Unbiased time-average estimators for Markov chains

Nabil Kahalé

ESCP Business School

MCM, June 2023

Table of Contents

1 Introduction

2 Conventional time-average estimators

3 Unbiased time-average estimators

4 Examples

5 Numerical experiments

6 Conclusion

Definition

- Consider a Markov chain $(X_i, i \geq 0)$ with state-space F
- X_0 is deterministic
- Let $f : F \to \mathbb{R}$ be a measurable function on F
- Assume that $E((f(X_i))^2) < \infty$ for $i \geq 0$.
- For $k \geq 1$, define the time-average estimator

$$f_k := \frac{1}{k - b(k)} \sum_{i=b(k)}^{k-1} f(X_i)$$

- $b(k)$ is a burn-in period with $0 \leq b(k) \leq k/2$

Definition

- Consider a Markov chain $(X_i, i \geq 0)$ with state-space F
- X_0 is deterministic
- Let $f : F \to \mathbb{R}$ be a measurable function on F
- Assume that $E((f(X_i))^2) < \infty$ for $i \geq 0$.
- For $k \geq 1$, define the time-average estimator

$$f_k := \frac{1}{k - b(k)} \sum_{i=b(k)}^{k-1} f(X_i)$$

- $b(k)$ is a burn-in period with $0 \leq b(k) \leq k/2$

Applications of time-average estimators

- f_k estimates the limit μ of $E(f(X_m))$ as $m \to \infty$, when it exists
- f_k is usually biased: $E(f_k) \neq \mu$

Time-average estimators have been used in

- sampling from a posterior distribution (Tierney 1994)
- computing the volume of a convex body (Cousins and Vempala 2016)
- estimating the steady-state metrics of time-dependent queues (Whitt and You 2019)

Long runs versus short runs

(Whitt 1991) finds that

- One long time-average estimator is more efficient than several independent replications of short time-average estimators
- If the simulation length is large enough to obtain reasonable estimates of μ, then several independent replications are almost as efficient as one longer run

Confidence intervals

- Because of the bias and since the $f(X_i)$'s are usually correlated, calculating confidence intervals from time-average estimators is challenging
- The method of batch means divides f_k into several consecutive batches
- It calculates an asymptotic confidence interval from the averages over each batch
- The quality of this confidence interval depends on the extent to which these averages are i.i.d. and Gaussian (Asmussen and Glynn 2007, p. 110)

Confidence intervals

- Because of the bias and since the $f(X_i)$'s are usually correlated, calculating confidence intervals from time-average estimators is challenging
- The method of batch means divides f_k into several consecutive batches
- It calculates an asymptotic confidence interval from the averages over each batch
- The quality of this confidence interval depends on the extent to which these averages are i.i.d. and Gaussian (Asmussen and Glynn 2007, p. 110)

Bias and mixing time

- The bias of f_k is related to the mixing time
- Explicit convergence rates to the steady-state distribution have been established for many Markov chains (Diaconis and Stroock 1991, Sinclair 1992, Cousins and Vempala 2016, Kahalé 2019, Barkhagen, Chau, Moulines, Rásonyi, Sabanis and Zhang 2021)
- The mixing time of other Markov chains arising in practice is not formally known (Diaconis 2009)

Overview of results

- Under suitable conditions, we first construct a Randomized Multilevel Monte Carlo (RMLMC) unbiased estimator Z_k of the bias
- That is, $E(Z_k) = \mu - E(f_k)$
- Combining Z_k with f_k yields an unbiased estimator $\hat{f}_k$ of μ
- That is, $E(\hat{f}_k) = \mu$

Properties of estimators

- $E({Z_k}^2) < \infty$
- $E(\hat{f}_k^2) < \infty$
- Z_k can be simulated in finite expected time
- $\hat{f}_k$ can be simulated in finite expected time $\hat{T}_k$
- Under certain assumptions, for sufficiently large k

$$\hat{T}_k \mathrm{Variance}(\hat{f}_k) \leq (1+\epsilon) k \mathrm{Variance}(f_k)$$

- Under suitable conditions, $\hat{f}_k$ can be constructed without any precomputations
- In our numerical experiments, f_k is about twice as efficient as $\hat{f}_k$ for large values of k

Properties of estimators

- $E({Z_k}^2) < \infty$
- $E(\hat{f}_k^2) < \infty$
- Z_k can be simulated in finite expected time
- $\hat{f}_k$ can be simulated in finite expected time $\hat{T}_k$
- Under certain assumptions, for sufficiently large k

$$\hat{T}_k \mathrm{Variance}(\hat{f}_k) \leq (1+\epsilon) k \mathrm{Variance}(f_k)$$

- Under suitable conditions, $\hat{f}_k$ can be constructed without any precomputations
- In our numerical experiments, f_k is about twice as efficient as $\hat{f}_k$ for large values of k

Relation with (Glynn and Rhee 2014)

- Our construction is based on
  - a coupling assumption and a time-reversal transformation inspired from (Glynn and Rhee 2014)
  - A RMLMC estimator introduced by (Rhee and Glynn 2015)
- Assuming that f is Lipschitz and that X is 'contractive on average', (Glynn and Rhee 2014) construct square-integrable unbiased RMLMC estimators for the steady-state expectation of Markov chain functionals
- Their method is not based on time-averaging
- Our approach does not require f to be Lipschitz or X to be contractive on average

Relation with (Glynn and Rhee 2014)

- Our construction is based on
  - a coupling assumption and a time-reversal transformation inspired from (Glynn and Rhee 2014)
  - A RMLMC estimator introduced by (Rhee and Glynn 2015)
- Assuming that f is Lipschitz and that X is 'contractive on average', (Glynn and Rhee 2014) construct square-integrable unbiased RMLMC estimators for the steady-state expectation of Markov chain functionals
- Their method is not based on time-averaging
- Our approach does not require f to be Lipschitz or X to be contractive on average

Relation with (Glynn and Rhee 2014)

- Our construction is based on
    - a coupling assumption and a time-reversal transformation inspired from (Glynn and Rhee 2014)
    - A RMLMC estimator introduced by (Rhee and Glynn 2015)
- Assuming that f is Lipschitz and that X is ‘contractive on average’, (Glynn and Rhee 2014) construct square-integrable unbiased RMLMC estimators for the steady-state expectation of Markov chain functionals
- Their method is not based on time-averaging
- Our approach does not require f to be Lipschitz or X to be contractive on average

Relation with (Jacob, OLeary and Atchadé 2020)

- (Jacob, OLeary and Atchadé 2020) study unbiased Markov Chain Monte Carlo methods that use time-averaging
- They use the exact coupling of two Markov chains together with a telescopic sum argument of (Glynn and Rhee 2014)
- In contrast, our approach uses approximate coupling

Notation

- Assume that there are i.i.d. random variables U_i, $i \geq 0$, and a measurable function g such that

$$X_{i+1} = g(X_i, U_i)$$

- Thus, there is a measurable function G_i with

$$X_i = G_i(X_0; U_0, \dots, U_{i-1})$$

- Extend $(U_i, i \geq 0)$ to $(U_i, i \in \mathbb{Z})$ and assume it is i.i.d.
- For $i, m \in \mathbb{Z}$ with $m \leq i$, let

$$X_{m:i} := G_{i-m}(X_0; U_m, U_{m+1}, \dots, U_{i-1})$$

- $X_{0:i} = X_i$ for $i \geq 0$
- For $i, m \in \mathbb{Z}$ with $m \leq i$, we have $X_{m:i} \sim X_{i-m}$

Notation

- Assume that there are i.i.d. random variables U_i, $i \geq 0$, and a measurable function g such that

$$X_{i+1} = g(X_i, U_i)$$

- Thus, there is a measurable function G_i with

$$X_i = G_i(X_0; U_0, \ldots, U_{i-1})$$

- Extend $(U_i, i \geq 0)$ to $(U_i, i \in \mathbb{Z})$ and assume it is i.i.d.
- For $i, m \in \mathbb{Z}$ with $m \leq i$, let

$$X_{m:i} := G_{i-m}(X_0; U_m, U_{m+1}, \ldots, U_{i-1})$$

- $X_{0:i} = X_i$ for $i \geq 0$
- For $i, m \in \mathbb{Z}$ with $m \leq i$, we have $X_{m:i} \sim X_{i-m}$

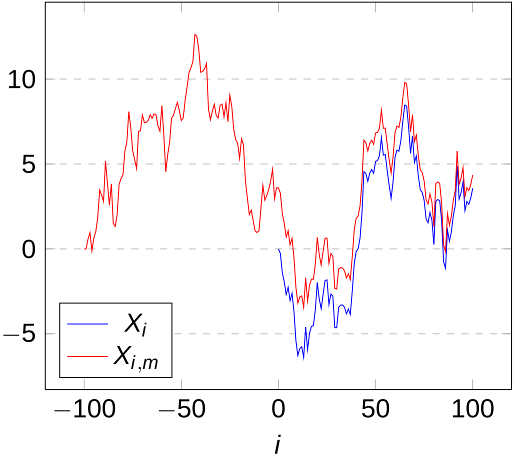
Example: Autoregressive sequence

- $X_0 = 0$
- For $i \geq 0$

$$X_{i+1} = \sqrt{\eta} X_i + U_i$$

- $\eta \in [0, 1)$
- $U_i, i \geq 0$, are i.i.d. with $E(U_i) = 0$ and $\mathrm{Variance}(U_i) = 1$

Autoregressive sequence, $\eta = 0.97, m = -100$



Main Assumption

- $(X_{m:i}, i \geq m)$ is a Markov chain that is a copy of $(X_i, i \geq 0)$, and is driven by $(U_i, i \geq m)$, for any given $m \in \mathbb{Z}$
- For $m \leq 0 \leq i$, the last i random variables driving the calculation of X_i and $X_{m:i}$, i.e., $U_0, \ldots, U_{i-1}$, are the same
- Assumption 1 (A1). There is a positive decreasing sequence $(\nu(i), i \geq 0)$ such that

$$\sum_{i=0}^{\infty} \sqrt{\frac{\nu(i)}{i+1}} < \infty,$$

and, for any $i, m \in \mathbb{Z}$ with $m \leq 0 \leq i$,

$$E((f(X_{m:i}) - f(X_i))^2) \leq \nu(i).$$

Main Assumption

- $(X_{m:i}, i \geq m)$ is a Markov chain that is a copy of $(X_i, i \geq 0)$, and is driven by $(U_i, i \geq m)$, for any given $m \in \mathbb{Z}$
- For $m \leq 0 \leq i$, the last i random variables driving the calculation of X_i and $X_{m:i}$, i.e., $U_0, \ldots, U_{i-1}$, are the same
- Assumption 1 (A1). There is a positive decreasing sequence $(\nu(i), i \geq 0)$ such that

$$\sum_{i=0}^{\infty} \sqrt{\frac{\nu(i)}{i+1}} < \infty,$$

and, for any $i, m \in \mathbb{Z}$ with $m \leq 0 \leq i$,

$$E((f(X_{m:i}) - f(X_i))^2) \leq \nu(i).$$

Autoregressive sequence Revisited

- When $f(x) = x$, Assumption A1 holds with

$$\nu(i) = \frac{\eta^i}{1-\eta}$$

for $i \geq 0$

Convergence properties of f_k

For $j \geq 0$, set

$$\overline{\nu}(j) := \sum_{i=j}^{\infty} \sqrt{\frac{\nu(i)}{i+1}}$$

Theorem

Under Assumption A1, $E(f(X_h))$ has a finite limit μ as $h \to \infty$. For $k > 0$,

$$|E(f_k) - \mu| \leq \frac{\overline{\nu}(\lfloor b(k)/2 \rfloor)}{\sqrt{k - b(k)}},$$

and

$$E((f_k - \mu)^2) \leq \frac{26(\overline{\nu}(0))^2}{k - b(k)}.$$

Assumptions

- Let $(Y_l, l \geq 0)$ with $E(Y_l^2) < \infty$
- Assume that $E(Y_l) \to \mu_Y$ as $l \to \infty$
- Intuitively, assume that $Y_l \approx Y_{l'}$ for $0 \leq l \leq l'$ and large l
- Let $(p_l, l \geq 0)$ be a probability distribution such that $p_l > 0$ for $l \geq 0$

- Let $N \in \mathbb{N}$ be a random variable independent of $(Y_l, l \geq 0)$ such that $\Pr(N = l) = p_l$ for $l \geq 0$

Theorem ((Rhee and Glynn 2015))

Set $Z := (Y_N - Y_{N-1})/p_N$, with $Y_{-1} := 0$. If

$$\sum_{l=0}^{\infty} \frac{E((Y_l - Y_{l-1})^2)}{p_l} < \infty$$

then $E(Z) = \mu_Y$, and

$$E(Z^2) = \sum_{l=0}^{\infty} \frac{E((Y_l - Y_{l-1})^2)}{p_l}.$$

Construction and properties of $f_{k,l}$

- For $k \geq 1$ and $l \geq 0$, let

$$f_{k,l} := \frac{1}{k - b(k)} \sum_{i=b(k)}^{k-1} f(X_{-k(2^l-1):i})$$

- In particular, $f_{k,0} = f_k$
- As $X_{-k(2^l-1):i} \sim X_{i+k(2^l-1)}$, Theorem 1 implies that

$$\lim_{l\to\infty} E(f_{k,l}) = \mu$$

- For $0 \leq l < l'$, the last $i + k(2^l - 1)$ copies of U_0 used to calculate $X_{-k(2^l-1):i}$ and $X_{-k(2^{l'}-1):i}$ are the same
- Intuitively speaking, $f_{k,l'} \approx f_{k,l}$ for large l

Construction and properties of $f_{k,l}$

- For $k \geq 1$ and $l \geq 0$, let

$$f_{k,l} := \frac{1}{k - b(k)} \sum_{i=b(k)}^{k-1} f(X_{-k(2^l-1):i})$$

- In particular, $f_{k,0} = f_k$
- As $X_{-k(2^l-1):i} \sim X_{i+k(2^l-1)}$, Theorem 1 implies that

$$\lim_{l\to\infty} E(f_{k,l}) = \mu$$

- For $0 \leq l < l'$, the last $i + k(2^l - 1)$ copies of U_0 used to calculate $X_{-k(2^l-1):i}$ and $X_{-k(2^{l'}-1):i}$ are the same
- Intuitively speaking, $f_{k,l'} \approx f_{k,l}$ for large l

Construction of Z_k

- For $l \geq 2$, set

$$p_l = \frac{\overline{\nu}(k2^{l-2}) - \overline{\nu}(k2^{l-1})}{2^l \overline{\nu}(k)},$$

- Let

$$p_1 = (1 - \sum_{l=2}^{\infty} p_l)/3 \text{ and } p_0 = 2p_1$$

- Let $N \in \mathbb{N}$ is a random variable independent of $(U_i, i \in \mathbb{Z})$ with $\Pr(N = l) = p_l$ for $l \geq 0$
- For $k \geq 1$, let

$$Z_k := \frac{f_{k,N+1} - f_{k,N}}{p_N}$$

Construction of Z_k

- For $l \geq 2$, set
$$p_l = \frac{\overline{\nu}(k2^{l-2}) - \overline{\nu}(k2^{l-1})}{2^l \overline{\nu}(k)},$$
- Let
$$p_1 = (1 - \sum_{l=2}^{\infty} p_l)/3 \text{ and } p_0 = 2p_1$$
- Let $N \in \mathbb{N}$ is a random variable independent of $(U_i, i \in \mathbb{Z})$ with $\Pr(N = l) = p_l$ for $l \geq 0$
- For $k \geq 1$, let
$$Z_k := \frac{f_{k,N+1} - f_{k,N}}{p_N}$$

Properties of Z_k

- Let T_k be the expected time required to simulate Z_k

Lemma

Suppose that A1 holds. For $k \geq 1$, we have $E(f_k + Z_k) = \mu$, $T_k \leq 9k$, and

$$kE(Z_k^2) \leq 20(\overline{\nu}(\lfloor b(k)/2 \rfloor))^2$$

Construction and properties of $\hat{f}_k$

- Let $Q \sim$ Bernoulli(q) independent of (f_k, Z'_k) with

$$q = \frac{\overline{\nu}(\lfloor b(k)/2 \rfloor)}{\overline{\nu}(0)}$$

- Let Z'_k be a copy of Z_k independent of f_k
- Set

$$\hat{f}_k := f_k + q^{-1} Q Z'_k$$

Theorem

Suppose A1 holds and $k \geq 1$. Then $E(\hat{f}_k^2) < \infty$ and $E(\hat{f}_k) = \mu$. Furthermore $\hat{T}_k \leq k + 9qk$, and

$$\hat{T}_k \mathrm{Variance}(\hat{f}_k) \leq k \mathrm{Variance}(f_k) + 8610(\overline{\nu}(0))^2 q$$

Construction and properties of $\hat{f}_k$

- Let $Q \sim \text{Bernoulli}(q)$ independent of (f_k, Z'_k) with

$$q = \frac{\overline{\nu}(\lfloor b(k)/2 \rfloor)}{\overline{\nu}(0)}$$

- Let Z'_k be a copy of Z_k independent of f_k
- Set

$$\hat{f}_k := f_k + q^{-1} Q Z'_k$$

Theorem

Suppose A1 holds and $k \geq 1$. Then $E(\hat{f}_k^2) < \infty$ and $E(\hat{f}_k) = \mu$. Furthermore $\hat{T}_k \leq k + 9qk$, and

$$\hat{T}_k \mathrm{Variance}(\hat{f}_k) \leq k \mathrm{Variance}(f_k) + 8610(\overline{\nu}(0))^2 q$$

GARCH volatility model

- The daily volatility σ_i satisfies, for $i \geq 0$,

$$\sigma_{i+1}{}^2 = w + \alpha \sigma_i{}^2 U_i^2 + \beta \sigma_i{}^2$$

- w, α and β are positive constants with $\alpha + \beta < 1$
- $U_i \sim N(0, 1)$ are i.i.d.
- Given $\sigma_0 \geq 0$ and $z \in \mathbb{R}$, we want to estimate $\lim_{i\to\infty} \Pr(\sigma_i^2 > z)$
- $X_i = \sigma_i^2$
- $f(u) = \mathbf{1}\{u > z\}$ for $u \in \mathbb{R}$
- Assumption A1 holds with $\nu(i) = c(\alpha + \beta)^{i/2}$ for some constant c

GARCH volatility model

- The daily volatility σ_i satisfies, for $i \geq 0$,

$$\sigma_{i+1}{}^2 = w + \alpha\sigma_i{}^2 U_i^2 + \beta\sigma_i{}^2$$

- w, α and β are positive constants with $\alpha + \beta < 1$
- $U_i \sim N(0, 1)$ are i.i.d.
- Given $\sigma_0 \geq 0$ and $z \in \mathbb{R}$, we want to estimate $\lim_{i\to\infty} \Pr(\sigma_i^2 > z)$
- $X_i = \sigma_i^2$
- $f(u) = \mathbf{1}\{u > z\}$ for $u \in \mathbb{R}$
- Assumption A1 holds with $\nu(i) = c(\alpha + \beta)^{i/2}$ for some constant c

$GI/G/1$ queue

- For $n \geq 0$, let A_n, V_n and X_n be the arrival time, service time and waiting time of customer n
- The waiting times satisfy the Lindley recursion

$$X_{i+1} = \max(0, X_i + U_i),$$

where $U_i := V_i - A_{i+1} + A_i$
- We want to estimate $\lim_{i\to\infty} E(X_i)$
- f is the identity function
- Under suitable conditions, Assumption A1 holds with $\nu(i) = \gamma\eta^i$ for $i \geq 0$, where γ and $\eta < 1$ are constants

$GI/G/1$ queue

- For $n \geq 0$, let A_n, V_n and X_n be the arrival time, service time and waiting time of customer n
- The waiting times satisfy the Lindley recursion

$$X_{i+1} = \max(0, X_i + U_i),$$

  where $U_i := V_i - A_{i+1} + A_i$
- We want to estimate $\lim_{i \to \infty} E(X_i)$
- f is the identity function
- Under suitable conditions, Assumption A1 holds with $\nu(i) = \gamma\eta^i$ for $i \geq 0$, where γ and $\eta < 1$ are constants

High-dimensional Gaussian vectors

- Let Σ be a $d \times d$ positive definite matrix with diagonal entries equal to 1
- (Kahalé 2019) approximately simulates $X \sim N(0, \Sigma)$
- Let j be a random integer uniformly distributed in $\{1, \ldots, d\}$
- Let $e \in \mathbb{R}^d$ be the vector whose j-th coordinate is 1 and remaining coordinates are 0
- Let $(e_i, i \geq 0)$ be i.i.d. copies of e
- Let $(g_i, i \geq 0)$ be i.i.d. with $g_i \sim N(0, 1)$
- Set $X_0 = 0$ and, for $i \geq 0$, let

$$X_{i+1} = X_i + (g_i - e_i^T X_i)(\Sigma e_i)$$

- For suitable functions f, Assumption A1 holds
- $\mu = E(f(X))$, with $X \sim N(0, \Sigma)$

High-dimensional Gaussian vectors

- Let Σ be a $d \times d$ positive definite matrix with diagonal entries equal to 1
- (Kahalé 2019) approximately simulates $X \sim N(0, \Sigma)$
- Let j be a random integer uniformly distributed in $\{1, \ldots, d\}$
- Let $e \in \mathbb{R}^d$ be the vector whose j-th coordinate is 1 and remaining coordinates are 0
- Let $(e_i, i \geq 0)$ be i.i.d. copies of e
- Let $(g_i, i \geq 0)$ be i.i.d. with $g_i \sim N(0, 1)$
- Set $X_0 = 0$ and, for $i \geq 0$, let

$$X_{i+1} = X_i + (g_i - e_i^T X_i)(\Sigma e_i)$$

- For suitable functions f, Assumption A1 holds
- $\mu = E(f(X))$, with $X \sim N(0, \Sigma)$

High-dimensional Gaussian vectors

- Let Σ be a $d \times d$ positive definite matrix with diagonal entries equal to 1
- (Kahalé 2019) approximately simulates $X \sim N(0, \Sigma)$
- Let j be a random integer uniformly distributed in $\{1, \ldots, d\}$
- Let $e \in \mathbb{R}^d$ be the vector whose j-th coordinate is 1 and remaining coordinates are 0
- Let $(e_i, i \geq 0)$ be i.i.d. copies of e
- Let $(g_i, i \geq 0)$ be i.i.d. with $g_i \sim N(0, 1)$
- Set $X_0 = 0$ and, for $i \geq 0$, let

$$X_{i+1} = X_i + (g_i - e_i^T X_i)(\Sigma e_i)$$

- For suitable functions f, Assumption A1 holds
- $\mu = E(f(X))$, with $X \sim N(0, \Sigma)$

Bias versus standard deviation

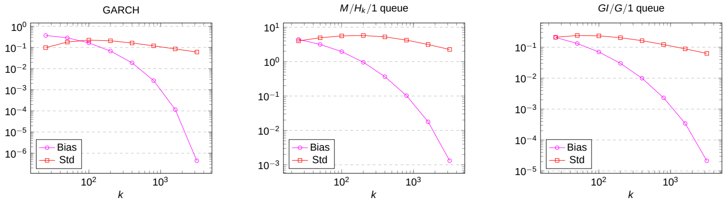


Figure: Absolute bias and standard deviation of time-average estimators with 10^6 independent replications and burn-in period $b(k) = \lfloor k/10 \rfloor$.

Performance of $\hat{f}_k$ in GARCH volatility model

Table: Estimation of $\lim_{i\to\infty} \Pr(\sigma_i^2 > z)$ in a GARCH volatility model with $b(k) = \lfloor k/10 \rfloor$, $\alpha = 0.05$, $\beta = 0.92$, ${\sigma_0}^2 = 2 \times 10^{-5}$, $w = 1.2 \times 10^{-6}$ and $z = 4 \times 10^{-5}$.

k	burn-in	Method	μ	Std	RMSE	Cost	Cost $\times$ MSE
50	5	LR	0.1126	1.8×10^{-1}	3.4×10^{-1}	5.00×10^{1}	5.8
		ULR	0.398 ± 0.003	1.4×10^{0}	1.4×10^{0}	1.02×10^{2}	200
		SULR	0.400 ± 0.002	1.2×10^{-3}	1.2×10^{-3}	1.01×10^{8}	140
200	20	LR	0.3319	2.1×10^{-1}	2.2×10^{-1}	2.00×10^{2}	9.5
		ULR	0.3991 ± 0.0008	4.2×10^{-1}	4.2×10^{-1}	4.03×10^{2}	71
		SULR	0.3993 ± 0.0007	3.8×10^{-4}	3.8×10^{-4}	3.96×10^{8}	58
800	80	LR	0.3969	1.2×10^{-1}	1.2×10^{-1}	8.00×10^{2}	11
		ULR	0.3996 ± 0.0002	1.2×10^{-1}	1.2×10^{-1}	1.59×10^{3}	23
		SULR	0.3996 ± 0.0002	1.2×10^{-4}	1.2×10^{-4}	1.59×10^{9}	23
3200	320	LR	0.39963	6.1×10^{-2}	6.1×10^{-2}	3.20×10^{3}	12
		ULR	0.39970 ± 0.0001	6.1×10^{-2}	6.1×10^{-2}	6.34×10^{3}	23
		SULR	0.39963 ± 0.0001	6.1×10^{-5}	6.1×10^{-5}	6.28×10^{9}	23

Performance of $\hat{f}_k$ in a $M/H_k/1$ queue

Table: Estimation of $E(X_i)$ in an $M/H_k/1$ queue with $b(k) = \lfloor k/10 \rfloor$, Poisson arrivals at rate $\lambda = 0.75$, service time distribution $\Pr(V_n \geq z) = pe^{-2pz} + (1-p)e^{-2(1-p)z}$ for $z \geq 0$, where $p = 0.8875$

k	burn-in	Method	μ	Std	RMSE	Cost	Cost $\times$ MSE
50	5	LR	4.35	4.9×10^0	5.9×10^0	5.00×10^1	1.7×10^3
		ULR	7.53 ± 0.07	3.8×10^1	3.8×10^1	1.00×10^2	1.5×10^5
		SULR	7.53 ± 0.07	3.8×10^{-2}	3.8×10^{-2}	9.87×10^7	1.4×10^5
200	20	LR	6.547	5.8×10^0	5.8×10^0	2.00×10^2	6.8×10^3
		ULR	7.51 ± 0.03	1.7×10^1	1.7×10^1	4.00×10^2	1.1×10^5
		SULR	7.50 ± 0.03	1.6×10^{-2}	1.6×10^{-2}	4.00×10^8	1.1×10^5
800	80	LR	7.412	4.2×10^0	4.2×10^0	8.00×10^2	1.4×10^4
		ULR	7.503 ± 0.01	5.3×10^0	5.3×10^0	1.59×10^3	4.4×10^4
		SULR	7.516 ± 0.01	5.4×10^{-3}	5.4×10^{-3}	1.60×10^9	4.7×10^4
3200	320	LR	7.512	2.3×10^0	2.3×10^0	3.20×10^3	1.6×10^4
		ULR	7.511 ± 0.004	2.3×10^0	2.3×10^0	6.29×10^3	3.2×10^4
		SULR	7.513 ± 0.004	2.3×10^{-3}	2.3×10^{-3}	6.70×10^9	3.4×10^4

Performance of $\hat{f}_k$ in a $GI/G/1$ queue

Table: Estimation of $\lim_{i\to\infty} \Pr(X_i > 1)$ in a $GI/G/1$ queue with $b(k) = \lfloor k/10 \rfloor$, interarrival time D_n and service time V_n distributions $\Pr(D_n \geq z) = (1+z)^{-7}$ and $\Pr(V_n \geq z) = (1+z/0.8)^{-7}$ for $z \geq 0$

k	burn-in	Method	μ	Std	RMSE	Cost	Cost $\times$ MSE
50	5	LR	0.2004	2.4×10^{-1}	2.7×10^{-1}	5.00×10^{1}	3.8
		ULR	0.33 ± 0.002	1.2×10^{0}	1.2×10^{0}	1.00×10^{2}	135
		SULR	0.334 ± 0.002	1.1×10^{-3}	1.1×10^{-3}	1.00×10^{8}	131
200	20	LR	0.302	2.0×10^{-1}	2.1×10^{-1}	2.00×10^{2}	8.4
		ULR	0.3327 ± 0.0008	4.0×10^{-1}	4.0×10^{-1}	3.93×10^{2}	64
		SULR	0.3323 ± 0.0008	4.0×10^{-4}	4.0×10^{-4}	3.97×10^{8}	63
800	80	LR	0.3298	1.2×10^{-1}	1.2×10^{-1}	8.00×10^{2}	12
		ULR	0.3321 ± 0.0003	1.3×10^{-1}	1.3×10^{-1}	1.64×10^{3}	29
		SULR	0.3321 ± 0.0003	1.3×10^{-4}	1.3×10^{-4}	1.63×10^{9}	28
3200	320	LR	0.33222	6.3×10^{-2}	6.3×10^{-2}	3.20×10^{3}	13
		ULR	0.33221 ± 0.0001	6.3×10^{-2}	6.3×10^{-2}	6.35×10^{3}	25
		SULR	0.33224 ± 0.0001	6.3×10^{-5}	6.3×10^{-5}	6.33×10^{9}	25

Conclusion (I)

- Under a coupling assumption, we have established bounds on the bias and mean square error of f_k
- We have built an unbiased estimator Z_k for the bias of f_k
- Combining Z_k with f_k yields an unbiased estimator $\hat{f}_k$ of μ
- Z_k and $\hat{f}_k$ are square-integrable with have finite expected running time

Conclusion (II)

- For a suitable choice of parameters, $\hat{f}_k$ is asymptotically at least as efficient as f_k
- Under certain conditions, Z_k and $\hat{f}_k$ can be built without any precomputations
- In several examples our approach is provably efficient
- Numerical experiments are consistent with theoretical findings

Asmussen, S. and Glynn, P. W. (2007). *Stochastic simulation: algorithms and analysis*, Vol. 57, Springer Science & Business Media.

Barkhagen, M., Chau, N. H., Moulines, É., Rásonyi, M., Sabanis, S. and Zhang, Y. (2021). On stochastic gradient Langevin dynamics with dependent data streams in the logconcave case, *Bernoulli* **27**(1): 1–33.

Cousins, B. and Vempala, S. (2016). A practical volume algorithm, *Mathematical Programming Computation* **8**(2): 133–160.

Diaconis, P. (2009). The Markov chain Monte Carlo revolution, *Bulletin of the American Mathematical Society* **46**(2): 179–205.

Diaconis, P. and Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains, *The annals of applied probability* pp. 36–61.

Glynn, P. W. and Rhee, C.-h. (2014). Exact estimation for Markov chain equilibrium expectations, *Journal of Applied Probability* **51**(A): 377–389.

Jacob, P. E., OLeary, J. and Atchadé, Y. F. (2020). Unbiased Markov chain Monte Carlo methods with couplings, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**(3): 543–600.

Kahalé, N. (2019). Efficient simulation of high dimensional Gaussian vectors, *Mathematics of Operations Research* **44**(1): 58–73.

Rhee, C.-h. and Glynn, P. W. (2015). Unbiased estimation with square root convergence for SDE models, *Operations Research* **63**(5): 1026–1043.

Sinclair, A. (1992). Improved bounds for mixing rates of Markov chains and multicommodity flow, *Combinatorics, probability and Computing* **1**(04): 351–370.

Tierney, L. (1994). Markov chains for exploring posterior distributions, *the Annals of Statistics* pp. 1701–1728.

Whitt, W. (1991). The efficiency of one long run versus independent replications in steady-state simulation, *Management Science* **37**(6): 645–666.

Whitt, W. and You, W. (2019). Time-varying robust queueing, *Operations Research* **67**(6): 1766–1782.