The BRAVO Effect in Queues

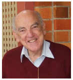
(and more stuff about variance of output counts)

Yoni Nazarathy

The University of Queensland

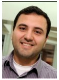
Presented in the Statistical Laboratory,
Cambridge, May 20, 2014

Variance Collaborators



Daryl Daley

Johan van Leeuwaarden



Ahmad Al-Hanbali

Michel Mandjes

Ward Whitt

Sophie Hautphenne

Yoav Kerner

Peter Taylor

Werner Scheinhardt

Brendan Patch

Thomas Taimre

Gideon Weiss

Queueing Models

$$M/M/1, \qquad M/M/1/K, \qquad M/M/s/K, \qquad M/M/s/K{+}M,$$

$$GI/G/1, \qquad GI/G/1/K, \quad \ldots$$

Basic conservation equation for a single queue

$$Q(t) = Q(0) + \Big(A(t) - L(t)\Big) - \Big(R(t) + D(t)\Big)$$

The Output Process $D(\cdot)$

$$D(t) = Q(0) + \Big(A(t) - L(t) - R(t)\Big) - Q(t)$$

The Output Process $D(\cdot)$

$$D(t) = Q(0) + \Big(A(t) - L(t) - R(t)\Big) - Q(t)$$

Why analyse $\{D(t), t \geq 0\}$?

- Orders
- Production
- Arrival process to a downstream queueing system

The Output Process $D(\cdot)$

$$D(t) = Q(0) + \Big(A(t) - L(t) - R(t)\Big) - Q(t)$$

Why analyse $\{D(t), t \geq 0\}$?

- Orders
- Production
- Arrival process to a downstream queueing system

The Output Process $D(\cdot)$

$$D(t) = Q(0) + \Big(A(t) - L(t) - R(t)\Big) - Q(t)$$

Why analyse $\{D(t), t \geq 0\}$?

- Orders
- Production
- Arrival process to a downstream queueing system

Some performance measures of interest

- The law of $\{D(t),\ t \geq 0\}$
- $\mathbb{E}[D(t)]$, $\mathrm{Var}\big(D(t)\big)$
- $\lambda^* := \lim_{t\to\infty} \frac{\mathbb{E}[D(t)]}{t}$, $\quad \overline{V} := \lim_{t\to\infty} \frac{\mathrm{Var}\big(D(t)\big)}{t}$, $\quad \mathcal{D} := \frac{\overline{V}}{\lambda^*}$
- Asymptotic normality: $D(t) \sim \mathcal{N}\Big(\lambda^* t,\ \overline{V} t\Big)$, large t
- Second order approximations, e.g.,
$$\mathrm{Var}\big(D(t)\big) = \overline{V} t + \overline{b} + o(1)$$
- Asymptotic covariances, etc...

Our Focus: Asymptotic Variance

Our Focus: Asymptotic Variance

- Reminder: Poisson processes:

$$\mathbb{E}[D(t)] = \mathrm{Var}\big(D(t)\big) = \lambda t$$

Our Focus: Asymptotic Variance

- Reminder: Poisson processes:

$$\mathbb{E}[D(t)] = \mathrm{Var}\big(D(t)\big) = \lambda t$$

- Reminder: Renewal processes:

$$\mathbb{E}[D(t)] \sim \lambda t \qquad \mathrm{Var}\big(D(t)\big) \sim \lambda c^2 t$$

Our Focus: Asymptotic Variance

- Reminder: Poisson processes:

$$\mathbb{E}[D(t)] = \mathsf{Var}\big(D(t)\big) = \lambda t$$

- Reminder: Renewal processes:

$$\mathbb{E}[D(t)] \sim \lambda t \qquad \mathsf{Var}\big(D(t)\big) \sim \lambda c^2 t$$

- What is $\mathcal{D} = \lim_{t\to\infty} \frac{\mathsf{Var}\big(D(t)\big)}{\mathbb{E}[D(t)]}$ for queues?

Our Focus: Asymptotic Variance

- Reminder: Poisson processes:

$$\mathbb{E}[D(t)] = \mathrm{Var}\big(D(t)\big) = \lambda t$$

- Reminder: Renewal processes:

$$\mathbb{E}[D(t)] \sim \lambda t \qquad \mathrm{Var}\big(D(t)\big) \sim \lambda c^2 t$$

- What is $\mathcal{D} = \lim_{t\to\infty} \frac{\mathrm{Var}\big(D(t)\big)}{\mathbb{E}[D(t)]}$ for queues?

- E.g. in stationary (and thus stable M/M/1): $\mathcal{D} = 1$

For illustration, consider M/M/1/K

- Let K be not so small, e.g. $K = 40$

For illustration, consider M/M/1/K

- Let K be not so small, e.g. $K = 40$
- Consider now $\lambda \ll \mu$, e.g. $\lambda = 0.5$, $\mu = 1$. What is $\mathcal{D}$?

For illustration, consider M/M/1/K

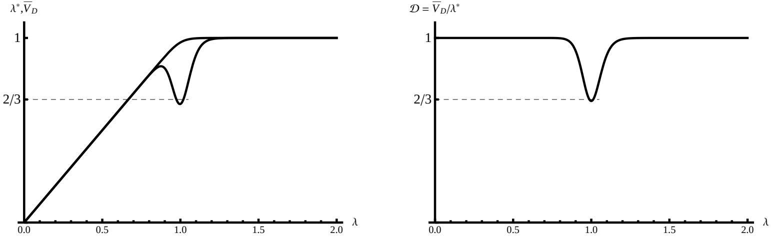
- Let K be not so small, e.g. $K = 40$
- Consider now $\lambda \ll \mu$, e.g. $\lambda = 0.5$, $\mu = 1$. What is $\mathcal{D}$?
- Consider now $\lambda \gg \mu$ e.g. $\lambda = 2.0$, $\mu = 1$. What is $\mathcal{D}$?

For illustration, consider M/M/1/K

- Let K be not so small, e.g. $K = 40$
- Consider now $\lambda \ll \mu$, e.g. $\lambda = 0.5$, $\mu = 1$. What is $\mathcal{D}$?
- Consider now $\lambda \gg \mu$ e.g. $\lambda = 2.0$, $\mu = 1$. What is $\mathcal{D}$?
- So how about $\mathcal{D}$ when $\lambda = \mu$ (e.g. $= 1$)?

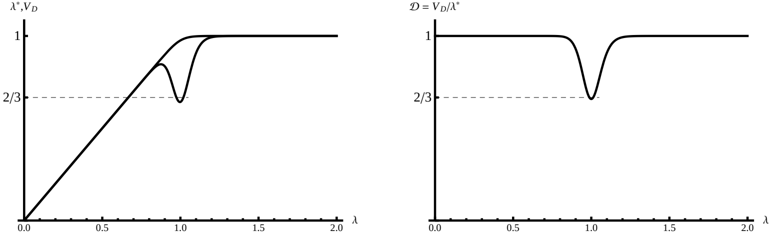
For illustration, consider M/M/1/K

- Let K be not so small, e.g. $K = 40$
- Consider now $\lambda \ll \mu$, e.g. $\lambda = 0.5$, $\mu = 1$. What is $\mathcal{D}$?
- Consider now $\lambda \gg \mu$ e.g. $\lambda = 2.0$, $\mu = 1$. What is $\mathcal{D}$?
- So how about $\mathcal{D}$ when $\lambda = \mu$ (e.g. $= 1$)?



For illustration, consider M/M/1/K

- Let K be not so small, e.g. $K = 40$
- Consider now $\lambda \ll \mu$, e.g. $\lambda = 0.5$, $\mu = 1$. What is $\mathcal{D}$?
- Consider now $\lambda \gg \mu$ e.g. $\lambda = 2.0$, $\mu = 1$. What is $\mathcal{D}$?
- So how about $\mathcal{D}$ when $\lambda = \mu$ (e.g. $= 1$)?



We call this **BRAVO**:

Balancing **R**educes **A**symptotic **V**ariance of **O**utputs

Finite Birth-Death Asymptotic Variance (and BRAVO)

Finite Birth-Death Setting

- Irreducible birth-death process on finite state space
- Birth rates: $\lambda_0, \dots, \lambda_{J-1}$
- Death rates: $\mu_1, \dots, \mu_J$
- Stationary distribution: $\pi_0, \dots, \pi_J$
- $D(t)$ is number of downward transitions (deaths) during $[0, t]$, each "filtered" independently with state-dependent probabilities, $q_1, \dots, q_J$.
- e.g. The output process (served customers) in M/M/s/K+M :

$$\lambda_i = \lambda, \quad \mu_i = \mu\,(i \wedge s) + \gamma\,(i - s)^+, \qquad q_i = \frac{\mu\,(i \wedge s)}{\mu\,(i \wedge s) + \gamma\,(i - s)^+}, \quad i = 0, 1, \dots, s+K$$

Of interest:

$$\mathcal{D} = \frac{\overline{V}}{\lambda^*} = \lim_{t \to \infty} \frac{\mathrm{Var}\big(D(t)\big)}{\mathbb{E}[D(t)]}$$

Finite Birth-Death Asymptotic Variance Formula

Theorem: Daryl Daley, Johan van Leeuwaarden, Y.N. 2014

$$\mathcal{D} := \lim_{t\to\infty} \frac{\mathrm{Var}\big(D(t)\big)}{\mathbb{E}[D(t)]} = 1 - 2\sum_{i=0}^{J}(P_i - \Lambda_i^*)\Big(q_{i+1} - \frac{\lambda^*}{\pi_i\lambda_i}(P_i - \Lambda_i^*)\Big),$$

with,

$$P_i := \sum_{j=0}^{i} \pi_j, \qquad \lambda^* := \sum_{j=1}^{J} \mu_j q_j \pi_j, \qquad \Lambda_i^* := \frac{\sum_{j=1}^{i} \mu_j q_j \pi_j}{\lambda^*}.$$

Note: In Y.N. and Weiss 2008, similar expression for case $q_i \equiv 1$

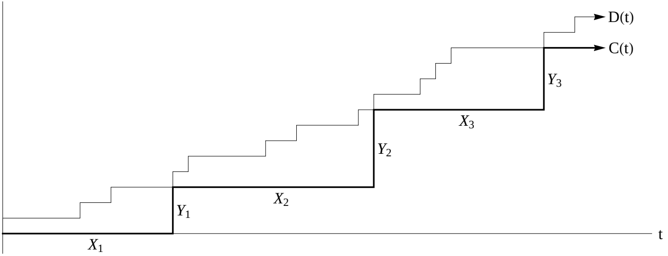
Note: In case $\lambda_i \equiv \lambda$, $q_i \equiv 1$:

$$\mathcal{D} = 1 - 2\frac{\pi_J}{1-\pi_J}\sum_{i=0}^{J} P_i\Big(1 - \pi_J\frac{P_i}{\pi_i}\Big)$$

Idea of Renewal Reward Derivation

"Embed" $D(t)$ in a Renewal-Reward Process, $C(t)$

1. $(X_n, Y_n) \equiv$ (busy cycle, number served) in cycle n
2. $N(t) = \sup\{n : \sum_{i=1}^{n} X_i \leq t\}$, $C(t) = \sum_{i=1}^{N(t)} Y_i$
3. Asymptotic variance rates of $C(t)$ and $D(t)$ are equal
4. Known:
   - Asymptotic variance rate of $C(t)$ is $\frac{1}{\mathbb{E}[X]}\mathrm{Var}\left(Y - \frac{\mathbb{E}[Y]}{\mathbb{E}[X]}X\right)$
   - Systems of equations for
     1'st, 2'nd and cross moments of X and Y



Back to the M/M/1/K Queue

Here π_i is truncated geometric distribution when $\lambda \neq \mu$ and a uniform distribution when $\lambda = \mu$

Using $\mathcal{D} = 1 - 2\frac{\pi_J}{1-\pi_J}\sum_{i=0}^{J} P_i\left(1 - \pi_J\frac{P_i}{\pi_i}\right)$:

Back to the M/M/1/K Queue

Here π_i is truncated geometric distribution when $\lambda \neq \mu$ and a uniform distribution when $\lambda = \mu$

Using $\mathcal{D} = 1 - 2\frac{\pi_J}{1-\pi_J}\sum_{i=0}^{J} P_i\left(1 - \pi_J\frac{P_i}{\pi_i}\right)$:

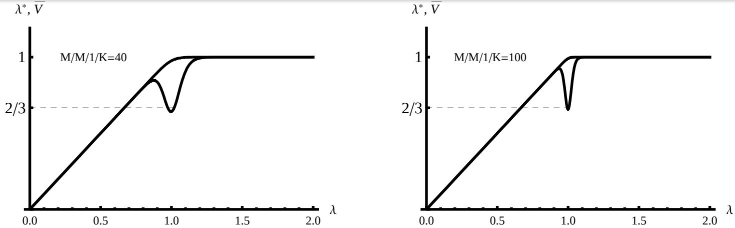
$$\mathcal{D} = \begin{cases} 1 + o_K(1), & \lambda \neq \mu, \\ \frac{2}{3} + o_K(1), & \lambda = \mu. \end{cases}$$

Back to the M/M/1/K Queue

Here π_i is truncated geometric distribution when $\lambda \neq \mu$ and a uniform distribution when $\lambda = \mu$

Using $\mathcal{D} = 1 - 2\frac{\pi_J}{1-\pi_J}\sum_{i=0}^{J} P_i\left(1 - \pi_J \frac{P_i}{\pi_i}\right)$:

$$\mathcal{D} = \begin{cases} 1 + o_K(1), & \lambda \neq \mu, \\ \frac{2}{3} + o_K(1), & \lambda = \mu. \end{cases}$$

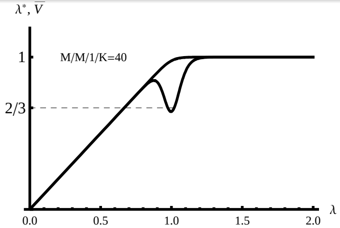


Back to the M/M/1/K Queue

Here π_i is truncated geometric distribution when $\lambda \neq \mu$ and a uniform distribution when $\lambda = \mu$

Using $\mathcal{D} = 1 - 2\frac{\pi_J}{1-\pi_J}\sum_{i=0}^{J} P_i\Big(1 - \pi_J \frac{P_i}{\pi_i}\Big)$:

$$\mathcal{D} = \begin{cases} 1 + o_K(1), & \lambda \neq \mu, \\ \frac{2}{3} + o_K(1), & \lambda = \mu. \end{cases}$$



In fact, for any λ, μ, we have an explicit expression for $\mathcal{D}$ (alt. $\overline{V}$, λ^*) and even for $\overline{b}$ in,

$$\mathrm{Var}\big(D(t)\big) = \overline{V}t + \overline{b} + o(1)$$

Multi-Server Systems in the Halfin-Whitt (QED) Regime

Quality and Efficiency Driven (QED) Scaling Regime

A sequence of systems

Consider a sequence of $M/M/s/K$ queues with increasing $s = 1, 2, \ldots$ and with $\rho_s := \frac{\lambda}{s\mu}$ and K_s such that,

$$(1-\rho_s)\sqrt{s} \to \beta \in (-\infty, \infty)$$
$$\frac{K_s}{\sqrt{s}} \to \eta \in (0, \infty)$$

So for large s:

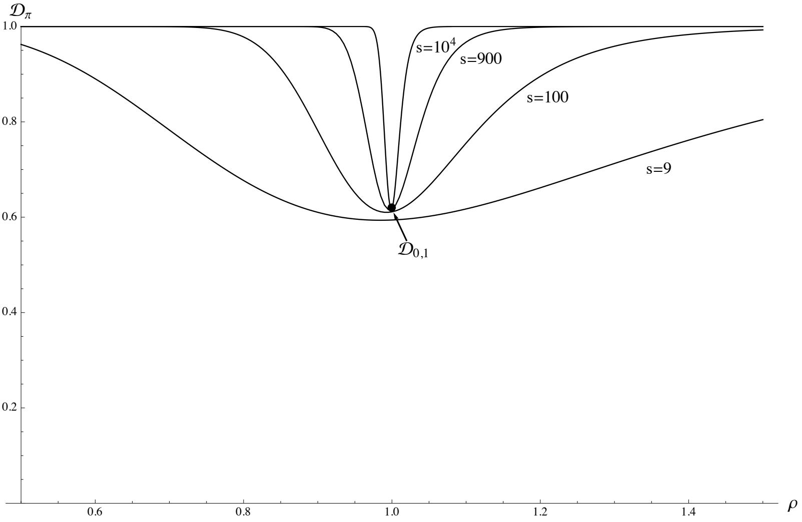
$$\rho_s \approx 1 - \beta/\sqrt{s}$$
$$K_s \approx \eta\sqrt{s}$$

Halfin, Whitt, 1981, Garnett, Mandelbaum, Reiman 2002, Borst, Mandelbaum, Reiman, 2004, Whitt, 2004, Pang, Talreja, Whitt, 2007, Janssen, van Leeuwaarden, Zwart, 2011, Kaspi, Ramanan 2011...

Favorable QED Properties

- Probability of delay converges to a value $\in (0, 1)$
- Mean waiting times are typically $O(s^{-1/2})$
- Large queue lengths almost never occur
- Quick mixing times
- In applications: Call-centers (etc...) describes behavior well and allows for asymptotic approximate optimization of staffing etc...
- How about BRAVO?

$M/M/s/\lfloor\sqrt{s}\rfloor$



M/M/s/K QED BRAVO

Theorem: Daryl Daley, Johan van Leeuwaarden, Y.N. 2013

Consider QED scaling with $\beta \neq 0$:

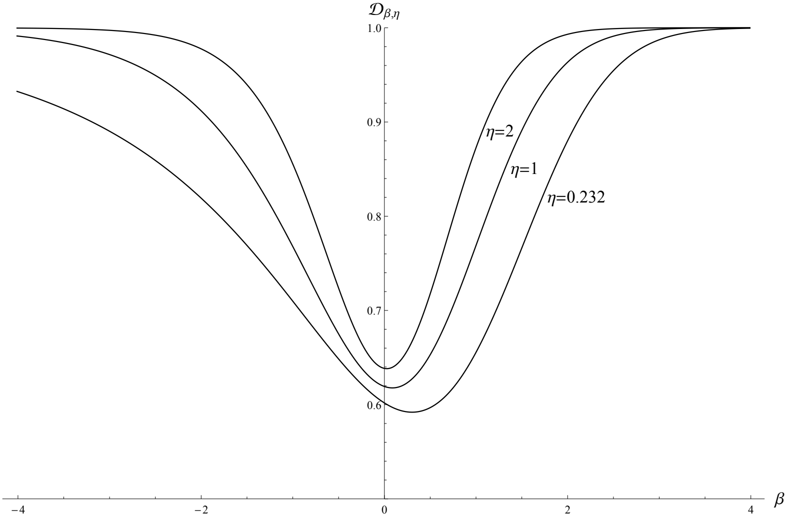
$$\mathcal{D}_{\beta,\eta} := \lim_{s,K\to\infty} \lim_{t\to\infty} \frac{Var\big(D(t)\big)}{\mathbb{E}\big(D(t)\big)},$$

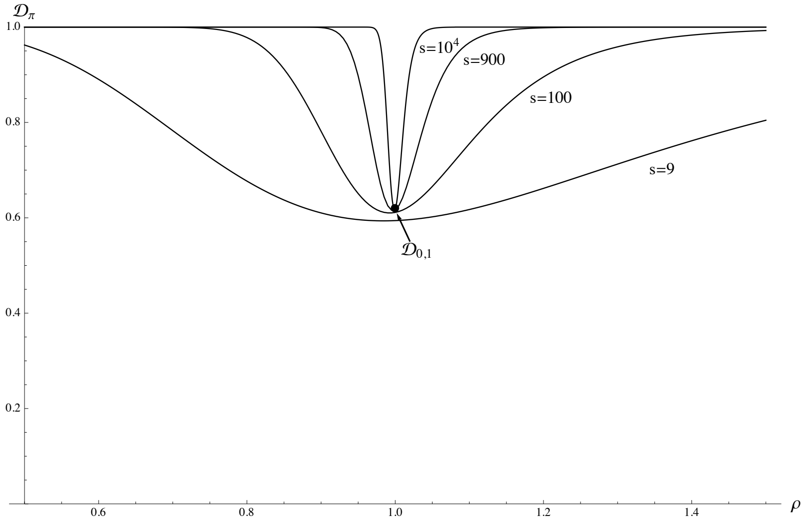
$$\mathcal{D}_{\beta,\eta} = 1 - \frac{2\beta^2 e^{-\beta\eta} h^2}{\phi(\beta)} \int_{-\beta}^{\infty} \left(1 - \beta e^{-\beta\eta} h \frac{\Phi(-u)}{\phi(u)}\right) \Phi(-u)\, du$$
$$+ 2e^{-\beta\eta} h(1 + e^{-\beta\eta} h)\Big(1 - \beta\eta - e^{-\beta\eta} + (1 - 2\beta\eta e^{-\beta\eta} - e^{-2\beta\eta})h\Big)$$

where

$$h = \lim_{s\to\infty} \frac{\mathbb{P}\big(Q_s \geq s\big)}{1 - e^{-\beta\eta}} = \frac{1}{1 - e^{-\beta\eta} + \frac{\beta\Phi(\beta)}{\phi(\beta)}}$$

BRAVO Viewed Through the QED Lens





$\mathcal{D}_\pi$
1.0
0.8
0.6
0.4
0.2
s=10^4
s=900
s=100
s=9
$\mathcal{D}_{0,1}$
0.6
0.8
1.0
1.2
1.4
ρ

M/M/s/K QED BRAVO with $\rho \equiv 1$ $(\beta = 0)$

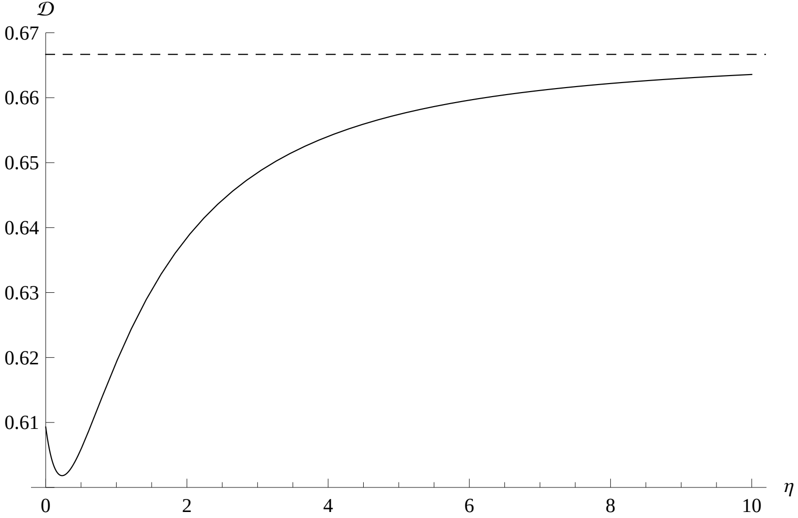
Theorem: Daryl Daley, Johan van Leeuwaarden, Y.N. 2013

Assume $\rho \equiv 1$ and $\frac{K_s}{\sqrt{s}} \to \eta \in (0, \infty)$. Then

$$\mathcal{D}_{0,\eta} := \lim_{s,K\to\infty} \lim_{t\to\infty} \frac{Var\big(D(t)\big)}{\mathbb{E}\big(D(t)\big)},$$

$$\mathcal{D}_{0,\eta} = \frac{2}{3} - \frac{\left(6 - \frac{3\pi}{2}\right)\eta - \frac{1}{2}\pi\sqrt{\frac{\pi}{2}} + 3\sqrt{2\pi}(1 - \log 2)}{3\left(\eta + \sqrt{\frac{\pi}{2}}\right)^3}.$$

M/M/s/$\lfloor \eta\sqrt{s} \rfloor$ $\quad s \to \infty \quad$ at $\rho \equiv 1$ ($\beta = 0$)



Idea of BRAVO QED Derivations

Use

$$\mathcal{D} = 1 - 2\frac{\pi_J}{1-\pi_J}\sum_{i=0}^{J} P_i\left(1 - \pi_J\frac{P_i}{\pi_i}\right).$$

Using QED scaling:

$$(1-\rho_s)\sqrt{s} \to \beta, \qquad \frac{K_s}{\sqrt{s}} \to \eta,$$

evaluate the limit,

$$\lim_{s,K\to\infty} \frac{\pi_J^{(s,K)}}{1-\pi_J^{(s,K)}}\sum_{i=0}^{J} P_i^{(s,K)}\left(1 - \pi_J^{(s,K)}\frac{P_i^{(s,K)}}{\pi_i^{(s,K)}}\right).$$

Beyond Finite Birth-Death Queues

M/M/1 Queue

When $K = \infty$, the birth-death $\mathcal{D}$ formula, generally does not hold. In this case,

$$\mathcal{D} = \begin{cases} 1, & \lambda \neq \mu, \\ ?, & \lambda = \mu. \end{cases}$$

A guess is $\frac{2}{3}$, since for $K < \infty$, $\mathcal{D} = \frac{2}{3} + o_K(1)$ . . .

M/M/1 Queue

When $K = \infty$, the birth-death $\mathcal{D}$ formula, generally does not hold. In this case,

$$\mathcal{D} = \begin{cases} 1, & \lambda \neq \mu, \\ ?, & \lambda = \mu. \end{cases}$$

A guess is $\frac{2}{3}$, since for $K < \infty$, $\mathcal{D} = \frac{2}{3} + o_K(1)$ . . .

Theorem:
Ahmad Al-Hanbali, Michel Mandjes, Y. N., Ward Whitt, 2011

For the M/M/1 queue with $\lambda = \mu$ and arbitrary initial conditions of $Q(0)$ (with finite second moments),

$$\mathcal{D} = 2\Big(1 - \frac{2}{\pi}\Big) \approx 0.727.$$

Proof based on analysis of classic Laplace transform of the generating function of $D(\cdot)$

M/M/1 Queue

$\mathrm{Var}\big(D(t)\big) =$ horrible expression involving integrals of Bessel functions

From it:

$$\mathrm{Var}\big(D(t)\big) = \begin{cases} \lambda t - \frac{\rho}{(1-\rho)^2} + o(1), & \text{if } \ \lambda < \mu, \\ 2(1 - \frac{2}{\pi})\lambda t - \sqrt{\frac{\lambda}{\pi}}\, t^{1/2} + \frac{\pi - 2}{4\pi} + o(1), & \text{if } \ \lambda = \mu, \\ \mu t - \frac{\rho}{(1-\rho)^2} + o(1), & \text{if } \ \lambda > \mu, \end{cases}$$

The Stable M/G/1 Queue

Theorem: Sophie Hautphenne, Yoav Kerner, Y. N., Peter Taylor, 2013

Consider the stable M/G/1 queue with finite third service moment, parameterized by (*arrival rate, load, scv, skewness*)$= (\lambda, \rho, c^2, \gamma)$.

Stationary version:

$$\mathrm{Var}\big(D(t)\big) = \lambda t + L_e \frac{\rho}{(1-\rho)^2} + o(1),$$

$$L_e = \frac{(3c^4 - 4\gamma c^3 + 6c^2 - 1)\rho^3 + (4\gamma c^3 - 12c^2 + 4)\rho^2 + (6c^2 - 6)\rho}{6}.$$

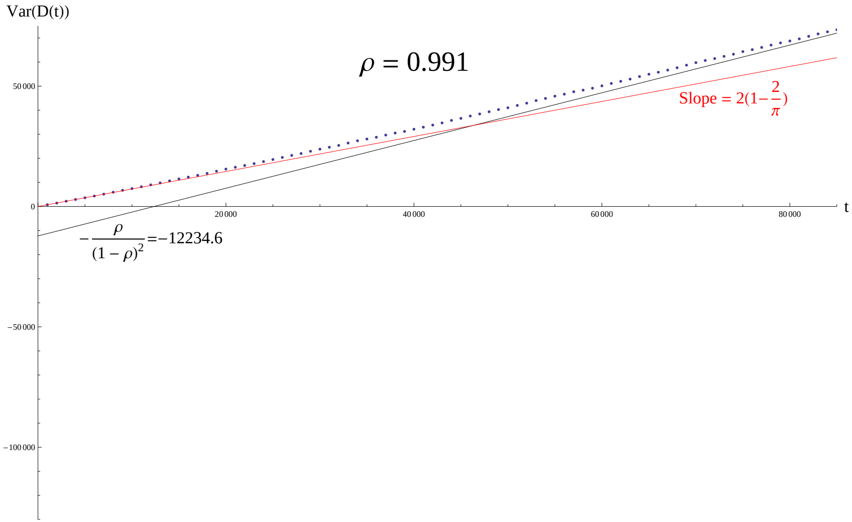
Starting empty version:

$$\mathrm{Var}\big(D(t)\big) = \lambda t - (1 - L_0) \frac{\rho}{(1-\rho)^2} + o(1),$$

$$L_0 = \frac{(3c^4 - 4\gamma c^3 + 6c^2 - 1)\rho^3 + (4\gamma c^3 - 6c^2 - 2)\rho^2 - (6c^2 - 6)\rho}{12}.$$

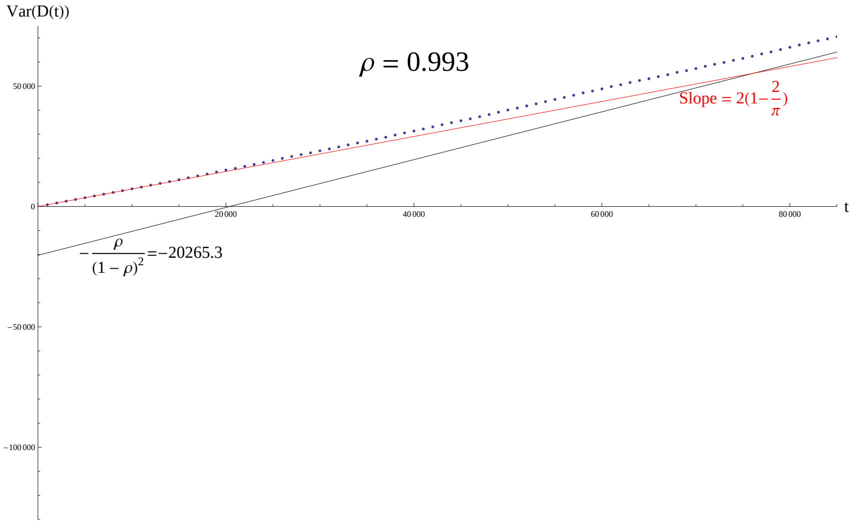
M/M/1: $c^2 = 1, \gamma = 2$. $L_e = 0$, $L_0 = 0$.

M/M/1 Queue



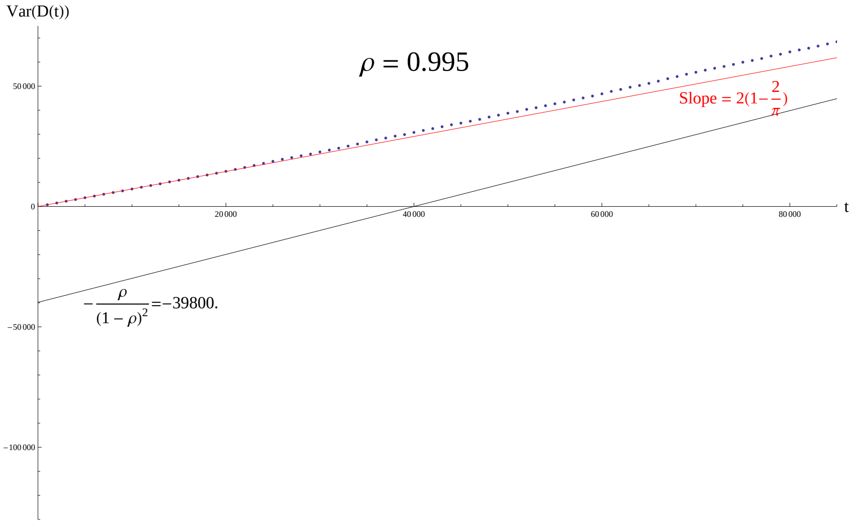
Var(D(t))
$\rho = 0.991$
50 000
0
−50 000
−100 000
20 000
40 000
60 000
80 000
t
Slope = $2(1-\frac{2}{\pi})$
$-\frac{\rho}{(1-\rho)^2}$=−12234.6

M/M/1 Queue

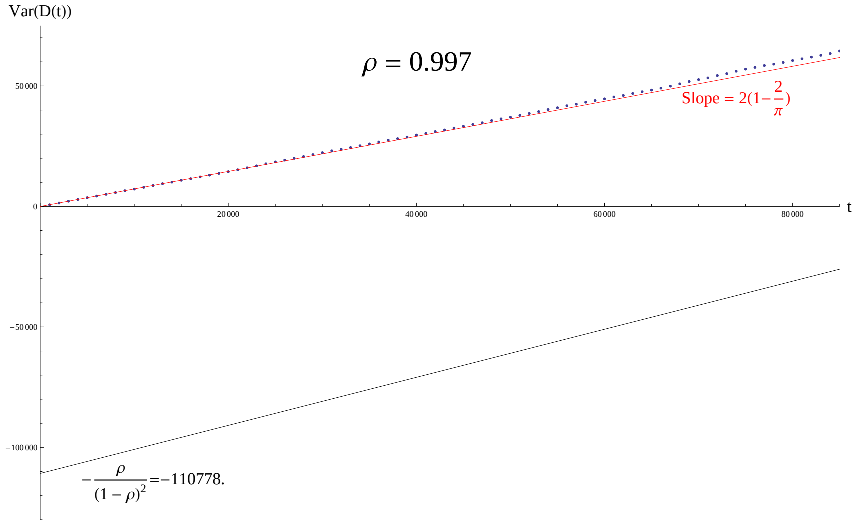


Var(D(t))
$\rho = 0.993$
50 000
0
−50 000
−100 000
20 000
40 000
60 000
80 000
t
Slope = $2(1-\frac{2}{\pi})$
$-\frac{\rho}{(1-\rho)^2}=-20265.3$

M/M/1 Queue


Var(D(t))
$\rho = 0.995$
Slope = $2(1-\frac{2}{\pi})$
$-\frac{\rho}{(1-\rho)^2}$=−39800.
t
50 000
0
−50 000
−100 000
20 000
40 000
60 000
80 000

M/M/1 Queue



Var(D(t))
$\rho = 0.997$
50 000
0
−50 000
−100 000
20 000
40 000
60 000
80 000
t
Slope = $2(1-\frac{2}{\pi})$
$-\frac{\rho}{(1-\rho)^2}$=−110778.

GI/G/1 Queue

Moving away from the memory-less assumptions,

$$\mathcal{D} = \begin{cases} c_a^2, & \lambda < \mu, \\ ?, & \lambda = \mu, \\ c_s^2, & \lambda > \mu. \end{cases}$$

For M/M/1 it was $2(1 - \frac{2}{\pi})$...

GI/G/1 Queue

Moving away from the memory-less assumptions,

$$\mathcal{D} = \begin{cases} c_a^2, & \lambda < \mu, \\ ?, & \lambda = \mu, \\ c_s^2, & \lambda > \mu. \end{cases}$$

For M/M/1 it was $2(1 - \frac{2}{\pi})$...

Theorem:
Ahmad Al-Hanbali, Michel Mandjes, Y.N., Ward Whitt, 2011

For the GI/G/1 queue with $\lambda = \mu$, arbitrary finite second moment initial conditions $\big(Q(0), V(0), U(0)\big)$, finite fourth moments of the inter-arrival and service times, and $\mathbb{P}(B > x) \sim L(x)x^{-1/2}$, where B denotes the busy period and $L(\cdot)$ is a slowly varying function,

$$\mathcal{D} = (c_a^2 + c_s^2)\Big(1 - \frac{2}{\pi}\Big).$$

Proof using diffusion limit of $(D(n\cdot) - \lambda n\cdot)/\sqrt{\lambda n\cdot}$ as $n \to \infty$ (Iglehart and Whitt 1971).

GI/G/1/K Queue

$$\mathcal{D} = \begin{cases} c_a^2 + o_K(1), & \lambda < \mu, \\ ?, & \lambda = \mu, \\ c_s^2 + o_K(1), & \lambda > \mu. \end{cases}$$

For $M/M/1/K$ it was $\frac{2}{3} + o_K(1)$, for $GI/G/1$ it was $(c_a^2 + c_s^2)(1 - \frac{2}{\pi})$...

GI/G/1/K Queue

$$\mathcal{D} = \begin{cases} c_a^2 + o_K(1), & \lambda < \mu, \\ ?, & \lambda = \mu, \\ c_s^2 + o_K(1), & \lambda > \mu. \end{cases}$$

For $M/M/1/K$ it was $\frac{2}{3} + o_K(1)$, for $GI/G/1$ it was $(c_a^2 + c_s^2)(1 - \frac{2}{\pi})$...

Conjecture (numerically tested): Y.N., 2011

For the GI/G/1/K queue with $\lambda = \mu$ and arbitrary initial conditions and light-tailed service and inter-arrival times,

$$\mathcal{D} = (c_a^2 + c_s^2)\frac{1}{3} + O\left(\frac{1}{K}\right).$$

Numerical verification done by representing the system as PH/PH/1/K MAPs

Wrap Up

Summary

Known BRAVO constants:

- Single server finite buffer: $2/3$
  (for GI/G replace 2 by $c_a^2 + c_s^2$)
- Single server infinite buffer $2(1 - 2/\pi)$:
  (for GI/G replace 2 by $c_a^2 + c_s^2$)
- Memoryless many servers finite buffer: $\mathcal{D}_{0,\eta} \in [0.6, 2/3]$

Not yet known:

- Formulas for asymptotic variance when $\rho \neq 1$ in other models
- Memoryless many servers infinite buffer (M/M/s)
- Many servers without memoryless assumptions (GI/G/s)
- Systems with reneging or other packet loss mechanisms
  (e.g. M/M/s/K+M in QED – work in progress)

Other questions: How can BRAVO be harnessed in practice?
Why does BRAVO occur?

References

- Brendan Patch, Thomas Taimre, Y.N., "*A Correction Term for the Covariance of Renewal-Reward Processes with Multivariate Rewards*", submitted.
- Sophie Hautphenne, Yoav Kerner, Y.N., Peter Taylor, "*The Second Order Terms of the Variance Curves for Some Queueing Output Processes*", submitted.
- Y. N., Werner Scheinhardt, "*Diffusion Parameters of Flows in Stable Queueing Networks*", submitted.
- Daryl J. Daley, Johan van Leeuwaarden and Y.N., "*BRAVO for QED Finite Birth-Death Queues*", Advances in Applied Probability, to appear.
- Y.N., "*The variance of departure processes: puzzling behavior and open problems*", Queueing Systems, 68, pp. 385–394, 2011.
- Ahmad Al-Hanbali, Michel Mandjes, Y.N. and Ward Whitt, "*The asymptotic variance of departures in critically loaded queues*", Advances in Applied Probability, 43, pp. 243–263, 2011.
- Y.N. and Gideon Weiss, "*The asymptotic variance rate of the output process of finite capacity birth-death queues*", Queueing Systems, 59, pp. 135–156, 2008.