Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

OPTIMAL BATCHING UNDER COMPUTATION BUDGET

Shengyi He
Henry Lam

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

ABSTRACT

Batching methods operate by dividing data into batches and conducting inference by aggregating estimates from batched data. These methods have been used extensively in simulation output analysis and, among other strengths, an advantage is the light computation cost when using a small number of batches. However, under computation budget constraints, it is open to our knowledge which batching approach among the range of alternatives is statistically optimal, which is important in guiding procedural configuration. We show that standard batching, but also certain carefully designed schemes using uneven-size batches or overlapping batches, are large-sample optimal in the sense of so-called uniformly most accurate unbiasedness from a dual view of hypothesis testing.

1 INTRODUCTION

Batching methods are commonly used in simulation analysis for confidence interval (CI) construction. These methods work by dividing data into batches, and suitably combining the batch estimates to cancel out the nuisance variability parameter. This approach thus enables the construction of CIs without the need to estimate standard error, which can be difficult for many problems such as quantile estimation (Nakayama 2014) whose variance estimation involves density estimation, and serially dependent and steady-state estimation (Asmussen and Glynn 2007; Nakayama 2007).

Compared with other resampling approaches such as the bootstrap, an advantage of batching is the light computation cost. Indeed, the bootstrap requires resampling and repeated model running for a sufficient number of times in order to approximate the resample distribution and hence the original sampling counterpart. In contrast, batching methods have the flexibility of dividing the data into only a small number of batches (as small as 2). This offers a significant advantage when the computation effort for even one model run is gigantic, such as in a big simulation or high-dimensional problems. On the other hand, the price to pay for such computation saving is the longer CI, stemming from the inflated uncertainty in the variance estimator and manifesting in the use of t instead of normal critical values in the CI.

In view of the above, it is natural to ask the question: *Given a fixed computation budget in terms of the number of model runs, do the classical batching methods give statistically optimal CIs?* To be concrete, we need to make clear the meaning of computation budget and define the optimality of a CI. For the former, we consider the setting where the target quantity is a functional of the input distribution. Given the input distribution, evaluating the target quantity is resource-consuming, but is not significantly affected by the choice of the input distribution, e.g., if the input distribution is an empirical distribution, the data size constructing this distribution does not significantly matter. An example is steady-state estimation under input uncertainty, in which case constructing a CI would involve capturing the statistical uncertainty of the input distributions, and each model run would be a steady-state simulation using fitted input distributions

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

from each data batch. Here, the runtime of the steady-state simulation could be insensitive to these input distributions and depend primarily on the system dynamic.

To define optimality, we use the so-called *uniformly most accurate unbiased* notion, in an asymptotic sense as the data size grows. This notion is based on the duality between CIs and hypothesis tests, and is equivalent to having the dual hypothesis test being uniformly most powerful (UMP) unbiased asymptotically. Note that there are other possible, perhaps even more straightforward, approaches to measure accuracy and define the optimality of CIs. These include the expected CI length, and the variance of the variance estimator (Alexopoulos et al. 2011). However, as we will see, our approach allows a systematic and clean analysis on batching CI comparisons by connecting to Neyman-Pearson-type arguments, and the leveraging of the extensive literature on UMP tests. Moreover, we will also see that the optimal CIs in the sense we define do exhibit the same expected length asymptotically under mild regularity conditions.

With these, we present several fundamental results in a roughly increasing level of sophistication. First, given a fixed computation budget and among any methods using equal-sized batches, standard batching gives an asymptotically uniformly most accurate unbiased CI. This ensures the reasonableness in using standard batching. Next, we consider more general schemes where batch sizes can vary and overlap with each other, and we derive the optimal way to combine the batch estimates among these schemes. In particular, we connect our derived optimal overlapping batching CIs to illustrate how these can coincide or differ from suggestions in the existing literature. Regarding the expected CI length, we show that the asymptotically most accurate unbiased CIs in all these comparisons are given by asymptotically equivalent t-based intervals. Lastly, we note that in the analysis of this paper, we assume i.i.d. samples for technical convenience. However, this assumption is not essential and can be properly extended to the dependent case using the approach outlined in Section 5.

We briefly review the literature on the analysis of batching methods. Schmeiser (1982) shows that when the number of data is sufficiently large, the expected length of standard batching would decrease as the number of batches increases, but the rate of decrease would become much slower when the number of batches is large. Song and Schmeiser (1995), Flegal and Jones (2010) derive batch sizes that minimize the mean squared errors of batching variance estimators for steady state estimation problems. Meketon and Schmeiser (1984) shows that for steady-state mean estimation, the variance estimator using the overlapping batching scheme they propose has a smaller variance than non-overlapping batching. Compared to these works, we are different in two aspects: 1) We consider a fixed computation budget, while these papers do not focus on the computational cost and allow different computation costs for each method. 2) Our criterion for comparing CIs is based on the power of hypothesis tests, and is arguably more directly related to CI accuracies than other criteria such as the variance of variance estimators.

The rest of this paper is organized as follows. Section 2 shows the optimality of standard batching among all equal-batch-size methods. Section 3 derives the optimal batching method when the batch sizes are unequal and compares it with standard batching. Section 4 derives the optimal batching method when the batches can overlap and introduces several concrete examples. Section 5 discusses the generalizations of our results to dependent data situations. Section 6 extends the analysis of our paper to the cheap bootstrap, a recently proposed computationally light bootstrap method that relates to batching. Section 7 concludes the paper. Some useful results in statistics and math details are presented in the appendix.

2 OPTIMALITY AMONG BATCHING WITH EQUAL BATCH SIZES

Suppose that we are interested in estimating $\psi(P)$ given samples $X_1,\ldots,X_n$ i.i.d. drawn from P. We consider the case where $\psi(\cdot)$ is a functional that is hard to compute, so the major computational cost is to evaluate $\psi(\cdot)$ at different distributions. We set the simulation budget as K, which means that we can only evaluate $\psi(\cdot)$ for at most K times. We assume that $\psi(\cdot)$ is smooth in the sense that the following central limit theorem (CLT) holds.

Assumption 1

$$\sqrt{n}\left(\psi(\hat{P}_n)-\psi(P)\right)\Rightarrow N(0,\sigma^2)$$

where $\hat{P}_n$ is the empirical distribution of $X_1,\ldots,X_n$ and σ is unknown.

To begin with, we consider the case where the batches are divided into equal length (for simplicity, assume that n is a multiple of K). Correspondingly, we have independent batch estimates $\psi_j := \psi(\hat{P}^{(j)}_{n/K}), j = 1,2,\ldots,K$ where $\hat{P}^{(j)}_{n/K}$ is the empirical distribution for the samples in the j-th batch (i.e., $X_{(j-1)n/K+1},\ldots,X_{jn/K}$).

By the standard batching we mean the following CI:

$$CI_B := \bar{\psi} \pm t_{K-1;1-\alpha/2} S/\sqrt{K}$$

where $\bar{\psi} = \sum_{j=1}^K \psi_j/K$, $t_{K-1;1-\alpha/2}$ is the $1-\alpha/2$ quantile of the t distribution with $K-1$ degrees of freedom, and

$$S^2 = \sum_{j=1}^K (\psi_j - \bar{\psi})^2/(K-1).$$

We want to understand, within the class of methods that construct CI by combining $\psi_1,\ldots,\psi_K$, whether the standard batching is optimal. To facilitate understanding, an example of an alternative candidate in this class of methods is the CI with variance estimator set as

$$S^2(\alpha) = \sum_{j=1}^K \gamma_j (\psi_j - \bar{\psi})^2.$$

where different batches can have different weights γ_j in the variance estimator, and we could calibrate the interval using an associated limiting distribution different from t_{K-1}.

To answer the above, we first define optimality:

Definition 1 Given data $\mathbf{Y}_n$, a CI $\mathscr{C}_n(\mathbf{Y}_n) = (L_n(\mathbf{Y}_n), U_n(\mathbf{Y}_n))$ is *asymptotically unbiased at level* $1-\alpha$ if

$$\lim_{n\to\infty} P(\psi(P) \in \mathscr{C}_n(\mathbf{Y}_n)) \geq 1-\alpha$$

and for any $\delta \neq 0$,

$$\limsup_{n\to\infty} P(\psi(P) + n^{-1/2}\delta \in \mathscr{C}_n(\mathbf{Y}_n)) \leq 1-\alpha.$$

$\mathscr{C}_n$ is *asymptotically uniformly most accurate unbiased at level* $1-\alpha$ if it minimizes the probabilities

$$\limsup_{n\to\infty} P(\psi(P) + n^{-1/2}\delta \in \mathscr{C}'_n(\mathbf{Y}_n))$$

among all asymptotically unbiased level $1-\alpha$ tests $\mathscr{C}'_n$ for any $\delta \neq 0$.

Definition 1 is an asymptotic version of the uniformly most accurate unbiased CI introduced in Lehmann et al. (2005) page 165, which is the dual of UMP unbiased hypothesis tests (a review of this is provided in Appendix A.1). Here, we use the alternative value $\psi(P) + n^{-1/2}\delta$ instead of a fixed $\psi_1 \neq \psi(P)$ because for most CIs, $P(\psi_1 \in \mathscr{C}_n(\mathbf{Y}_n)) \to 0$ as $n \to \infty$.

Theorem 1 Suppose that Assumption 1 holds. When the batch estimates are given by $\mathbf{Y}_{n,B} = (\psi_1,\ldots,\psi_K)^T$, CI_B is asymptotically uniformly most accurate unbiased at level $1-\alpha$.

Sketch of Proof. With our CLT assumption, we have that $\sqrt{n/K}(\mathbf{Y}_{n,B} - \psi(P)1_K) \Rightarrow N(0,\sigma^2 I)$ where 1_K is a column K-dimensional vector whose entries are all 1. Therefore, asymptotically, the problem is equivalent to constructing a CI for μ given $Z_1,\ldots,Z_K \sim N(\mu,\sigma^2)$ where μ and σ^2 are both unknown. Asymptotically, CI_B can be written in terms of Z as $\bar{Z} \pm t_{K-1;1-\alpha/2} S/\sqrt{K}$ where $S^2 = \sum_{j=1}^K (Z_j - \bar{Z})^2/(K-1)$. The corresponding hypothesis test is the standard two-sided t-test, which rejects μ if

$$\frac{\sqrt{K}|\bar{Z}-\mu|}{S} \geq t_{K-1;1-\alpha/2}.$$

It is shown in Section 5.2 of Lehmann et al. (2005) that this standard t-test is UMP unbiased, which is equivalent to saying that the CI is uniformly most accurate unbiased (as claimed on page 165 of Lehmann et al. (2005), a CI constructed based on a UMP unbiased test is uniformly most accurate). Hence we have shown the desired result. □

3 OPTIMALITY AMONG BATCHING WITH GENERAL BATCH SIZES

To further our investigation, we consider potential alternative schemes where different batches can have different sizes. More precisely, suppose that the j-th batch has $\gamma_j n$ samples (assume for simplicity that $\gamma_j n, j = 1,2,\ldots$ are integers) where $\gamma_1 + \gamma_2 + \cdots + \gamma_K = 1$. Correspondingly, the batch estimates are given by $\psi_j^{(\gamma)} = \psi(\hat{P}_{\gamma_j n})$ where $\hat{P}_{\gamma_j n}$ is the empirical distribution of $X_{\left(\sum_{i=1}^{j-1}\gamma_i n\right)+1}, \ldots, X_{\sum_{i=1}^{j}\gamma_i n}$ The comparison between this setting and the equal-size case cannot be covered by the notion of optimality specified in Definition 1 because the batch estimates are different. Therefore, to start with, we study the optimal CI using batch estimates $\psi_1^{(\gamma)}, \ldots, \psi_K^{(\gamma)}$ when γ is fixed.

Theorem 2 Suppose that Assumption 1 holds. When the batch estimates are given by $\mathbf{Y}_{n,B}^{(\gamma)} = (\psi_1^{(\gamma)}, \ldots, \psi_K^{(\gamma)})^T$, an asymptotically uniformly most accurate unbiased CI at level $1-\alpha$ is given by

$$CI_B^{(\gamma)} := \bar{\psi}^{(\gamma)} \pm t_{K-1;1-\alpha/2} \sqrt{\sum_{j=1}^{K} \gamma_j (\psi_j^{(\gamma)} - \bar{\psi}^{(\gamma)})^2} / \sqrt{K-1}$$

where $\bar{\psi}^{(\gamma)} = \sum_{j=1}^{K} \gamma_j \psi_j^{(\gamma)}$.

Sketch of Proof. We have that $\sqrt{n}(\mathbf{Y}_{n,B}^{(\gamma)} - \psi(P)1_K) \Rightarrow N(0, \sigma^2 \mathrm{Diag}\{1/\gamma_1, \ldots, 1/\gamma_K\})$. Therefore, the problem is asymptotically equivalent to constructing a CI for μ given $Z_j \sim N(\mu, \sigma^2/\gamma_j)$ and $Z_1, \ldots, Z_K$ are independent. The corresponding hypothesis test is to test for a given null hypothesis μ_0 against the two-sided alternative $\mu \neq \mu_0$. Without loss of generality, we suppose that $\mu_0 = 0$ (otherwise, we can subtract μ_0 in every Z_j and reduce to the case where $\mu_0 = 0$).

The joint density of $Z = \{Z_1, \ldots, Z_K\}$ can be written as

$$\begin{aligned} p(z, \mu, \sigma, \gamma) &= C(\sigma^2, \gamma) \exp\left\{ -\sum_{j=1}^{K} \frac{\gamma_j}{2\sigma^2} (z_j - \mu)^2 \right\} \\ &= C(\sigma^2, \gamma) \exp\left\{ -\frac{\mu^2}{2\sigma^2} \right\} \exp\left\{ \frac{\sum_{j=1}^{K} \gamma_j z_j \mu}{\sigma^2} \right\} \exp\left\{ -\frac{\sum_{j=1}^{K} \gamma_j z_j^2}{2\sigma^2} \right\} \end{aligned}$$

Note that, when $\mu_0 = 0$, the test is equivalent to the test with null hypothesis $\theta := \frac{\mu}{\sigma^2} = 0$ and we can regard $\vartheta := -\frac{1}{2\sigma^2}$ as the nuisance parameter. Then, we have that $U(Z) = \sum_{j=1}^{K} \gamma_j Z_j$ and $T(Z) = \sum_{j=1}^{K} \gamma_j Z_j^2$ in the notation of (5.1) of Lehmann et al. (2005).

Let

$$V := \frac{\sum_{j=1}^{K} \gamma_j Z_j}{\sqrt{\sum_{i=1}^{K} \gamma_i \left(Z_i - \sum_{j=1}^{K} \gamma_j Z_j\right)^2}} = \frac{U(Z)}{\sqrt{T(Z) - U(Z)^2}}$$

where we used the decomposition $\sum_{j=1}^{K} \gamma_j Z_j^2 = \sum_{i=1}^{K} \gamma_i \left(Z_i - \sum_{j=1}^{K} \gamma_j Z_j\right)^2 + \left(\sum_{j=1}^{K} \gamma_j Z_j\right)^2$ in the second equality. Note that under the null hypothesis, V and T are independent as a result of Corollary 5.1.1 of Lehmann et al. (2005). Therefore, from Theorem 7 in Appendix A.2 and the fact that the distribution of V is symmetric about 0, we have that the test which rejects the null hypothesis when $|V|$ is larger than its $1-\alpha$

quantile under the null hypothesis (denoted by $v_{1-\alpha}$) is UMP unbiased. When $\mu_0 \neq 0$, the aforementioned subtraction argument will give a test which rejects μ_0 when $\left|\frac{\sum_{j=1}^K \gamma_j Z_j - \mu_0}{\sqrt{\sum_{i=1}^K \gamma_i \left(Z_i - \sum_{j=1}^K \gamma_j Z_j\right)^2}}\right| > v_{1-\alpha}$. This justifies the form of the claimed CI (by replacing Z_j with $\psi_j^{(\gamma)}$ to go back to the finite-sample version).

It remains to show that $v_{1-\alpha} = t_{K-1;1-\alpha/2}/\sqrt{K-1}$. Indeed, we will show a stronger claim: the distribution of $\sqrt{K-1}V$ is exactly t_{K-1} when $\mu = 0$. From the property of normal distribution, we know that $\sum_{j=1}^K \gamma_j Z_j$ is independent of $Z_i - \sum_{j=1}^K \gamma_j Z_j$ (since their covariance is 0). It is not hard to see that the numerator of V has distribution $N(0,\sigma^2)$, so it suffices to show that the denominator of V has distribution $\sqrt{\sigma^2 \chi^2_{K-1}}$. Denote $\tilde{Z}_j := \sqrt{\gamma_j} Z_j/\sigma$ which has distribution $N(0,1)$. Then the denominator of V can be expressed as the square root of $\sigma^2 \sum_{i=1}^K \left(\tilde{Z}_i - \sqrt{\gamma_i}\sum_{j=1}^K \sqrt{\gamma_j}\tilde{Z}_j\right)^2$. We have the sum-of-squares decomposition written in terms of $\tilde{Z}_j, j = 1,2,\ldots,K$:

$$\sum_{i=1}^K \tilde{Z}_i^2 = \sum_{i=1}^K \left(\tilde{Z}_i - \sqrt{\gamma_i}\sum_{j=1}^K \sqrt{\gamma_j}\tilde{Z}_j\right)^2 + \left(\sum_{j=1}^K \sqrt{\gamma_j}\tilde{Z}_j\right)^2$$

Therefore, by applying Cochran's theorem (Theorem 8 in Appendix A.3), we have that $\sum_{i=1}^K \left(\tilde{Z}_i - \sqrt{\gamma_i}\sum_{j=1}^K \sqrt{\gamma_j}\tilde{Z}_j\right)^2 \sim \chi^2_{K-1}$, which finishes the proof. □

The idea arguing Theorem 2 is to reduce the problem to finding a uniformly most accurate CI for the mean using Gaussian data with equal mean but unequal variances. From the duality between CI and hypothesis testing, this is equivalent to finding a UMP unbiased hypothesis test, which can be done via the test representation for exponential families in Theorem 7. Next, with the optimal interval we found in Theorem 2, we can compare different choices of γ by considering the expected length of the optimal intervals under these batch sizes.

Theorem 3 Suppose that the conditions in Theorem 2 hold. For any $\gamma = (\gamma_1,\ldots,\gamma_K)$ such that $\gamma_i > 0$ and $\gamma_1 + \cdots + \gamma_K = 1$, we have that

$$\lim_{n\to\infty} \frac{\text{expected length of } CI_B^{(\gamma)}}{\text{expected length of } CI_B} = 1.$$

Sketch of Proof. From the CLT assumption, we have that

$$n\sum_{j=1}^K \gamma_j(\psi_j^{(\gamma)} - \bar{\psi}^{(\gamma)})^2 \Rightarrow \sum_{i=1}^K \gamma_i \left(Z_i - \sum_{j=1}^K \gamma_j Z_j\right)^2 \tag{1}$$

where $Z_j \sim N(0,\sigma^2/\gamma_j)$. From the proof sketch of Theorem 2, the RHS of (1) has distribution $\sigma^2\chi^2_{K-1}$ (regardless of the choice of γ). Therefore, for any γ, the expected length is asymptotically equivalent to $n^{-1/2}\sigma t_{K-1;1-\alpha/2} E\sqrt{\chi^2_{K-1}/(K-1)}$. This gives the desired result. □

Theorem 3 implies that regardless of the way to divide into batches, an asymptotically uniformly most accurate unbiased CI has the same asymptotic length.

4 OPTIMALITY AMONG BATCHING WITH OVERLAPPING BATCHES

The initial overlapping batching proposed in Meketon and Schmeiser (1984) considers the case where the distance between the starting indexes of adjacent batches is only 1. For the mean estimation example studied in Meketon and Schmeiser (1984), one can use the trick of computing the next batch mean by updating from the previous batch mean. However, when the output is a general functional, such a trick

is not available. Therefore, for a general output functional, to study overlapping batching under a fixed simulation budget, we consider the setting where we have K overlapping batches and correspondingly K batch estimates $\psi_{1,OB},\ldots,\psi_{K,OB}$. Suppose that the j-th batch has batch size $\gamma_j n$ and for $1\le i,j\le K$, batch i and batch j share $\beta_{ij}n$ samples. We assume that the batch estimates converge in distribution to a multivariate normal distribution.

Assumption 2 Let $\mathbf{Y}_{n,OB}:=(\psi_{1,OB},\ldots,\psi_{K,OB})^T$. $\sqrt{n}(\mathbf{Y}_{n,OB}-\psi(P)1_K)\Rightarrow N(0,\sigma^2V)$ where $V_{ii}=1/\gamma_i$ and $V_{ij}=\frac{\beta_{ij}}{\gamma_i\gamma_j}$ for $i\neq j, 1\le i,j,\le K$. Moreover, V is invertible.

It can be checked that the CLT in Assumption 2 holds under regularity conditions. The assumption that V is invertible guarantees that the asymptotic distribution of $\mathbf{Y}_{n,OB}$ is not degenerate. Otherwise, some coordinates of $\mathbf{Y}_{n,OB}$ can be asymptotically represented as a linear combination of other components and the problem can be reduced to another case with smaller K.

Theorem 4 Suppose that Assumption 2 holds. When the batch estimates are given by $\mathbf{Y}_{n,OB}$, an asymptotically uniformly most accurate unbiased CI is given by

$$CI_{OB}:=\frac{1_K^TV^{-1}\mathbf{Y}_{n,OB}}{\lambda}\pm\frac{t_{K-1;1-\alpha/2}}{\sqrt{K-1}}\sqrt{\frac{1}{\lambda}\left(\mathbf{Y}_{n,OB}-\frac{1_K^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda}1_K\right)^\top V^{-1}\left(\mathbf{Y}_{n,OB}-\frac{1_K^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda}1_K\right)}$$

where $\lambda=1_K^\top V^{-1}1_K$.

Sketch of Proof. The proof follows a similar structure as the proof of Theorem 2 with more general algebra. With Assumption 2, the problem can be reduced to the construction of CI for μ using data $Z\sim N(\mu 1_K,\sigma^2V)$ where μ and σ^2 are unknown. We study the hypothesis test for $H_0:\mu=\mu_0$ against $H_1:\mu\neq\mu_0$ given these data and consider the case where $\mu_0=0$ first. The density function of Z can be written as

$$\begin{aligned}p(z,\mu,\sigma)&=C(\sigma)\exp\left\{-\frac{1}{2\sigma^2}(z-\mu 1_K)^\top V^{-1}(z-\mu 1_K)\right\}\\&=C(\sigma)\exp\left\{-\frac{z^\top V^{-1}z}{2\sigma^2}\right\}\exp\left\{\frac{\mu 1_K^\top V^{-1}z}{2\sigma^2}\right\}\exp\left\{-\frac{1}{2\sigma^2}\right\}\end{aligned}$$

Therefore, similar to the proof of Theorem 2, letting $\theta:=\mu/\sigma^2$ and regarding $\vartheta:=-\frac{1}{2\sigma^2}$ as the nuisance parameter, we have that $U(Z)=1_K^\top V^{-1}Z$ and $T(Z)=Z^TV^{-1}Z$. Let

$$V:=\frac{1_K^\top V^{-1}Z/\lambda}{\sqrt{1/\lambda}\sqrt{\left(Z-\frac{1_K^\top V^{-1}Z}{\lambda}1_K\right)^\top V^{-1}\left(Z-\frac{1_K^\top V^{-1}Z}{\lambda}1_K\right)}}=\frac{U(Z)/\lambda}{\sqrt{1/\lambda}\sqrt{T(Z)-U(Z)^2/\lambda}}$$

Again, it follows from Corollary 5.1.1 of Lehmann et al. (2005) that V and T are independent. As a result, Theorem 7 in Appendix A.2 gives that the UMP unbiased test rejects the null hypothesis when $|V|$ is larger than its $1-\alpha$ quantile under $\mu=0$ (denoted by $v_{1-\alpha,OB}$). For general $\mu_0\neq 0$, subtracting Z by $\mu_0 1_K$ gives the rejection region

$$\left\{Z:\frac{\left|1_K^\top V^{-1}Z/\lambda-\mu_0\right|}{\sqrt{1/\lambda}\sqrt{\left(Z-\frac{1_K^\top V^{-1}Z}{\lambda}1_K\right)^\top V^{-1}\left(Z-\frac{1_K^\top V^{-1}Z}{\lambda}1_K\right)}}>v_{1-\alpha,OB}\right\}$$

which is the dual of CI_{OB} after replacing $\mathbf{Y}_{OB}$ with Z. Next, we show that $v_{1-\alpha;OB}=t_{K-1,1-\alpha/2}$ by showing that $\sqrt{K-1}V$ has distribution t_{K-1} when $\mu=0$. Suppose in what follows $\mu=0$. The numerator of

V has distribution $N(0,\sigma^2/\lambda)$. Therefore, it suffices to show that $\sqrt{\lambda}$ times the denominator of V has distribution $\sqrt{\chi^2_{K-1}}$. Denote $\tilde{Z} = V^{-1/2}Z/\sigma$ which has standard K-dimensional normal distribution. We have the sum-of-squares decomposition

$$\tilde{Z}^\top \tilde{Z} = \left(\tilde{Z} - \frac{V^{-1/2}1_K 1_K^\top V^{-1/2}\tilde{Z}}{\lambda}\right)^\top \left(\tilde{Z} - \frac{V^{-1/2}1_K 1_K^\top V^{-1/2}\tilde{Z}}{\lambda}\right) + \left(1_K^\top V^{-1/2}\tilde{Z}\right)^2/\lambda \tag{2}$$

Therefore, by Cochran's theorem and Lemma 1, we have that the first term on the RHS in (2) has distribution χ^2_{K-1}. This is equivalent to the claim that $\sqrt{\lambda}$ times the denominator of V has distribution $\sqrt{\chi^2_{K-1}}$. Hence, the proof is concluded. □

Similar to Theorem 3, from the proof of Theorem 4, we have the asymptotic comparison of the lengths of CI_{OB} and CI_B.

Theorem 5 Under the same condtions as in Theorem 4, we have that

$$\lim_{n\to\infty} \frac{\text{expected length of } CI_{OB}}{\text{expected length of } CI_B} = \sqrt{\frac{1}{\lambda}} \geq 1.$$

The inequality holds as an equality if and only if there exists a vector u such that $1^\top u = 1$ and $u^\top V u = 1$.

Sketch of Proof. From Assumption 2, we have the asymptotic for the length:

$$\begin{aligned} &\sqrt{n}\frac{1}{\sqrt{K-1}}\sqrt{\frac{1}{\lambda}\left(\mathbf{Y}_{n,OB} - \frac{1_K^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda}1_K\right)^\top V^{-1}\left(\mathbf{Y}_{n,OB} - \frac{1_K^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda}1_K\right)} \\ \Rightarrow &\sqrt{1/\lambda}\sqrt{\frac{1}{K-1}\left(Z - \frac{1_K^\top V^{-1}Z}{\lambda}1_K\right)^\top V^{-1}\left(Z - \frac{1_K^\top V^{-1}Z}{\lambda}1_K\right)} \end{aligned} \tag{3}$$

The proof sketch of Theorem 4 implies that the limit (3) has distribution $\sqrt{\frac{\chi^2_{K-1}}{\lambda(K-1)}}$. Therefore, comparing with Theorem 3, we get the equality part of the claim.

To see the inequality, note that $1/\lambda$ is the optimal value of the optimization problem

$$\min w^\top V w \quad s.t. \quad 1_K^\top w = 1.$$

In other words, σ^2/λ is the smallest possible variance among all possible affine combinations of Z. On the other hand, by the construction of this problem, we know that for any affine combination of $\mathbf{Y}_{n,OB}$, its asymptotic variance should be larger than or equal to the asymptotic variance of $\psi(\hat{P}_n)$ (which is σ^2). Therefore, $\lambda \leq 1$, which gives the inequality of the claim. The claimed condition for the inequality to be held as equality is also clear from the above analysis. □

Note that a sufficient condition for the inequality to be held as equality in Theorem 5 is that $\mathbf{Y}_{n,OB}$ contains $\psi(\hat{P}_n)$ as one of its coordinate (in which case u in Theorem 5 can be set as e_i where the i-th entry of e_i is 1, all other entries of e_i are 0, and i is the coordinate of $\psi(\hat{P}_n)$ in $\mathbf{Y}_{n,OB}$). The condition also holds if $\psi(\hat{P}_n)$ can be asymptotically written as an affine combination of the entries of $\mathbf{Y}_{n,OB}$ (in which case u can be chosen as the vector of the coefficients in this affine combination). Theorefore, Theorem 5 indicates that the asymptotically uniformly most accurate unbiased CI has the same asymptotic length in the more general setting with overlapping batches, provided that the aforementioned condition holds. We discuss a few examples and make a comparision with the literature.

Example 1 (batched jackknife) For batched jackknife, we have $\psi_{1,SJ},\dots,\psi_{K,SJ}$ where $\psi_{j,SJ}$ is given by the estimate using leave-one-batch-out data $\{X_1,\dots,X_n\}\backslash\{X_{(j-1)n/K+1},\dots,X_{jn/K}\}$. In this case, we have that $\gamma_j=(K-1)/K$ and $\beta_{ij}=(K-2)/K$. Therefore, $V=\frac{I+(K-2)1_K1_K^\top}{K}$, $V^{-1}=K\left(I-\frac{K-2}{(K-1)^2}1_K1_K^\top\right)$, $\lambda=\frac{K^2}{(K-1)^2}$. Based on these, it can be checked that CI_{OB} has point estimator $\frac{1_K^TV^{-1}\mathbf{Y}_{n,OB}}{\lambda}=\frac{1_K^T\mathbf{Y}_{n,OB}}{K}$ and variance estimator

$$\begin{aligned}&\frac{1}{\lambda}\left(\mathbf{Y}_{n,OB}-\frac{1_K^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda}1_K\right)^\top V^{-1}\left(\mathbf{Y}_{n,OB}-\frac{1_K^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda}1_K\right)\\ =&\frac{1}{K}\left(\left(1_K1_K^T-(K-1)I-\frac{1}{K}1_K1_K^\top\right)\mathbf{Y}_{n,OB}\right)^\top\left(1_K1_K^T-(K-1)I-\frac{1}{K}1_K1_K^\top\right)\mathbf{Y}_{n,OB}\end{aligned}\tag{4}$$

A proof of (4) is provided in Appendix B. Note that $\frac{1}{K}1_K1_K^\top\mathbf{Y}_{n,OB}$ is the average of the batch estimates and the i-th coordinate of $\left(1_K1_K^\top-(K-1)I\right)\mathbf{Y}_{OB}$ is given by $\sum_{j=1}^K\psi_{j,SJ}-(K-1)\psi_{i,SJ}$. Therefore, the asymptotically most accurate unbiased CI CI_{OB} derived in our paper is equivalent to the construction in Section III.5b of Asmussen and Glynn (2007).

Example 2 We consider the overlapping batching studied in Su et al. (2023). It considers a class of overlapping schemes where each batch has equal sizes denoted by γn and adjacent batches have equal distances, or more concretely, batch j starts from $(j-1)\frac{n-\gamma n}{K-1}+1$ and ends at $(j-1)\frac{n-\gamma n}{K-1}+\gamma n$. As can be checked, a batch could overlap with other batches when $\gamma>1/K$. Let ψ_j denote the batch estimate using the empirical distribution of batch j and let $\psi_{K+1}:=\hat{\psi}$ denote $\psi(\hat{P}_n)$, which is the estimate using the entire empirical distribution. Therefore, we have that $\gamma_i=\gamma$ and $\beta_{ij}=\left(\gamma-|i-j|\frac{1-\gamma}{K-1}\right)_+$ for $1\le i,j\le K$. Also we have that $\gamma_{K+1}=1$ and $\beta_{i,K+1}=\gamma$. Therefore, an asymptotically uniformly most accurate unbiased CI can be constructed using Theorem 4 (we do not provide the explicit expression for the CI here for simplicity). However, we note that this will be different from the CI proposed in Su et al. (2023). Indeed, the CI constructed in Theorem 4 is calibrated based on a pivotal statistic with an asymptotic t_K distribution, while the CI in Su et al. (2023) is calibrated based on a more general distribution (which they denote by OB-I). This suggests that we can potentially obtain a better CI (in the notion of our Definition 1) using the construction in Theorem 4. A further understanding of the improvement, e.g., comparisons in terms of the length of the CI, is left for future work.

Finally, note that there are other variants of batching such as sectioning, which works similarly as standard batching, but its point estimate is based on the entire empirical distribution (i.e., $\psi(\hat{P}_n)$) as the point estimator). In the notion of this section, we have that $\mathbf{Y}_{n,OB}=\left(\psi(\hat{P}_n),\psi(\hat{P}_{n/K}^{(1)}),\dots,\psi(\hat{P}_{n/K}^{(K)})\right)$ where $\hat{P}_{n/K}^{(j)}$ is the j-th batch estimate defined as in Section 2. However, the asymptotic distribution of $\mathbf{Y}_{n,OB}$ would be degenerate since

$$\sqrt{n}\left(\psi(\hat{P}_n)-(\psi(\hat{P}_{n/K}^{(1)})+\cdots+\psi(\hat{P}_{n/K}^{(K)}))/K\right)\to 0,\quad a.s.$$

under regularity conditions. In other words, replacing $\psi(\hat{P}_n)$ with the average of the batch estimates (which gives the standard batching) would lead to the same CI asymptotically as $n\to\infty$. Therefore, the difference between sectioning and standard batching lies only in the finite-sample case, or in the higher-order terms of the asymptotic.

5 EXTENSIONS TO DEPENDENT DATA

While we have assumed i.i.d. samples in our setup described at the beginning of Section 2, our theoretical developments discussed in Sections 2–4 can be generalized to the dependent data case by properly replacing (if needed) and verifying the required CLT assumptions in the latter case. To explain, first we can check that Theorem 4 holds as long as the joint CLT in the form specified in Assumption 2 holds for the batch

estimates. To generalize Theorems 1–3, we can relax Assumption 1 to a multivariate CLT assumption where asymptotic independence holds among different batch estimates. More precisely, from the proof sketches, we see that Theorem 1 holds when $\sqrt{n/K}(\mathbf{Y}_{n,B} - \psi(P)1_K) \Rightarrow N(0,\sigma^2 I)$ for some σ^2, and that Theorems 2–3 hold when $\sqrt{n}(\mathbf{Y}^{(\gamma)}_{n,B} - \psi(P)1_K) \Rightarrow N(0,\sigma^2 \text{Diag}\{1/\gamma_1,\dots,1/\gamma_K\})$, and these CLTs can serve as the assumptions for these theorems. Second, we can establish the validity of these CLTs in the dependent case by leveraging existing results in the literature. For example, it is shown in Glynn and Iglehart (1990) that when $\{X_1, X_2, \dots\}$ is stationary satisfying standard mixing conditions, we have that $\sqrt{n/K}(\mathbf{Y}_{n,B} - \psi(P)1_K)$ converges to a K-dimensional normal distribution whose components are i.i.d.. More generally, when $\psi(\cdot)$ is a functional of dependent data, a CLT for the joint distribution of the batch estimates is also available, which essentially requires that a normalized and interpolated version of $Z_t := \psi(X_1,\dots,X_t)$ converges in distribution to a Brownian motion (Su et al. 2023).

6 EXTENSIONS TO GENERAL t-DISTRIBUTION-BASED CONFIDENCE INTERVALS

The analysis in our paper can be generalized to settings even beyond batching, although in this case the extension to dependent data might become less straightforward. An example is the cheap bootstrap recently proposed in Lam (2022b) and further developed in Lam and Liu (2023) and Lam (2022a). The CI is constructed as $CI_{CB} := \psi(\hat{P}_n) \pm t_{K,1-\alpha/2} S$ where $\hat{P}_n$ is the empirical distribution and

$$S^2 = \frac{1}{K}\sum_{b=1}^{K} (\psi(P_n^{*b}) - \psi(\hat{P}_n))^2.$$

Here $P_n^{*b}, b = 1,2,\dots,K$ is the empirical distribution of $\{X_1^{*b},\dots,X_n^{*b}\}$ that are sampled independently and uniformly with replacement from $\{X_1,\dots,X_n\}$.

Under regularity conditions, it is shown in Proposition 1 of Lam (2022b) that the following joint CLT holds:

$$\left(\sqrt{n}(\psi(\hat{P}_n) - \psi(P)), \sqrt{n}(\psi(P_n^{*1}) - \psi(\hat{P}_n)), \dots, \sqrt{n}(\psi(P_n^{*K}) - \psi(\hat{P}_n))\right)^\top \Rightarrow N(0,\sigma^2 I).$$

This implies that

$$\left(\sqrt{n}(\psi(\hat{P}_n) - \psi(P)), \sqrt{n}(\psi(P_n^{*1}) - \psi(P)), \dots, \sqrt{n}(\psi(P_n^{*K}) - \psi(P))\right)^\top \Rightarrow N(0,\sigma^2 V).$$

where $V_{00} = 1, V_{0i} = 1, V_{ii} = 2, V_{ij} = 1$ for $1 \le i \ne j \le K$ and it can be calculated that

$$V^{-1} = \begin{bmatrix} K+1 & -1 & -1 & \dots & -1 \\ -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -1 & 0 & 0 & \dots & 1 \end{bmatrix}$$

From this, it can be calculated that the CI_{OB} estimator derived in Theorem 4 corresponding to this V and

$$\mathbf{Y}_{n,OB} = (\psi(\hat{P}_n), \psi(P_n^{*1}), \dots, \psi(P_n^{*K}))$$

has point estimator $\frac{1_K^T V^{-1}\mathbf{Y}_{n,OB}}{\lambda} = \psi(\hat{P}_n)$ and variance estimator

$$\frac{1}{\lambda}\left(\mathbf{Y}_{n,OB} - \frac{1_{K+1}^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda} 1_1\right)^\top V^{-1}\left(\mathbf{Y}_{n,OB} - \frac{1_{K+1}^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda} 1_{K+1}\right)$$
$$= \sum_{j=1}^{K} (\psi(P_n^{*j}) - \psi(P))^2$$

which is the same as CI_{CB}. Therefore, with the same proof as Theorem 4 (note that it applies as long as there is a joint CLT for $\mathbf{Y}_{n,OB}$), we have that CI_{CB} is asymptotically uniformly most accurate unbiased given data $\psi(\hat{P}_n), \psi(P_n^{*1}), \ldots (P_n^{*b})$.

7 CONCLUSION AND DISCUSSION

In this paper, we studied optimal batching methods given fixed computation budgets using our notion of an asymptotically uniformly most accurate unbiased CI that connects to the powers of dual hypothesis tests. We showed that the standard batching is optimal among the class of equal-sized batches. We also derived the corresponding optimal CIs among general-sized and overlapping batching schemes. In particular, we showed how our derived optimal overlapping batching CIs coincide but also differ from some existing suggestions. Moreover, we showed that these optimal CIs have asymptotically equivalent expected lengths. We discussed how our results apply to general dependent data situations. Finally, we discussed extensions of our study to show the optimality of the cheap bootstrap.

In addition, we provide numerical experiments on the CIs studied in this paper in He and Lam (2023). We tested the empirical coverage and half length for standard batching, batching with general batch sizes, overlapping batching, cheap bootstrap, and OB-I as proposed in Su et al. (2023), under the same computational budget on a quantile estimation problem. The experimental results show that the empirical half lengths of batching, batching with general batch sizes, and cheap bootstrap are quite close, which appear consistent with Theorem 3 as well as our deduction in the end of Section 6. They also reveal that the construction in Theorem 4 gives shorter CI compared to OB-I. Nonetheless, further and more complete analyses and numerical experiments are needed to justify our claims more generally and in the dependent data case. We view our study as the starting point in building a general analyzable framework to systematically understand the performance comparisons and optimality of inference schemes under computation budget constraints.

A REVIEW OF SOME USEFUL RESULTS

We review some results in statistics used in our paper.

A.1 Duality Between Uniformly Most Accurate Unbiased CI and UMP Unbiased Hypothesis Test

We consider two-sided CI $\mathscr{C}(\mathbf{X}) = (L(\mathbf{X}), U(\mathbf{X}))$ for some quantity θ given data $\mathbf{X}$ (whose distribution depends on θ and nuisance parameters ϑ but otherwise general). Its dual is hypothesis testing for $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta \neq \theta_0$ with the rejection region given by $\{\mathbf{x} : \theta_0 \notin \mathscr{C}(\mathbf{x})\}$.

$\mathscr{C}(\mathbf{X})$ is called unbiased at confidence level $1-\gamma$ if $P_{\theta_0,\vartheta}(\theta_0 \in \mathscr{C}(\mathbf{X})) \geq 1-\gamma$ and $P_{\theta_0,\vartheta}(\theta' \in \mathscr{C}(\mathbf{X})) \leq 1-\gamma$ for any $\theta' \neq \theta_0$ and all ϑ. $\mathscr{C}(\mathbf{X})$ is called uniformly most accurate unbiased at confidence level $1-\gamma$ if it minimizes $P_{\theta_0,\vartheta}(\theta' \in \mathscr{C}'(\mathbf{X}))$ for all $\theta' \neq \theta_0$ and all ϑ, θ_0 among unbiased CIs at level $1-\gamma$.

A hypothesis test with rejection region $\mathscr{R}$ is called unbiased with size γ if $P_{\theta_0}(\mathbf{X} \in \mathscr{R}) \leq \gamma$ and $P_\theta(\mathbf{X} \in \mathscr{R}) \geq \gamma$ for any $\theta \neq \theta_0$. It is UMP unbiased if it maximizes $P_\theta(\mathbf{X} \in \mathscr{R})$ for any $\theta \neq \theta_0$.

The relation between a uniformly most accurate unbiased CI and a UMP unbiased hypothesis test is described by the following theorem.

Theorem 6 If $\mathscr{R}_{\theta_0}$ is a family of UMP unbiased tests with size γ for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$, then $\mathscr{C}(\mathbf{X}) := \{\theta : \mathbf{X} \in \mathscr{R}_\theta\}$ is a uniformly most accurate unbiased CI at level $1-\gamma$.

A.2 Constructing UMP Unbiased Tests for Exponential Families

The following theorem is adapted from Theorem 5.1.1 of Lehmann et al. (2005).

Theorem 7 Consider testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ using data $\mathbf{X}$. Suppose that the distribution of $\mathbf{X}$ is parameterized by θ and nuisance parameter ϑ with density given by

$$p(x, \theta, \vartheta) = C(\theta, \vartheta) \exp\{\theta U(x) + \sum \vartheta_i T_i(x)\}$$

Suppose that there exists a continuous random variable $V(\mathbf{X}) = h(U(\mathbf{X}), T(\mathbf{X}))$ that is independent of $T(\mathbf{X})$ when $\theta = \theta_0$. Moreover, suppose that when $\theta = \theta_0$, the distribution of $V(\mathbf{X})$ is symmetric about 0 and $P_{\theta_0}(|V(\mathbf{X})| \geq v_\gamma) = 1 - \gamma$. Then, the test with rejection region $\{x : |V(x)| > v_\gamma\}$ is UMP unbiased provided that $h(u,t) = a(t)u + b(t)$ with $a(t) > 0$.

A.3 Cochran's Theorem

Cochran's theorem provides a way to conclude that a quadratic form of multivariate standard normal follows a chi-squared distribution using a sum-of-squares decomposition.

Theorem 8 (Cochran's theorem) Suppose that $U_1, \ldots, U_K$ are i.i.d. standard normals and denote $U = (U_1, \ldots, U_K)$. Suppose that we have the sum-of-squares decomposition

$$U^\top U = \sum_{i=1}^{m} U^\top B_i U$$

where B_i is a square matrix with rank r_i. Then $U^\top B_i U$ follows a chi-square distribution with r_i degrees of freedom.

B MATH DETAILS

B.1 A Lemma on Linear Algebra

Lemma 1 Let V and λ be defined as in Theorem 4. The rank of matrix $I - \frac{V^{-1/2} 1_K 1_K^\top V^{-1/2}}{\lambda}$ is $K-1$.

Proof. From the property that $rank(A) + rank(B) \geq rank(A+B)$, we have that $rank\left(I - \frac{V^{-1/2} 1_K 1_K^\top V^{-1/2}}{\lambda}\right) \geq rank(I) - rank(\frac{V^{-1/2} 1_K 1_K^\top V^{-1/2}}{\lambda}) = K-1$.

On the other hand, since $\left(I - \frac{V^{-1/2} 1_K 1_K^\top V^{-1/2}}{\lambda}\right) 1_K = 0$, we have that $rank\left(I - \frac{V^{-1/2} 1_K 1_K^\top V^{-1/2}}{\lambda}\right) \geq rank(I) \leq K-1$. Combining this with the lower bound, we get the desired result. □

B.2 Derivation of (4)

$$\begin{aligned}
&\frac{1}{\lambda}\left(\mathbf{Y}_{n,OB} - \frac{1_K^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda} 1_K\right)^\top V^{-1}\left(\mathbf{Y}_{n,OB} - \frac{1_K^\top V^{-1}\mathbf{Y}_{n,OB}}{\lambda} 1_K\right) \\
=&\frac{(K-1)^2}{K}\left(\mathbf{Y}_{n,OB} - \frac{1_K 1_K^\top \mathbf{Y}_{n,OB}}{K}\right)^\top \left(I - \frac{K-2}{K-1} 1_K 1_K^\top\right)\left(\mathbf{Y}_{n,OB} - \frac{1_K 1_K^\top \mathbf{Y}_{n,OB}}{K}\right) \\
=&\frac{(K-1)^2}{K}\mathbf{Y}_{n,OB}^\top \left(I - \frac{1}{K} 1_K 1_K^\top\right)\left(I - \frac{K-2}{K-1} 1_K 1_K^\top\right)\left(I - \frac{1}{K} 1_K 1_K^\top\right)\mathbf{Y}_{n,OB} \\
=&\frac{(K-1)^2}{K}\left(\left(I - \frac{1}{K} 1_K 1_K^\top\right)\mathbf{Y}_{n,OB}\right)^\top \left(I - \frac{1}{K} 1_K 1_K^\top\right)\mathbf{Y}_{n,OB} \\
=&\frac{1}{K}\left(\left(1_K 1_K^T - (K-1)I - \frac{1}{K} 1_K 1_K^\top\right)\mathbf{Y}_{n,OB}\right)^\top \left(1_K 1_K^T - (K-1)I - \frac{1}{K} 1_K 1_K^\top\right)\mathbf{Y}_{n,OB}
\end{aligned}$$

Here, in the second equality, we used the fact that $\left(I - \frac{1}{K} 1_K 1_K^\top\right) 1_K = 0$.

ACKNOWLEDGEMENTS

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

REFERENCES

Alexopoulos, C., D. Goldsman, and J. R. Wilson. 2011. "Overlapping Batch Means: Something More for Nothing?". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. Creasey, J. Himmelspach, K. White, and M. Fu, 401–411. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. New York, NY: Springer.

Flegal, J. M., and G. L. Jones. 2010. "Batch Means and Spectral Variance Estimators in Markov Chain Monte Carlo". *The Annals of Statistics* 38(2):1034–1070.

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

He, S., and H. Lam. 2023. Numerical Experiments for "Optimal Batching under Computational Budget", https://github.com/Shengyi98/wsc23experiments/blob/main/wsc23report.pdf, assessed 14th September 2022.

Lam, H. 2022a. "Cheap Bootstrap for Input Uncertainty Quantification". In *Proceedings of the 2022 Winter Simulation Conference*, edited by B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, E. Chew, T. Roeder, and P. Lendermann, 2318–2329. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lam, H. 2022b. "A Cheap Bootstrap Method for Fast Inference". *arXiv:2202.00090*.

Lam, H., and Z. Liu. 2023. "Bootstrap in High Dimension With Low Computation". In *Proceedings of the 40th International Conference on Machine Learning*. July 23th-29th, Honolulu, Hawaii, USA.

Lehmann, E. L., J. P. Romano, and G. Casella. 2005. *Testing Statistical Hypotheses*, Volume 3. New York, NY: Springer.

Meketon, M. S., and B. Schmeiser. 1984. "Overlapping Batch Means: Something for Nothing?". In *Proceedings of the 1984 Winter Simulation Conference*, edited by S. Sheppard, U. Pooch, and D. Pegden, 226–230. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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 Transactions on Modeling and Computer Simulation* 24(4):1–21.

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

Song, W. T., and B. W. Schmeiser. 1995. "Optimal Mean-squared-error Batch Sizes". *Management Science* 41(1):110–123.

Su, Z., R. Pasupathy, Y. Yeh, and P. W. Glynn. 2023. "Overlapping Batch Confidence Intervals on Statistical Functionals Constructed from Time Series: Application to Quantiles, Optimization, and Estimation". *arXiv:2307.08609*.

AUTHOR BIOGRAPHIES

SHENGYI HE is a Ph.D. 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. His research interests include Monte Carlo methods, uncertainty quantification, risk analysis and data-driven optimization. He serves as the area editor for Stochastic Models and Data Science in *Operations Research Letters*, and on the editorial boards of *Operations Research*, *INFORMS Journal on Computing*, *Applied Probability Journals*, *Stochastic Models*, *Manufacturing and Service Operations Management*, and *Queueing Systems*. His email address is henry.lam@columbia.edu and his website is http://www.columbia.edu/~khl2114/.