Discounting in Markov Chain Estimation

Ethan Che and Jing Dong

Abstract

Discounting can be viewed as a perturbation to improve the ergodicity of the Markov chain by imposing more regular regenerations. It can improve the estimation efficiency in Markov chain estimation tasks. On the other hand, the perturbation can also lead to estimation bias, which imposes an efficiency-accuracy tradeoff. In this paper, we apply the Wasserstein ergodicity framework to investigate the efficiency-accuracy tradeoff for discounting in two important estimation tasks: steady-state estimation and estimating the solution to the Poisson equation. Our results quantifies the overall benefit of discounting and provide guidance on choosing the appropriate discount factors in these estimation tasks.

1 Introduction

Consider a discrete-time Markov chain with transition kernel P and a properly defined stationary measure π. For some measurable function f, we consider the following estimation tasks:
1. Estimating the steady-state performance metric πf.
2. Estimating the relative value function $h(x)$, i.e. a solution to the Poisson equation

$$(P - I)h = -f + \pi f \tag{1}$$

These estimation tasks arise in many applications such as Markov chain Monte Carlo (MCMC) in Bayesian statistics [Smith and Roberts, 1993], performance evaluation in queueing models [Asmussen and Glynn, 2007], and policy evaluation in reinforcement learning with long-run average cost criteria [Meyn, 1997].

In most practical settings, we cannot compute these quantities explicitly or draw samples from π directly. Instead, we can simulate the Markov chain following the transition kernel P and use sampling based estimators. An important consideration when using such estimators is the ergodicity of the Markov chain. If the Markov chain converges to stationarity very slowly, the corresponding estimator can have a very large variance and poor estimation efficiency.

To illustrate, consider the sample average estimate of the steady-state mean πf. Under proper ergodicity conditions, the sample average satisfies a central limit theorem (CLT) with an asymptotic variance $\sigma^2(f)$,which takes the form (Propositon 4.2.2 in Asmussen and Glynn [2007])

$$\sigma^2(f) = \mathrm{var}_\pi(f(X_0)) + 2\sum_{k=1}^{\infty} \mathrm{cov}_\pi(f(X_0), f(X_k)) \tag{2}$$

Note that the asymptotic variance is determined not only by the variance of f under the stationary measure, i.e., $\mathrm{var}_\pi(f(X_0))$, but also by the sum of autocovariances, i.e., $\sum_{k=1}^{\infty} \mathrm{cov}_\pi(f(X_0), f(X_k))$. If correlations decay very slowly, the latter term can be exceedingly large.

Next, consider the relative value function in Markov decision processes (MDP), which can be obtained as a solution to the Poisson equation (1). Under suitable ergodicity conditions, the Poisson equation admits a solution of the form (Proposition A.3.1 in Meyn [2011])

$$h^{x^*}(x) = \mathbb{E}\left[\sum_{t=0}^{\tau-1}(f(X_t) - \pi f)\middle| X_0 = x\right],$$

where τ is the first hitting time of a regeneration state x^*, When using simulation to estimate h^{x^*}, if the Markov chain mixes slowly, the trajectories of the Markov chain can exhibit large fluctuations, in which case the corresponding sums can vary across different magnitude over different sample paths. In many applications, the large number of samples required to control the variance of these estimators is the main computational bottleneck.

When the Markov chain lacks sufficient ergodicity, a potential modification is to apply discounting. The idea of discounting has been widely used in the MDP literature [Kakade, 2003]. In many applications, the discounted cost is a natural objective. However, there are also numerous examples where discounting is imposed mainly for analytical and computational tractability. We consider a "discounted" chain with transition kernel

$$P_{\gamma,\nu}(x,\cdot) = \gamma P(x,\cdot) + (1-\gamma)\nu(\cdot) \tag{3}$$

where $\gamma \in (0,1)$ is the discount factor and ν is the initial distribution.

According to the transition kernel $P_{\gamma,\nu}$, at each step, with probability γ, the modified chain follows the transition kernel of the original chain, P; with probability $(1-\gamma)$, the chain regenerates and draws a sample from the initial distribution ν. Intuitively, the random regeneration "breaks" the autocorrelation. This can ensure faster convergence to stationarity in steady-state estimation and more regular trajectories when estimating the Poisson equation solution. However, the stationary measure and the Poisson solutions associated with discounted chain may be different from the original Markov chain, which introduces estimation bias. The different effects of discounting on variance and bias pose an interesting tradeoff between efficiency and accuracy. In this paper, we study this efficiency-accuracy tradeoff in Markov chain estimation from a statistical perspective.

We apply the Wasserstein ergodicity framework, which allows us to quantify the estimation bias and variance. Based on these quantifications, we further characterize the overall benefit of discounting and how to choose the appropriate discount factor to balance the efficiency-accuracy tradeoff. In particular, we characterize how the optimal discount factor scales with the sampling budget and the ergodicity of the Markov chain.

1.1 Related literature

Using the discounted cost as an approximation for the long-run average cost has a long history in the MDP literature. It is well known that for a given stationary policy, the normalized discounted cost and the long-run average cost are equivalent in the limit as the discount factor γ approaches 1 [Blackwell, 1962]. Note that the expected normalized cumulative discounted cost can be viewed as the long-run average cost of a modified Markov chain as defined in (3) [Kakade, 2003]. Many papers use discounting to improve the efficiency of simulation-based algorithms in policy evaluation related tasks (see, for example, Jaakkola et al. [1994], Baxter and Bartlett [2001], Marbach and Tsitsiklis [2001]). In those settings, to balance the bias-variance tradeoff, it is argued that the choice of γ should depend on the mixing time of the underlying Markov process. Our analysis confirms this insight and provides further statistical justifications for it. The most relevant papers to ours are those that establish upper bounds for bias or variance of the discounted estimators [Jaakkola et al., 1994, Petrik and Scherrer, 2008, Jiang et al., 2016, Dai and Gluzman, 2021]. We extend existing results by developing new upper bounds for both the bias and the variance, which allows us to better quantify the tradeoff and provide guidance on the choice of the discount factor.

There are some works that quantify the optimality gap of algorithms learned under the discounted cost criteria [Kakade, 2001, Thomas, 2014]. There are also works on Blackwell optimality, which seek to obtain policies that are optimal for all discount factors γ above some cutoff γ^* [Schneckenreither, 2020, Perotto and Vercouter, 2018]. The results are very relevant for MDPs but is beyond the scope of this paper.

Analyzing the convergence rate and approximation accuracy are fundamental problems in MCMC [Roberts and Tweedie, 1996, Hairer et al., 2014]. There are many works that study how perturbation due to numerical errors affect the convergence rate and approximation accuracy of MCMC algorithms (see, for example, Hervé and Ledoux [2014], Dwivedi et al. [2019], Rudolf and Schweizer [2018]). These works often use the ergodicity framework to quantify the convergence. We apply a similar ergodicity framework, which allows us to develop finite-sample performance bounds [Joulin and Ollivier, 2010]. Recently, Wang et al. [2021] propose a regeneration-enriched Markov chain for steady-state estimation. There, the original Markov chain dynamics are augmented with random regenerations occurring according to a time-inhomogeneous Poisson process with a state-dependent rate and the rate function is carefully calibrated to maintain the same stationary measure. These calibrations require knowledge of the stationary distribution up to a normalizing constant. The discounting scheme considered in this paper uses a fixed regeneration rate. The choice of a good discount rate in this case requires much less knowledge of the target stationary measure.

Lastly, our work is related to the rich literature on steady-state simulation (see Ni and Henderson [2015] and Chapter IV in Asmussen and Glynn [2007] for a comprehensive review). Using regeneration for steady-state estimation is one of the key strategies in simulation output analysis [Glynn and Iglehart, 1993]. To improve the estimation efficiency, various variance reduction techniques has been proposed in the literature, including control variates [Yang and Nelson, 1992], approximating martingales [Henderson and Glynn, 2002], importance sampling [Blanchet and Lam, 2012], etc. Discounting can also be viewed as a variance reduction technique, but it introduces substantial extra bias. To the best of our knowledge, how to strike a balance between the variance and the bias in this case has not been well studied in the literature and is one of the main contributions of this paper.

1.2 Notations

The following notations are used throughout the paper. Let $\mathcal{X}$ denote a Polish space and $\mathcal{B}(\mathcal{X})$ denote the corresponding Borel σ-algebra. Let $\mathcal{P}$ denote the set of all Borel probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. For $\mu \in \mathcal{P}$ and a measurable function $f : \mathcal{X} \to \mathbb{R}$, define $\mu f = \mathbb{E}_\mu[f(X)] = \int_\mathcal{X} f(x)\mu(dx)$. If $\mu f^2 < \infty$, we also define $\text{var}_\mu f = \text{var}_\mu(f(X)) = \int_\mathcal{X} (f(x) - \mu f)^2 \mu(dx)$.

Let P be a transition kernel on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. In particular, $P : \mathcal{P} \to \mathcal{P}$, is defined as

$$\mu P(A) = \int_\mathcal{X} P(x, A)\mu(dx), \text{ for any } \mu \in \mathcal{P} \text{ and } A \in \mathcal{B}(\mathcal{X}).$$

Let δ_x denote a delta measure on x. Then, $\delta_x P(A) = P(x, A)$. For a measurable function f, $Pf(x) = \int_\mathcal{X} f(y)P(x, dy)$. We write $(X_n)_{n\in\mathbb{Z}}$ as the Markov chain with transition kernel P. If P has a unique stationary distribution, we denote it as π. Denote

$$\mathbb{E}_\mu[f(X_k)] = \mu P^k f \text{ and } \text{var}_\mu(f(X_k)) = \int_\mathcal{X} (f(x) - \mu P^k f)^2 \mu P^k(dx)$$

For simplicity, we also denote $\mathbb{E}_x[f(X_k)] = \mathbb{E}[f(X_k)|X_0 = x] = P^k f(x)$ and $\text{var}_x(f(X_k)) = \text{var}(f(X_k)|X_0 = x)$. We distinguish between two important variances: the steady-state variance $\text{var}_\pi f$ and the asymptotic variance $\sigma(f)$ (see (2)). Note that under suitable regularity conditions,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n f(X_i) - \pi f\right) \Rightarrow N(0, \sigma^2(f)) \text{ as } n \to \infty.$$

Let $P_{\gamma,\nu}$ (see (3)) and $\pi_{\gamma,\nu}$ denote the transition kernel and stationary measure of the discounted modification of $(X_n)_{n\in\mathbb{Z}}$, which is also referred to as the discounted chain and we write it as $(X_n^{\gamma,\nu})_{n\in\mathbb{Z}}$. To keep the notations more concise, we sometimes suppress the dependence on the initial distribution ν when it is clear from the context.

Lastly, given two sequences of nonnegative real numbers $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$, we define $b_n = O(a_n)$ and $b_n = \Omega(a_n)$ if there exist some constants $C, C' > 0$ such that $b_n \leq Ca_n$ and $C'a_n \leq b_n \leq Ca_n$ respectively. We define $b_n = o(a_n)$ and $b_n \sim a_n$ if $\lim_{n\to\infty} b_n/a_n \to 0$ and $\lim_{n\to\infty} b_n/a_n = C$ for some $C \in (0, \infty)$ respectively.

2 Preliminaries

In this section, we introduce some basic properties of the discounted chain and our analysis framework – Wasserstein ergodicity.

2.1 Properties of the discounted Markov chain

A key property of the discounted chain is that it is a regenerative process. Regenerations occur when the state is drawn from ν instead of following the transition kernel P. In this case, the excursion length follows a Geometric distribution with success probability $1-\gamma$. Due to the regeneration structure, the discounted chain is uniformly ergodic, even if the original Markov chain is not.

Proposition 1. *For an irreducible Markov chain with transition kernel P, the discounted chain with transition kernel $P_{\gamma,\nu}$ is uniformly ergodic. That is, there exists $\rho \in (0,1)$ and $C > 0$ such that*

$$\sup_{x\in\mathcal{X}} \|\delta_x P^n_{\gamma,\nu}, \pi_\gamma\|_{tv} \leq C\rho^n$$

where $||\cdot||_{tv}$ is total variation distance.

This is a strong regularity property, which guarantees that the sample average of the discounted chain satisfies a law of large numbers, a CLT, and a concentration inequality for bounded functions (see Appendix B.1). Although the discounted chain has this appealing property, if the original Markov chain does not have a well-defined stationary distribution or does not satisfy a CLT, then it is unclear how to make a proper comparison between the two. Thus, in what follows, we assume that the original Markov chain is suitably ergodic as quantified in the next subsection.

2.2 Wasserstein ergodicity

In order to quantify the magnitude of the costs and benefits of estimators based on the discounted chain, we consider a class of Markov chains that satisfy Wasserstein ergodicity.

Let d be a metric which is assumed to be lower semi-continuous with respect to the product topology of $\mathcal{X}$. For a measurable function $f : \mathcal{X} \to \mathbb{R}$, let $\|f\|_{\text{Lip}} = \sup_{x,y\in\mathcal{X},x\neq y} |f(x)-f(y)|/d(x,y)$ denote its Lipschitz constant with respect to d. We also denote $\text{Lip}(\mathcal{X}, d)$ as the set of 1-Lipschitz functions on $(\mathcal{X}, d)$. For any $\mu, \nu \in \mathcal{P}$, the Wasserstein-1 distance is defined as

$$W(\mu,\nu) := \inf_{M\in\mathcal{C}(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d(x,y)M(dx,dy) = \sup_{f\in\text{Lip}(\mathcal{X},d)} |\mu f - \nu f|$$

where $\mathcal{C}(\mu,\nu)$ is the set of all couplings of μ and ν, i.e., all probability measures with marginals ν and μ. We restrict our analysis to Markov chains that satisfy the Wasserstein ergodicity condition as defined in Assumption 1. This notion of ergodicity is used in [Rudolf and Schweizer, 2018] and is similar to Wasserstein contraction used in [Joulin and Ollivier, 2010].

Assumption 1 (Wasserstein ergodicity)**.** *A Markov chain with transition kernel P satisfies the Wasserstein ergodicity condition if there exist constants $C \in [1,\infty)$ and $\kappa \in [0,1)$, such that*

$$\sup_{x\neq y} \frac{W(\delta_x P^n, \delta_y P^n)}{d(x,y)} < C\kappa^n \text{ for all } n \in \mathbb{N}.$$

As we will show in Proposition 2, for a transition kernel P with stationary distribution π, if it satisfies Assumption 1, then for any $\mu \in \mathcal{P}$, $W(\mu P^n, \pi) \leq C\kappa^n W(\mu,\pi)$. Thus, κ measures how fast the Markov chain converges to stationarity. Smaller κ implies faster convergence. In what follows, we shall refer to κ as the **ergodicity constant** of the Markov chain.

Wasserstein ergodicity is closely related to other definitions of ergodicity. When $d(x,y) = 2 \times \mathbf{1}_{x\neq y}$, where $\mathbf{1}$ denote an indicator function, $W(\mu,\nu) = \|\mu-\nu\|_{\text{tv}}$ and Wasserstein ergodicity coincides with uniform ergodicity. More generally, when $d(x,y) = (V(x)+V(y))\mathbf{1}_{x\neq y}$ for a measurable function $V:\mathcal{X}\to[1,\infty]$, Wasserstein ergodicity coincides with V-uniform ergodicity. In particular, Lemma 3.1 in [Rudolf and Schweizer, 2018] shows that

$$||\mu-\nu||_V \equiv \sup_{|f|\leq V}\left|\int_{\mathcal{X}} f(y)(\mu(dy)-\nu(dy))\right| = W(\mu,\nu)$$

Furthermore, if P is ϕ-irreducible and aperiodic, then geometric ergodicity is equivalent to V-uniform ergodicity [Roberts and Rosenthal, 1997]. This implies that geometrically ergodic chain is almost equivalent to Wasserstein ergodic with a proper defined metric.

Another framework to characterize the convergence rate to stationarity is the spectral gap. The spectral gap measures the convergence rate of μP^n to π in χ^2-distance. For reversible Markov chains, under suitable regularity conditions, exponential ergodicity leads to the existence of a spectral gap (see, e.g., Proposition 2.8 in Hairer et al. [2014]). On the other hand, convergence under the χ^2-distance often leads to convergence under the total variation distance.

The next proposition summarizes basic properties of Markov chains under Wasserstein ergodicity. Define

$$\mathrm{V}_\pi := \sup_{f\in\mathrm{Lip}(\mathcal{X},d)} \mathrm{var}_\pi f \quad \text{and} \quad \mathrm{V}_P(x) := \sup_{f\in\mathrm{Lip}(\mathcal{X},d)} \mathrm{var}_{\delta_x P} f \text{ for } x\in\mathcal{X}.$$

Proposition 2. *For a transition kernel P with stationary distribution π, if Assumption 1 holds, P satisfies the following properties.*

(1) For any measurable and Lipchitz continuous function f,

$$|P^n f(x) - P^n f(y)| \leq C\|f\|_{Lip}\kappa^n d(x,y),$$

i.e., $P^n f(x)$ is $C\|f\|_{Lip}\kappa^n$-Lipschitz.

(2) For any measures μ,ν, $W(\mu P^n,\nu P^n)\leq C\kappa^n W(\mu,\nu)$. This implies that $W_1(\mu P^n,\pi)\leq C\kappa^n W(\mu,\pi)$.

(3) For any measurable and Lipchitz continuous function f with $\pi f^2<\infty$, the steady-state variance satisfies

$$var_\pi f \leq C^2\|f\|_{Lip}^2 \frac{1}{1-\kappa^2}\pi V_P.$$

The asymptotic variance satisfies

$$\sigma^2(f) \leq C\|f\|^2_{Lip}\frac{1}{1-\kappa}V_\pi \leq C^3\|f\|^2_{Lip}\left(\frac{1}{1-\kappa}\right)\left(\frac{1}{1-\kappa^2}\right)\pi V_P.$$

Item (3) in Proposition 2 quantifies how the ergodicity constant affect both the steady-state variance and the asymptotic variance. Even though the Wasserstein ergodicity constant κ will affect the value of πV_P, it is in general bounded, i.e., it does not blow up as κ approaches 1. Thus,

$$\mathrm{var}_\pi(f) = O\left(\frac{1}{1-\kappa}\right) \text{ and } \sigma^2(f) = O\left(\frac{1}{(1-\kappa)^2}\right).$$

These bounds are often sharp as we demonstrate through the following three examples.

Example 2.1. *Consider the linear autoregressive process of order one, AR(1),*

$$X_{n+1} = \phi X_n + \epsilon_{n+1}$$

where $\phi\in(0,1)$ and ϵ_n's are independent $N(0,\sigma^2)$. The stationary distribution of the Markov chain is $N(0,\sigma^2/(1-\phi^2))$. Under the metric $d(x,y) = (V(x)+V(y))\mathbf{1}_{x\neq y}$ with $V(x) = (1-\phi^2)x^2+1$, the chain *satisfies Assumption 1 with* $\kappa = \phi^2$*. Note that* κ *approaches 1 as* ϕ *approaches* 1*. Moreover,* $\pi(V_P) \le 21\sigma^4 + 2$*. For* $f(x) = x$*, we have [Spitzner and Boucher, 2007]*

$$var_\pi(f) = \frac{\sigma^2}{1-\phi^2} \sim \frac{1}{1-\kappa} \text{ and } \sigma^2(f) = \frac{\sigma^2}{1-\phi^2}\frac{1+\phi}{1-\phi} \sim \frac{1}{(1-\kappa)^2}.$$

Example 2.2. *Consider a discretized* $M/M/1$ *queue with transition probability*

$$P(x, x+1) = \lambda \text{ and } P(x, (x-1)^+) = \mu, \quad x = 0, 1, \dots,$$

where $\lambda, \mu > 0$ *with* $\lambda + \mu = 1$*. Let* $\rho = \lambda/\mu$*. The stationary distribution of the Markov chain is Geometric*$(1-\rho)$*. Under the metric* $d(x,y) = \left|\sum_{k=1}^{x} (\mu/\lambda)^{k/3} - \sum_{k=1}^{y} (\mu/\lambda)^{k/3}\right|$ *with* $\sum_{k=1}^{0} (\mu/\lambda)^{k/3} \equiv 1$*, the chain satisfies Assumption 1 with* $\kappa = \mu(\rho^{2/3} + \rho^{1/3})$ *[Joulin, 2009]. Note that* κ *approaches 1 as* ρ *approaches 1. Moreover,* $\pi V_P \le 3$*. For* $f(x) = x$*, we have*

$$var_\pi(f) = \frac{\rho}{(1-\rho)^2} \sim \frac{1}{1-\kappa} \text{ and } \sigma^2(f) = \frac{2\rho(1+\rho)}{(1-\rho)^4} \sim \frac{1}{(1-\kappa)^2}.$$

Example 2.3. *Consider a Binomial Markov chain with* $N \in \mathbb{N}$ *and* $\lambda = aN$ *for some fixed* $a \in (0,1)$*. Then, the transition probability takes the form*

$$P(x, x+1) = \frac{\lambda}{N}\left(1 - \frac{x}{N}\right), \; x = 0, ..., N-1, \quad P(x, x-1) = \left(1 - \frac{\lambda}{N}\right)\left(\frac{x}{N}\right), \; x = 1, ..., N$$
$$P(x, x) = \left(\frac{\lambda}{N}\right)\left(\frac{x}{N}\right) + \left(1 - \frac{\lambda}{N}\right)\left(1 - \frac{x}{N}\right), \; x = 0, ..., N$$

The stationary distribution of the Markov chain is Binomial$(N, \lambda/N)$*. Under the Euclidean metric, the chain satisfies Assumption 1 with* $\kappa = 1 - 1/N$ *[Joulin and Ollivier, 2010]. Note that* κ *approaches 1 as* N *approaches* ∞*. Moreover,* $\pi V_P \le 4\lambda/N = 4a$*. For* $f(x) = x$*, we have*

$$var_\pi(f) = Na(1-a) \sim \frac{1}{1-\kappa} \text{ and } \sigma^2(f) = N(2N-1)a(1-a) \sim \frac{1}{(1-\kappa)^2}.$$

3 Steady-state estimation

In this section, we study the effect of discounting on the efficiency and accuracy in steady-state estimation.

3.1 Bias and Variance quantification

Our first result quantifies the effect of discounting on the ergodicity constant.

Lemma 1. *Under Assumption 1, for any* $n \in \mathbb{N}$ *and* $x, y \in \mathcal{X}$*,* $x \neq y$*,*

$$W(\delta_x P_\gamma^n, \delta_y P_\gamma^n) = \gamma^n W(\delta_x P^n, \delta_y P^n) \le C(\gamma\kappa)^n d(x,y).$$

Lemma 1 indicates that if P has an ergodicity constant of κ, the discounted chain P_γ will have an improved ergodicity constant of $\gamma\kappa$. We next study the implications the improved ergodicity on the bias and variance of the discounted estimator. Let $\sigma_\gamma^2(f)$ denote the asymptotic variance of the discounted chain. We also define

$$D(\mu, \nu) := \sup_{f \in \text{Lip}(\mathcal{X}, d)} (\mu f - \nu f)^2.$$

Theorem 1. *Suppose Assumption 1 holds and* f *is Lipschitz continuous. Assume* $W(\nu, \pi)$*,* $V_\nu < \infty$*, and* *for any* $\gamma \in [0,1]$, $\pi_\gamma V_P$, $\pi_\gamma D(\delta_x P, \nu) < \infty$. *For the discounted chain, the bias satisfies*

$$|\pi f - \pi_\gamma f| \le C\|f\|_{Lip} \frac{1-\gamma}{1-\gamma\kappa} W(\nu, \pi).$$

The asymptotic variance satisfies

$$\begin{aligned}\sigma_\gamma^2(f) &\le C\|f\|_{Lip}^2 \left(\frac{1}{1-\gamma\kappa}\right) V_{\pi_\gamma} \\ &\le C^3 ||f||_{Lip}^2 \frac{1}{(1-\gamma\kappa)^2(1+\gamma\kappa)} \left(\gamma\pi_\gamma V_P + (1-\gamma) V_\nu + \gamma(1-\gamma)\pi_\gamma D(\delta_x P, \nu)\right).\end{aligned}$$

We make several observations from Theorem 1. First, as a sanity check, in the extreme case where $\nu = \pi$ and $\gamma = 1$, we obtain $|\pi f - \pi_\gamma f| = 0$ and $\sigma_\gamma^2(f) = \mathrm{var}_\pi(f) \le C\|f\|_{\mathrm{Lip}}^2 V_\pi$, which satisfy the bounds in Theorem 1. Second, consider a fixed distribution ν with $W(\nu,\pi) > 0$ and $V_\nu < \infty$. As demonstrated in Examples 2.1 – 2.3, $\pi_\gamma V_P$ and $\pi_\gamma D(\delta_x P, \nu)$ are likely to be bounded, i.e., they do not blow up as κ approaches 1 or γ approaches 1. Then, we have

$$|\pi f - \pi_\gamma f| = O\left(\frac{1-\gamma}{1-\gamma\kappa}\right), \quad \sigma_\gamma^2(f) = O\left(\frac{1}{(1-\gamma\kappa)^2}\right).$$

Note that to achieve a small bias, we want γ to be large. On the other hand, to achieve a small variance, we want γ to be small. Third, for a fixed value of $\gamma \in (0,1)$, to achieve a small bias, we want to choose ν close to π, i.e., $W(\nu,\pi)$ is small. For the variance, the terms V_ν and $\pi_\gamma D(\delta_x P, \nu)$ suggest that a good choice of ν should be 1) highly concentrated, with tails that decay at least as fast as as π, and 2) not too far from the stationary distribution π.

3.2 Efficiency and accuracy tradeoff in steady-state estimation

In this subsection, we study the efficiency-accuracy tradeoff using the bounds characterized in Theorem 1. We assume ν is fixed with $W(\nu,\pi)$, V_ν, $\pi_\gamma \mathrm{V}_P$, and $\pi_\gamma D(\delta_x P, \nu)$ bounded. Define

$$M(\nu, P) = \sup_{\gamma\in[0,1]} \left\{\gamma\pi_\gamma \mathrm{V}_P + (1-\gamma)\mathrm{V}_\nu + \gamma(1-\gamma)\pi_\gamma D(\delta_x P, \nu)\right\}. \tag{4}$$

When using the sample average $\bar{f}_n = \frac{1}{n}\sum_{t=0}^{n-1} f(X_t^\gamma)$ to approximate πf, we want to find γ that minimizes the MSE of $\bar{f}_n$. Let $\mathrm{MSE}(\gamma)$ denote the the MSE when applying the discount factor γ and the chain is initialized from stationarity (i.e., we ignore the transient bias, which is in general of a smaller order). Then $\mathrm{MSE}(\gamma) = |\pi_\gamma f - \pi f|^2 + \sigma_\gamma^2(f)/n$. Since the bias and the asymptotic variance are not known explicitly, we approximate them using the upper bounds developed in Theorem 1. In particular, we consider the following approximated MSE:

$$\widehat{\mathrm{MSE}}(\gamma) := \left(\frac{1-\gamma}{1-\gamma\kappa} W(\nu,\pi)\right)^2 + \frac{1}{n(1-\gamma\kappa)^2} M(\nu, P).$$

The optimal solution of $\min_\gamma \widehat{\mathrm{MSE}}(\gamma)$ takes the form $\gamma^* = 1 - n^{-1}M(\nu,P)(1-\kappa)^{-1}W(\nu,\pi)^{-2}$. In practice, $M(\nu,P)/(1-\kappa)$ and $W(\nu,\pi)$ can be hard to calculate. Using the fact that $M(\nu,P)/(1-\kappa) \ge \pi V_P/(1-\kappa) = \mathrm{var}_\pi(f)$ and $W(\nu,\pi) \ge |\pi f - \nu f|$, we obtain the following heuristic

$$\widehat{\gamma}^* = 1 - \frac{c}{n}\frac{\mathrm{var}_\pi(f)}{|\pi f - \nu f|^2},$$

where c is a constant independent of n and P. In our numerical experiments, setting $c = 4$ leads to good performance (see Appendix A.1.2).

We make several important observations. First, $\hat{\gamma}^*$ converges to 1 as the sample size n increases to infinity. This is expected since as the sample size increases, the variance of the estimator decreases while the intrinsic bias from discounting stays the same. Second, $(1-\hat{\gamma}^*)^{-1} \sim n$. This indicates that the number of regenerations remains of constant order as the sample size increases. Moreover, $\text{var}_\pi(f)/|\pi f - \nu f|^2$ is of a constant order as $\kappa \to 1$, so the ergodicity of the chain has a small effect on the optimal discount factor as n grows large (see, for example, a case study for the $M/M/1$ queue in Appendix C, where the stationary measure and asymptotic variance of the discounted estimator are known explicitly). Lastly, discounting does not improve how the MSE scales with n and κ. In particular, based on $\widehat{\text{MSE}}(\gamma^*)$, we have

$$\min_{\gamma} \text{MSE}(\gamma) = O\left(n^{-1}(1-\kappa)^{-2}\right).$$

This indicates that discounting with a properly chosen γ can only improve the efficiency of the estimator by a constant factor, not by order of magnitude. Our numerical experiments in Appendix A.1.2 confirms this observation.

4 Poisson equation estimation

In this section, we study another important estimation problem – the solution to the Poisson equation. This problem arises in MDPs with long-run average reward.

For a Markov chain with transition kernel P, stationary distribution π, and cost function f, the Poisson equation is defined as

$$(P-I)h = -f + \pi f. \tag{5}$$

A function $h : \mathcal{X} \to \mathbb{R}$ that satisfies (5) is referred to as a solution to the Poisson equation. Under suitable ergodicity conditions, the solution exist and is unique up to a constant shift. A solution is called fundamental if $\pi h = 0$. If the Markov chain is V-uniformly ergodic, then for any cost function f with $|f| \le V$, the Poisson equation (5) admits a fundamental solution of the form (Proposition A.3.11 in Meyn [2011])

$$\bar{h}(x) = \mathbb{E}_x\left[\sum_{k=0}^{\infty}(f(X_k) - \pi f)\right].$$

We also introduce a solution based on the regeneration idea, which is widely used when applying Monte Carlo simulation to estimate the solution [Dai and Gluzman, 2021]. Fix $x^* \in \mathcal{X}$ and define $\tau = \min\{t \ge 1 : X_t = x^*\}$. Suppose $\mathbb{E}_{x^*}[\tau] < \infty$, i.e., x^* is a regeneration state. Then, we have the solution:

$$h^{x^*}(x) = \mathbb{E}_x\left[\sum_{t=0}^{\tau-1}(f(X_t) - \pi f)\right].$$

Note that $h^{x^*}(x^*) = 0$. When using simulation to estimate $h^{x^*}(x)$, we can generate independent and identically distributed (iid) copies of $H_x(i) = \sum_{t=0}^{\tau-1}(f(X_t) - \pi f)$ given $X_0 = x$ [1]. Let n denote the total number of steps we generate the Markov chain and $N(n)$ denote the number of $H_x(i)$'s generated in n steps. Then, we have [Glynn and Whitt, 1992]

$$\sqrt{N(n)}\left(\sum_{i=1}^{N(n)} H_x(i) - h^{x*}(x)\right) \Rightarrow N\left(0, \mathbb{E}_x[\tau]\text{var}_x\left(\sum_{t=0}^{\tau-1}(f(X_t) - \pi f)\right)\right) \text{ as } n \to \infty.$$

We refer to

$$\eta(f)(x) := \mathbb{E}_x[\tau]\text{var}_x\left(\sum_{t=0}^{\tau-1}(f(X_t) - \pi f)\right)$$

as the asymptotic variance of the estimator. A key challenge in implementation is that it can take a long time for the Markov chain to regenerate, which can lead to a large value of $\eta(f)(x)$.

[1] In practice, we may not know πf. Thus, we need to plug in a properly constructed estimate of πf.

We next introduce two Poisson equation solutions based on the discounted chain. First, the discounted chain $P_{\gamma,\nu}$, regenerates every time we draw samples from ν. Thus, we can consider:

$$h_\gamma(x) = \mathbb{E}_x\left[\sum_{t=0}^{\tau_\gamma-1}(f(X_t)-\pi f)\right],$$

where $\tau_\gamma \sim$ Geometric$(1-\gamma)$ and is independent of the Markov chain $\{X_t\}_{t\geq 0}$. Note that we drop the dependence on ν as it has no effect on $h_\gamma(x)$.

Second, we consider using the geometric regeneration time to augment the existing regenerative structure of P. This gives rise to the approximated solution:

$$h_\gamma^{x^*}(x) = \mathbb{E}_x\left[\sum_{t=0}^{\tau\wedge\tau_\gamma-1}(f(X_t)-\pi_{\gamma,\delta_{x^*}} f)\right].$$

Lemma 2. *Consider the Poisson equation for the discounted chain $P_{\gamma,\nu}$ with stationary distribution $\pi_{\gamma,\nu}$, given by $(P_{\gamma,\nu}-I)h = -f + \pi_{\gamma,\nu}f$. The function h_γ is a solution when $\nu = \pi$ and $\pi h_\gamma = 0$. $h_\gamma^{x^*}$ is a solution when $\nu = \delta_{x^*}$ and $h_\gamma^{x^*}(x^*) = 0$.*

4.1 Estimation based on h_γ

For estimation based on h_γ, we have the following quantification of its accuracy and efficiency:

Theorem 2. *Suppose Assumption 1 holds and f is Lipschitz continuous. Assume for any $x \in \mathcal{X}$ and $\gamma \in [0,1]$, $W(\delta_x, \pi)$, $\pi_{\gamma,\delta_x} V_P$, $\pi_{\gamma,\delta_x} D(\delta_z P, \delta_x) < \infty$ Then, the bias of the discounted estimator satisfies*

$$\left|h_\gamma(x) - \bar{h}(x)\right| \leq C\|f\|_{Lip}\frac{\kappa(1-\gamma)}{(1-\kappa)(1-\gamma\kappa)}W(\delta_x,\pi).$$

The asymptotic variance satisfies

$$\mathbb{E}_x[\tau_\gamma] var_x\left(\sum_{k=0}^{\tau_\gamma-1}[f(X_k)-\pi f]\right) \leq 2C^3\|f\|_{Lip}^2\frac{M(\delta_x,P)}{(1-\gamma)^2(1-\gamma\kappa)^2} + 2C^2\frac{1}{(1-\gamma\kappa)^3}\|f\|_{Lip}^2 W(\delta_x,\pi)^2.$$

We make several remarks about the bounds in Theorem 2. First, the bounds for the bias and the variance have similar dependence on κ and γ as those developed in Lemmas 7 and 8 in [Dai and Gluzman, 2021]. The key difference is that for the variance bound, we do not require f^2 to be strongly dominated by a Lyapunov function V. We only require f to be Lipschitz and $\pi_{\gamma,\delta_x}(\mathrm{V}_P)$ to be bounded. Moreover, we are able to characterize the the dependence of the bias and the variance on κ more explicitly. These differences are due to a different analysis framework we use. Second, there is again an efficiency-accuracy tradeoff. The bias decreases while the variance increases as γ increases. Lastly, let $\eta_\gamma(f)$ denote the asymptotic variance of the discounted estimator. Then,

$$|h_\gamma(x) - \bar{h}(x)| = O\left(\frac{1-\gamma}{(1-\kappa)(1-\gamma\kappa)}\right) \text{ and } \eta_\gamma(f)(x) = O\left(\frac{1}{(1-\gamma)^2(1-\gamma\kappa)^2}\right).$$

Let $\mathrm{MSE}_p(\gamma)(x) = |h_\gamma(x) - \bar{h}(x)|^2 + \eta_\gamma(f)(x)/n$ be the MSE of the sample-average estimator based on h_γ. Utilizing the upper bounds for the bias and variance developed in in Theorem 2, we consider the following approximated MSE,

$$\widehat{\mathrm{MSE}}_p(\gamma) := W(\delta_x,\pi)^2\left(\frac{(1-\gamma)}{(1-\kappa)(1-\gamma\kappa)}\right)^2 + \frac{M(\delta_x,P)}{n(1-\gamma)^2(1-\gamma\kappa)^2} + \frac{W(\delta_x,\pi)^2}{(1-\gamma\kappa)^3}.$$

Let γ^* be the minimizer of $\widehat{\mathrm{MSE}}_p(\gamma)$. Then, as n grows large

$$\frac{1}{1-\gamma^*} \sim \left(\frac{W(\delta_x,\pi)^2}{M(\delta_x,P)(1-\kappa)^2}\right)^{1/4} n^{1/4}.$$

Based on $\widehat{\mathrm{MSE}}_p(\gamma^*)$, we also obtain an upper bound on the optimal MSE.

$$\min_\gamma \mathrm{MSE}_p(\gamma) = O\left(\frac{W(\delta_x,\pi)\sqrt{M(\delta_x,P)}}{n^{1/2}(1-\kappa)^3}\right).$$

From the analysis above, we note that for Poisson equation estimation based on h_γ, γ controls the bias, the variance, and the effective sample size. To minimize the MSE, we want to balance $(1-\gamma)$ (the bias) and $(1-\gamma)^{-2}n^{-1}$ (the variance). This leads to $\mathbb{E}\tau_{\gamma^*} = n^{1/4}$ and a convergence rate of $n^{-1/2}$ for the MSE, which is slower than the canonical Monte Carlo rate.

4.2 Estimation based on $h_\gamma^{x^*}$

The solution $h_\gamma^{x^*}(x)$ uses the regeneration times $\tau \wedge \tau_\gamma$, which provides a better control over the regeneration time than using τ or τ_γ alone. However, this introduces greater dependence on the state x. Through an analysis of the regeneration times, we argue that for the optimal discount factor, $(1-\gamma^*)^{-1}$ should scale at least as fast as $\mathbb{E}_x[\tau]$ as x grows large.

Consider a bounded 1-Lipschitz function $f \in \mathrm{Lip}(\mathcal{X}, d)$ with $\sup_{x\in\mathcal{X}} |f(x)| \le \alpha \in (0,\infty)$. We have the following bounds for the bias and asymptotic variance:

$$\left|h_\gamma^{x^*}(x) - h^{x^*}(x)\right|^2 \le \left(\frac{\alpha}{1-\kappa}\mathbb{E}_x[\tau\wedge\tau_\gamma](1-\gamma) + \alpha|\mathbb{E}_x[\tau] - \mathbb{E}_x[\tau\wedge\tau_\gamma]|\right)^2,$$
$$\mathbb{E}_x[\tau\wedge\tau_\gamma]\mathrm{var}_x\left(\sum_{t=0}^{\tau\wedge\tau_\gamma-1} f(X_t) - \pi_{\gamma,\delta_{x^*}} f\right) \le 4\alpha^2\mathbb{E}_x[\tau\wedge\tau_\gamma]\mathbb{E}_x[(\tau\wedge\tau_\gamma)^2].$$

The next proposition characterizes the moments of $\tau \wedge \tau_\gamma$.

Proposition 3. *Let $\tau \ge 1$ be a random variable with a finite moment generating function in a neighborhood of the origin. For τ_γ independent of τ, we have:*

$$\mathbb{E}_x[\tau\wedge\tau_\gamma] = \frac{1-\mathbb{E}_x[\gamma^\tau]}{1-\gamma}, \quad \mathbb{E}_x[(\tau\wedge\tau_\gamma)^2] = \frac{(1+\gamma)(1-\mathbb{E}_x[\gamma^\tau])}{(1-\gamma)^2} - 2\frac{\mathbb{E}_x[\tau\gamma^\tau]}{1-\gamma},$$
$$\textit{and} \quad \mathbb{E}_x[e^{\lambda(\tau\wedge\tau_\gamma)}] = \begin{cases} \frac{e^\lambda-1}{\gamma e^\lambda-1}\left(\mathbb{E}_x\left[e^{(\lambda+\log\gamma)\tau}\right]-1\right)+1 & \textit{if } \lambda \ne -\log\gamma \\ \frac{1-\gamma}{\gamma}\mathbb{E}_x[\tau]+1 & \textit{if } \lambda = -\log\gamma. \end{cases}$$

We note from Proposition 3 that since τ has a finite moment generating function in a neighborhood of the origin, i.e., $\mathbb{E}_x[e^{\lambda\tau}] < \infty$ for some $\lambda > 0$, $\mathbb{E}_x[e^{\lambda(\tau\wedge\tau_\gamma)}] < \infty$ for some $\lambda > -\log(\gamma)$. This suggests that $\tau\wedge\tau_\gamma$ can have a lighter tail than τ or τ_γ. To gain more insights into the tail behavior, define the Orlicz norm of a random variable X as

$$\|X\|_o := \inf\left\{c > 0 : \exp\left(\frac{|X|}{c}\right) \le 2\right\}.$$

The Orlicz norm provides a standardized way to compare the tail behavior of random variables through the Chernoff bound. For the discounted estimator, we have

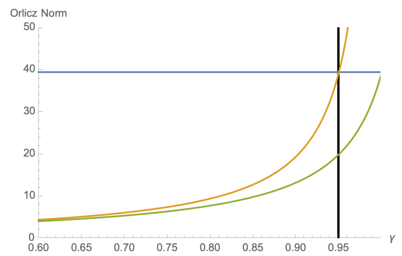
$$\mathbb{P}\left(\left|\sum_{t=0}^{\tau\wedge\tau_\gamma-1} f(X_t) - \pi_\gamma f\right| > t\right) \le \mathbb{P}\left(2\alpha(\tau\wedge\tau_\gamma) > t\right) \le 2e^{-\frac{t}{2\alpha\|\tau\wedge\tau_\gamma\|_o}}.$$

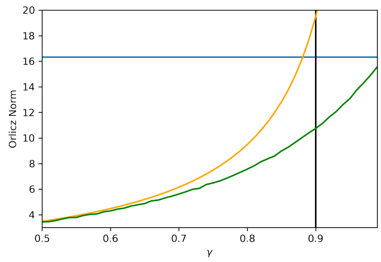
If we set $\mathbb{E}[\tau_\gamma] = \mathbb{E}_x[\tau]$, $\gamma = 1 - 1/\mathbb{E}_x[\tau]$. For $\lambda = -\log\gamma$, from Proposition 3, we have

$$\mathbb{E}_x[e^{\lambda(\tau\wedge\tau_\gamma)}] = \frac{2\mathbb{E}_x[\tau]-1}{\mathbb{E}_x[\tau]-1} < 2.$$

This indicates that $\|\tau\wedge\tau_\gamma\|_o < -1/\log\gamma$. Thus, $\mathbb{P}_x(\tau\wedge\tau_\gamma > t) = o(\gamma^t)$. Figure 1 compares the Orlicz norm of τ, τ_γ and $\tau\wedge\tau_\gamma$ when we assume τ has some specific distributions. We note that when $\mathbb{E}[\tau_\gamma] \approx \mathbb{E}_x[\tau]$, $\|\tau\wedge\tau_\gamma\|_o$ can be much smaller than $\|\tau\|_o \wedge \|\tau_\gamma\|_o$. To balance the efficiency-accuracy tradeoff, we suggest setting $(1-\gamma^*(x))^{-1}$ at least as large as $\mathbb{E}_x[\tau]$. And the value of $\gamma^*(x)$ should depend on x through $\mathbb{E}_x[\tau]$.

Figure 1: Comparison of $\|\tau\wedge\tau_\gamma\|_o$ (green) to $\|\tau\|_o$ (blue) and $\|\tau_\gamma\|_o$ (orange) when $\tau\sim$ Geometric$(1-0.95)$ (left) and $\tau\sim$Poisson(10) (right) for different values of γ





5 Conclusion

In this paper, we apply the Wasserstein ergodicty framework to quantify the accuracy-efficiency tradeoff when using discouting in Markov chain estimation. We study two estimation tasks: steady-state estimation and estimating the solution to the Poisson equation. Let γ^* denote the optimal discount factor. In steady-state estimation, we show that $(1-\gamma^*)^{-1} \sim n$ and discounting does not improve how the MSE scales with n and κ (the ergodicity constant). In Poisson equation estimation, discounting can help control the tail behavior of the regeneration time. In our numerical experiments, we see large efficiency gains when applying proper discounting schemes.

References

Søren Asmussen. *Applied probability and queues.* Springer Science & Business Media, 2008. B.1, C

Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007. 1, 1.1

Jonathan Baxter and Peter L. Bartlett. Infinite-horizon policy-gradient estimation. *Journal of Artificial Intelligence Research*, 15(1):319–350, 2001. 1.1

David Blackwell. Discrete dynamic programming. *The Annals of Mathematical Statistics*, pages 719–726, 1962. 1.1

Jose Blanchet and Henry Lam. State-dependent importance sampling for rare-event simulation: An overview and recent advances. *Surveys in Operations Research and Management Science*, 17(1):38–59, 2012. 1.1

A. Di Crescenzo, V. Giorno, A.G. Nobile, and L.M. Ricciardi. On the m/m/1 queue with catastrophes and its continuous approximation. *Queuing Systems*, 43:329–347, 2003. C

Jim G Dai and Mark Gluzman. Queueing network controls via deep reinforcement learning. *Stochastic Systems*, 2021. 1.1, 4, 4.1

Raaz Dwivedi, Yuansi Chen, Martin J. Wainwright, and Bin Yu. Log-concave sampling: Metropolis-hastings algorithms are fast. *Journal of Machine Learning Research*, 20(183):1–42, 2019. URL http://jmlr.org/papers/v20/19-306.html. 1.1

Peter W Glynn and Donald L Iglehart. Conditions for the applicability of the regenerative method. *Management Science*, 39(9):1108–1111, 1993. 1.1

Peter W. Glynn and Dirk Ormoneit. Hoeffding's inequality for uniformly recurrent markov chains. *Statistics and Probability Letters*, 56:143–146, 2002. B.1

Peter W Glynn and Ward Whitt. The asymptotic efficiency of simulation estimators. *Operations research*, 40(3):505–520, 1992. 4

Martin Hairer, Andrew M Stuart, and Sebastian J Vollmer. Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *The Annals of Applied Probability*, 24(6):2455–2490, 2014. 1.1, 2.2

Shane G Henderson and Peter W Glynn. Approximating martingales for variance reduction in markov process simulation. *Mathematics of Operations Research*, 27(2):253–271, 2002. 1.1

Loïc Hervé and James Ledoux. Approximating markov chains and v-geometric ergodicity via weak perturbation theory. *Stochastic Processes and their Applications*, 124(1):613–638, 2014. 1.1

Tommi Jaakkola, Satinder P. Singh, and Michael I. Jordan. Reinforcement learning algorithm for partially observable markov decision problems. In *NIPS '94*, 1994. 1.1

Nan Jiang, Satinder Singh, and Ambuj Tewari. On structural properties of mdps that bound loss due to shallow planning. In *IJCAI'16*, 2016. 1.1

Alderic Joulin. A new poisson-type deviation inequality for markov jump processes with positive wasserstein curvature. *Bernoulli*, 15(532-549), 2009. 2.2

Alderic Joulin and Yann Ollivier. Curvature, concentration, and error estimates for markov chain monte carlo. *The Annals of Probability*, 38(6):2418–2442, 2010. 1.1, 2.2, 2.3

Sham Kakade. Optimizing average reward using discounted rewards. In *COLT '01*, 2001. 1.1

Sham Kakade. *On the sample complexity of reinforcement learning*. PhD thesis, 2003. 1, 1.1

B. Krishna Kumar and D. Arivudainambi. Transient solution of an m/m/1 queue with catastrophes. *Computers & Mathematics with Applications*, 40(10-11):1233–1240, 2000. C

Peter Marbach and John N. Tsitsiklis. Simulation-based optimization of markov reward processes. *IEEE Transactions on Automatic Control*, 46(2):191–209, 2001. 1.1

Sean Meyn. *Control techniques for complex networks*. Cambridge University Press, 2011. 1, 4

Sean Meyn and Richard Tweedie. *Markov chains and stochastic stability*. Springer-Verlag, 1993. B.1

Sean P Meyn. The policy iteration algorithm for average reward markov decision processes with general state space. *IEEE Transactions on Automatic Control*, 42(12):1663–1680, 1997. 1

Eric C Ni and Shane G Henderson. How hard are steady-state queueing simulations? *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 25(4):1–21, 2015. 1.1

Filipo Studzinski Perotto and Laurent Vercouter. Tuning the discount factor in order to reach average optimality on deterministic mdps. In *SGAI '18*, 2018. 1.1

Marek Petrik and Bruno Scherrer. Biasing approximate dynamic programming with a lower discount factor. In *NIPS '08*, 2008. 1.1

Gareth O. Roberts and Jeffrey Rosenthal. Geometric ergodicity and hybrid markov chains. *Electronic Commu*, 1997. 2.2

Gareth O Roberts and Richard L Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. 1.1

Daniel Rudolf and Nikolaus Schweizer. Perturbation theory for markov chains via wasserstein distance. *Bernoulli*, 24(4):2610–2639, 2018. 1.1, 2.2, 2.2

Manuel Schneckenreither. Average reward adjusted discounted reinforcement learning near-blackwell-optimal policies for real-world applications. arXiv:2004.00857, 2020. 1.1

Adrian FM Smith and Gareth O Roberts. Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(1): 3–23, 1993. 1

Dan Spitzner and Thomas Boucher. Asymptotic variance of functionals of discrete-time markov chains via the drazin inverse. *Electronic Communications in Probability*, 12:120–133, 2007. 2.1

Philip S. Thomas. Bias in natural actor-critic algorithms. In *ICML '14*, 2014. 1.1

Andi Q. Wang, Murray Pollock, Gareth O. Roberts, and David Steinsaltz. Regeneration-enriched markov processes with application to monte carlo. *Annals of Applied Probability*, 31(2):703–735, 2021. 1.1

Wei-Ning Yang and Barry L Nelson. Multivariate batch means and control variates. *Management science*, 38(10):1415–1431, 1992. 1.1

A Numerical Experiments

A.1 Steady-state estimation

In section, we numerically investigate the effect of discounting on the bias and variance of the sample average approximation. We consider the Markov chains in Examples 2.1 – 2.3:

- AR(1): $f(x) = |x|$ and $\nu = \delta_0$.
- Binomial Chain: $f(x) = |x|$ and $\nu = \delta_{\lceil aN \rceil}$.
- $M/M/1$ queue: $f(x) = x$ and $\nu = \delta_0$.

A.1.1 Bias and Variance Quantification

We first demonstrate that the upper bounds for the bias and the asymptotic variance are tight in terms of its dependence on the discount rate. Since the constants in our bounds can be off, we consider the normalized bias $b(\gamma) \equiv |\pi f - \pi_\gamma f|/|\pi f - \nu f|$ and variance $v(\gamma) \equiv \sigma_\gamma^2(f)/\sigma^2(f)$. Under such normalization, the bounds derived from Theorem 1 give

$$b(\gamma) = O(\frac{1-\gamma}{1-\gamma\kappa}) \text{ and } v(\gamma) = O\left(\left(\frac{1-\kappa}{1-\gamma\kappa}\right)^2\right).$$

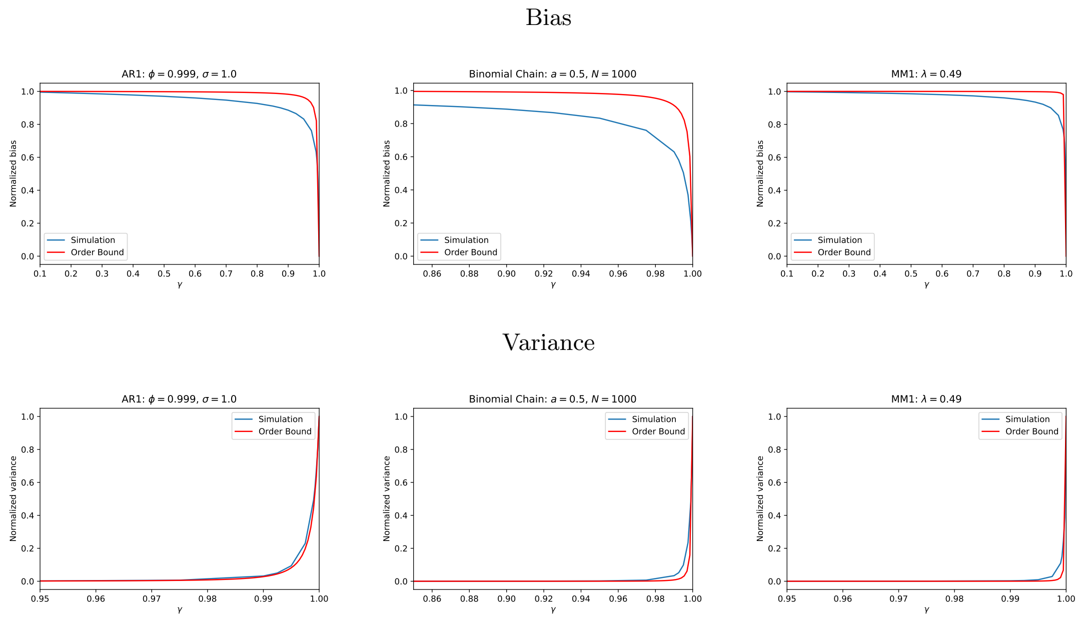
Figure 2 compares the normalized bias and variance from simulations of the discounted estimator with the bounds derived from Theorem 1. We observe that these bounds accurately capture the dependence of the bias and the variance on γ.

A.1.2 Optimal Discount Factor for Steady-State Estimation

We next calculate the MSE (using simulation) for a grid of discount factors. Figure 3 compares the heuristic discount factor

$$\widehat{\gamma}^* = 1 - \frac{4}{n}\frac{\mathrm{var}_\pi(f)}{|\pi f - \nu f|^2},$$

Figure 2: Bias and asymptotic variance comparison



(denoted by the black line in Figure 3) with the optimal discount factor γ^* through grid search. We observe that the heuristic is able to accurately locate the optimal discount factor. In Table 1, we also present the reduction and percentage reduction in the mean squared error at the optimal discount factor γ^* compared to the non-discounted case ($\gamma = 1$), i.e., ΔMSE and $\%\Delta$MSE respectively. As our analysis has suggested, discounting does not change how the MSE scales with the ergodicity of the Markov chain, but it still can result in constant-order reduction in MSE.

A.2 Poisson equation estimation

A.2.1 Estimation based on h_γ

In this section, we run simulation to investigate the scaling of the optimal discount factor and the corresponding MSE on the sampling budget n as $n \to \infty$ for Poisson solution estimation. We again consider the Markov chains in Examples 2.1 – 2.3 with $f(x) = x$. We perform stochastic bisection search to find the discount factor that minimizes the MSE given a particular Markov chain for a range of sample sizes. We then compute $-\log(1-\gamma^*)$ and $\log \mathrm{MSE}(\gamma^*)$ and regress them against $\log n$. Our prediction for the coefficients for $\log n$ is $1/4$ and $-1/2$ respectively.

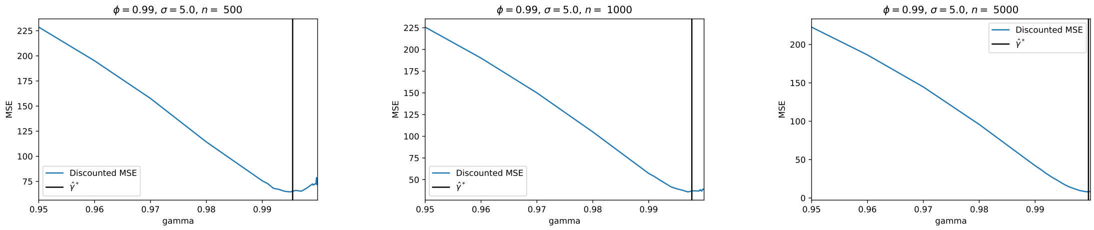
Figure 4 plots $-\log(1-\gamma^*)$ and $\log \mathrm{MSE}(\gamma^*)$ against $\log n$ (The red line is the predicted line with slope $1/4$ and $-1/2$ respectively; the green line is the best fitted regression line). We observe that the predicted coefficients are close to the fitted coefficient.

A.2.2 Estimation based on $h_\gamma^{x^*}$

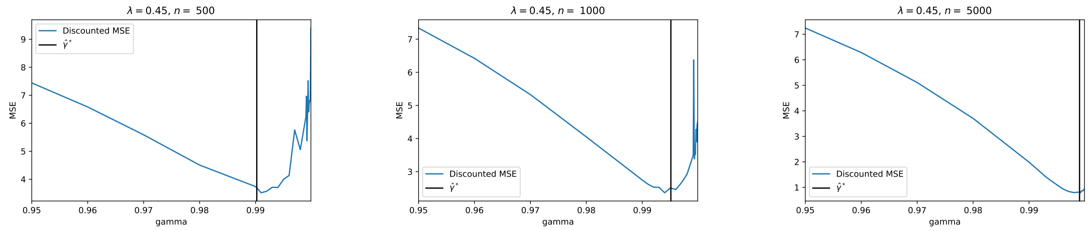
In this section, we run simulation to study the efficiency gain from discounting. We consider the discounted estimator h_γ^0, i.e., $x^* = 0$. We use a grid search to find the optimal discount factor γ^*. Table 2 compares the value of $\mathrm{MSE}(\gamma)$ and $\mathrm{MSE}(1)$ (the MSE of the estimator without discounting) for an M/M/1 queue with traffic intensities (which leads to different ergodicity constants) and different values of x. We observe that proper discounting can achieve order of magnitude performance improvements. In addition, the optimal discount factor increases as x increases.

Figure 3: MSE of the discounted estimator of πf

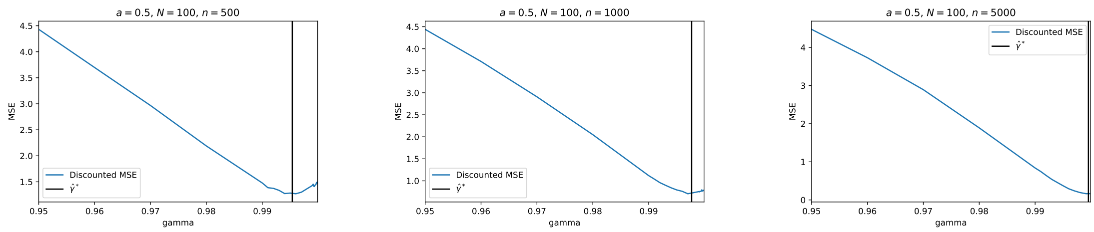
$AR(1) : \phi = 0.9, \sigma = 5$



$M/M/1 : \lambda = 0.45$



$Binomial : a = 0.5, N = 100$



B Proofs of the technical results

B.1 Proof of Proposition 1

Proof. By construction of the discounted Markov chain, for any set A with $\nu(A) > \epsilon$, we have

$$\inf_{x \in \mathcal{X}} P(x, A) \geq (1-\gamma)\nu(A) > (1-\gamma)\epsilon$$

Thus, the Markov chain is uniformly ergodic (Theorem 16.0.2 in Meyn and Tweedie [1993]). □

As a result of uniform ergodicity, the discounted Markov chain is positive Harris and satisfies a Law of Large Numbers along with a Central Limit Theorem (Theorem 17.0.1 in Meyn and Tweedie [1993]),

$$\sqrt{n}\left(\frac{1}{n}\sum_{t=0}^{n-1} f(X_t) - \pi_{\gamma,\nu} f\right) \Rightarrow N(0, \sigma^2_{\gamma,\nu}(f))\text{as } n \to \infty.$$

By the Central Limit Theorem for regenerative processes (Theorem 6.3.2 in Asmussen [2008]), the asymp-

Table 1: Comparison of $\hat{\gamma}$ with optimal discount factor γ^*

AR(1)

ϕ	0.99	0.99	0.99	0.99	0.999	0.999	0.999	0.999
σ	1	1	1	1	1	1	1	1
n	500	1000	2000	5000	3000	5000	7000	10000
γ^*	0.996	0.997	0.9991	0.9994	0.9994	0.9996	0.9998	0.9997
$\hat{\gamma}$	0.995	0.998	0.999	0.9995	0.9992	0.9995	0.9997	0.9998
ΔMSE	-0.5	-0.2	-0.0	-0.0	-11.0	-5.2	-3.1	-1.5
%ΔMSE	-17%	-10%	-6	-3%	-23%	-17%	-14%	-9%

Binomial Chain

N	100	100	100	100	1000	1000	1000	1000
a	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
n	500	1000	2000	5000	3000	5000	7000	10000
γ^*	0.996	0.998	0.9996	0.9995	0.9993	0.9996	0.9997	0.9998
$\hat{\gamma}$	0.995	0.998	0.999	0.9995	0.9992	0.9995	0.9997	0.9998
ΔMSE	-1.5	-0.8	-0.2	-0.2	-5.2	-2.4	-1.4	-0.8
%ΔMSE	-15%	-9%	-5%	-3%	-22%	-16%	-12%	-11%

$M/M/1$

λ	0.45	0.45	0.45	0.45	0.49	0.49	0.49	0.49
n	500	1000	2000	5000	3000	5000	7000	10000
γ^*	0.991	0.994	0.997	0.998	0.9991	0.9994	0.9994	0.9995
$\hat{\gamma}$	0.990	0.995	0.998	0.999	0.999	0.9991	0.9994	0.9996
ΔMSE	-4.00	-1.97	-0.81	-0.16	-239.8	-191.4	-156.6	-121.0
%ΔMSE	-53%	-45%	-35%	-17%	-60%	-59%	-57%	-55%

totic variance $\sigma^2_{\gamma,\nu}(f)$ can be expressed as,

$$\sigma^2_{\gamma,\nu}(f) = \frac{1}{\mathbb{E}[\tau_\gamma]}\left[\mathrm{var}_\nu\left(\sum_{k=0}^{\tau_\gamma-1} f(X_k)\right) + \frac{\mathbb{E}_\nu\left[\sum_{k=0}^{\tau_\gamma-1} f(X_k)\right]^2}{\mathbb{E}[\tau_\gamma]^2}\mathrm{var}(\tau_\gamma)\right.$$
$$\left. -2\frac{\mathbb{E}_\nu\left[\sum_{k=0}^{\tau_\gamma-1} f(X_k)\right]}{\mathbb{E}[\tau_\gamma]}\mathrm{Cov}_\nu(\tau_\gamma, \sum_{k=0}^{\tau_\gamma-1} f(X_k))\right].$$

We next note that

$$\mathrm{Cov}_\nu\left(\tau_\gamma, \sum_{k=0}^{\tau_\gamma-1} f(X_k)\right) = \mathbb{E}_\nu\left[\tau_\gamma \sum_{k=0}^{\tau_\gamma-1} f(X_k))\right] - \mathbb{E}[\tau_\gamma]\mathbb{E}_\nu\left[\sum_{k=0}^{\tau_\gamma-1} f(X_k))\right].$$

Table 2: Comparison of MSE for the discounted and standard Poisson solution estimators for the $M/M/1$ queue

λ	x	n	γ^*	MSE(γ^*)	MSE(1)
0.4	1	100	0.8	2.88 ± 0.26	20.18 ± 0.88
0.4	1	500	0.95	1.26 ± 0.08	4.69 ± 0.29
0.45	1	100	0.85	18.09 ± 2.96	595.85 ± 45.64
0.45	1	500	0.85	14.49 ± 1.33	120.76 ± 7.48
0.49	1	100	0.9	$10,152 \pm 2,819$	$1,589,682 \pm 214,295$
0.49	1	500	0.99	$1,966 \pm 456$	$269,053 \pm 15,662$
0.495	1	100	0.9	$338,292 \pm 83,284$	$41,170,786 \pm 7,740,090$
0.496	1	500	0.95	$86,426 \pm 24,454$	$10,868,799 \pm 1,082,131$

λ	x	n	γ^*	MSE(γ^*)	MSE(1)
0.4	3	100	0.99	70.67 ± 1.72	91.22 ± 3.72
0.4	3	500	0.9999	18.41 ± 0.61	19.39 ± 0.59
0.45	3	100	0.99	682.35 ± 30	$2,402 \pm 161$
0.45	3	500	0.999	357.03 ± 12	421.21 ± 18
0.49	3	100	0.99	$64,075 \pm 22,312$	$3,927,396 \pm 425,485$
0.49	3	500	0.999	$59,433 \pm 4,797$	$955,687 \pm 50,221$
0.495	3	100	0.99	$474,387 \pm 149,954$	$108,732,453 \pm 10,901,321$
0.495	3	500	0.999	$322,360 \pm 62,828$	$28,844,200 \pm 1,679,721$

Since

$$\begin{aligned}\mathbb{E}_\nu\left[\tau_\gamma \sum_{k=0}^{\tau_\gamma-1} f(X_k))\right] &= \sum_{t=1}^{\infty}(1-\gamma)\gamma^{t-1}t\sum_{k=0}^{t-1}\nu P^k f \\ &= \sum_{k=0}^{\infty}(1-\gamma)\nu P^k f\sum_{t=k}^{\infty}(t+1)\gamma^t \text{ by Fubini's theorem} \\ &= \frac{1}{1-\gamma}\sum_{k=0}^{\infty}\gamma^k\nu P^k f + \sum_{k=0}^{\infty}k\gamma^k\nu P^k f\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\tau_\gamma]\mathbb{E}_\nu\left[\sum_{k=0}^{\tau_\gamma-1} f(X_k))\right] &= \frac{1}{1-\gamma}\sum_{t=1}^{\infty}(1-\gamma)\gamma^{t-1}\sum_{k=0}^{t-1}\nu P^k f \\ &= \frac{1}{1-\gamma}\sum_{k=0}^{\infty}(1-\gamma)\nu P^k f\sum_{t=k}^{\infty}\gamma^t \text{ by Fubini's theorem} \\ &= \frac{1}{1-\gamma}\sum_{k=0}^{\infty}\gamma^k\nu P^k(f),\end{aligned}$$

$$\mathrm{Cov}_\nu\left(\tau_\gamma, \sum_{k=0}^{\tau_\gamma-1} f(X_k)\right) = \sum_{k=0}^{\infty}k\gamma^k\nu P^k f.$$

Then,

$$\sigma^2_{\gamma,\nu}(f) = (1-\gamma)\mathrm{var}_\nu\left(\sum_{k=0}^{\tau_\gamma-1} f(X_k)\right) + \gamma(1-\gamma)\left(\sum_{k=0}^{\infty}\gamma^k \nu P^k f\right)^2 - 2(1-\gamma)^2\left(\sum_{k=0}^{\infty}\gamma^k \nu P^k f\right)\left(\sum_{k=0}^{\infty}k\gamma^k \nu P^k f\right).$$

Finally, the chain satisfies a concentration inequality for bounded functions. For any measurable function f with $\sup_{x\in\mathcal{X}}|f(x)| \le \alpha$, we have

$$\mathbb{P}_\nu\left(\frac{1}{n}\sum_{t=0}^{n-1} f(X_t) - \mathbb{E}_\nu\left[\frac{1}{n}\sum_{t=0}^{n-1} f(X_t)\right] \ge \epsilon\right) \le \exp\left(-\frac{(1-\gamma)^2(n\epsilon - \frac{2\alpha}{1-\gamma})^2}{2n\alpha^2}\right)$$

for $n > 2\alpha/(\epsilon(1-\gamma))$. Since $P(x, A) \ge (1-\gamma)\nu(A)$ for all $x \in \mathcal{X}$, the result follows from Theorem 2 in [Glynn and Ormoneit, 2002].

B.2 Proof of Proposition 2

Proof. (1) Recall that $||f||_{\mathrm{Lip}}$ is the Lipchitz constant of f. Then $f/||f||_{\mathrm{Lip}} \in \mathrm{Lip}(\mathcal{X}, d)$.

$$\begin{aligned}|P^n f(x) - P^n f(y)| &\le ||f||_{\mathrm{Lip}} \sup_{g\in\mathrm{Lip}(\mathcal{X},d)} |P^n g(x) - P^n g(y)| \\ &= ||f||_{\mathrm{Lip}} W(\delta_x P^n, \delta_y P^n)\frac{d(x,y)}{d(x,y)} \le \|f\|_{\mathrm{Lip}} C\kappa^n d(x,y)\end{aligned}$$

(2) By the definition of Wasserstein distance,

$$\begin{aligned}W(\mu P^n, \nu P^n) = \sup_{f\in\mathrm{Lip}(\mathcal{X},d)} |\mu P^n f - \nu P^n f| &= \sup_{f\in\mathrm{Lip}(\mathcal{X},d)} |\mu(P^n f) - \nu(P^n f)| \\ &\le ||P^n f||_{\mathrm{Lip}} W(\mu,\nu) \le C\|f\|_{\mathrm{Lip}}\kappa^n W(\mu,\nu).\end{aligned}$$

(3) Since $\pi P = \pi$, we have

$$\begin{aligned}\mathrm{var}_\pi(f) = \mathrm{var}_{\pi P}(f) &= \pi P f^2 - (\pi P f)^2 \\ &= \pi P f^2 - \int_{\mathcal{X}} (Pf(x))^2 \pi(dx) + \int_{\mathcal{X}} (Pf(x))^2\pi(dx) - (\pi P f)^2 \\ &= \int_{\mathcal{X}} \mathrm{var}_{\delta_x P}(f)\pi(dx) + \mathrm{var}_\pi(Pf) \\ &= \sum_{t=0}^{\infty}\int_{\mathcal{X}} \mathrm{var}_{\delta_x P}(P^t f)\pi(dx) \text{ by induction} \\ &= \sum_{t=0}^{\infty} C^2\|f\|^2_{\mathrm{Lip}}\kappa^{2t}\int_{\mathcal{X}} \mathrm{var}_{\delta_x P}\left(\frac{P^t f}{C\|f\|_{\mathrm{Lip}}\kappa^t}\right)\pi(dx) \\ &\le \sum_{t=0}^{\infty} C^2\|f\|^2_{\mathrm{Lip}}\kappa^{2t}\int_{\mathcal{X}} \sup_{g\in\mathrm{Lip}(\mathcal{X},d)} \mathrm{var}_{\delta_x P}(g)\,\pi(dx) \text{ from part (1)} \\ &= C^2\|f\|^2_{\mathrm{Lip}}\frac{1}{1-\kappa^2}\pi \mathrm{V}_P.\end{aligned}$$

This implies that

$$V_\pi \le C^2\|f\|^2_{\mathrm{Lip}}\frac{1}{1-\kappa^2}\pi \mathrm{V}_P. \tag{6}$$

Since $P^n f$ is $C||f||_{\text{Lip}}\kappa^n$-Lipschitz,

$$\begin{aligned}
\text{Var}_\pi(P^n f) &= \frac{1}{2}\int_{\mathcal{X}}\int_{\mathcal{X}}(P^n f(x) - P^n f(y))^2 \pi(dx)\pi(dy) \\
&\le \frac{1}{2}C^2\|f\|_{\text{Lip}}^2\kappa^{2n} \sup_{g\in\text{Lip}(\mathcal{X},d)} \int_{\mathcal{X}}\int_{\mathcal{X}}(g(x)-g(y))^2\pi(dx)\pi(dy) \\
&= \frac{1}{2}C^2\|f\|_{\text{Lip}}^2\kappa^{2n}\text{V}_\pi.
\end{aligned}$$

Then,

$$\begin{aligned}
\sigma^2(f) &= \text{var}_\pi(f) + 2\sum_{k=1}^{\infty}\text{cov}_\pi(f, P^k f) \\
&\le \text{var}_\pi(f) + 2\sum_{k=1}^{\infty}\text{var}_\pi(f)^{1/2}\text{var}_\pi(P^k f)^{1/2} \\
&\le \frac{1}{2}\|f\|_{\text{Lip}}^2\text{V}_\pi + 2\sum_{k=1}^{\infty}\left(\frac{1}{2}\|f\|_{\text{Lip}}^2\text{V}_\pi\right)^{1/2}\left(\frac{1}{2}C^2||f||_{\text{Lip}}^2\kappa^{2k}\text{V}_\pi\right)^{1/2} \\
&\le C||f||_{\text{Lip}}^2\text{V}_\pi\sum_{k=0}^{\infty}\kappa^k = C\|f\|_{\text{Lip}}^2\frac{1}{1-\kappa}\text{V}_\pi.
\end{aligned}$$

Plugging in the bound for V_π in (6), we have

$$\sigma^2(f) \le C^3||f||_{\text{Lip}}^2\left(\frac{1}{1-\kappa}\right)\left(\frac{1}{1-\kappa^2}\right)\pi\text{V}_P.$$

□

B.3 Proof of Lemma 1

Proof. We prove by induction that for any 1-Lipschitz function f,

$$P_\gamma^n f(x) - P_\gamma^n f(y) = \gamma^n P^n f(x) - \gamma^n P^n f(y). \tag{7}$$

When $n = 1$,

$$P_\gamma f(x) - P_\gamma f(x) = \gamma P f(x) + (1-\gamma)\nu f - \gamma P f(y) - (1-\gamma)\nu f = \gamma P f(x) - \gamma P f(y).$$

Suppose $P_\gamma^k f(x) - P_\gamma^k f(y) = \gamma^k P^k f(x) - \gamma^k P^k f(y)$ for $k \ge 1$. Then,

$$\begin{aligned}
P_\gamma^{k+1} f(x) - P_\gamma^{k+1} f(y) &= \gamma P_\gamma^k(Pf)(x) + (1-\gamma)\nu f - \gamma P_\gamma^k(Pf)(y) - (1-\gamma)\nu f \\
&= \gamma^{k+1}P^k(Pf)(x) - \gamma^{k+1}P^k(Pf)(y) = \gamma^{k+1}P^{k+1}f(x) - \gamma^{k+1}P^{k+1}f(y)
\end{aligned}$$

Thus, we have proved (7).

Next, note that

$$\begin{aligned}
W(\delta_x P_\gamma^n, \delta_y P_\gamma^n) &= \sup_{f\in\text{Lip}(\mathcal{X},d)} |P_\gamma^n f(x) - P_\gamma^n f(y)| \\
&= \sup_{f\in\text{Lip}(\mathcal{X},d)} |\gamma^n P^n f(x) - \gamma^n P^n f(y)| \\
&= \gamma^n \sup_{f\in\text{Lip}(\mathcal{X},d)} |P^n f(x) - P^n f(y)| = \gamma^n W(\delta_x P^n, \delta_y P^n).
\end{aligned}$$

□

B.4 Proof of Theorem 1

Proof. We first bound the bias

$$
\begin{aligned}
|\pi^\gamma f - \pi f| &= \left|(1-\gamma)\sum_{t=0}^{\infty}\gamma^t \nu P^t f - (1-\gamma)\sum_{t=0}^{\infty}\gamma^t \pi P^t f\right| \\
&= (1-\gamma)\sum_{t=0}^{\infty}\gamma^t \left|\nu P^t f - \pi P^t f\right| \\
&\leq (1-\gamma)\sum_{t=0}^{\infty}\gamma^t C\|f\|_{\mathrm{Lip}}\kappa^t W(\nu,\pi) \text{ by Lemma 1} \\
&\leq C\|f\|_{\mathrm{Lip}}\frac{1-\gamma}{1-\gamma\kappa}W(\nu,\pi).
\end{aligned}
$$

For the variance, from Proposition 2, we have

$$
\sigma_\gamma^2(f) \leq C\|f\|_{\mathrm{Lip}}^2\left(\frac{1}{1-\gamma\kappa}\right)\mathrm{V}_{\pi_\gamma} \leq C^3\|f\|_{\mathrm{Lip}}^2\left(\frac{1}{1-\gamma\kappa}\right)\left(\frac{1}{1-(\gamma\kappa)^2}\right)\pi_\gamma \mathrm{V}_{P_\gamma},
$$

where recall that $\mathrm{V}_{\pi_\gamma} = \sup_{f\in\mathrm{Lip}(\mathcal{X},d)}\mathrm{var}_{\pi_\gamma}(f)$ and $\mathrm{V}_{P_\gamma}(x) = \sup_{f\in\mathrm{Lip}(\mathcal{X},d)}\mathrm{var}_{\delta_x P_\gamma}(f)$.

We next establish bound for $V_{P_\gamma}(x)$. First, by the law of total variance we have

$$
\mathrm{var}_{\delta_x P_\gamma}(f) = \gamma\mathrm{var}_{\delta_x P}(f) + (1-\gamma)\mathrm{var}_\nu(f) + [\gamma(Pf(x))^2 + (1-\gamma)(\nu f)^2 - (\gamma Pf(x) + (1-\gamma)\nu f)^2]
$$

Then,

$$
\mathrm{V}_{P_\gamma}(x) = \sup_{f\in\mathrm{Lip}(\mathcal{X},d)}\mathrm{var}_{\delta_x P_\gamma}(f) \leq \gamma\mathrm{V}_P(x) + (1-\gamma)V_\nu + D_\gamma(\delta_x P,\nu).
$$

and we obtain

$$
\pi_\gamma \mathrm{V}_{P_\gamma} = \gamma\pi_\gamma\mathrm{V}_P + (1-\gamma)\mathrm{V}_\nu + \pi_\gamma(D_\gamma(\delta_x P,\nu)).
$$

□

B.5 Proof of Lemma 2

Proof. The Poisson equation for the discounted Markov chain can be written as

$$
f(x) - \pi_{\gamma,\nu}f + \gamma Ph(x) + (1-\gamma)\nu h - h(x) = 0. \tag{8}
$$

Note that $h_\gamma(x)$ is defined independent of the initial distribution ν. We take $\nu = \pi$, in which case $\pi_{\gamma,\nu} = \pi$. Note that

$$\begin{aligned} h_\gamma(x) &= \mathbb{E}_x\left[\sum_{t=0}^{\tau_\gamma-1}(f(X_t)-\pi f)\right] \\ &= \sum_{t=1}^{\infty}\sum_{k=0}^{t-1}(\delta_x P^k f - \pi f)(1-\gamma)\gamma^{t-1} \\ &= \sum_{k=0}^{\infty}\sum_{t=k}^{\infty}(1-\gamma)\gamma^t(\delta_x P^k - \pi f) \quad \text{by Fubini's theorem} \\ &= \sum_{k=0}^{\infty}\gamma^k(\delta_x P^k - \pi f) = \sum_{k=0}^{\infty}\gamma^k\delta_x P^k f - \frac{1}{1-\gamma}\pi f. \end{aligned}$$

Then,

$$\pi h_\gamma = \pi\left(\sum_{k=0}^{\infty}\gamma^k(\delta_x P^k - \pi f)\right) = \sum_{k=0}^{\infty}\gamma^k(\pi f - \pi f)) = 0.$$

Moreover,

$$\gamma P h_\gamma(x) - h_\gamma(x) = -f(x) + \pi f.$$

Thus, h_γ is the solution to (8).

For $h_\gamma^{x^*}$, let $\nu = \delta_{x^*}$. Note that $\tau \wedge \tau_\gamma$ is a regeneration time for P_γ with regeneration state x^*.

$$\begin{aligned} h_\gamma^{x^*} &= \mathbb{E}_x\left[\sum_{t=0}^{\tau\wedge\tau_\gamma-1}(f(X_t)-\pi_{\gamma,\delta_{x^*}}f)\right] \\ &= f(x) - \pi_{\gamma,\delta_{x^*}} + \mathbb{E}_x\left[\mathbb{E}_{X_1}\left[\sum_{t=1}^{\tau\wedge\tau_\gamma-1}(f(X_t)-\pi_{\gamma,\delta_{x^*}}f)\right]\right] \\ &= f(x) - \pi_{\gamma,\delta_{x^*}} + P_\gamma h_\gamma^{x^*} \end{aligned}$$

which implies that $h_\gamma^{x^*}$ is the solution to (8). □

B.6 Proof of Theorem 2

Proof. From the proof of Lemma 2, we have the following bound for the bias:

$$\begin{aligned} \left|h_\gamma(x) - \bar{h}(x)\right| &= \left|\sum_{t=0}^{\infty}\gamma^t(\delta_x P^t - \pi)f - \sum_{t=0}^{\infty}(\delta_x P^t - \pi)f\right| \\ &= \left|\sum_{t=0}^{\infty}(\gamma^t - 1)((\delta_x P^t)f - \pi f)\right| \\ &\le \sum_{t=0}^{\infty}(1-\gamma^t)\|f\|_{\mathrm{Lip}}C\kappa^t W(\delta_x,\pi) = C\|f\|_{\mathrm{Lip}}\frac{\kappa(1-\gamma)}{(1-\kappa)(1-\gamma\kappa)}W(\delta_x,\pi). \end{aligned}$$

Next, we develop a bound for $\mathrm{Var}_x\left(\sum_{k=0}^{\tau_\gamma-1}(f(X_k)-\pi f)\right)$. Consider a discounted chain with initial distribution δ_x, i.e., P^{γ,δ_x}. By Proposition 1, we have

$$\begin{aligned}\sigma^2_{\gamma,\delta_x}(f-\pi f) =&(1-\gamma)\mathrm{var}_x\left(\sum_{k=0}^{\tau_\gamma-1}(f(X_k)-\pi f)\right)+\gamma(1-\gamma)\left(\sum_{k=0}^{\infty}\gamma^k\delta_x P^k(f-\pi f)\right)^2\\ &-2(1-\gamma)^2\left(\sum_{k=0}^{\infty}\gamma^k\delta_x P^k(f-\pi f)\right)\left(\sum_{k=0}^{\infty}k\gamma^k\delta_x P^k(f-\pi f)\right).\end{aligned}$$

By rearrange the equation above, we have

$$\begin{aligned}\mathrm{var}_x\left(\sum_{k=0}^{\tau_\gamma-1}(f(X_k)-\pi f)\right) =&\frac{1}{1-\gamma}\sigma^2_{\gamma,\delta_x}(f-\pi f)-\gamma\left(\sum_{k=0}^{\infty}\gamma^k\delta_x P^k(f-\pi f)\right)^2\\ &+2(1-\gamma)\left(\sum_{k=0}^{\infty}\gamma^k\delta_x P^k(f-\pi f)\right)\left(\sum_{k=0}^{\infty}k\gamma^k\delta_x P^k(f-\pi f)\right)\\ \le&\underbrace{\frac{1}{1-\gamma}\sigma^2_{\gamma,\delta_x}(f-\pi f)}_{(A)}\\ &+\underbrace{2(1-\gamma)\left(\sum_{k=0}^{\infty}\gamma^k\delta_x P^k(f-\pi f)\right)\left(\sum_{k=0}^{\infty}k\gamma^k\delta_x P^k(f-\pi f)\right)}_{(B)}\end{aligned}$$

We next develop bounds for (A) and (B) respectively. For (A), from Theorem 1, we have

$$\begin{aligned}\frac{1}{1-\gamma}\sigma^2_{\gamma,\delta_x}(f-\pi f)\le& C^3\|f\|^2_{\mathrm{Lip}}\frac{1}{1-\gamma\kappa}\frac{1}{1-(\gamma\kappa)^2}\frac{1}{1-\gamma}\\ &\times\left(\gamma\pi_{\gamma,\delta_x}\mathrm{V}_P+(1-\gamma)\mathrm{V}_\nu+\pi_{\gamma,\delta_x}(D_\gamma(\delta_x P,\nu))\right).\end{aligned}$$

For (B), we first note that

$$\begin{aligned}\left|\sum_{k=0}^{\infty}k\gamma^k\left(\delta_x P^k-\pi\right)f\right| &=\left|\sum_{k=0}^{\infty}k\gamma^k\left((\delta_x P^k-\pi)f\right)\right|\\ &\le\sum_{k=0}^{\infty}k\gamma^k C\kappa^k\|f\|_{\mathrm{Lip}}W(\delta_x,\pi)=C\frac{\gamma\kappa}{(1-\gamma\kappa)^2}\|f\|_{\mathrm{Lip}}W(\delta_x,\pi).\end{aligned}$$

Similarly,

$$\left|\sum_{k=0}^{\infty}\gamma^k\left(\delta_x P^k-\pi\right)f\right|\le\sum_{k=0}^{\infty}\gamma^k C\kappa^k\|f\|_{\mathrm{Lip}}W(\delta_x,\pi)=C\frac{1}{1-\gamma\kappa}\|f\|_{\mathrm{Lip}}\left(W(\delta_x,\pi)\right).$$

Then,

$$\begin{aligned}&\left|2(1-\gamma)\left(\sum_{k=0}^{\infty}\gamma^k\delta_x P^k(f-\pi f)\right)\left(\sum_{k=0}^{\infty}k\gamma^k\delta_x P^k(f-\pi f)\right)\right|\\ \le&2C^2\frac{(1-\gamma)}{(1-\gamma\kappa)^3}\|f\|^2_{\mathrm{Lip}}\left(W(\delta_x,\pi)\right)^2\end{aligned}$$

Putting the bounds for (A) and (B) together, we have

$$\begin{aligned}
&\mathrm{Var}_x\left(\sum_{k=0}^{\tau_\gamma-1}[f(X_k)-\pi_\gamma f]\right) \\
\leq & 2C^3\|f\|_{\mathrm{Lip}}^2\frac{1}{(1-\gamma)(1-\gamma\kappa)^2}\left(\gamma\pi_{\gamma,\delta_x}\mathrm{V}_P+(1-\gamma)\mathrm{V}_\nu+\pi_{\gamma,\delta_x}(D_\gamma(\delta_x P,\nu))\right) \\
&+2C^2\frac{(1-\gamma)}{(1-\gamma\kappa)^3}\|f\|_{\mathrm{Lip}}^2\left(W(\delta_x,\pi)\right)^2.
\end{aligned}$$

□

B.7 Proof of Proposition 3

Proof. We first note that

$$\begin{aligned}
\mathbb{E}_x[\tau\wedge\tau_\gamma] &= \sum_{t=1}^\infty \mathbb{P}_x(\tau\wedge\tau_\gamma\geq t) \\
&= \sum_{t=1}^\infty \gamma^{t-1}\sum_{k=t}^\infty \mathbb{P}_x(\tau=k) \quad \text{as } \tau \text{ and } \tau_\gamma \text{ are independent} \\
&= \sum_{k=1}^\infty\left(\sum_{t=0}^{k-1}\gamma^t\right)\mathbb{P}_x(\tau=k) \quad \text{by Fubini's theorem} \\
&= \sum_{t=1}^\infty \frac{1-\gamma^k}{1-\gamma}\mathbb{P}_x(\tau=k) = \frac{1-\mathbb{E}_x[\gamma^\tau]}{1-\gamma}
\end{aligned}$$

We next establish the expression for the moment generating function. Note that both the mean and the second moment can be derived by taking the (first and second) derivative of the moment generating function and evaluate the corresponding derivatives at $\lambda=0$. First, note that

$$\begin{aligned}
\sum_{t=1}^\infty e^{\lambda t}\mathbb{P}_x(\tau\wedge\tau_\gamma\geq t) &= \sum_{k=1}^\infty\left(\sum_{t=0}^k e^{\lambda t}\right)\mathbb{P}_x(\tau\wedge\tau_\gamma=k) \\
&= \sum_{k=1}^\infty \frac{e^\lambda(e^{\lambda k}-1)}{e^\lambda-1}\mathbb{P}_x(\tau\wedge\tau_\gamma=k) \\
&= \frac{e^\lambda}{e^\lambda-1}\left(\mathbb{E}_x\left[e^{\lambda(\tau\wedge\tau_\gamma)}\right]-1\right)
\end{aligned}$$

Thus,

$$\mathbb{E}_x[e^{\lambda(\tau\wedge\tau_\gamma)}] = \frac{e^\lambda-1}{e^\lambda}\sum_{t=0}^\infty e^{\lambda t}\mathbb{P}_x(\tau\wedge\tau_\gamma\geq t)+1.$$

Next, note that when $\lambda+\log\gamma\neq 0$,

$$\begin{aligned}
\sum_{t=1}^\infty e^{\lambda t}\mathbb{P}_x(\tau\wedge\tau_\gamma\geq t) &= \sum_{t=1}^\infty \gamma^{t-1}e^{\lambda t}\mathbb{P}_x(\tau\geq t) \\
&= \frac{1}{\gamma}\sum_{t=1}^\infty e^{(\lambda+\log\gamma)t}\mathbb{P}_x(\tau\geq t) \\
&= \frac{1}{\gamma}\frac{e^{\lambda+\log\gamma}}{e^{\lambda+\log\gamma}-1}\left(\mathbb{E}_x\left[e^{(\lambda+\log\gamma)\tau}\right]-1\right)
\end{aligned}$$

Then,

$$\mathbb{E}_x[e^{\lambda(\tau\wedge\tau_\gamma)}] = \frac{e^\lambda - 1}{\gamma e^\lambda - 1}\left(\mathbb{E}_x\left[e^{(\lambda+\log\gamma)\tau}\right] - 1\right) + 1$$

When $\lambda + \log\gamma \neq 0$,

$$\sum_{t=1}^{\infty} e^{\lambda t}\mathbb{P}_x(\tau\wedge\tau_\gamma \geq t) = \frac{1}{\gamma}\sum_{t=1}^{\infty}\mathbb{P}_x(\tau \geq t) = \frac{1}{\gamma}\mathbb{E}_x[\tau].$$

Then

$$\mathbb{E}_x[e^{\lambda(\tau\wedge\tau_\gamma)}] = \frac{1-\gamma}{\gamma}\mathbb{E}_x[\tau] + 1.$$

Lastly, for second moment we have:

$$\mathbb{E}_x[(\tau\wedge\tau_\gamma)^2] = \frac{\partial^2}{\partial\lambda^2}\mathbb{E}_x[e^{\lambda(\tau\wedge\tau_\gamma)}]\Big|_{\lambda=0} = \frac{(1-\mathbb{E}[\gamma^\tau])(1+\gamma)}{(1-\gamma)^2} - 2\frac{\mathbb{E}_x[\tau\gamma^\tau]}{1-\gamma}$$

□

C Discounting for the $M/M/1$ queue

In this section, we describe the stationary distribution, the solution to the Poisson equation, and the asymptotic variance of the discounted Markov chain P_γ for the $M/M/1$ queue with $f(x) = x$ and $\nu = \delta_0$. Note that the transient and steady-state distribution of the chain have been fully derived in [Kumar and Arivudainambi, 2000] and [Crescenzo et al., 2003]. However, to the best of our knowledge, the expression for the asymptotic variance is new. For completeness, we include derivations of all the results mentioned.

First, we can characterize the steady-state distribution and moments of the discounted chain. As it turns out, the stationary distribution of the discounted chain remains a geometric distribution with an altered rate. Note here, we normalize $\lambda + \mu = 1$.

Proposition 4. *Let π_γ be the stationary distribution of P_γ, and let η_γ and V_{π_γ} be the steady state mean and variance:*

$$\begin{aligned}
\pi_\gamma(n) &= \frac{2(1-\gamma)}{1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right)^n \\
\eta_\gamma &= \frac{-1+2\gamma\lambda+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)} \\
V_{\pi_\gamma} &= \frac{\gamma\left(1-(\mu-\lambda)\sqrt{1-4\gamma^2\lambda\mu}-4\gamma\lambda\mu\right)}{2(1-\gamma)^2}
\end{aligned}$$

Consider the Poisson equation for the discounted Markov chain:

$$(P_\gamma - I)h_\gamma = -x + \eta_\gamma$$

We can directly solve for h_γ, and in this case we compute the solution such that $h_\gamma(0) = 0$. Using the solution we can then compute the asymptotic variance of discounted estimator as $\sigma^2_\gamma(f) = \pi_\gamma(h^2_\gamma) - \pi_\gamma((Ph_\gamma)^2)$ [Asmussen, 2008, Theorem 1.7.2]

Proposition 5. *The solution to the Poisson equation and the asymptotic variance of the discounted Markov* *chain are:*

$$
\begin{aligned}
h_\gamma(x) =& \frac{x}{1-\gamma} + \left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right)\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right)^x \\
&- \left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right),
\end{aligned}
$$

$$
\begin{aligned}
\sigma_\gamma^2 =& \frac{32\gamma^2\mu^2}{\left(\sqrt{1-4\gamma^2\lambda\mu}-2\gamma\mu+1\right)\left(\sqrt{1-4\gamma^2\lambda\mu}+2\gamma\mu-1\right)^3\left(1-4\gamma^2\lambda\mu-\sqrt{1-4\gamma^2\lambda\mu}\right)^2} \\
&\times\Big[-4\gamma^5\lambda^2\mu^2\left(\sqrt{1-4\gamma^2\lambda\mu}-2\mu-2\right)-\sqrt{1-4\gamma^2\lambda\mu}+1 \\
&+\gamma^3\lambda\mu\left(2\mu\left(3\sqrt{1-4\gamma^2\lambda\mu}-5\right)+\sqrt{1-4\gamma^2\lambda\mu}+1\right) \\
&+\gamma^2\lambda\mu\left(\sqrt{1-4\gamma^2\lambda\mu}-3\right)-\gamma(\mu-\lambda)\left(\sqrt{1-4\gamma^2\lambda\mu}-1\right)\Big].
\end{aligned}
$$

We can then solve for the discount factor that minimizes $\mathrm{MSE}(\gamma) = |\eta - \eta_\gamma|^2 + \sigma_\gamma^2/n$. There is no explicit closed-form solution for the optimal discount factor, so we use a numerical solver. Figure 5 plots $n^{-1}(1-\gamma^*)^{-1}$ for the optimal discount factor as n grows large. We can observe that this converges as n grows large and that the limit is similar across queues with different traffic intensities. This confirms the $\frac{1}{1-\gamma^*} \sim n$ scaling that was predicted for the optimal discount factor. This also confirms the fact that asymptotically, the ergodicity of the chain has a small effect on the optimal discount factor.

C.1 Proof of Proposition 4

We can verify that the probability mass function $\pi_\gamma(n)$ satisfies the steady-state equations:

$$
\pi_\gamma(n) = \frac{2(1-\gamma)}{1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right)^n
$$

First, we can observe that for any $n > 0$:

$$
\begin{aligned}
&\gamma\lambda\pi_\gamma(n-1)+\gamma\mu\pi_\gamma(n+1) \\
&= \frac{2(1-\gamma)}{1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}\left(\gamma\lambda\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right)^{n-1}+\gamma\mu\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right)^{n+1}\right) \\
&= \frac{2(1-\gamma)}{1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right)\left(\frac{2\gamma^2\lambda\mu}{1-\sqrt{1-4\gamma^2\lambda\mu}}+\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2}\right) \\
&= \frac{2(1-\gamma)}{1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right)\left(\frac{4\gamma^2\lambda\mu+(1-\sqrt{1-4\gamma^2\lambda\mu})^2}{2(1-\sqrt{1-4\gamma^2\lambda\mu})}\right) \\
&= \frac{2(1-\gamma)}{1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right) = \pi_\gamma(n)
\end{aligned}
$$

For $\pi_\gamma(0)$ we have:

$$
\begin{aligned}
&\gamma\mu(\pi_\gamma(1)+\pi_\gamma(0))+(1-\gamma)\\
=&\gamma\mu\left(\frac{2(1-\gamma)}{1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}\right)\left(1+\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right)+(1-\gamma)\\
=&\gamma\mu\left(\frac{2(1-\gamma)}{1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}\right)\left(\frac{1+2\gamma\mu-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}\right)\\
&+\frac{2\gamma\mu(1-\gamma)(1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu})}{2\gamma\mu(1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu})}\\
=&\frac{(1-\gamma)(1+2\gamma\mu-\sqrt{1-4\gamma^2\lambda\mu})}{(1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu})}+\frac{(1-\gamma)(1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu})}{(1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu})}\\
=&\frac{2(1-\gamma)}{(1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu})}=\pi_\gamma(0).
\end{aligned}
$$

Thus $\pi_\gamma(n)$ solves the stationary equations and is a Geometric distribution with the observation that since $\lambda+\mu=1$,

$$
1-\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}=\frac{2\gamma\mu-1+\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}=\frac{2(1-\gamma)}{(1-2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu})}.
$$

The stationary mean and variance arrive as a result of the mean and variance for Geometric random variables:

$$
\begin{aligned}
\eta_\gamma &= \frac{\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}}{1-\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}}=\frac{-1+2\gamma\lambda+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)},\\
\mathrm{V}_{\pi_\gamma} &= \frac{\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu}}{(1-\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\mu})^2}=\frac{\gamma\left(1-(\mu-\lambda)\sqrt{1-4\gamma^2\lambda\mu}-4\gamma\lambda\mu\right)}{2(1-\gamma)^2}.
\end{aligned}
$$

C.2 Proof of Proposition 5

The Poisson equation is $(P_\gamma - I)h_\gamma = -f + \eta_\gamma$ or more explicitly:

$$
\begin{aligned}
\gamma(\mu h_\gamma(x-1)+\lambda h_\gamma(x+1))+(1-\gamma)h_\gamma(0)-h_\gamma(x)&=-x+\eta_\gamma\\
\gamma(\mu h_\gamma(0)+\lambda h_\gamma(1))+(1-\gamma)h_\gamma(0)-h_\gamma(0)&=-x+\eta_\gamma
\end{aligned}
$$

We can verify that the following function solves the Poisson equation:

$$
\begin{aligned}
h_\gamma(x)=&\frac{x}{1-\gamma}+\left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right)\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right)^x\\
&-\left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right).
\end{aligned}
$$

First, we can see that $h_\gamma(0) = 0$ so:

$$
\begin{aligned}
\gamma\lambda h_\gamma(1) =& \gamma\lambda\left[\frac{1}{1-\gamma} - \left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right)\right.\\
&\left.+\left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right)\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right)\right]\\
=&\gamma\lambda\frac{1}{1-\gamma} + \gamma\lambda\frac{2\lambda\gamma - 4\gamma^2\lambda\mu - 2\gamma\lambda\sqrt{1-4\gamma^2\lambda\mu}}{4\gamma\lambda(1-\gamma)^2}\\
&+\gamma\lambda\frac{(-1+2\gamma\mu+2\sqrt{1-4\gamma^2\lambda\mu} - 2\gamma\mu\sqrt{1-4\gamma^2\lambda\mu} - (1-4\gamma^2\lambda\mu)}{4\gamma\lambda(1-\gamma)^2}\\
=&\gamma\lambda\left(\frac{1}{1-\gamma} + \frac{-2+2\gamma(\mu+\lambda)+(2-2\gamma\lambda-2\gamma\mu)\sqrt{1-4\gamma^2\lambda\mu}}{4\gamma\lambda(1-\gamma)^2}\right)\\
=&\gamma\lambda\left(\frac{4\gamma\lambda(1-\gamma) - 2(1-\gamma) + 2(1-\gamma)\sqrt{1-4\gamma^2\lambda\mu}}{4\gamma\lambda(1-\gamma)^2}\right)\\
=&\left(\frac{-1+2\gamma\lambda+\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda(1-\gamma)^2}\right).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
&\gamma\mu h_\gamma(x-1) + \gamma\lambda h_\gamma(x+1) - h_\gamma(x)\\
=& -x + \gamma\frac{\lambda-\mu}{1-\gamma} + (1-\gamma)\left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right)\\
&+\gamma\left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right)\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right)^x\\
&\times\left(\mu\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right)^{-1} + \lambda\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right) - \frac{1}{\gamma}\right)\\
=& -x - \gamma\frac{\mu-\lambda}{1-\gamma} + (1-\gamma)\left(\frac{-1+2\gamma\mu+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right)\\
=& -x + \left(\frac{-1+2\gamma\lambda+\sqrt{1-4\gamma^2\lambda\mu}}{2(1-\gamma)^2}\right)
\end{aligned}
$$

Since:

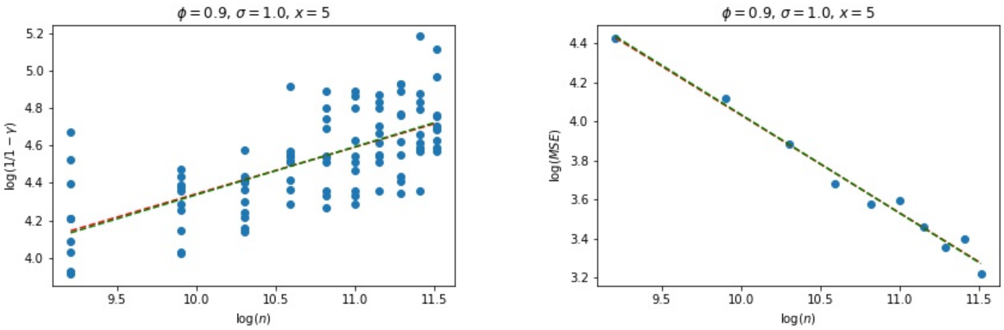
$$
\begin{aligned}
&\mu\left(\frac{2\gamma\lambda}{1-\sqrt{1-4\gamma^2\lambda\mu}}\right)+\lambda\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right)-\frac{1}{\gamma}\\
=&\frac{\gamma\mu(4\gamma^2\lambda^2)+\gamma\lambda(1-\sqrt{1-4\gamma^2\lambda\mu})^2-2\gamma\lambda(1-\sqrt{1-4\gamma^2\lambda\mu})}{2\gamma\lambda(1-\sqrt{1-4\gamma^2\lambda\mu})}\\
=&\frac{4\gamma^3\lambda^2\mu+\gamma\lambda-2\gamma\lambda\sqrt{1-4\gamma^2\lambda\mu}+\gamma\lambda(1-4\gamma^2\lambda\mu)-2\gamma\lambda+2\gamma\lambda\sqrt{1-4\gamma^2\lambda\mu}}{\gamma2\gamma\lambda(1-\sqrt{1-4\gamma^2\lambda\mu})}=0
\end{aligned}
$$

Thus $h_\gamma(x)$ solves the Poisson equation. Since it solves the Poisson equation, we can calculate the asymptotic variance as:

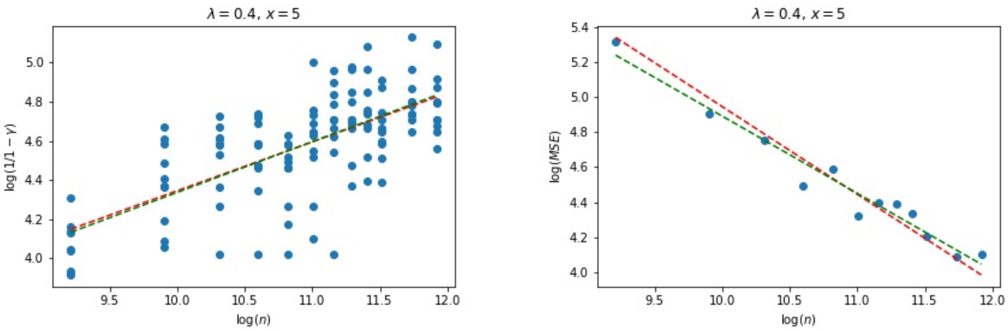
$$
\begin{aligned}
\sigma^2_\gamma(f) =&\pi_\gamma(h_\gamma^2)-\pi_\gamma((P_\gamma h_\gamma)^2)\\
=&\pi_\gamma(h_\gamma^2)-\pi_\gamma((h_\gamma-(x-\eta_\gamma))^2)\\
=&\pi_\gamma(2h_\gamma(x-\eta_\gamma))-\pi_\gamma((x-\eta_\gamma)^2)\\
=&2\pi_\gamma\left((x-\eta_\gamma)\left[\frac{x-\alpha_\gamma}{1-\gamma}+\frac{\alpha_\gamma}{1-\gamma}\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right)^x\right]\right)-V_{\pi_\gamma}\\
=&2\pi_\gamma\left(\frac{(x-\eta_\gamma)(x-\alpha_\gamma)}{1-\gamma}+\frac{\alpha_\gamma(x-\eta_\gamma)}{1-\gamma}\left(\frac{1-\sqrt{1-4\gamma^2\lambda\mu}}{2\gamma\lambda}\right)^x\right)-V_{\pi_\gamma}\\
=&\frac{16\gamma^2\mu^2\left(1-\sqrt{1-4\gamma^2\lambda\mu}\right)}{\left(\sqrt{1-4\gamma^2\lambda\mu}-2\gamma\mu+1\right)\left(\sqrt{1-4\gamma^2\lambda\mu}+2\gamma\mu-1\right)^3}\\
&+\frac{8\gamma^3\lambda\mu^2\left(2\gamma^2(\mu-1)\mu-\gamma(1-\mu)\left(1-\sqrt{1-4\gamma^2\lambda\mu}\right)+1-\sqrt{1-4\gamma^2\lambda\mu}\right)}{(1-\gamma)\left(\sqrt{1-4\gamma^2\lambda\mu}-2\gamma\mu+1\right)\left(4\gamma^2\lambda\mu+\sqrt{4g^2(\mu-1)\mu+1}-1\right)^2}\\
&-\frac{\gamma\left(1-(\mu-\lambda)\sqrt{1-4\gamma^2\lambda\mu}-4\gamma\lambda\mu\right)}{2(1-\gamma)^2}\\
=&\frac{32\gamma^2\mu^2}{\left(\sqrt{1-4\gamma^2\lambda\mu}-2\gamma\mu+1\right)\left(\sqrt{1-4\gamma^2\lambda\mu}+2\gamma\mu-1\right)^3\left(1-4\gamma^2\lambda\mu-\sqrt{1-4\gamma^2\lambda\mu}\right)^2}\\
&\times\left[-4\gamma^5\lambda^2\mu^2\left(\sqrt{1-4\gamma^2\lambda\mu}-2\mu-2\right)\right.\\
&+\gamma^3\lambda\mu\left(2\mu\left(3\sqrt{1-4\gamma^2\lambda\mu}-5\right)+\sqrt{1-4\gamma^2\lambda\mu}+1\right)\\
&\left.+\gamma^2\lambda\mu\left(\sqrt{1-4\gamma^2\lambda\mu}-3\right)-\gamma(\mu-\lambda)\left(\sqrt{1-4\gamma^2\lambda\mu}-1\right)-\sqrt{1-4\gamma^2\lambda\mu}+1\right]
\end{aligned}
$$

Figure 4: $\log \frac{1}{1-\gamma^*}$ (left) and $\log MSE(\gamma^*)$ (right) with $\log n$

$$AR(1) : \phi = 0.9, \sigma = 1, x = 5$$



$$M/M/1 : \lambda = 0.4, x = 5$$



Binomial:$a = 0.5, N = 10, x = 5$

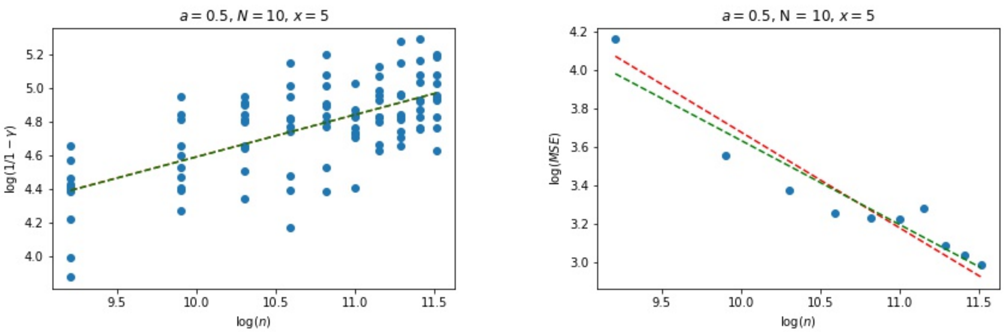


Figure 5: $\frac{1}{n(1-\gamma^*)}$ of the discounted estimator for the $M/M/1$ queue