (6) $<$ RHS of (5) < 0. Thus, batching has the smallest higher-order coverage error and SB has the largest higher-order coverage error.

However, if the underlying distribution is not normal, then we also need to consider the error induced by that. The joint density of $(\sqrt{n}\bar{X}_1, \sqrt{n}\bar{X}_2)$ admits an Edgeworth expansion where the coefficients are determined by the cumulants of X. For simplicity, consider the case when $EX^3 = EX = 0$ and $VarX = 1$. Let $\kappa_4 = EX^4 - 3$ be the 4-th cumulant. Then the density of $\sqrt{n}X_i$ has Edgeworth expansion $p_{X_i}(x) =$

$\phi(x)(1+\frac{1}{24n}\kappa_4 He_4(x))+O(n^{-3/2})$. Here He_4 is the 4-th Hermit polynomial given by $He_4(x)=x^4-6x^2+3$. Noting that all of W_B, W_S and W_{SB} can be expressed as $\frac{\bar{X}_1+\bar{X}_2}{|\bar{X}_1-\bar{X}_2|}+O_p(n^{-1/2})$, we have that the contribution of the error term in the Edgeworth expansion to the coverage error is given by (for both sectioning and batching):

$$P\left(-q\le\frac{\bar{X}_1+\bar{X}_2}{|\bar{X}_1-\bar{X}_2|}\le q\right)-P\left(-q\le\frac{U_1+U_2}{|U_1-U_2|}\le q\right)=2n^{-1}\int_{-q\le f(\mathbf{z})\le q}\phi(z_1)\phi(z_2)\frac{1}{24}\kappa_4 He_4(z_1)d\mathbf{z}+O(n^{-3/2}) \tag{9}$$

where $f(z_1,z_2)=\frac{z_1+z_2}{|z_1-z_2|}$. Therefore, the coverage errors become the RHS of (6), (7) and (5) plus the above term. Noting that κ_4 can be positive or negative, we have that after adding the above term, it could be the case that $0<$ RHS of (7) + (9) $<$ RHS of (6) + (9) $<$ RHS of (5) + (9). If this is the case, then the coverage error of SB is the smallest.

5.2 Analysis of Coverage Error When K Increases

For a general $\psi(\cdot)$ and distribution P, we have the following theorem regarding the asymptotic as $K\to\infty$:
Theorem 3 Suppose that $E\psi(\hat{P}_1)-\psi\neq 0$ (which happens when ψ is nonlinear), $\mathrm{Var}(\psi(\hat{P}_1))<\infty$, and $\sqrt{nK}\left(\psi(\hat{P})-\psi\right)\Rightarrow N(0,\sigma^2)$ as $K\to\infty$ for some $\sigma<\infty$. Fix n and let $K\to\infty$. Then for any $q>0$, $P(-q\le W_B\le q)\to 0$ and $P(-q\le W_S\le q)\to\Phi\left(q\sqrt{nE(\psi(\hat{P}_1)-\psi)^2}/\sigma\right)-\Phi\left(-q\sqrt{nE(\psi(\hat{P}_1)-\psi)^2}/\sigma\right)$.

Proof. As $K\to\infty$, $S_{\text{batch}}\to\sqrt{n\mathrm{Var}\psi(\hat{P}_1)}$. But $\sqrt{nK}\left(\frac{1}{K}\sum_i\psi\left(\hat{P}_i\right)-\psi\right)\to\mathrm{sign}(E\psi(\hat{P}_1)-\psi)\cdot\infty$. Thus W_B either converges to ∞ or $-\infty$ which implies $P(-q\le W_B\le q)\to 0$. Similarly, since

$$\left|S_{\text{sec}}^2-\frac{1}{K-1}\sum_{i=1}^K\left(\psi\left(\hat{P}_i\right)-\psi\right)^2\right|\le\frac{K}{K-1}(\psi(\hat{P})-\psi)^2=o_p(1),$$

by the strong law of large numbers and Slutsky's theorem we have $S_{\text{sec}}\to\sqrt{E(\psi(\hat{P}_1)-\psi)^2}$ (as $K\to\infty$), using the asymptotic result $\sqrt{nK}\left(\psi(\hat{P})-\psi\right)\Rightarrow N(0,\sigma^2)$ we can prove the claim for $P(-q\le W_S\le q)$. □

Theorem 3 implies that when K is large, batching has a significant bias which could lead to an extremely small coverage, and sectioning performs better. This observation is also consistent with the numerical findings in Nakayama (2014), which considers $K=10$ and $K=20$ and observes severe undercoverage for batching when $K=20$ in some examples.

We can also draw a similar conclusion by studying the coverage error in the specific example $\psi(P)=E_PX+\lambda\left(E_PX\right)^2$ and $X\sim N(0,1)$ under P. In this case, we have that (to simplify the computation, here we focus on the one-sided CI, so the leading term for the coverage error is of order $O(n^{-1/2})$ instead of $O(n^{-1})$)

$$P\left(W_S\le q\right)=P\left(A_0\le q\sqrt{\frac{1}{K-1}\sum_{k=1}^K\left(\frac{A_k}{\sqrt{K}}\right)^2}+\frac{\lambda q^2}{\sqrt{nK}}\frac{1}{K-1}\sum_{k=1}^K\left(\frac{A_k}{\sqrt{K}}\right)^2\right)+O(n^{-1})$$

Here A_i is defined as in Section 3. Note that the same expansion holds for SB, since the difference between the coverage of SB and sectioning is $O(n^{-1})$ (as can be seen in the proof of Theorem 2). For batching, the coverage error is given by

$$P(W_B\le q)=P\left(A_0\le q\sqrt{\frac{1}{K-1}\sum_{k=1}^K\left(\frac{A_k}{\sqrt{K}}\right)^2}+\frac{\lambda}{\sqrt{nK}}\left(\frac{q^2}{K-1}-1\right)\sum_{k=1}^K\left(\frac{A_k}{\sqrt{K}}\right)^2\right)+O(n^{-1}).$$

So when $\frac{q^2}{K-1} > \frac{1}{2}$ (so that $\left|\frac{q^2}{K-1} - 1\right| < \frac{q^2}{K-1}$), batching has a smaller one-sided coverage error, otherwise sectioning has a smaller error (similar to the $K = 2$ case, this is only true when the underlying distribution is normal. If the underlying distribution deviates from the standard normal, this relation may not hold). As K becomes larger, $\frac{q^2}{K-1} > \frac{1}{2}$ becomes harder to be satisfied, so it is more likely that sectioning has a smaller coverage error when K is large.

6 NUMERICAL EXPERIMENTS

We study the coverage probability using batching and sectioning. We consider the example $\psi(P) = E_P X + \lambda(E_P X)^2$ and $K = 2$ as in Section 5. We do experiments with different sample sizes and true distribution.

Table 1: Coverage accuracy when $P_0 = N(0,1), \lambda = 5$

Nominal Level	80%	90%	95%
Batching, $n = 200$	$77.07\% \pm 0.08\%$	$86.78\% \pm 0.07\%$	$92.37\% \pm 0.05\%$
Sectioning, $n = 200$	$77.68\% \pm 0.08\%$	$87.65\% \pm 0.06\%$	$93.30\% \pm 0.05\%$
SB, $n = 200$	$75.68\% \pm 0.08\%$	$86.10\% \pm 0.07\%$	$92.05\% \pm 0.05\%$
Batching, $n = 400$	$78.33\% \pm 0.08\%$	$88.36\% \pm 0.06\%$	$93.80\% \pm 0.05\%$
Sectioning, $n = 400$	$78.16\% \pm 0.08\%$	$88.51\% \pm 0.06\%$	$94.00\% \pm 0.05\%$
SB, $n = 400$	$77.74\% \pm 0.08\%$	$88.16\% \pm 0.06\%$	$93.73\% \pm 0.05\%$
Batching,$n = 800$	$79.15\% \pm 0.08\%$	$89.28\% \pm 0.06\%$	$94.57\% \pm 0.04\%$
Sectioning, $n = 800$	$79.02\% \pm 0.08\%$	$89.26\% \pm 0.06\%$	$94.58\% \pm 0.04\%$
SB, $n = 800$	$78.90\% \pm 0.08\%$	$89.22\% \pm 0.06\%$	$94.56\% \pm 0.04\%$

For the first experiment, we set $P = N(0,1)$ and $\lambda = 5$. We replicate the data generation and CI construction 10^6 times to estimate the coverage probabilities. In Table 1, we report the the estimated coverage probabilities with their 95% confidence intervals (the 95% CI is given as $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$ where $\hat{p}$ is the empirical coverage probability and N is the number of replications). We observe that 1) All of batching, sectioning and SB have empirical coverage probabilities smaller than the nominal level. For example, when the nominal level is 90% and $n = 200$, each empirical coverage probability is smaller than 88%. 2) When n is small ($n = 200$), sectioning has the smallest coverage error, but when n is larger ($n = 800$), sectioning is comparable with or worse than batching. And in each case, SB has the largest coverage error. For example, when the nominal level is 80% and $n = 200$, the coverage error of SB is about 4.3%, which is larger than batching ($\approx 3.0\%$), and sectioning has the smallest coverage error ($\approx 2.3\%$). But when $n = 800$, the coverage error of batching ($\approx 0.85\%$) is close to or slightly smaller than the coverage error of sectioning ($\approx 1.0\%$).

Table 2: Coverage accuracy when $P_0 = N(0,1)^7, \lambda = 0.015$

Nominal Level	80%	90%	95%
Batching, $n = 100$	$82.97\% \pm 0.23\%$	$91.19\% \pm 0.18\%$	$95.34\% \pm 0.13\%$
Sectioning, $n = 100$	$84.04\% \pm 0.23\%$	$92.07\% \pm 0.17\%$	$95.89\% \pm 0.12\%$
SB, $n = 100$	$81.23\% \pm 0.24\%$	$90.37\% \pm 0.18\%$	$94.96\% \pm 0.14\%$
Batching, $n = 200$	$81.88\% \pm 0.24\%$	$90.72\% \pm 0.18\%$	$95.06\% \pm 0.13\%$
Sectioning, $n = 200$	$82.55\% \pm 0.24\%$	$91.02\% \pm 0.18\%$	$95.48\% \pm 0.13\%$
SB, $n = 200$	$80.52\% \pm 0.25\%$	$90.36\% \pm 0.18\%$	$94.77\% \pm 0.14\%$
Batching, $n = 400$	$81.47\% \pm 0.24\%$	$90.61\% \pm 0.18\%$	$95.12\% \pm 0.13\%$
Sectioning, $n = 400$	$81.75\% \pm 0.24\%$	$90.72\% \pm 0.18\%$	$95.37\% \pm 0.13\%$
SB, $n = 400$	$80.52\% \pm 0.25\%$	$90.43\% \pm 0.18\%$	$94.95\% \pm 0.14\%$

For the second experiment, we set $P = N(0,1)^7$ (i.e., $X \sim P$ has the same distribution as Z^7 where $Z \sim N(0,1)$) and we again estimate the coverage probabilities for each method. This time, the required computation is heavier and we only replicate 10^5 times. The result is reported in Table 2. Unlike the previous example, we see that the empirical coverage probabilities are larger than the nominal levels. The relation among the coverage errors is reversed compared to the previous example: This time SB has the smallest coverage error, batching has a smaller coverage error than sectioning when n is small. When n is large, the coverage errors of batching and sectioning are close. For example, in Table 2, when the nominal level is 80% and $n = 400$, all of the methods have empirical coverage probabilities greater than 80%. SB has the smallest coverage error ($\approx 0.5\%$), and the coverage error of batching ($\approx 1.5\%$) is slightly smaller than the coverage error of sectioning ($\approx 1.8\%$).

These simulation results are also consistent with our Lemma 1 and the discussion after that. For the first experiment (normal case), the comparisons of the coverage errors among the different methods are consistent with the comparisons among the RHS of (5), (6) and (7) as discussed after Lemma 1. For the second experiment, since the underlying distribution has a large 4th cumulant, (9) is large and as a result, the situation mentioned at the end of Section 5.1 unfolds, so the relation among the coverage errors is reversed.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Science Foundation under grants CAREER CMMI-1834710 and IIS-1849280.

REFERENCES

Asmussen, S., and P. W. Glynn. 2007. *Steady-State Simulation*, 96–125. New York, NY: Springer New York.

Bhattacharya, R. N., and J. K. Ghosh. 1978. "On the Validity of the Formal Edgeworth Expansion". *The Annals of Statistics* 6(2):434 – 451.

Bhattacharya, R. N., and R. R. Rao. 2010. *4. Asymptotic Expansions—Nonlattice Distributions*, 188–222.

DiCiccio, T., P. Hall, and J. Romano. 1991. "Empirical Likelihood Is Bartlett-correctable". *Annals of statistics* 19(2):1053–1061.

Glynn, P. W., and D. L. Iglehart. 1990. "Simulation Output Analysis Using Standardized Time Series". *Mathematics of Operations Research* 15(1):1–16.

Glynn, P. W., and H. Lam. 2018. "Constructing Simulation Output Intervals Under Input Uncertainty via Data Sectioning". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1551–1562. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hall, P. 1992. *The Bootstrap and Edgeworth Expansion*. Springer.

Lewis, P. A. W., and E. J. Orav. 1989. *Simulation Methodology for Statisticians, Operations Analysts, and Engineers: Vol. 1*. USA: Wadsworth Publ. Co.

Nakayama, M. K. 2007. "Fixed-width Multiple-comparison Procedures Using Common Random Numbers for Steady-state Simulations". *European Journal of Operational Research* 182(3):1330–1349.

Nakayama, M. K. 2014, November. "Confidence Intervals for Quantiles Using Sectioning When Applying Variance-Reduction Techniques". *ACM Trans. Model. Comput. Simul.* 24(4):1–21.

Pope, A. 1995. "Improving Confidence Intervals Obtained by Sectioning in Monte Carlo Analysis". In *MODSIM 1995 International Congress on Modelling and Simulation*. November 27th-30th, Newcastle, NSW, Australia, 150-154.

Schmeiser, B. 1982. "Batch Size Effects in the Analysis of Simulation Output". *Operations Research* 30(3):556–568.

AUTHOR BIOGRAPHIES

SHENGYI HE is a PhD student in the Department of Industrial Engineering and Operations Research at Columbia University. He received his B.S. degree in statistics from Peking University in 2019. His research interests include variance reduction and uncertainty quantification via stochastic and robust optimization. His email address is sh3972@columbia.edu.

HENRY LAM is an associate professor in the Department of Industrial Engineering and Operations Research at Columbia University. He received his Ph.D. degree in statistics from Harvard University in 2011. His research interests include efficient methodologies and statistical uncertainty quantification for Monte Carlo computation, predictive modeling and data-driven optimization. His email address is khl2114@columbia.edu. His website is http://www.columbia.edu/ khl2114/.