Online Supplement for "Critically Loaded Time-Varying Multi-Server Queues: Computational Challenges and Approximations"

Young Myoung Ko
Sabre Holdings, 3150 Sabre Drive, Southlake, Texas 76092, USA, YoungMyoung.Ko@sabre.com
Natarajan Gautam
Department of Industrial and Systems Engineering, Texas A&M University, 3131 TAMU, College Station, Texas 77843-3131, USA, gautam@tamu.edu

---

1. Derivation of $g_i^\eta(\cdot,\cdot,\cdot)$'s in Section 5

For a fixed η, suppose $x_1^\eta(t) \sim N(E[x_1^\eta(t)], \sigma_1^\eta(t)^2)$. For $x = (x_1, x_2)'$, we have

$$\begin{aligned} g_3^\eta(t,x) &= \eta E\Big[\mu_t^1\Big(\Big(\frac{x_1^\eta(t)}{\eta} - \frac{E[x_1^\eta(t)]}{\eta} + \frac{x_1}{\eta}\Big) \wedge n_t\Big)\Big] \\ &= E\big[\mu_t^1\big((x_1^\eta(t) - E[x_1^\eta(t)] + x_1) \wedge \eta n_t\big)\big] \\ &= \mu_t^1\Big\{E\big[(x_1^\eta(t) - E[x_1^\eta(t)] + x_1)\mathbb{I}_{x_1^\eta(t) - E[x_1^\eta(t)] + x_1 \le \eta n_t}\big] \\ &\quad + \eta n_t Pr[x_1^\eta(t) - E[x_1^\eta(t)] + x_1 > \eta n_t]\Big\}. \end{aligned}$$

Let $y_1(t) = x_1^\eta(t) - E[x_1^\eta(t)] + x_1$. Then,

$$\begin{aligned} g_3^\eta(t,x) &= \mu_t^1\Bigg[\int_{-\infty}^{\eta n_t} \frac{y_1(t)}{\sqrt{2\pi}\sigma_1^\eta(t)} \exp\Big(-\frac{(y_1(t)-x_1)^2}{2\sigma_1^\eta(t)^2}\Big) dy_1(t) + \eta n_t Pr[y_1(t) > \eta n_t]\Bigg] \\ &= \mu_t^1\Bigg[\frac{-\sigma_1^\eta(t)}{\sqrt{2\pi}} \int_{-\infty}^{\eta n_t} -\frac{y_1(t) - x_1}{\sigma_1^{\eta 2}} \exp\Big(-\frac{(y_1(t)-x_1)^2}{2\sigma_1^\eta(t)^2}\Big) dy_1(t) \\ &\quad + x_1 Pr[y_1(t) \le \eta n_t] + \eta n_t Pr[y_1(t) > \eta n_t]\Bigg] \\ &= \mu_t^1\Bigg[-\sigma_1^\eta(t)^2 \frac{1}{\sqrt{2\pi}\sigma_1^\eta(t)} \exp\Big(-\frac{(\eta n - x_1)^2}{2\sigma_1^\eta(t)^2}\Big) \\ &\quad + (x_1 - \eta n_t) Pr[y_1(t) \le \eta n_t] + \eta n_t\Bigg]. \end{aligned}$$

Then, we have an expression of $g_3^\eta(t,x)$. In order to obtain the unknown $\sigma_1^\eta(t)$ for computation, we replace $\sigma_1^\eta(t)$ with $\sqrt{u_1}$ as described in equations (22) and (23) in the paper and obtain $g_3^\eta(t,x,u)$. Note $g_4^\eta(\cdot,\cdot,\cdot)$ and $g_5^\eta(\cdot,\cdot,\cdot)$ are the same except a constant part with respect to x. Therefore, it is enough to derive $g_5^\eta(\cdot,\cdot)$. We can show that

$$\begin{aligned}
g_5^\eta(t,x) &= \eta E\Big[\beta_t p_t\Big(\frac{x_1^\eta(t)}{\eta} - \frac{E[x_1^\eta(t)]}{\eta} + \frac{x_1}{\eta} - n_t\Big)^+\Big] \\
&= \beta_t p_t\Big\{E\big[\big((x_1^\eta(t) - E[x_1^\eta(t)] + x_1) \vee \eta n_t\big)\big] - \eta n_t\Big\} \\
&= \beta_t p_t\Big\{E\big[(x_1^\eta(t) - E[x_1^\eta(t)] + x_1)\mathbb{I}_{x_1^\eta(t)-E[x_1^\eta(t)]+x_1>\eta n_t}\big] \\
&\quad +\eta n_t Pr[x_1^\eta(t) - E[x_1^\eta(t)] + x_1 \le \eta n_t] - \eta n_t\Big\}.
\end{aligned}$$

Let $y_1(t) = x_1^\eta(t) - E[x_1^\eta(t)] + x_1$. Then,

$$\begin{aligned}
g_5^\eta(t,x) &= \beta_t p_t\Bigg[\int_{\eta n_t}^{\infty} \frac{y_1(t)}{\sqrt{2\pi}\sigma_1^\eta(t)} \exp\Big(-\frac{(y_1(t)-x_1)^2}{2\sigma_1^\eta(t)^2}\Big)dy_1(t) \\
&\quad +\eta n_t Pr[y_1(t) \le \eta n_t] - \eta n_t\Bigg] \\
&= \beta_t p_t\Bigg[\frac{-\sigma_1^\eta(t)}{\sqrt{2\pi}} \int_{\eta n_t}^{\infty} -\frac{y_1(t)-x_1}{\sigma_1^\eta(t)^2} \exp\Big(-\frac{(y_1(t)-x_1)^2}{2\sigma_1^\eta(t)^2}\Big)dy_1(t) \\
&\quad +x_1 Pr[y_1(t) > \eta n_t] + \eta n_t Pr[y_1(t) \le \eta n_t] - \eta n_t\Bigg] \\
&= \beta_t p_t\Bigg[\sigma_1^\eta(t)^2 \frac{1}{\sqrt{2\pi}\sigma_1^\eta(t)} \exp\Big(-\frac{(\eta n - x_1)^2}{2\sigma_1^\eta(t)^2}\Big) \\
&\quad +(x_1 - \eta n_t) Pr[y_1(t) > \eta n_t]\Bigg].
\end{aligned}$$

Then, we have an expression of $g_5^\eta(t,x)$. Just like $g_3^\eta(t,x,u)$, we obtain $g_5^\eta(t,x,u)$ by replacing $\sigma_1^\eta(t)$ with $\sqrt{u_1}$.

2. Numerical studies for multi-class preemptive queues

We provide two numerical results comparing standard and adjusted limits to approximate multi-class preemptive queues. Figure 1 illustrates a two-class multi-server queue we consider. Customers in class 1 and 2 arrive to the queue with rate λ_t^1 and λ_t^2 respectively. Service rates are μ_t^1 for class 1 customers and μ_t^2 for class 2 customers. Class 1 customers have higher priority and preemptive discipline applies for serving customers. We show numerical results first and then provide g functions that we used for those queues in the following sections.

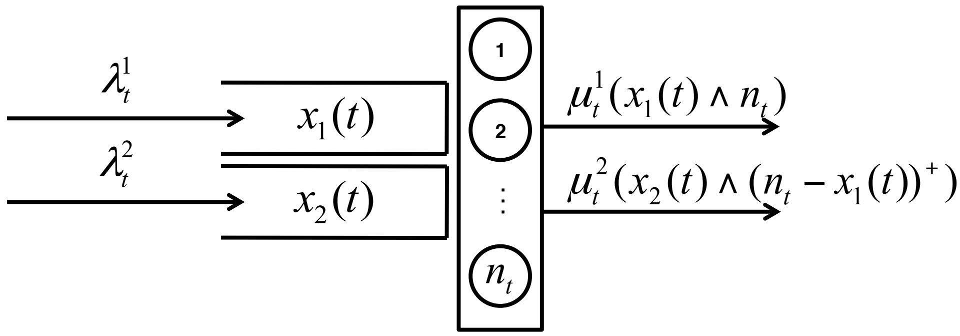


Figure 1: Multi-class preemptive queue

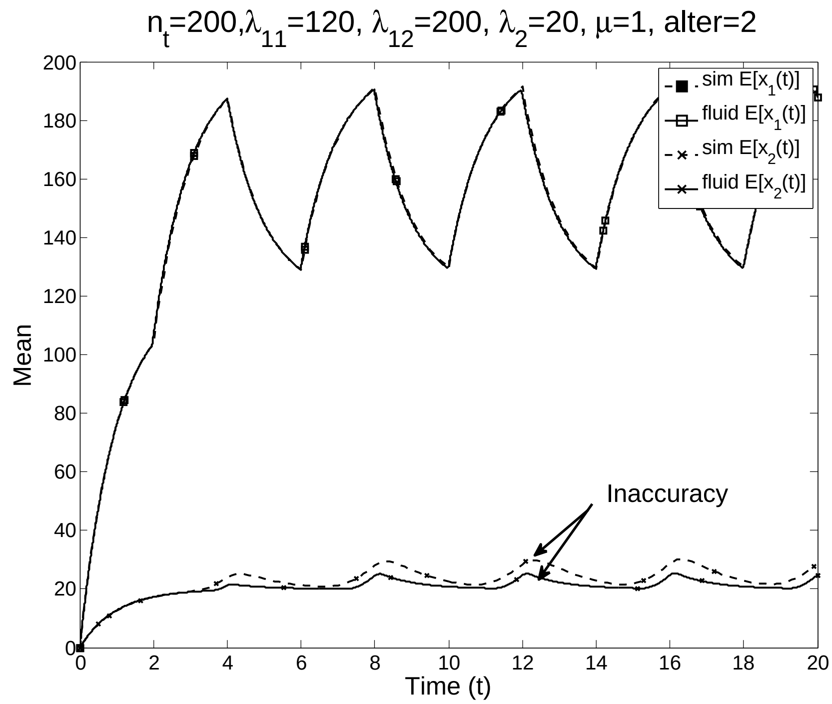
2.1. Numerical results

Table 1 describes the setting of each experiment. In Table 1, "svrs" is the number of servers (n_t), "λ_{11}" and "λ_{12}" are alternating arrival rates of class 1 customers (λ_t^1), "λ_2" is the arrival rate of class 2 customers (λ_t^2), "alter" is the time length for which each class 1 arrival rate lasts, and "time" is the end time of our analysis. We conduct $10,000$ independent simulation runs for each experiment. Figures 2 and 5 compare standard and adjusted fluid limits.

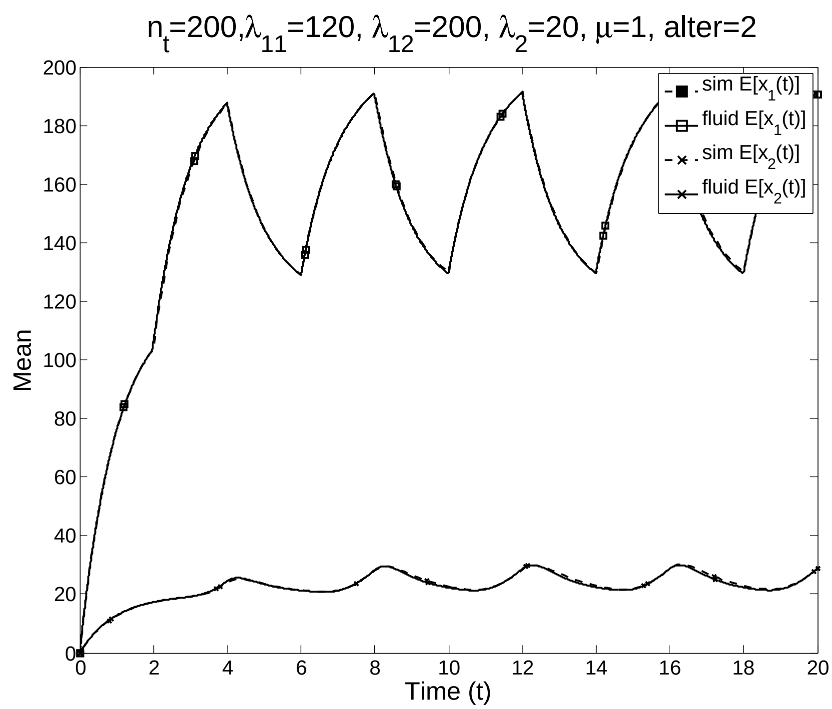
Table 1: Experiment setting

exp	svrs	λ_{11}	λ_{12}	λ_2	$\mu_1 = \mu_2$	alter	time
1	200	120	200	20	1	2	20
2	300	190	205	100	1	2	10

We notice that in both experiments, the adjusted fluid limit provides excellent approximation results and outperforms the standard fluid limit. For the diffusion limits, as seen in Figures 3 and 6, both approaches show non-trivial inaccuracy especially for the estimation of $Var[x_2(t)]$. However, we still observe that the adjusted diffusion limit provides better estimation results than the standard diffusion limit. The reason why the adjusted diffusion limit shows inaccuracy for $Var[x_2(t)]$ is that the empirical density is not close to Gaussian density. In Figures 4 and 7, we can see that the empirical density functions of $x_2(t)$ in both experiments do not match with Gaussian PDF well while those of $x_1(t)$ do match well.

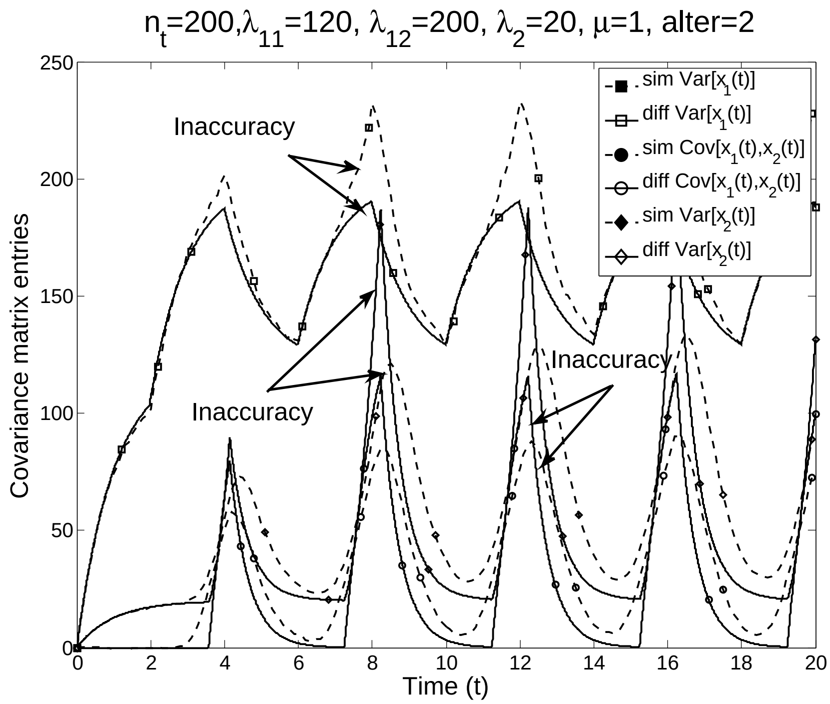


(a) Simulation vs Standard fluid limit

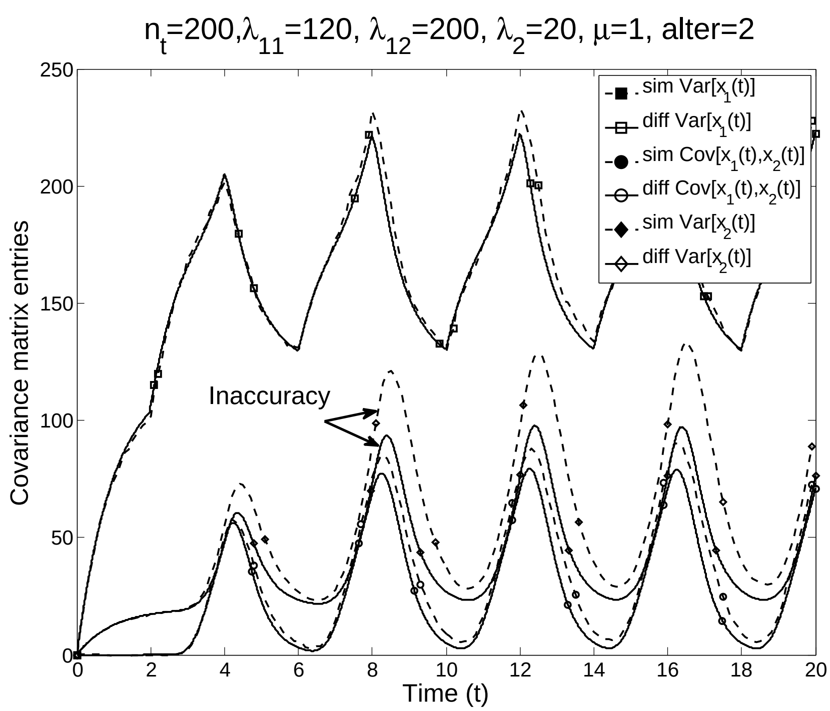


(b) Simulation vs Adjusted fluid limit

Figure 2: Comparison between fluid limits: exp. 1

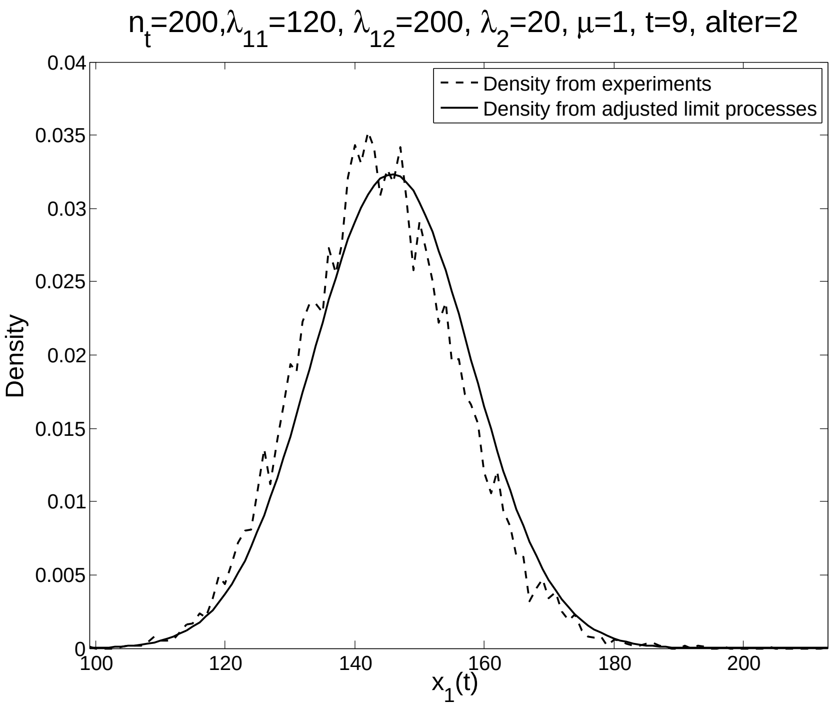


(a) Simulation vs Standard diffusion limit



(b) Simulation vs Adjusted diffusion limit

Figure 3: Comparison between diffusion limits: exp. 1



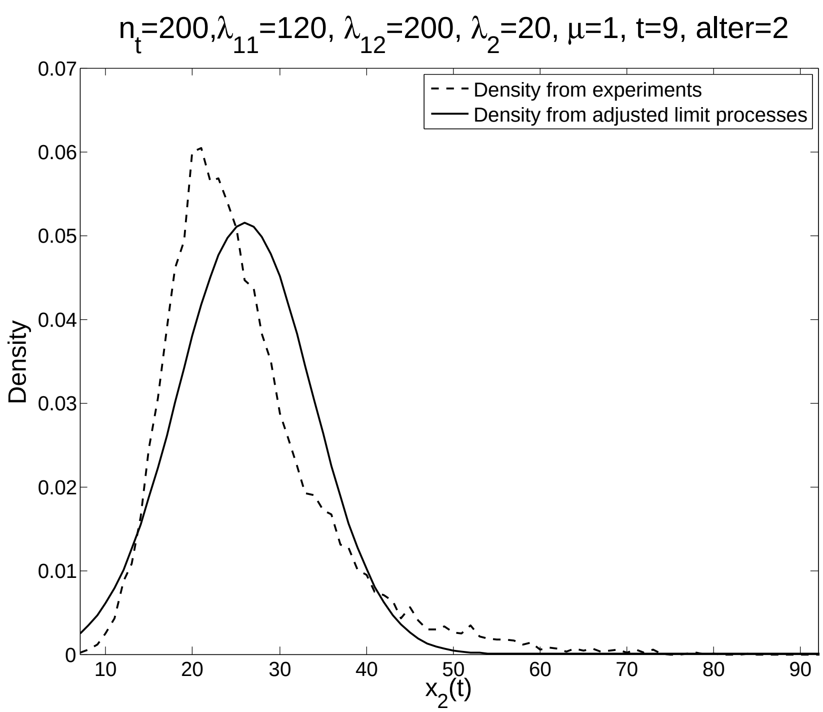
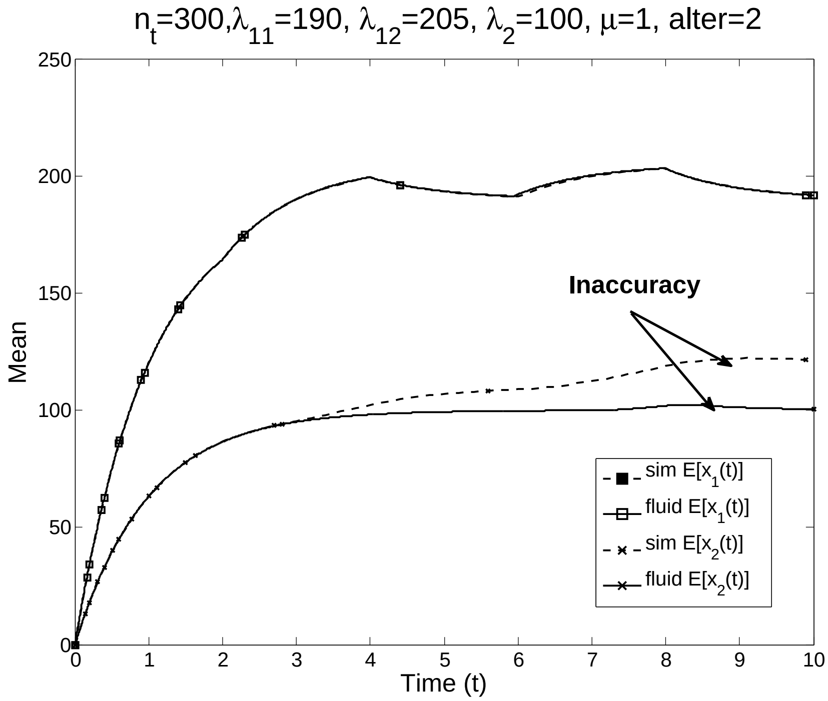
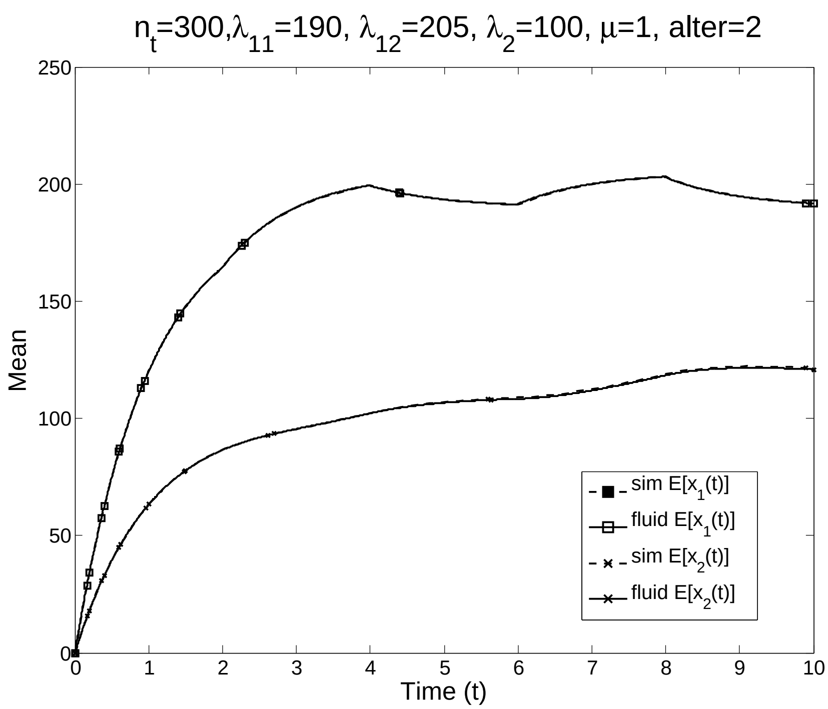


Figure 4: Empirical density vs Gaussian density at $t = 9$: exp. 1

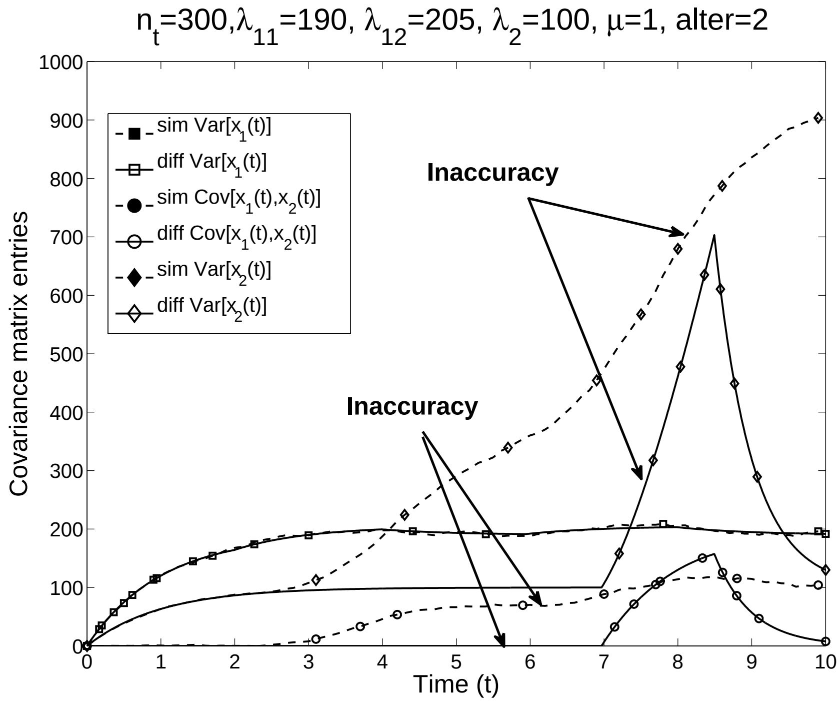


(a) Simulation vs Standard fluid limit

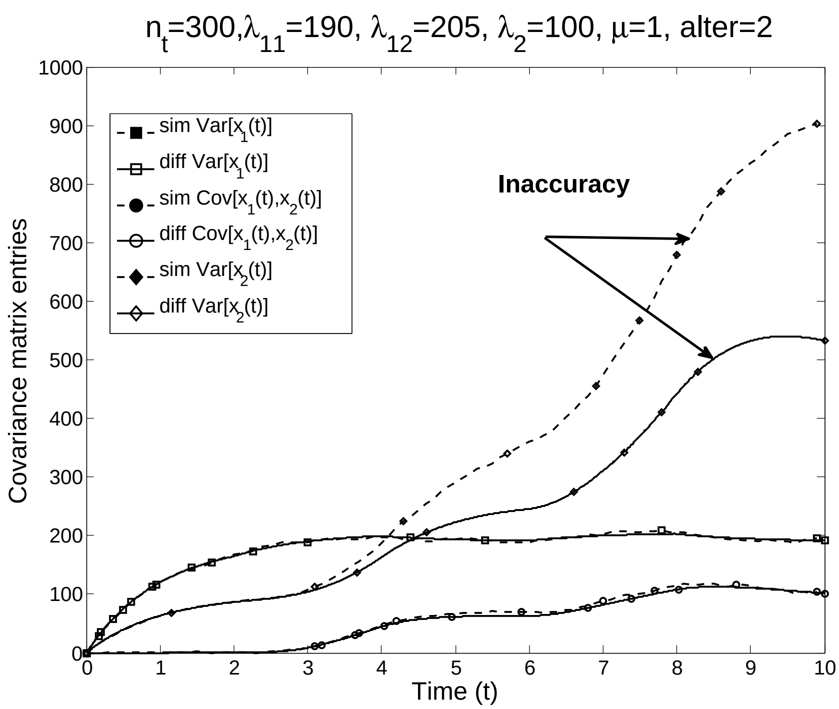


(b) Simulation vs Adjusted fluid limit

Figure 5: Comparison between fluid limits: exp. 2



(a) Simulation vs Standard diffusion limit



(b) Simulation vs Adjusted diffusion limit

Figure 6: Comparison between diffusion limits: exp. 2

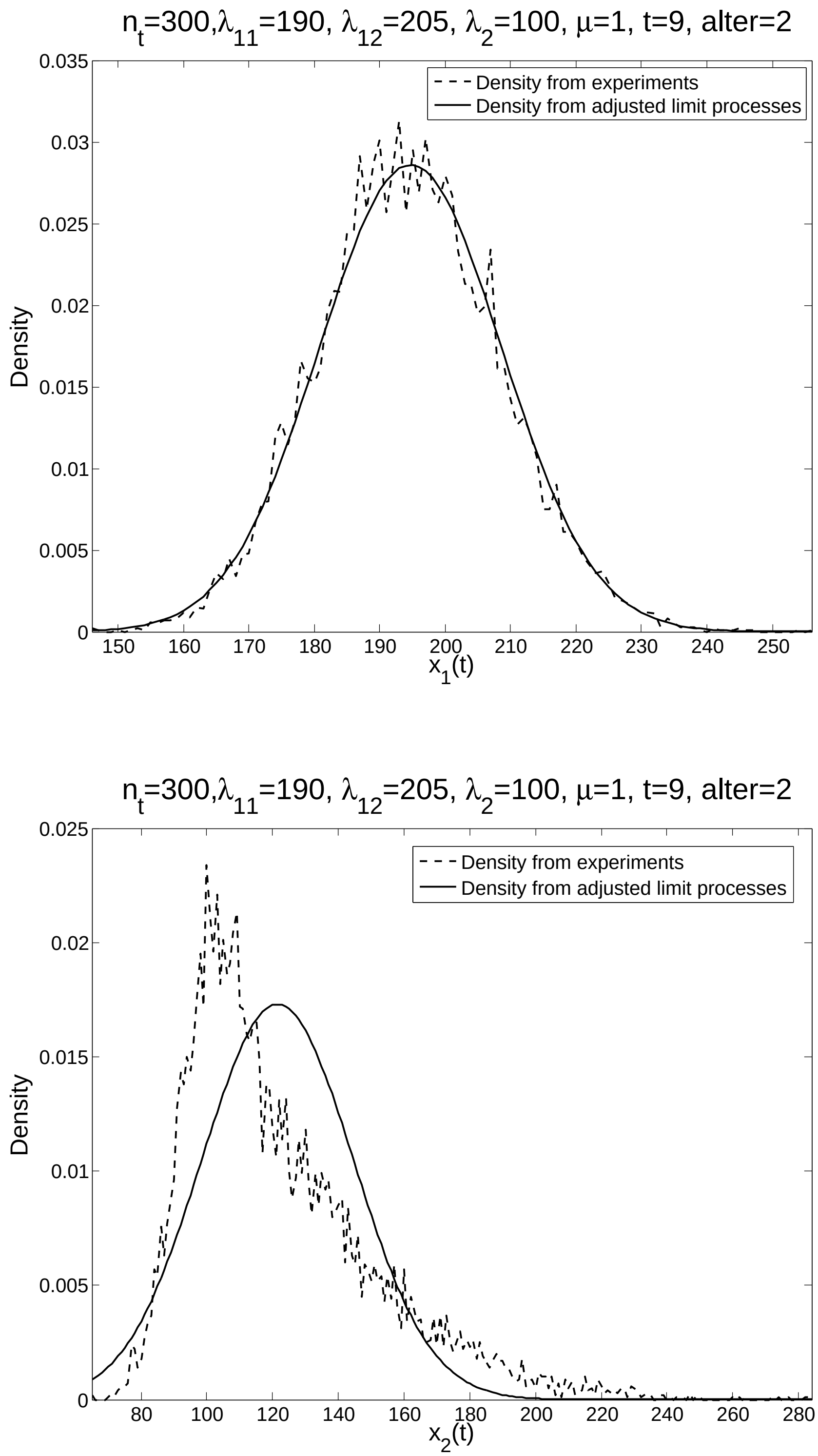


Figure 7: Empirical density vs Gaussian density at $t = 9$: exp. 2

2.2. Rate functions for adjusted limits ($g(\cdot,\cdot,\cdot)$'s)

For constant rates, i.e., λ_t^1 and λ_t^2, g functions are the same as original rate functions (f). The one corresponding to $\mu_t^1(x_1(t) \wedge n_t)$ is the same as the g_3 function in the main paper. We, therefore, just provide the g function for $\mu_t^2(x_2(t) \wedge (n_t - x_1(t))^+)$ which is new and indeed complicated. Without loss of generality, we assume $\eta = 1$ for the sake of simplicity; for any fixed η, we can easily find an equivalent formulation of which η value is 1 using substitution.

Let $\bar{Z}(t) = (\bar{z}_1(t), \bar{z}_2(t))'$ be the adjusted fluid limit and $\Sigma(t) = \begin{pmatrix} \sigma_1(t)^2 & \text{cov}(t) \\ \text{cov}(t) & \sigma_2(t)^2 \end{pmatrix}$ be the covariance matrix of the adjusted diffusion limit. In addition, define the followings:

$$\begin{aligned}
w(t) &= n_t - \bar{z}_1(t), \\
v(t) &= \bar{z}_1(t) + \bar{z}_2(t) - n_t, \\
\sigma_w(t)^2 &= \sigma_1(t)^2, \\
\sigma_v(t)^2 &= \sigma_1(t)^2 + \sigma_2(t)^2 + 2\text{cov}(t), \\
\text{cov}_{wv}(t) &= -\sigma_1(t)^2 - \text{cov}(t), \\
\rho_{wv}(t) &= \frac{\text{cov}_{wv}(t)}{\sigma_w(t) \cdot \sigma_v(t)}, \\
\text{cov}_{w2}(t) &= -\text{cov}(t), \\
\rho_{w2}(t) &= \frac{\text{cov}_{w2}(t)}{\sigma_2(t) \cdot \sigma_w(t)},
\end{aligned}$$

$$\begin{aligned}
\phi_v(t) &= \phi(0, v(t), \sigma_v(t)), \\
\phi_w(t) &= \phi(0, w(t), \sigma_w(t)), \\
\phi_2(t) &= \phi(0, \bar{z}_2(t), \sigma_2(t)), \\
\Phi_w(t) &= \Phi(0, w(t), \sigma_w(t)), \\
\Phi_v(t) &= \Phi(0, v(t), \sigma_v(t)),
\end{aligned}$$

$$\begin{aligned}
\phi_{wv}(t) &= \phi(0, w(t) - \sigma_w(t) \cdot \rho_{wv}(t) \cdot v(t)/\sigma_v(t), \sigma_w(t)\sqrt{1 - \rho_{wv}(t)^2}), \\
\Phi_{wv}(t) &= \Phi(0, w(t) - \sigma_w(t) \cdot \rho_{wv}(t) \cdot v(t)/\sigma_v(t), \sigma_w(t)\sqrt{1 - \rho_{wv}(t)^2}), \\
\phi_{vw}(t) &= \phi(0, v(t) - \sigma_v(t) \cdot \rho_{wv}(t) \cdot w(t)/\sigma_w(t), \sigma_v(t)\sqrt{1 - \rho_{wv}(t)^2}), \\
\Phi_{vw}(t) &= \Phi(0, v(t) - \sigma_v(t) \cdot \rho_{wv}(t) \cdot w(t)/\sigma_w(t), \sigma_v(t)\sqrt{1 - \rho_{wv}(t)^2}), \\
\Psi_{wv}(t) &= \Psi\left((0,0)', (w(t), v(t))', \begin{pmatrix} \sigma_w(t)^2 & \text{cov}_{wv}(t) \\ \text{cov}_{wv}(t) & \sigma_v(t)^2 \end{pmatrix}\right),
\end{aligned}$$

$$
\begin{aligned}
\phi_{w2}(t) &= \phi(0, w(t) - \sigma_w(t) \cdot \rho_{w2}(t) \cdot \bar{z}_2(t)/\sigma_2(t), \sigma_w(t)\sqrt{1-\rho_{w2}(t)^2}), \\
\Phi_{w2}(t) &= \Phi(0, w(t) - \sigma_w(t) \cdot \rho_{w2}(t) \cdot \bar{z}_2(t)/\sigma_2(t), \sigma_w(t)\sqrt{1-\rho_{w2}(t)^2}), \\
\phi_{2w}(t) &= \phi(0, \bar{z}_2(t) - \sigma_2(t) \cdot \rho_{w2}(t) \cdot w(t)/\sigma_w(t), \sigma_2(t)\sqrt{1-\rho_{w2}(t)^2}), \\
\Phi_{2w}(t) &= \Phi(0, \bar{z}_2(t) - \sigma_2(t) \cdot \rho_{w2}(t) \cdot w(t)/\sigma_w(t), \sigma_2(t)\sqrt{1-\rho_{w2}(t)^2}), \\
\Psi_{w2}(t) &= \Psi\left((0,0)', (w(t), \bar{z}_2(t))', \begin{pmatrix} \sigma_w(t)^2 & \mathrm{cov}_{w2}(t) \\ \mathrm{cov}_{w2}(t) & \sigma_2(t)^2 \end{pmatrix}\right),
\end{aligned}
$$

where $\Psi(a, b, c)$ is the function values at point a of the multivariate Gaussian CDF with mean b and covariance matrix c. Note that $\phi(\cdot, \cdot, \cdot)$ and $\Phi(\cdot, \cdot, \cdot)$ are defined in Section 5. Then, the $g(\cdot, \cdot, \cdot)$ function corresponding to $\mu_t^2(x_2(t) \wedge (n_t - x_1(t))^+)$ is

$$
\begin{aligned}
g(t, \bar{Z}(t), \Sigma(t)) = \mu_t^2\Big(& w(t) - w(t) \cdot \Phi_w(t) + \sigma_w(t)^2 \cdot \phi_w(t) + v(t) \cdot \Phi_v(t) - \sigma_v(t)^2 \cdot \phi_v(t) \\
& - v(t) \cdot \Psi_{wv}(t) + \sigma_v(t) \cdot (\Phi_{wv}(t) \cdot \phi_v(t) \cdot \sigma_v(t) + \rho_{wv}(t) \cdot \Phi_{vw}(t) \cdot \phi_w(t) \cdot \sigma_w(t)) \\
& + \bar{z}_2(t) \cdot \Psi_{w2}(t) - \sigma_2(t) \cdot (\Phi_{w2}(t) \cdot \phi_2(t) \cdot \sigma_2(t) + \rho_{w2}(t) \cdot \Phi_{2w}(t) \cdot \phi_w(t) \cdot \sigma_w(t)\Big).
\end{aligned}
$$