Predictable Queues

Fluid Models and Diffusion Approximations

for Time-Varying Queues with Abandonment and Retrials

with

Bill Massey

Marty Reiman

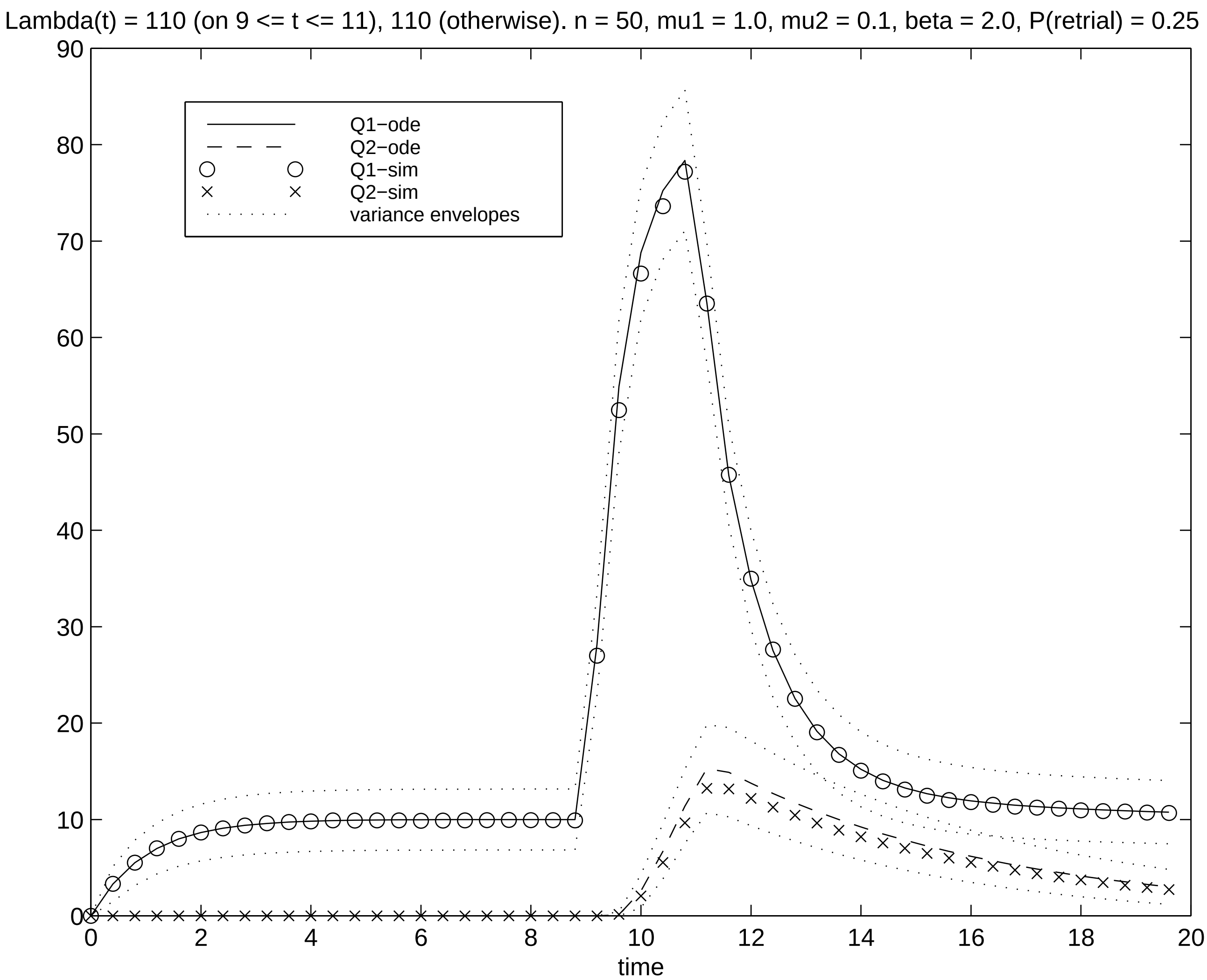
Brian Rider

Sasha Stolyar

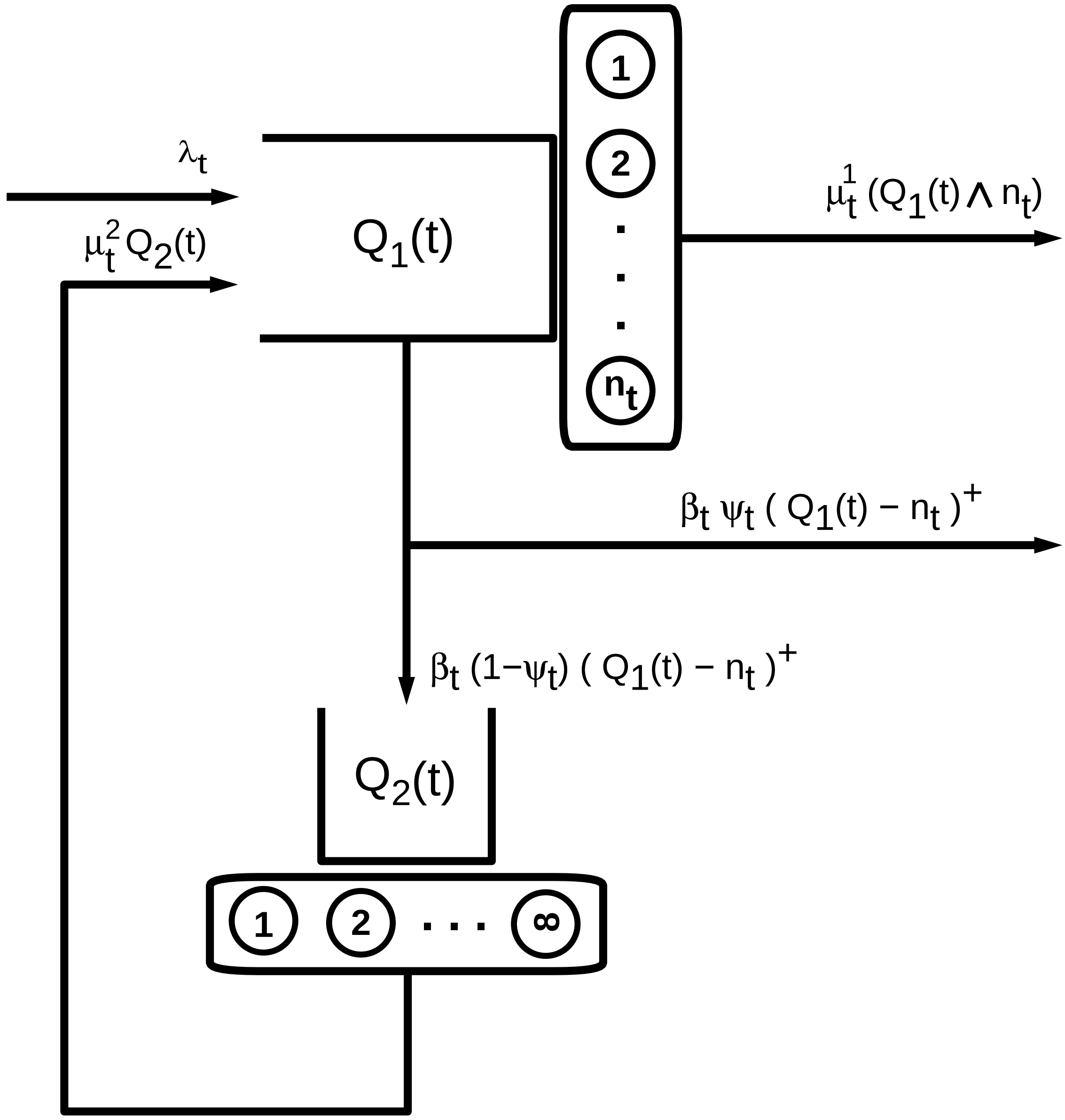
Sudden Rush Hour

$$n = 50 \text{ servers}; \qquad \mu = 1$$

$$\lambda_t = 110 \quad \text{for } 9 \le t \le 11, \quad \lambda_t = 10 \quad \text{otherwise}$$



Call Center: A Multiserver Queue with Abandonment and Retrials



Primitives (Time-Varying Predictably)

λ_t	exogenous arrival rate e.g., continuously changing, sudden peak
μ_t^1	service rate e.g., change in nature of work or fatigue
n_t	number of servers e.g., in response to predictably varying workload
β_t	abandonment rate while waiting e.g., in response to IVR discouragement at predictable overloading
ψ_t	probability of no retrial
$1/\mu_t^2$	average time to retry

Large system: $\eta \uparrow \infty$ scaling parameter. Now define

$Q^\eta(\cdot)$ via $\lambda_t \longrightarrow \eta\lambda_t$

$$n_t \longrightarrow \eta n_t$$

What do we get, as $\eta \uparrow \infty$?

Fluid Model

Replacing random processes by their rates yields

$$Q^{(0)}(t) = (Q_1^{(0)}(t),\ Q_2^{(0)}(t))$$

Solution to nonlinear differential balance equations

$$\begin{aligned}
\frac{d}{dt}\, Q_1^{(0)}(t) &= \lambda_t - \ \mu_t^1\, (Q_1^{(0)}(t) \wedge n_t) \\
&\quad + \mu_t^2\, Q_2^{(0)}(t) - \beta_t\, (Q_1^{(0)}(t) - n_t)^+ \\
\frac{d}{dt}\, Q_2^{(0)}(t) &= \beta_1 (1 - \psi_t)(Q_1^{(0)}(t) - n_t)^+ \\
&\quad - \mu_t^2\, Q_2^{(0)}(t)
\end{aligned}$$

Justification: **Functional Strong Law of Large Numbers**,
with $\lambda_t \longrightarrow \eta\lambda_t,\ n_t \longrightarrow \eta n_t$.

As $\eta \uparrow \infty$,

$$\frac{1}{\eta}\, Q^{\eta}(t) \longrightarrow Q^{(0)}(t)\,, \quad \text{uniformly on compacts, a.s.}$$

given convergence at $t = 0$

Diffusion Refinement

$$Q^\eta(t) \stackrel{d}{=} \eta\, Q^{(0)}(t) + \sqrt{\eta}\, Q^{(1)}(t) + o(\sqrt{\eta}\,)$$

Justification: **Functional Central Limit Theorem**

$$\sqrt{\eta}\, \left[\frac{1}{\eta}\, Q^\eta(t) - Q^{(0)}(t)\right] \stackrel{d}{\to} Q^{(1)}(t), \quad \text{in } D[0,\infty)\,,$$

given convergence at $t = 0$.

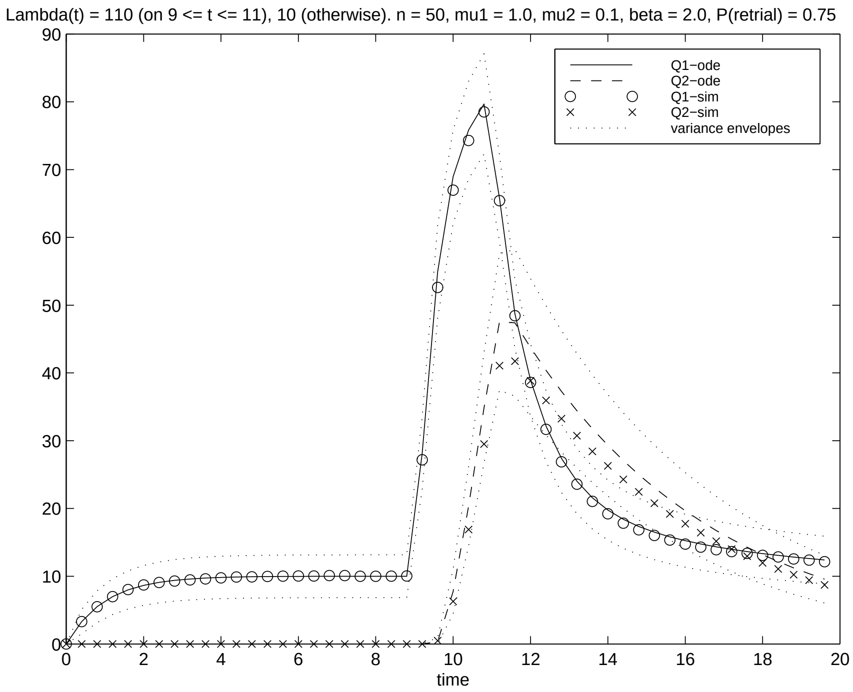
$Q^{(1)}$ solution to stochastic differential equation.

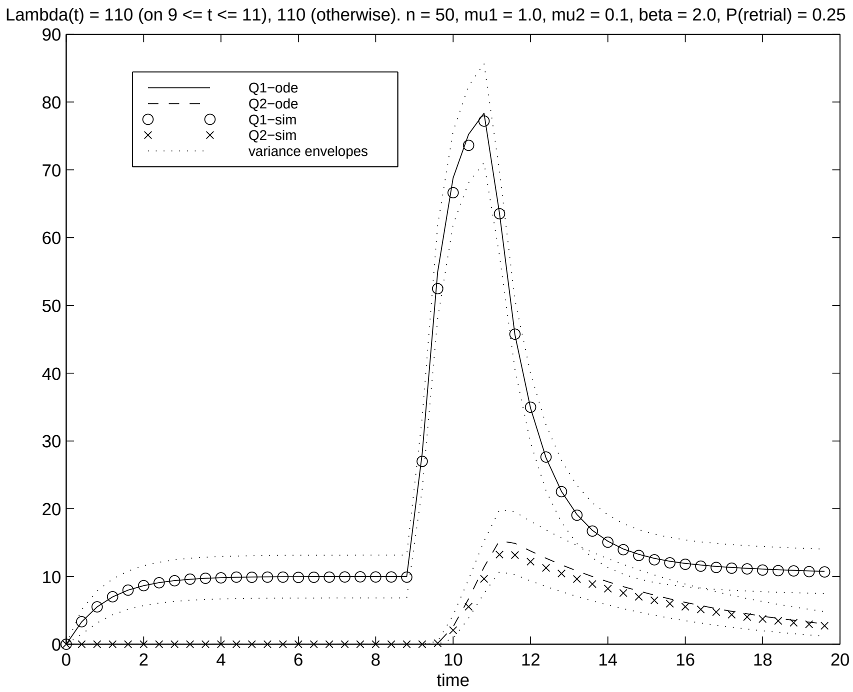
If the set of critical times $\{t \geq 0 : Q_1^{(0)}(t) = n_t\}$ has Lebesque measure zero, then $Q^{(1)}$ is a Gaussian process. In this case, one can deduce ordinary differential equations for

$$EQ_i^{(1)}(t)\,, \quad \text{Var}\; Q_i^{(1)}(t) : \quad \text{confidence envelopes}$$

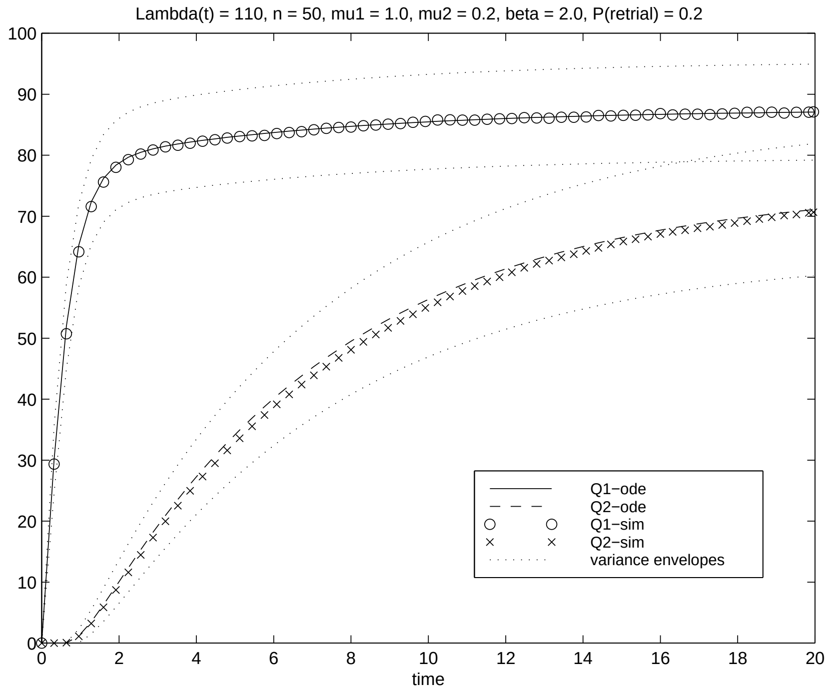
These ode's are easily solved numerically (in a spreadsheet, via forward differences).

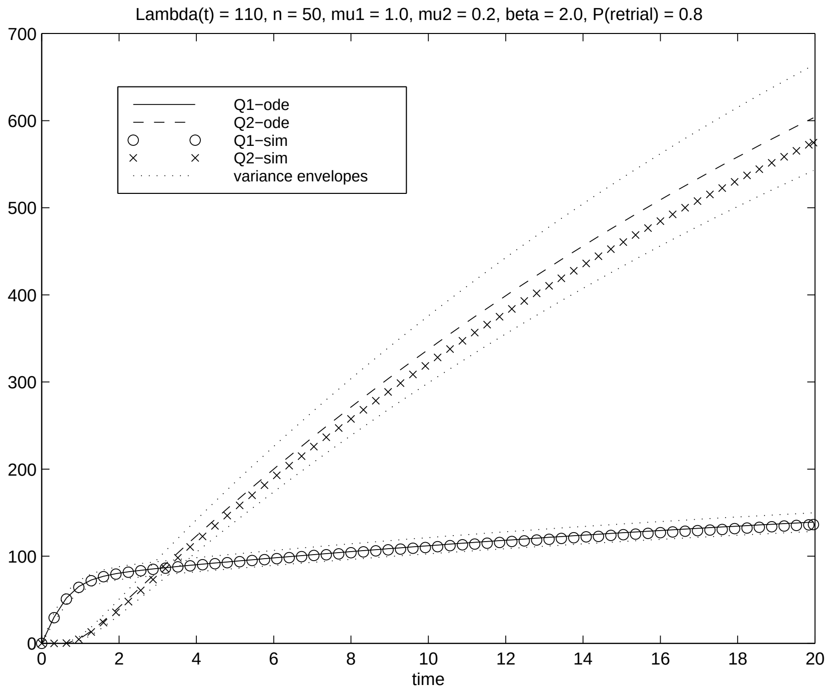
What if $P_r\{$Retrial$\}$ increases to 0.75 from 0.25 ?





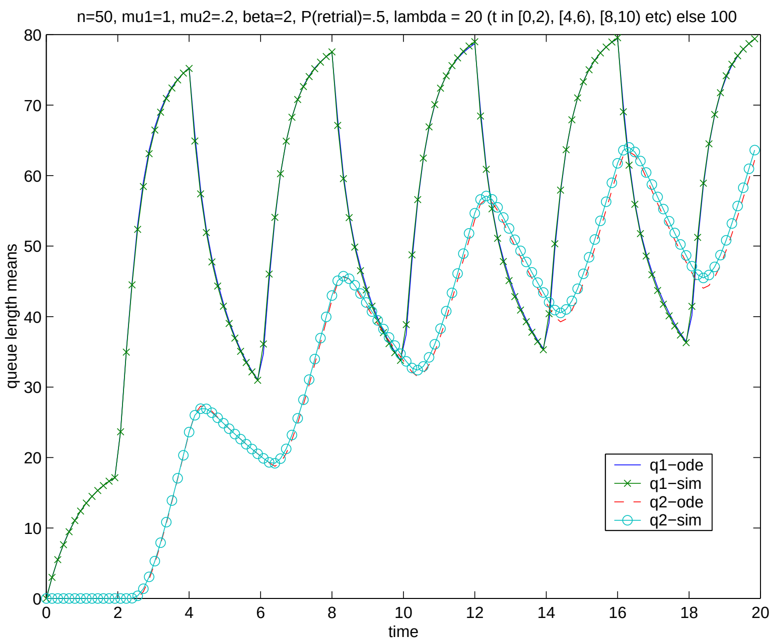
Starting Empty and Approaching Stationarity

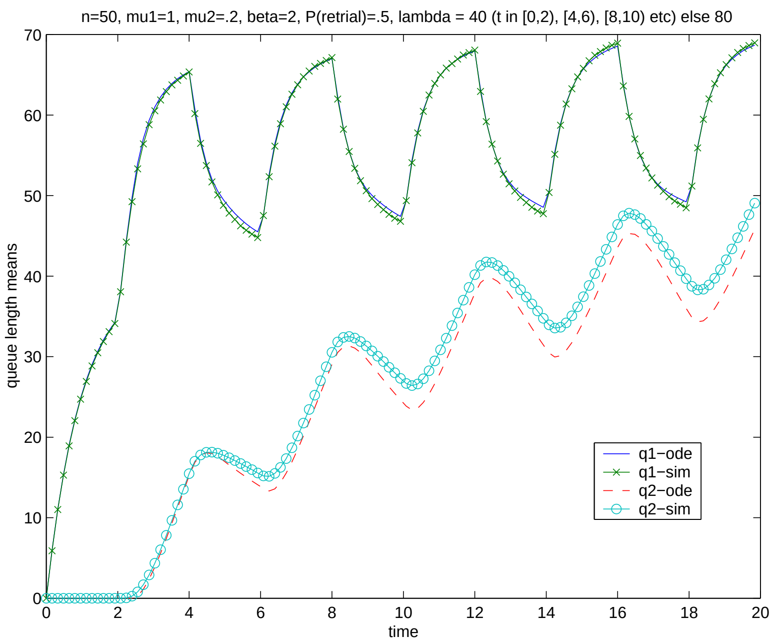




Sample Mean vs. Fluid Approximation

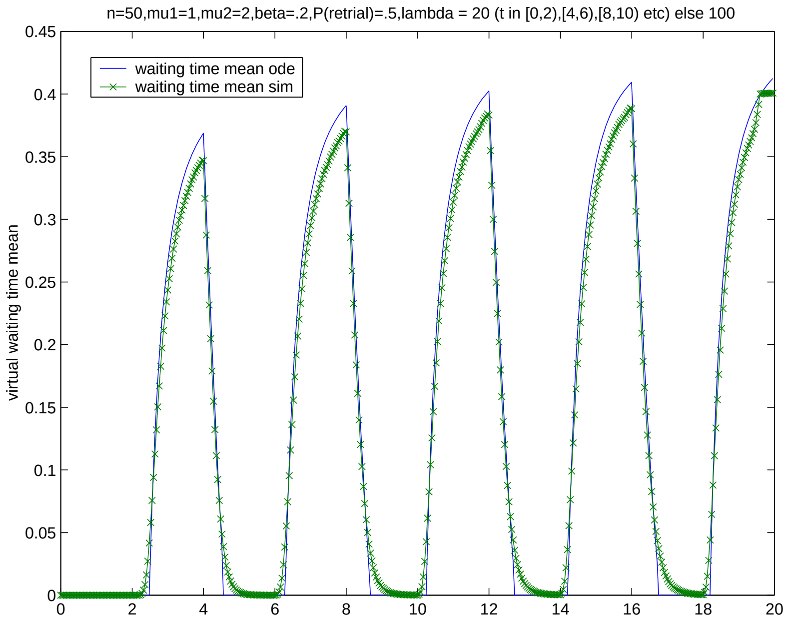
Queue Lengths (λ_t = 20 or 100)

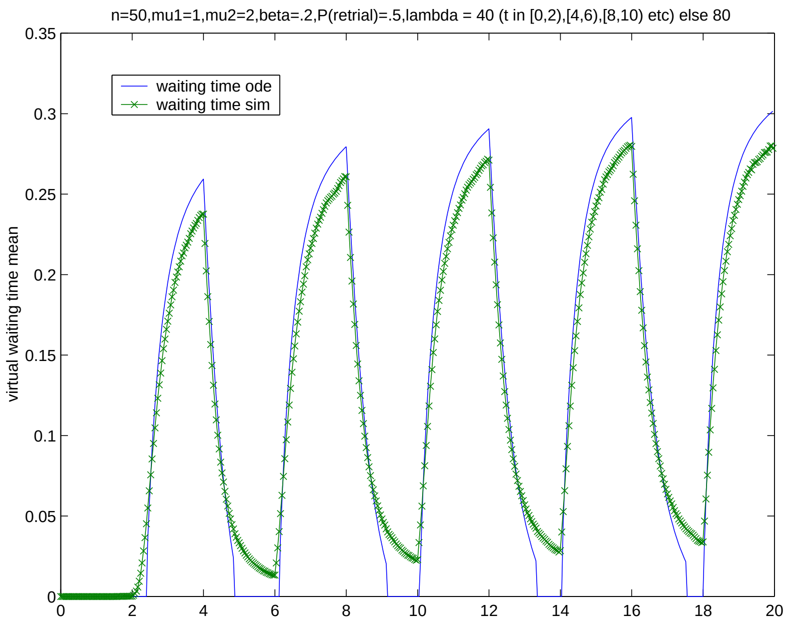




Sample Mean vs. Fluid Approximation

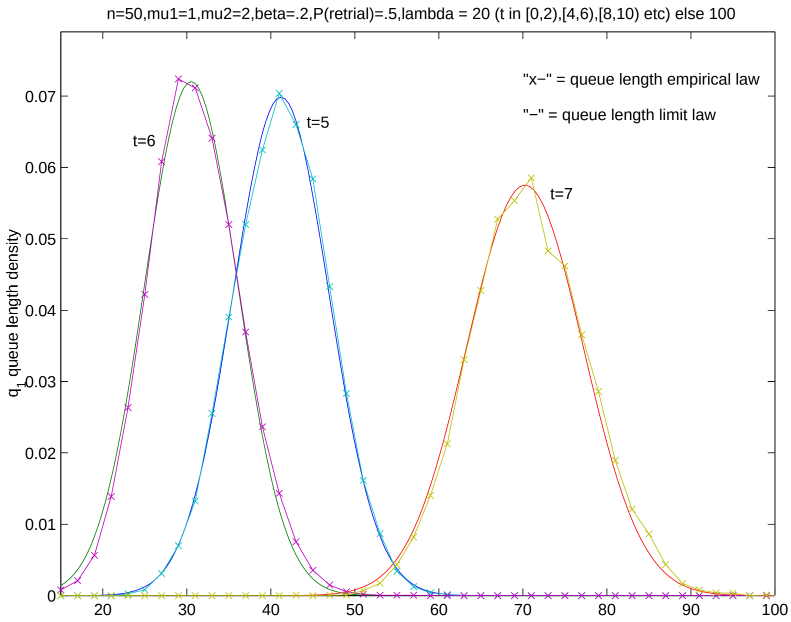
Virtual Waiting Time

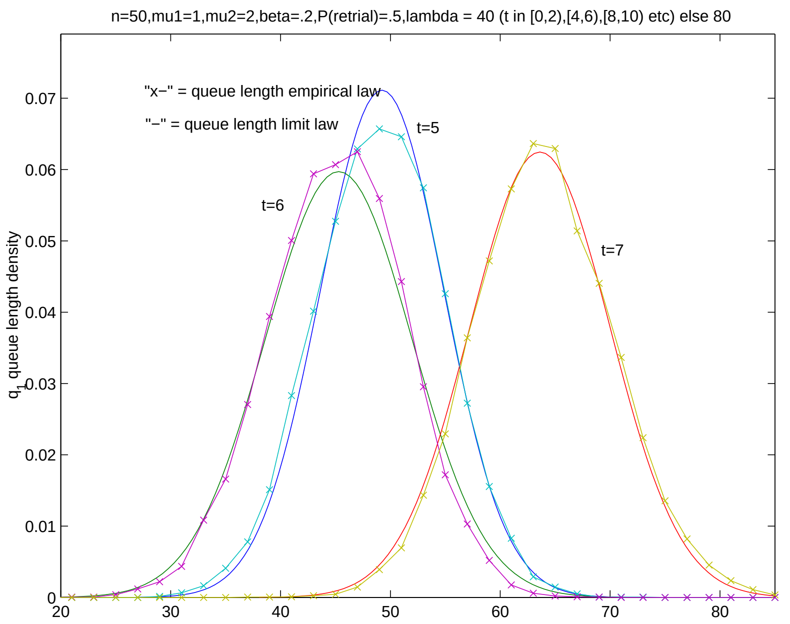




Sample Density vs. Gaussian Approximation

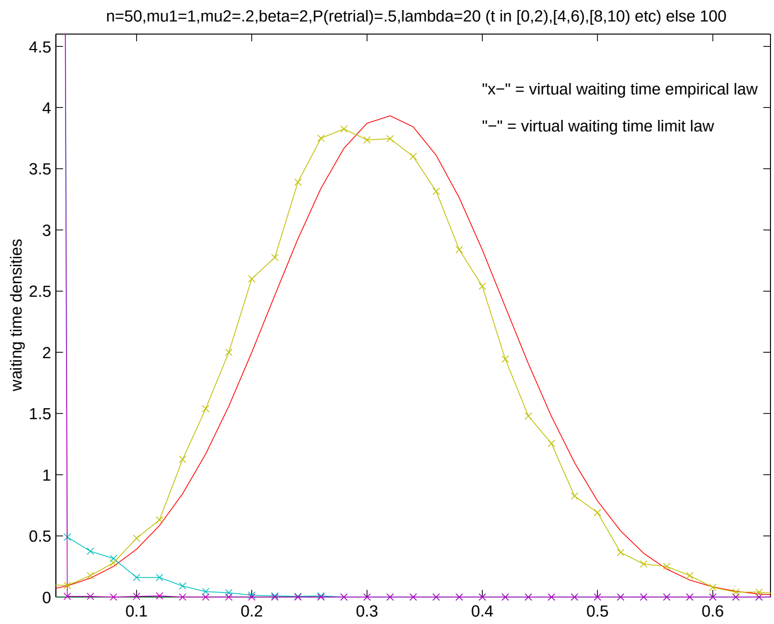
Multi-Server Queue

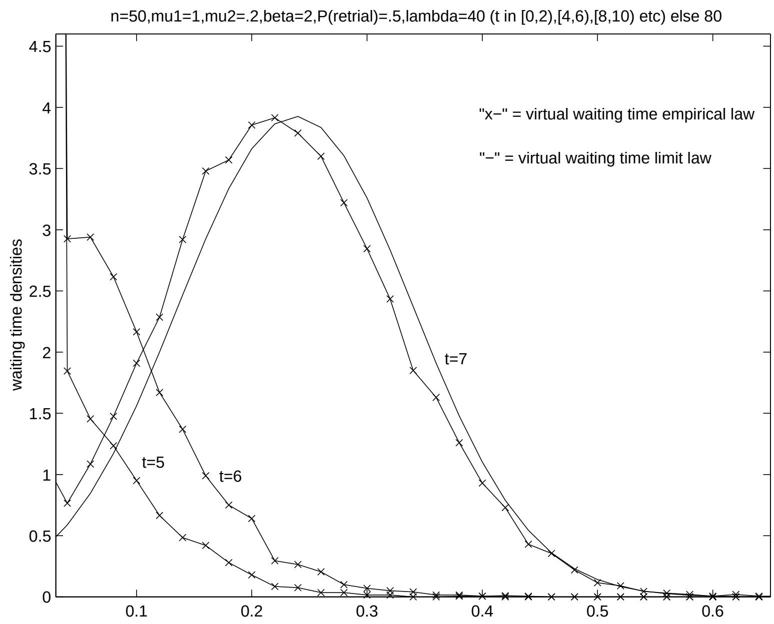




Sample Density vs. Gaussian Approximation

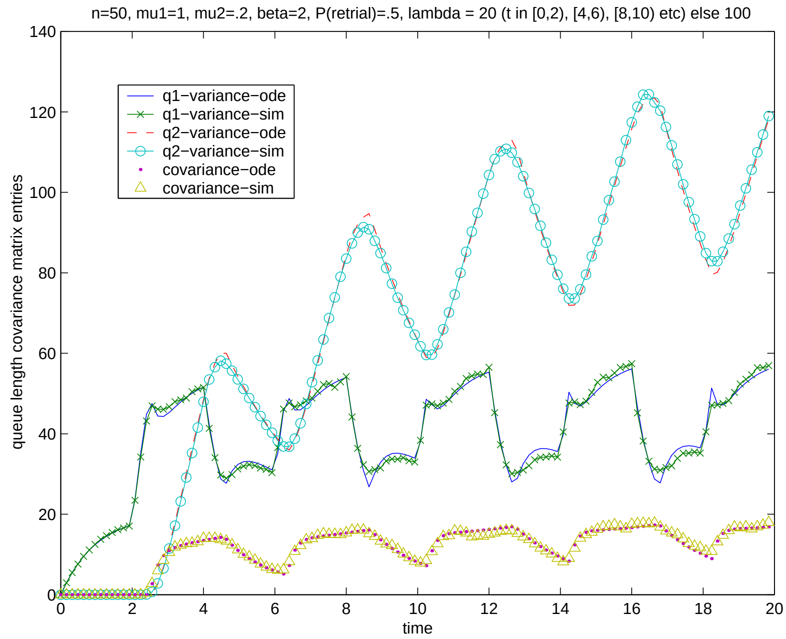
Virtual Waiting Time

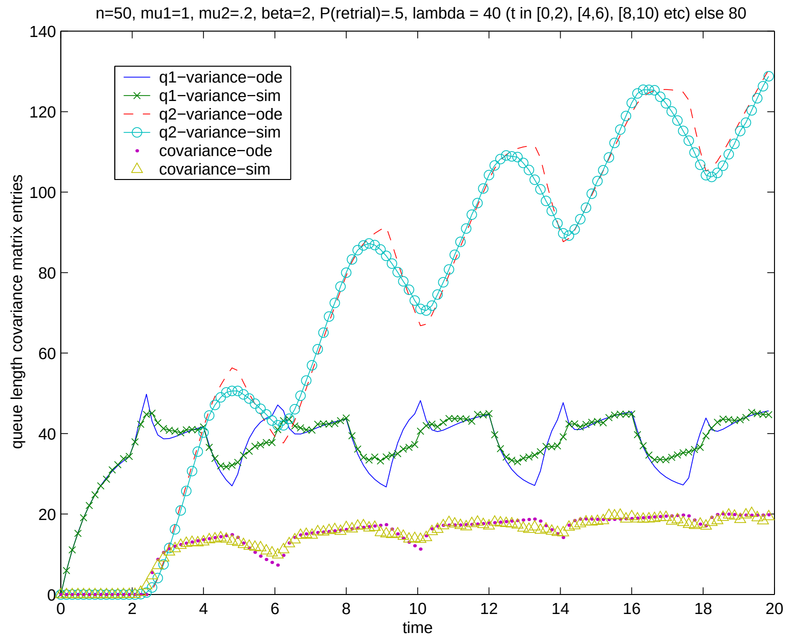




Variances and Covariances

Queue Lengths





Sample Variance vs. Diffusion Variance

Virtual Waiting Time ⇒ Future Research