Stein's Method for Steady-State Approximations: Error Bounds and Engineering Solutions

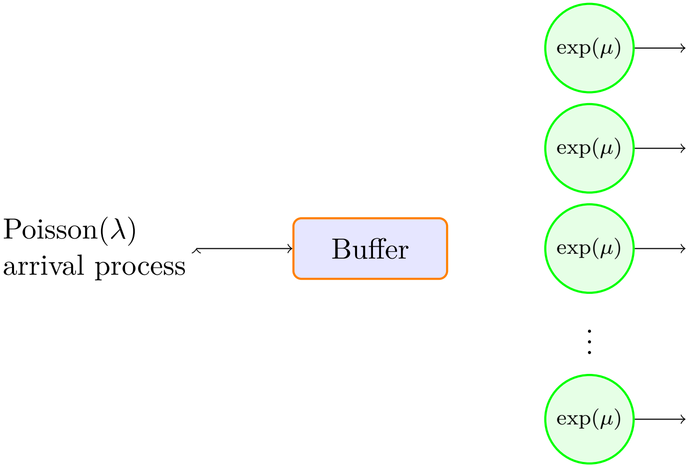
Jim Dai

Joint work with Anton Braverman and Jiekun Feng

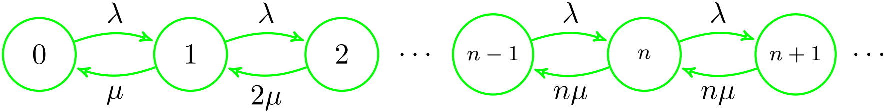
Cornell University

Workshop on Congestion Games
Institute for Mathematical Sciences, Singapore
December 16, 2015

$M/M/n$ queue – Erlang-C model



Markov chain and its transitions



- $X = \{X(t), t \geq 0\}$ is a CTMC on $\mathbb{Z}_+ = \{0, 1, \ldots, \}$.
- Generator

$$G_X f(i) = \lambda\Big(f(i+1) - f(i)\Big) + \min(i,n)\mu\Big(f(i-1) - f(i)\Big)$$

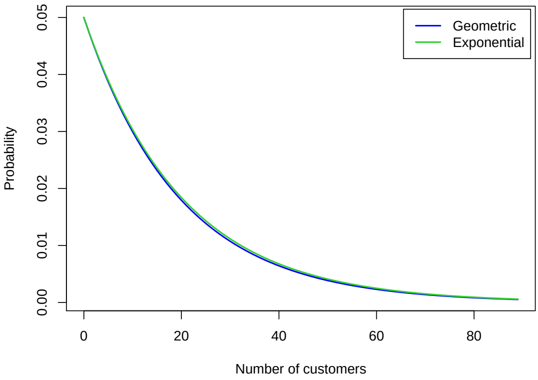
  for $i \in \mathbb{Z}_+$
- Assume

$$R \equiv \lambda/\mu < n.$$

- Random variable $X(\infty)$ has the stationary distribution.

$M/M/1$ queue: $R = .95$

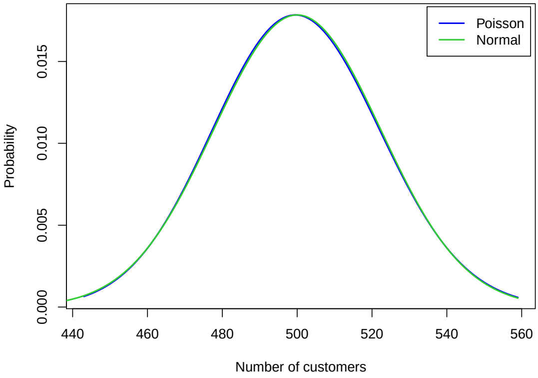
- $X(\infty)$ is geometric: $\mathbb{P}\{X(\infty) = i\} = (1-R)R^i$, $i \in \mathbb{Z}_+$



- Continuous random variable $Y(\infty) \sim \exp(.05)$

$M/M/\infty$ queue: $R = 500$

- $X(\infty)$ is Poisson(500).



- Continuous random variable $Y(\infty) \sim N(500, 500)$

$M/M/n$: distribution of $X(\infty)$?

- Engineering solution: identify a continuous random variable $Y(\infty)$
- Error bound: bound distance between $X(\infty)$ and $Y(\infty)$
- Stein's method: is able to achieve both
- Known results:
  - $M/M/n+M$ (Braverman, Dai & Feng '15)
  - $M/Ph/n+M$ (Braverman, & Dai '15)
  - A discrete time queue arising from hospital patient flow (Dai & Shi '15)
  - Many systems having mean field limits (Ying '15)
  - ... (your papers)

Outline

- Error Bounds: Erlang-C model (Braverman-D-Feng 2015)
- Stein's Method: proof framework and solution technique
- Engineering Solution: High order approximations (Braverman & Dai 2015)

Erlang-C Model – $M/M/n$ Queue

- Steady-state number of customers in system $X(\infty)$. Define

$$\tilde{X}(\infty) = \frac{X(\infty) - R}{\sqrt{R}}.$$

- $\tilde{X}(\infty)$ lives on grid $\{x = \delta(i - R), i \in \mathbb{Z}_+\}$, $\delta = 1/\sqrt{R}$.

Theorem 1 (Braverman-D-Feng 2015)

For all $n \geq 1$, $\lambda > 0$, $\mu > 0$ with $1 \leq R < n$,

$$\sup_{h \in \text{Lip}(1)} \left| \mathbb{E}h(\tilde{X}(\infty)) - \mathbb{E}h(Y(\infty)) \right| \leq \frac{157}{\sqrt{R}},$$

where

$$\text{Lip}(1) = \{h : \mathbb{R} \to \mathbb{R}, |h(x) - h(y)| \leq |x - y|\}.$$

The continuous random variable $Y(\infty)$

- $Y(\infty)$ has density

$$\kappa \exp\Big(\frac{1}{\mu}\int_0^x b(y)dy\Big), \tag{1}$$

where

$$b(x) = \begin{cases} -\mu x, & x \le |\zeta|, \\ \mu\zeta, & x \ge |\zeta| \end{cases} \tag{2}$$

and

$$\zeta = \frac{R-n}{\sqrt{R}} < 0.$$

Discussions

Corollary

For all $n \geq 1$, $\lambda > 0$ and $\mu > 0$ with $1 \leq \lambda/\mu < n$,

$$\left| \mathbb{E}X(\infty) - R - \sqrt{R}\mathbb{E}Y(\infty) \right| \leq 157.$$

- *Not* a limit theorem
- For $\mu = 1$.

$n = 5$			$n = 500$		
λ	$\mathbb{E}X(\infty)$	Error	λ	$\mathbb{E}X(\infty)$	Error
3	3.35	0.10	300	300.00	6×10^{-14}
4	6.22	0.20	400	400.00	2×10^{-6}
4.9	51.47	0.28	490	516.79	0.24
4.95	101.48	0.29	495	569.15	0.28
4.99	501.49	0.29	499	970.89	0.32

- Universal

Universal approximation

- $Y(\infty)$ depends on system parameters λ, n and μ:

$$Y(\infty) \sim f(x) \sim \begin{cases} N(0,1) & \text{if } x < |\zeta|, \\ \text{Exponential}(|\zeta|) & \text{if } x \geq |\zeta|. \end{cases}$$

  $X(\infty) < n$ behaves like a normal, and $X(\infty) \geq n$ behaves like an exponential.
- Gurvich, Huang, Mandelbaum (2014), *Mathematics of Operations Research*
- Glynn, Ward (2003), *Queueing Systems*

Approximating Diffusion Process

A diffusion process $Y = \{Y(t) \in \mathbb{R},\ t \geq 0\}$ satisfies the stochastic differential equation

$$Y(t) = Y(0) + \int_0^t b(Y(s))ds + \int_0^t \sigma(Y(s))dB(s).$$

- $Y(\infty)$ in (1) is the stationary distribution of diffusion process with drift $b(x)$ in (2) and diffusion coefficient

$$\sigma^2(x) \equiv 2\mu \quad \text{for now.}$$

- Identifying $Y(\infty)$ is equivalent to identifying the generator of a diffusion process

$$Gf(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x) \quad \text{for } f \in C^2(\mathbb{R}).$$

A Proof Outline for Theorem 1

Stein Framework for Proofs

- Poisson (Stein) equation and gradient bounds
- Basic Adjoint Relationship (BAR) and the generator coupling
- State space collapse (SSC)
- Moment bounds

- Stein ('72, '86)
- Louis Chen (75')
- Andrew Barbour (88')
- Chen, Goldstein, and Shao (2011)

Gurvich (2014), Diffusion models and steady-state approximations for exponentially ergodic Markovian queues, *Annals of Applied Probability.*

Poisson Equation

- Generator of diffusion process $Y = \{Y(t), t \geq 0\}$

$$G_Y f(x) = \mu f''(x) + b(x) f'(x)$$

- Given an $h \in \text{Lip}(1)$, solve $f = f_h$ from the Poisson equation

$$G_Y f(x) = h(x) - \mathbb{E}[h(Y(\infty))], \quad x \in \mathbb{R}.$$

- On each sample path of $\tilde{X}(\infty)$

$$G_Y f(\tilde{X}(\infty)) = h(\tilde{X}(\infty)) - \mathbb{E}[h(Y(\infty))].$$

- Key identity

$$\mathbb{E}[h(\tilde{X}(\infty))] - \mathbb{E}[h(Y(\infty))] = \mathbb{E}[G_Y f(\tilde{X}(\infty))]$$

Basic Adjoint Relationship

- Suppose $X = \{X(t), t \geq 0\}$ is a CTMC on $S = \{1, 2, 3\}$ with generator

$$G = \begin{pmatrix} -3 & 2 & 1 \\ 1 & -2 & 1 \\ 1 & 0 & -1 \end{pmatrix}.$$

- Unique $\pi = (\pi(1), \pi(2), \pi(3))$ that satisfies

$$\pi G \begin{pmatrix} f(1) \\ f(2) \\ f(3) \end{pmatrix} = 0 \quad \text{for each } f = \begin{pmatrix} f(1) \\ f(2) \\ f(3) \end{pmatrix} \in \mathbb{R}^3.$$

- Alternatively,

$$\mathbb{E}\big[Gf(X(\infty))\big] = 0 \quad \text{for each } f \in \mathbb{R}^3. \tag{3}$$

- When the state space is infinite, Glynn and Zeevi (2008, *Kurtz Festschrift*) provides conditions on f for (3) to hold.

Generator Coupling

From the Stein equation,

$$\begin{aligned}\mathbb{E}[h(\tilde{X}(\infty))] - \mathbb{E}[h(Y(\infty))] &= \mathbb{E}[G_Y f_h(\tilde{X}(\infty))] \\ &= \mathbb{E}[G_Y f_h(\tilde{X}(\infty))] - \mathbb{E}[G_{\tilde{X}} f_h(\tilde{X}(\infty))] \\ &= \mathbb{E}\Big[G_Y f_h(\tilde{X}(\infty)) - G_{\tilde{X}} f_h(\tilde{X}(\infty))\Big].\end{aligned}$$

- $\tilde{X}(\infty)$ lives on grid $\{x = \delta(i - R), i \in \mathbb{Z}_+\}$, $\delta = 1/\sqrt{R}$.
- The generator of birth-death process $\tilde{X}$ is

$$G_{\tilde{X}} f_h(x) = \lambda\Big(f_h(x+\delta) - f_h(x)\Big) + \mu(i \wedge n)\Big(f_h(x-\delta) - f_h(x)\Big).$$

Taylor Expansion

- To bound

$$\mathbb{E}\Big[G_Y f_h(\tilde{X}(\infty)) - G_{\tilde{X}} f_h(\tilde{X}(\infty))\Big]$$

one bounds

$$\big|G_Y f_h(x) - G_{\tilde{X}} f_h(x)\big| \qquad \text{for } x = \delta(i - R) \text{ with } i \in \mathbb{Z}_+.$$

- Conduct Taylor expansion

$$\begin{aligned} G_{\tilde{X}} f_h(x) &= f_h'(x)\delta\Big(\lambda - \mu(i \wedge n)\Big) + \frac{1}{2} f_h''(x)\delta^2\Big(\lambda + \mu(i \wedge n)\Big) \\ &\quad + \text{ higher order term} \\ &= f_h'(x)\delta\Big(\lambda - \mu(i \wedge n)\Big) + \frac{1}{2} f_h''(x)\delta^2(2\lambda) \\ &\quad - \frac{1}{2} f_h''(x)\delta^2\Big(\lambda - \mu(i \wedge n)\Big) + \text{ higher order term} \end{aligned}$$

Approximation

- One can check that $\delta^2\lambda = \mu$ and

$$\delta\Big(\lambda - \mu(i \wedge n)\Big) = \mu\Big((x+\zeta)^- + \zeta\Big) = b(x).$$

- From the CTMC generator we extract

$$\begin{aligned} G_{\tilde{X}} f_h(x) = &\ f_h'(x)b(x) + \mu f_h''(x) \\ &- \frac{1}{2}\delta f_h''(x)b(x) + \text{higher order terms} \\ = &\ G_Y f_h(x) - \frac{1}{2}\delta f_h''(x)b(x) + \text{higher order terms.} \end{aligned}$$

- Typical error term

$$\delta \mathbb{E}\big|f_h''(\tilde{X}(\infty))b(\tilde{X}(\infty))\big|, \qquad |b(x)| \leq \mu\,|x|$$

Gradient Bounds

For all λ, n, and μ satisfying $n \geq 1$ and $0 < \lambda < n\mu$,

$$|f_h''(x)| \leq \begin{cases} \frac{18}{\mu}(1 + 1/|\zeta|), & x \leq -\zeta, \\ \frac{1}{\mu|\zeta|}, & x \geq -\zeta, \end{cases}$$

$$|f_h'''(x)| \leq \begin{cases} \frac{1}{\mu}(23 + 13/|\zeta|), & x < -\zeta, \\ 2/\mu, & x > -\zeta. \end{cases}$$

Recall

$$\zeta = \frac{R - n}{\sqrt{R}}.$$

Moment Bounds

For all λ, n, and μ satisfying $n \geq 1$ and $0 < R < n$,

$$\mathbb{E}\Big[(\tilde{X}(\infty))^2 1(\tilde{X}(\infty) \leq -\zeta)\Big] \leq \frac{4}{3} + \frac{\delta^2}{3} + \frac{\delta}{3},$$

$$\mathbb{E}\Big[\big|\tilde{X}(\infty) 1(\tilde{X}(\infty) \leq -\zeta)\big|\Big] \leq \sqrt{\frac{4}{3} + \frac{\delta^2}{3} + \frac{\delta}{3}},$$

$$\mathbb{E}\Big[\big|\tilde{X}(\infty) 1(\tilde{X}(\infty) \leq -\zeta)\big|\Big] \leq 2\,|\zeta|$$

$$\mathbb{E}\Big[\big|\tilde{X}(\infty) 1(\tilde{X}(\infty) \geq -\zeta)\big|\Big] \leq \frac{1}{|\zeta|} + \frac{\delta^2}{4\,|\zeta|} + \frac{\delta}{2},$$

$$\mathbb{P}(\tilde{X}(\infty) \leq -\zeta) \leq (2 + \delta)\,|\zeta|\,.$$

Recall

$$\delta^2 = 1/R.$$

Moment Version of Theorem 1

Theorem 1 (b)

Recall that $\zeta = (R-n)/\sqrt{R}$. There exists a constant $C = C(m)$ such that for all $n \geq 1$, $\lambda > 0$ and $\mu > 0$ with $1 \leq R < n$,

$$\sup_{x\in\mathbb{R}} \left| \mathbb{E}(\tilde{X}(\infty))^m - \mathbb{E}(Y(\infty))^m \right| \leq \left(1 + \frac{1}{|\zeta|^{m-1}}\right)\frac{C(m)}{\sqrt{R}}.$$

- For $n = 500$, $\mu = 1$.

λ	$\mathbb{E}(\tilde{X}(\infty))^2$	Error	$\mathbb{E}(\tilde{X}(\infty))^{10}$	Error
300	1	4.55×10^{-15}	9.77×10^2	31.58
400	1	5.95×10^{-7}	9.70×10^2	24.44
490	6.96	0.11	7.51×10^9	7.01×10^8
495	31.56	0.27	9.10×10^{12}	4.34×10^{11}
499	9.47×10^2	1.59	1.07×10^{20}	1.03×10^{18}
499.9	9.94×10^4	16.50	1.13×10^{30}	1.09×10^{27}

- *Not* universal

Kolmogorov Metric Version of Theorem 1

Theorem 1 (c)

For all $n \geq 1$, $\lambda > 0$ and $\mu > 0$ with $1 \leq R < n$,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\{\tilde{X}(\infty) \leq x\} - \mathbb{P}\{Y(\infty) \leq x\} \right| \leq \frac{190}{\sqrt{R}}.$$

An $M/M/250$ System

Consider $x = (i - x(\infty))/\sqrt{R}$, where $i = 0, 1, 2, 3, \ldots$

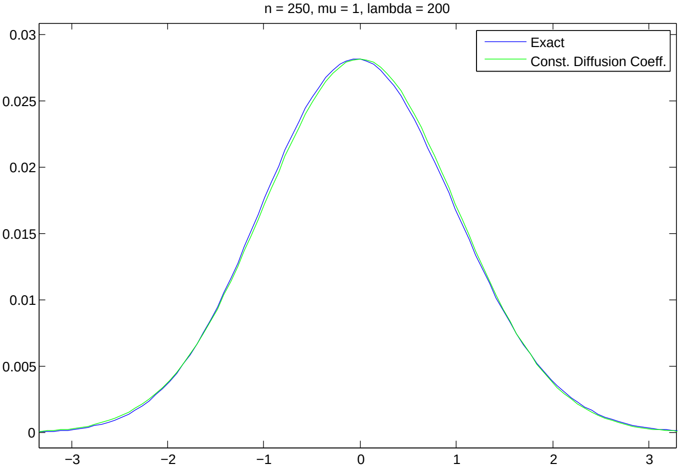


Figure: $P(\tilde{X}(\infty) = x)$, $\mathbb{P}(x - 0.5 \leq Y(\infty) \leq x + 0.5)$

An $M/M/5$ System

- With only 5 servers, diffusion approximation not as good.

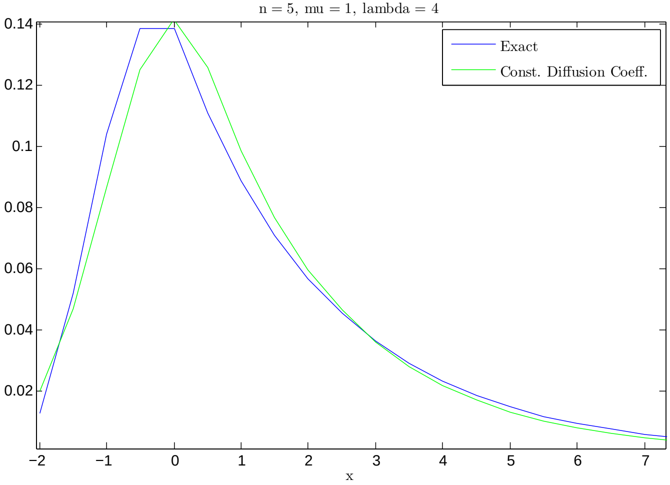


Figure: $P(\tilde{X}(\infty) = x)$, $\mathbb{P}(x - 0.5 \leq Y(\infty) \leq x + 0.5)$

High Order Approximation – $M/M/5$

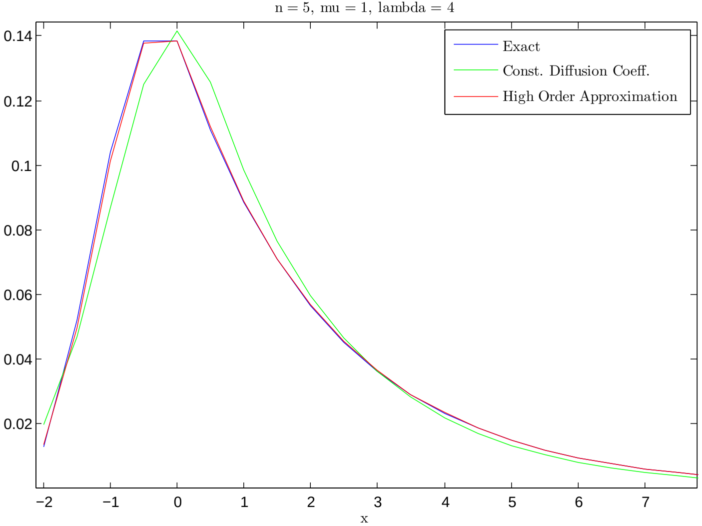
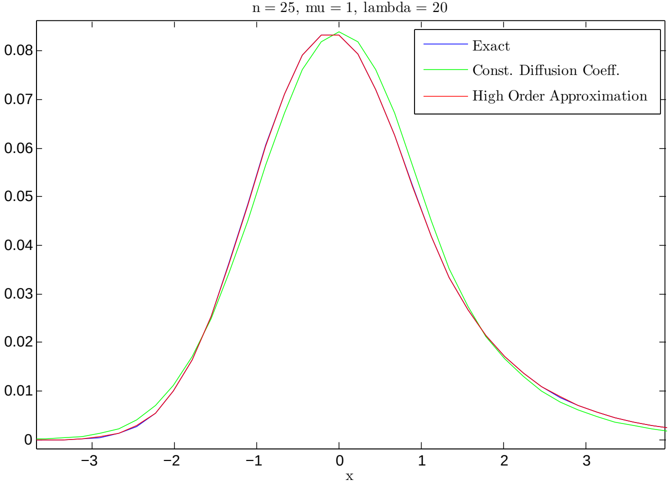


Figure: $\mathbb{P}(x - 0.5 \leq Y_H(\infty) \leq x + 0.5)$

High Order Approximation – $M/M/25$



High order results

	$n = 5$			$n = 500$	
λ	$\mathbb{E}X(\infty)$	Error	λ	$\mathbb{E}X(\infty)$	Error
3	3.35	1.62×10^{-2}	300	300.00	2.86×10^{-13}
4	6.22	2.39×10^{-2}	400	400.00	1.06×10^{-7}
4.9	51.47	2.85×10^{-2}	490	516.79	2.79×10^{-3}
4.95	101.48	2.87×10^{-2}	495	569.15	3.13×10^{-3}
4.99	501.49	2.89×10^{-2}	499	970.89	3.38×10^{-3}

High order results – Higher Moments

- $n = 500, \mu = 1$

λ	$\mathbb{E}(\tilde{X}(\infty))^2$	Error	$\mathbb{E}(\tilde{X}(\infty))^{10}$	Error
300	1	4.63×10^{-14}	9.77×10^{2}	5.39
400	1	2.88×10^{-8}	9.70×10^{2}	4.05
490	6.96	6.20×10^{-4}	7.51×10^{9}	2.78×10^{6}
495	31.56	1.3×10^{-3}	9.10×10^{12}	9.93×10^{8}
499	9.47×10^{2}	6.8×10^{-3}	1.07×10^{20}	1.06×10^{15}
499.9	9.94×10^{4}	6.9×10^{-2}	1.13×10^{30}	8.04×10^{23}

Fixed $\zeta = -1/2$

n	λ	$\mathbb{E}(\tilde{X}(\infty))^2$	Error	High Order Approx. Error
5	4	6.54	1.00	6×10^{-2}
50	46.59	5.84	0.30	5.7×10^{-3}
500	488.94	5.63	0.092	5.6×10^{-4}
5000	4965	5.57	0.029	5.5×10^{-5}

- $\lambda = R$ increases by a factor of 10
- Error decreases by a factor of $\sqrt{10}$
- High order approx. error decreases by a factor of 10.

Deriving High order approximation

- Recall the Taylor expansion

$$G_{\tilde{X}} f_h(x) = f_h'(x)b(x) + \mu f_h''(x) - \frac{1}{2}\delta b(x) f_h''(x) + \text{higher order terms}$$

- Recall the Taylor expansion

$$\begin{aligned} G_{\tilde{X}} f_h(x) &= f_h'(x)\delta\Big(\lambda - \mu(i \wedge n)\Big) + \frac{1}{2} f_h''(x)\delta^2\Big(\lambda + \mu(i \wedge n)\Big) \\ &\quad + \frac{1}{6} f_h'''(x)\delta^3\Big(\lambda - \mu(i \wedge n)\Big) + \text{ fourth order term} \\ &= b(x)f'(x) + \Big(\mu - \delta b(x)/2\Big) f''(x) + \frac{1}{6}\delta^3 f_h'''(x) b(x) \\ &\quad + \text{ fourth order term} \end{aligned}$$

- Use entire second order term and bound

$$\delta \mathbb{E}\Big|f_h'''(\tilde{X}(\infty)) b(\tilde{X}(\infty))\Big|, \qquad |b(x)| \leq \mu\,|x|$$

High Order Approximation

- $Y_H(\infty)$ – corresponds to diffusion process with generator

$$G_{Y_H} f(x) = \frac{1}{2}\sigma^2(x) f''(x) + b(x) f'(x), \quad f \in C^2(\mathbb{R}),$$
$$\sigma^2(x) = \mu + (\mu - \delta b(x)) 1(x \geq -\sqrt{R}) \geq \mu, \quad x \in \mathbb{R}.$$

- Previously, used $\sigma^2(x) = 2\mu$.

Theorem 2 (High Order Approximation, Braverman-D 2015)

$\exists\ C_{W_2} > 0$ (explicit) such that for all $n \geq 1$, $1 \leq R < n$,

$$\sup_{h \in W_2} \left| \mathbb{E} h(\tilde{X}(\infty)) - \mathbb{E} h(Y_H(\infty)) \right| \leq C_{W_2} \frac{1}{R},$$

$$W_2 = \{h : \mathbb{R} \to \mathbb{R},\ |h(x) - h(y)| \leq |x - y|,$$
$$|h'(x) - h'(y)| \leq |x - y|\}.$$

Other Works

- $M/Ph/n + M$ systems with phase-type service time distribution and customer abandonment – Halfin-Whitt or QED regime (Braverman-D 2015)
- Erlang-A system (Braverman-D-Feng 2015)
- A discrete time queue arising from hospital patient flow management (Dai & Shi 15')
- Mean field analysis (Ying 2015)

References

- Braverman, Dai, Stein's method for steady-state diffusion approximations of $M/Ph/n+M$ systems, `http://arxiv.org/abs/1503.00774`
- Braverman, Dai, Feng, Stein's method for steady-state diffusion approximations: an introduction, working paper.
- Gurvich 2014, Diffusion models and steady-state approximations for exponentially ergodic Markovian queues, *Annals of Applied Probability*, Vol. 24, 2527-2559.
- Gurvich, Huang, Mandelbaum, 2014, Excursion-Based Universal Approximations for the Erlang-A Queue in Steady-State, *Mathematics of Operations Research*, Vol. 39(2), 325-373

References

- Ying, On the Rate of Convergence of Mean-Field Models: Stein's Method Meets the Perturbation Theory, `http://arxiv.org/abs/1510.00761`
- J. G. Dai and Pengyi Shi (2015), "A Two-Time-Scale Approach to Time-Varying Queues in Hospital Inpatient Flow Management," Submitted to *Operations Research*.