Rare Event Simulation for Stochastic Networks

Jose Blanchet

Columbia University

June 2010

What is this talk about?

- **General theme of this line of research**

What is this talk about?

- **General theme of this line of research**
- **Design: Light Tails**

What is this talk about?

- **General theme of this line of research**
- **Design: Light Tails**
- **Design: Heavy Tails**

Optimal design of rare event simulation algorithms

- Prof. Varadhan's Abel prize citation on large deviations theory: **"...It has greatly expanded our ability to use computers to analyze rare events."**

Optimal design of rare event simulation algorithms

- Prof. Varadhan's Abel prize citation on large deviations theory: **"...It has greatly expanded our ability to use computers to analyze rare events."**

- **Goal of this line of research: To investigate exactly HOW?**

Before I Answer How: Why Would Anybody Care?

A fast computational engine
enhances our ability to quantify uncertainty
via sensitivity analysis & stress tests...

Performance analysis of rare event simulation algorithms

- $P(A_n) = \exp(-nI + o(n))$ as $n \nearrow \infty$ for $I > 0$.

Performance analysis of rare event simulation algorithms

- $P(A_n) = \exp(-nI + o(n))$ as $n \nearrow \infty$ for $I > 0$.
- **Asymptotic optimality:** Z_n satisfies $EZ_n = P(A_n)$ and

$$\frac{EZ_n^2}{P(A_n)^2} = \mathbf{Com}(n) = \exp(o(n)).$$

Performance analysis of rare event simulation algorithms

- $P(A_n) = \exp(-nI + o(n))$ as $n \nearrow \infty$ for $I > 0$.
- **Asymptotic optimality:** Z_n satisfies $EZ_n = P(A_n)$ and

$$\frac{EZ_n^2}{P(A_n)^2} = \textbf{Com}(n) = \exp(o(n)).$$

- Can still get $\exp(o(n)) = \exp(n^{1/2})$...

Performance analysis of rare event simulation algorithms

- $P(A_n) = \exp(-nI + o(n))$ as $n \nearrow \infty$ for $I > 0$.
- **Asymptotic optimality:** Z_n satisfies $EZ_n = P(A_n)$ and

$$\frac{EZ_n^2}{P(A_n)^2} = \mathbf{Com}(n) = \exp(o(n)).$$

- Can still get $\exp(o(n)) = \exp(n^{1/2})$...
- Sufficient number of replications of Z_n to get ε relative error with $1 - \delta$ confidence:

$$\varepsilon^{-2}\delta^{-1}\mathbf{Com}(n) \quad \text{—> } \textbf{subexponential complexity in } n$$

Performance analysis of rare event simulation algorithms

- $P(A_n) = \exp(-nI + o(n))$ as $n \nearrow \infty$ for $I > 0$.
- **Asymptotic optimality:** Z_n satisfies $EZ_n = P(A_n)$ and

$$\frac{EZ_n^2}{P(A_n)^2} = \mathbf{Com}(n) = \exp(o(n)).$$

- Can still get $\exp(o(n)) = \exp(n^{1/2})$...
- Sufficient number of replications of Z_n to get ε relative error with $1 - \delta$ confidence:

$$\varepsilon^{-2}\delta^{-1}\mathbf{Com}(n)$$ —> **subexponential complexity in** n

- Total cost $\mathbf{TC}(n) = \mathbf{Com}(n) \times$ *Cost per replication*

Tandem Jackson networks

- Two node tandem network (λ, μ_1, μ_2)

Tandem Jackson networks

- Two node tandem network (λ, μ_1, μ_2)
- Exponential inter-arrivals with rate λ

Tandem Jackson networks

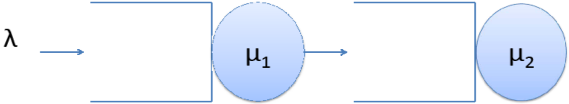
- Two node tandem network (λ, μ_1, μ_2)
- Exponential inter-arrivals with rate λ
- Exponential service times with rate μ_1, μ_2

Tandem Jackson networks

- Two node tandem network (λ, μ_1, μ_2)
- Exponential inter-arrivals with rate λ
- Exponential service times with rate μ_1, μ_2
- Traffic intensity $\rho_1 = \lambda/\mu_1 \in (0, 1)$, $\rho_2 = \lambda/\mu_2 \in (0, 1)$

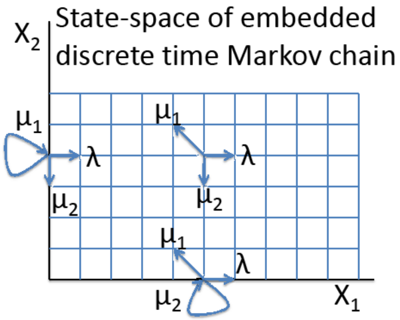
Tandem Jackson networks

- Two node tandem network (λ, μ_1, μ_2)
- Exponential inter-arrivals with rate λ
- Exponential service times with rate μ_1, μ_2
- Traffic intensity $\rho_1 = \lambda/\mu_1 \in (0,1)$, $\rho_2 = \lambda/\mu_2 \in (0,1)$
- P_0[Total population reaches n in a busy period]



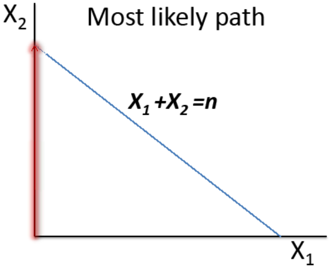
Tandem Jackson networks

- $X_1(k) = \#$ in station 1 at transition k in embedded discrete time Markov chain
- $X_2(k) = \#$ in station 2 at transition k in embedded discrete time Markov chain
- Assume $\lambda + \mu_1 + \mu_2 = 1$



Intuitive sampling & counter-examples

- Assume that $\rho_1 < \rho_2 < 1$ –> 2nd is bottleneck –> $\lambda < \mu_2 < \mu_1$
- Large deviations theory says: **"Most likely path in fluid scale looks like that of system** (μ_2, μ_1, λ)"



Intuitive sampling & counter-examples

- Natural importance sampling —> simulate (μ_2, μ_1, λ) system

Intuitive sampling & counter-examples

- Natural importance sampling —> simulate (μ_2, μ_1, λ) system
- Let $\widetilde{K}\left(\left(x_1, x_2\right), \left(y_1, y_2\right)\right)$ be the transition matrix from (μ_2, μ_1, λ) system

Intuitive sampling & counter-examples

- Natural importance sampling —> simulate (μ_2, μ_1, λ) system
- Let $\widetilde{K}\left((x_1, x_2), (y_1, y_2)\right)$ be the transition matrix from (μ_2, μ_1, λ) system
- Let $K\left((x_1, x_2), (y_1, y_2)\right)$ be the transition matrix from (λ, μ_1, μ_2) system

Intuitive sampling & counter-examples

- Natural importance sampling —> simulate (μ_2, μ_1, λ) system
- Let $\widetilde{K}\left(\left(x_1, x_2\right), \left(y_1, y_2\right)\right)$ be the transition matrix from (μ_2, μ_1, λ) system
- Let $K\left(\left(x_1, x_2\right), \left(y_1, y_2\right)\right)$ be the transition matrix from (λ, μ_1, μ_2) system
- Importance sampling estimator

$$I\left(T_n < T_{\{0\}}\right) \prod_{k=0}^{T_n-1} \frac{K\left(\left(X_1(k), X_2(k)\right), \left(X_1(k+1), X_2(k+1)\right)\right)}{\widetilde{K}\left(\left(X_1(k), X_2(k)\right), \left(X_1(k+1), X_2(k+1)\right)\right)}$$

Intuitive sampling & counter-examples

- Natural importance sampling —> simulate (μ_2, μ_1, λ) system
- Let $\widetilde{K}((x_1, x_2), (y_1, y_2))$ be the transition matrix from (μ_2, μ_1, λ) system
- Let $K((x_1, x_2), (y_1, y_2))$ be the transition matrix from (λ, μ_1, μ_2) system
- Importance sampling estimator

$$I\left(T_n < T_{\{0\}}\right) \prod_{k=0}^{T_n - 1} \frac{K\left(\left(X_1(k), X_2(k)\right), \left(X_1(k+1), X_2(k+1)\right)\right)}{\widetilde{K}\left(\left(X_1(k), X_2(k)\right), \left(X_1(k+1), X_2(k+1)\right)\right)}$$

- $T_n =$ 1st step k such that $X_1(k) + X_2(k) = n$.

Intuitive sampling & counter-examples

- Natural importance sampling —> simulate (μ_2, μ_1, λ) system
- Let $\widetilde{K}((x_1, x_2), (y_1, y_2))$ be the transition matrix from (μ_2, μ_1, λ) system
- Let $K((x_1, x_2), (y_1, y_2))$ be the transition matrix from (λ, μ_1, μ_2) system
- Importance sampling estimator

$$I\left(T_n < T_{\{0\}}\right) \prod_{k=0}^{T_n - 1} \frac{K\left(\left(X_1(k), X_2(k)\right), \left(X_1(k+1), X_2(k+1)\right)\right)}{\widetilde{K}\left(\left(X_1(k), X_2(k)\right), \left(X_1(k+1), X_2(k+1)\right)\right)}$$

- T_n = 1st step k such that $X_1(k) + X_2(k) = n$.
- $T_{\{0\}}$ = 1st **return** time to origin.

Intuitive sampling & counter-examples

- Glasserman & Kou '95:

  Previous sampler can even yield infinite variance!

Intuitive sampling & counter-examples

- Glasserman & Kou '95:

  Previous sampler can even yield infinite variance!

- Reason: Likelihood ratio very poorly behaved when process reaches T_n OUTSIDE most likely region!

Summary of the introduction

- **Large deviations helps computers BUT much more than direct interpretation is needed!**

Importance Sampling for Jackson Networks

- Dupuis, Sezer & Wang '07: First asymptotically optimal importance sampling for total population overflow in tandem networks

Importance Sampling for Jackson Networks

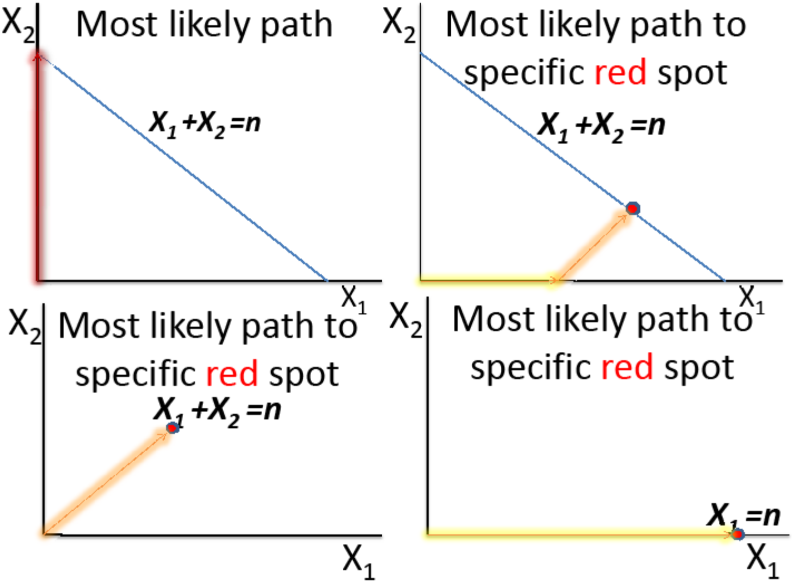
- Dupuis, Sezer & Wang '07: First asymptotically optimal importance sampling for total population overflow in tandem networks
- Dupuis & Wang '09: First asymptotically optimal importance sampling for **any overflow** event in **any open Jackson network**

Importance Sampling for Jackson Networks

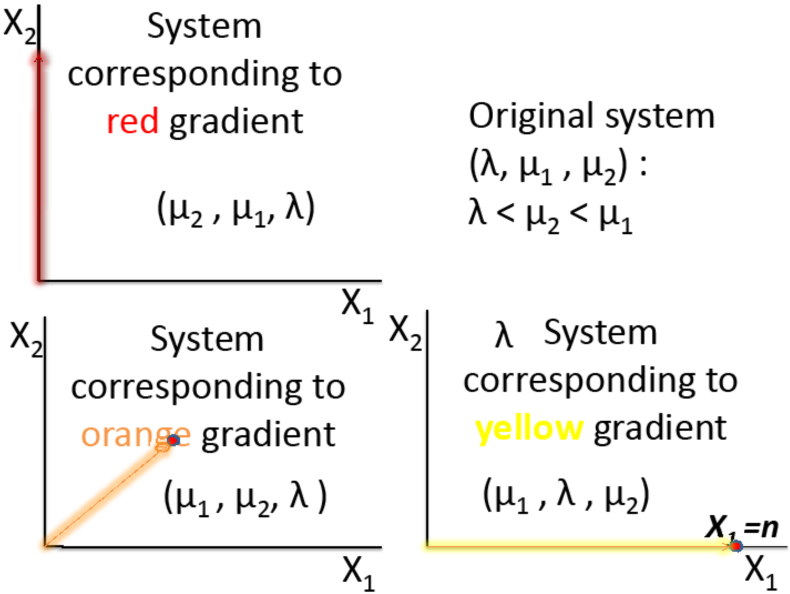
- Dupuis, Sezer & Wang '07: First asymptotically optimal importance sampling for total population overflow in tandem networks
- Dupuis & Wang '09: First asymptotically optimal importance sampling for **any overflow** event in **any open Jackson network**

- Note: these results guarantee only **subexponential complexity** (i.e. **Com**$(n) = \exp(o(n))$ replications)

Finitely many gradients



Systems corresponding to gradients



Randomized selection of gradients at each step

- At each step randomize selection of gradient depending on current state

Randomized selection of gradients at each step

- At each step randomize selection of gradient depending on current state
- The *i*-th gradient at position y is selected with a probability proportional to

$$w_i\left(y\right) \approx \exp\left(-n[a_i - \theta_i^T y]\right)$$

Randomized selection of gradients at each step

- At each step randomize selection of gradient depending on current state
- The i-th gradient at position y is selected with a probability proportional to

$$w_i(y) \approx \exp\left(-n[a_i - \theta_i^T y]\right)$$

- Weight corresponds to the probability of "exiting" through the corresponding point induced point by the gradient assuming one is on the corresponding fluid path at position y

Randomized selection of gradients at each step

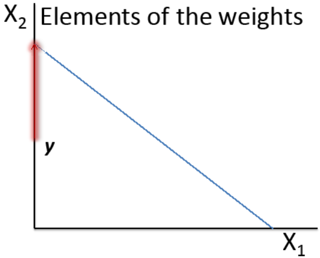
- At each step randomize selection of gradient depending on current state
- The i-th gradient at position y is selected with a probability proportional to

$$w_i\,(y) \approx \exp\left(-n[a_i - \theta_i^T y]\right)$$

- Weight corresponds to the probability of "exiting" through the corresponding point induced point by the gradient assuming one is on the corresponding fluid path at position y
- NOTE: One applies the same rule EVEN if we are not in the corresponding fluid path!

Randomized selection of gradients at each step

$$w_i(y) \approx \exp\left(-n[a_i - \theta_i^T y]\right)$$



Importance Sampling for Jackson Networks

Theorem (B., Glynn and Leder 2010)

A slight variation of the algorithm by Dupuis, Sezer & Wang '07 satisfies

$$\textit{Total cost} = \mathbf{TC}\,(n) = O\left(n^{2(d+1-\beta)}\right),$$

where $\beta = \#$ *of bottleneck stations.and* $d = \#$ *of stations.*

- **Highlights of proof:**

Importance Sampling for Jackson Networks

Theorem (B., Glynn and Leder 2010)

A slight variation of the algorithm by Dupuis, Sezer & Wang '07 satisfies

$$\textit{Total cost} = \mathbf{TC}(n) = O\left(n^{2(d+1-\beta)}\right),$$

where $\beta = \#$ *of bottleneck stations.and* $d = \#$ *of stations.*

- **Highlights of proof:**
- **A) Translate subsolution into an appropriate Lyapunov inequality**

Importance Sampling for Jackson Networks

Theorem (B., Glynn and Leder 2010)

A slight variation of the algorithm by Dupuis, Sezer & Wang '07 satisfies

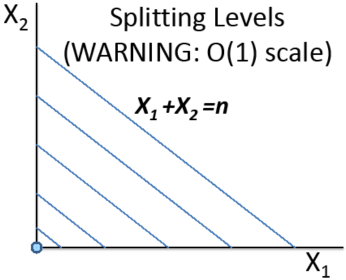
$$\textit{Total cost} = \mathbf{TC}\left(n\right) = O\left(n^{2(d+1-\beta)}\right),$$

where $\beta = \#$ *of bottleneck stations.and* $d = \#$ *of stations.*

- **Highlights of proof:**
- **A) Translate subsolution into an appropriate Lyapunov inequality**
- **B) Insight into selection of various "mollification" parameters**

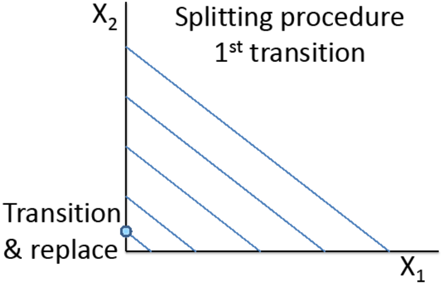
Splitting for Jackson Networks

- Splitting levels



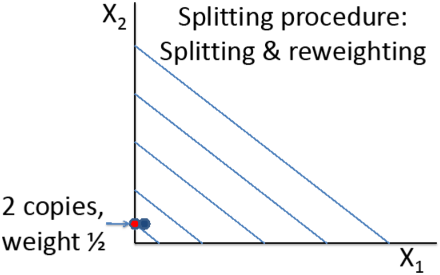
Splitting for Jackson Networks

- First transition



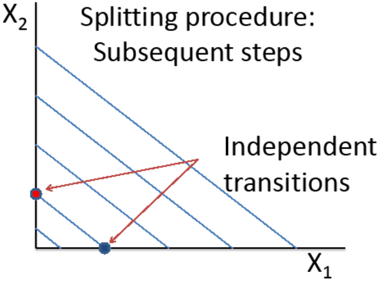
Splitting for Jackson Networks

- Replacement by identical copies & reweighting



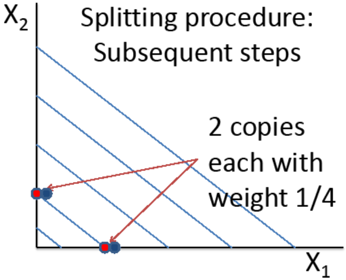
Splitting for Jackson Networks

- Subsequent transitions: advancing



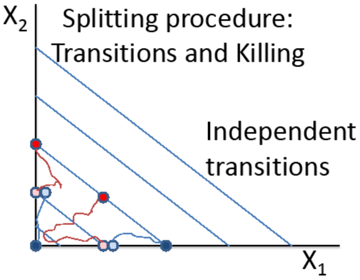
Splitting for Jackson Networks

- Subsequent transitions: reweighting



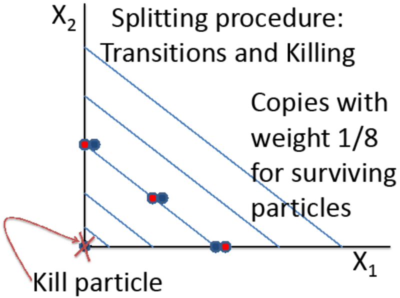
Splitting for Jackson Networks

- One more transition and killing



Splitting for Jackson Networks

- One more transition and killing



Splitting for Jackson Networks

- Say want to compute $P_x(T_n < T_0)$

Splitting for Jackson Networks

- Say want to compute $P_x\left(T_n < T_0\right)$
- Single replication of estimator, starting at x (previous illustration $x = 0$)

$$N_n\left(x\right) / r^{l_n(x)}$$

Splitting for Jackson Networks

- Say want to compute $P_x\left(T_n < T_0\right)$
- Single replication of estimator, starting at x (previous illustration $x = 0$)

$$N_n\left(x\right)/r^{l_n(x)}$$

- $N_n\left(x\right) = \#$ of particles that make it to total population $= n$

Splitting for Jackson Networks

- Say want to compute $P_x(T_n < T_0)$
- Single replication of estimator, starting at x (previous illustration $x = 0$)

$$N_n(x) / r^{l_n(x)}$$

- $N_n(x) = \#$ of particles that make it to total population $= n$
- $l_n(x) = \#$ of levels to reach total population $= n$

How to select the levels

- Dupuis & Dean '09 (general), Villen-Altamirano '09 (conjecture Jackson)

How to select the levels

- Dupuis & Dean '09 (general), Villen-Altamirano '09 (conjecture Jackson)
- Heuristic: Controlling total expected # of particles yields

$$r^{l_n(x)} P_x \left(T_n < T_0 \right) = \exp \left(o \left(n \right) \right)$$

How to select the levels

- Dupuis & Dean '09 (general), Villen-Altamirano '09 (conjecture Jackson)
- Heuristic: Controlling total expected # of particles yields

$$r^{l_n(x)} P_x \left(T_n < T_0\right) = \exp\left(o\left(n\right)\right)$$

- If $P_x \left(T_n < T_0\right) \leq \exp\left(-n[a - x^T \theta_*]\right)$

$$l_n\left(x\right) \log r - n[x^T \theta_* - a] = o\left(n\right)$$

How to select the levels

- Dupuis & Dean '09 (general), Villen-Altamirano '09 (conjecture Jackson)
- Heuristic: Controlling total expected # of particles yields

$$r^{l_n(x)} P_x \left(T_n < T_0 \right) = \exp \left(o \left(n \right) \right)$$

- If $P_x \left(T_n < T_0 \right) \leq \exp \left(-n[a - x^T \theta_*] \right)$

$$l_n \left(x \right) \log r - n[x^T \theta_* - a] = o \left(n \right)$$

- Pick $l_n \left(x \right) = n[x^T \theta_* - a] / \log(r)$

How to select the levels

- Dupuis & Dean '09 (general), Villen-Altamirano '09 (conjecture Jackson)
- Heuristic: Controlling total expected # of particles yields

$$r^{l_n(x)} P_x \left(T_n < T_0 \right) = \exp\left(o\left(n \right) \right)$$

- If $P_x \left(T_n < T_0 \right) \leq \exp\left(-n[a - x^T \theta_*] \right)$

$$l_n\left(x \right) \log r - n[x^T \theta_* - a] = o\left(n \right)$$

- Pick $l_n\left(x \right) = n[x^T \theta_* - a] / \log(r)$

- Conclusion: Select splitting levels according to level curves of a VERY SPECIFIC linear function

Particles and Large Deviations Analysis

Theorem (B., Leder and Shi '09)

The large deviations splitting rule of Dupuis & Dean '09 has complexity

$$\mathbf{TC}\,(n) = O\left(n^{2\beta+1}\right)$$

$\beta = \#$ *of bottleneck stations.*

- Key idea: understand conditional distribution of network at subsequent milestone events!

Summary

- **Guaranteed cost of well selected state-dependent samplers (tandem)** $O\left(n^{2(d-\beta+1)}\right)$

Summary

- **Guaranteed cost of well selected state-dependent samplers (tandem)** $O\left(n^{2(d-\beta+1)}\right)$
- **Guaranteed cost of splitting (general)** $O\left(n^{2\beta+1}\right)$

Summary

- **Guaranteed cost of well selected state-dependent samplers (tandem)** $O\left(n^{2(d-\beta+1)}\right)$
- **Guaranteed cost of splitting (general)** $O\left(n^{2\beta+1}\right)$
- Keep in mind that benchmark is solving linear system with $O\left(n^{d}\right)$ unknowns!

$$(Ku)(x) = u(x) \quad \text{subject to} \quad u(x) = 1 \quad x \in A_n$$

Generalities

- Nicola & Juneja ‘05, Anantharam et al ‘90

Generalities

- Nicola & Juneja '05, Anantharam et al '90
- Keep in mind for the moment 2 node tandem network

Generalities

- Nicola & Juneja '05, Anantharam et al '90
- Keep in mind for the moment 2 node tandem network
- Notation $x = (x(1), x(2))$, $x_0 = (x_0(1), x_0(2))$, similarly x_k & y...

Generalities

- Nicola & Juneja '05, Anantharam et al '90
- Keep in mind for the moment 2 node tandem network
- Notation $x = (x(1), x(2))$, $x_0 = (x_0(1), x_0(2))$, similarly x_k & y...
- Underlying Markov transition matrix $K(x, y)$, steady-state distribution $\pi(x)$

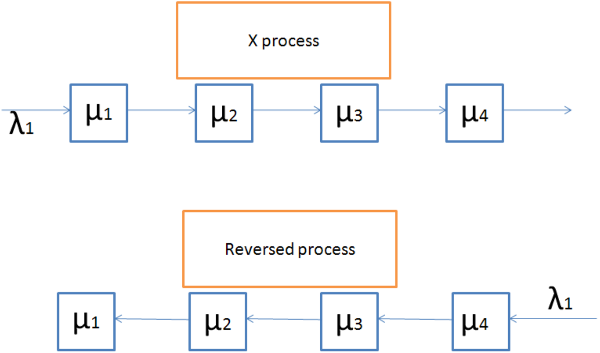
Generalities

- Nicola & Juneja '05, Anantharam et al '90
- Keep in mind for the moment 2 node tandem network
- Notation $x = (x(1), x(2))$, $x_0 = (x_0(1), x_0(2))$, similarly x_k & y...
- Underlying Markov transition matrix $K(x, y)$, steady-state distribution $\pi(x)$
- Time-reversed Markov chain

$$K^{\leftarrow}(y, x) = \frac{\pi(x) K(x, y)}{\pi(y)}$$

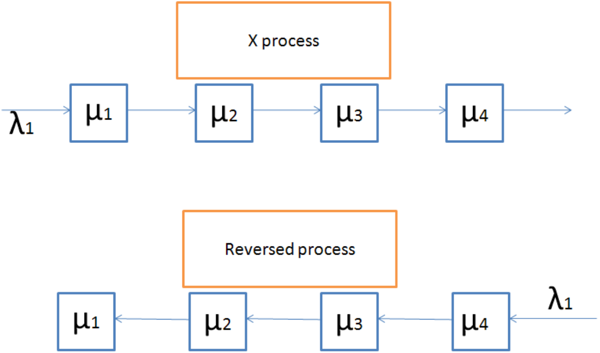
Time reversed characteristics for tandem networks

- Time reversed Jackson networks (example tandem)



Time reversed characteristics for tandem networks

- Time reversed Jackson networks (example tandem)



- Steady-state numbers are independent Geometrics with mean $\rho_i/(1-\rho_i)$

A general identity

- T_n = 1st time to reach total population = n

A general identity

- T_n = 1st time to reach total population = n
- $T_{\{0\}}$ = 1st return time to origin

A general identity

- T_n = 1st time to reach total population = n
- $T_{\{0\}}$ = 1st return time to origin
- $T\{x\}$ = 1st time to hit x

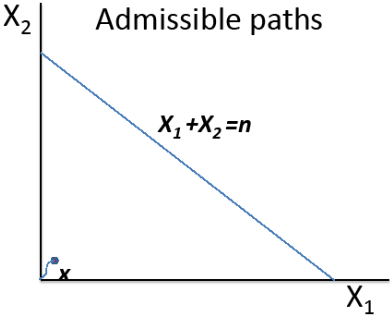
$$P_0\left(T_n < T_{\{0\}}, T_{\{x\}} < T_n \wedge T_{\{0\}}\right)$$

$$= \sum_{\substack{admissible \\ paths:\{x_0 \to x_1 \to \ldots \to x_{T_n}\}}} \{ \underbrace{K\left(x_0, x_1\right)}_{x_0 = 0} \underset{\text{region \& NOT} = x}{\underset{\text{continuation}}{\times \ldots \times}}$$

$$\underbrace{K\left(x_{T_{\{x\}-1}}, x\right) K\left(x, x_{T_{\{x\}+1}}\right)}_{x = x_{T_{\{x\}}}} \underset{\text{region}}{\underset{\text{cont.}}{\times \ldots \times}} \underbrace{K\left(x_{T_n - 1}, x_{T_n}\right)}_{x_{T_n} \in \text{population} = n}$$

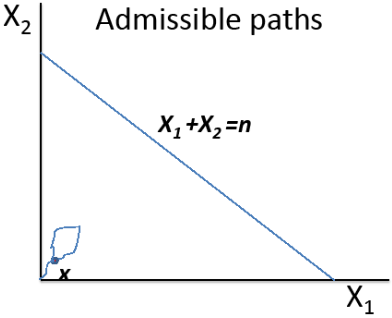
Admissible paths

- Section of admissible path reaching x



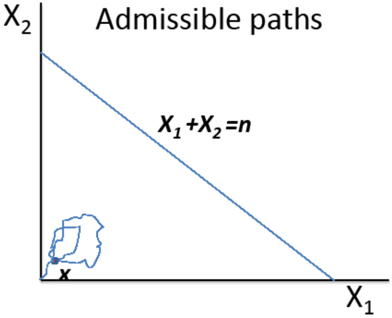
Admissible paths

- Section of admissible path after reaching x



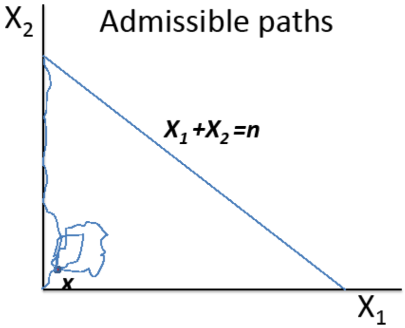
Admissible paths

- Section of admissible path after reaching x



Admissible paths

- Section of admissible path after reaching x



Towards a useful identity

$$
\begin{aligned}
& P_0\left(T_n < T_{\{0\}}, T_{\{x\}} < T_n \wedge T_{\{0\}}\right) \\
= & \frac{1}{\pi(x_0)} \sum_{\substack{\text{admissible} \\ \text{paths}}} \underbrace{\pi\left(x_0\right) K\left(x_0, x_1\right)}_{K^{\leftarrow}\left(x_1, x_0\right) \pi\left(x_1\right)} \times \ldots \times K\left(x_{T_n-1}, x_{T_n}\right) \\
= & \frac{1}{\pi(x_0)} \sum_{\substack{\text{admissible} \\ \text{paths}}} K^{\leftarrow}\left(x_1, x_0\right) \times \ldots \times K^{\leftarrow}\left(x_{T_n-1}, x_{T_n-1}\right) \pi\left(x_{T_n}\right) \\
= & \frac{1}{\pi(x_0)} \sum_{\substack{\text{admissible} \\ \text{paths}}} \pi\left(x_0^{\leftarrow}\right) K^{\leftarrow}\left(x_0^{\leftarrow}, x_1^{\leftarrow}\right) \times \ldots \times K^{\leftarrow}\left(x_{T_n-1}^{\leftarrow}, x_{T_n}^{\leftarrow}\right)
\end{aligned}
$$

- Where $x_k^{\leftarrow} = x_{T_n-k}$ for $k = 0, 1, \ldots, T_n$

Towards a useful identity

- What are the admissible paths in terms of the $\overleftarrow{x_k}$'s?

Towards a useful identity

- What are the admissible paths in terms of the $x_k^{\leftarrow}$'s?
- $T_{\{0\}}^{\leftarrow}$ = 1st time to hit the origin

Towards a useful identity

- What are the admissible paths in terms of the $x_k^{\leftarrow}$'s?
- $T_{\{0\}}^{\leftarrow} =$ 1st time to hit the origin
- $T_n^{\leftarrow} =$ 1st return time to get total population $= n$

Towards a useful identity

- What are the admissible paths in terms of the $x_k^{\leftarrow}$'s?
- $T_{\{0\}}^{\leftarrow}$ = 1st time to hit the origin
- $T_n^{\leftarrow}$ = 1st return time to get total population $= n$
- $T_{\{x\}}^{\leftarrow}$ = 1st hitting time to x

Towards a useful identity

- What are the admissible paths in terms of the $x_k^{\leftarrow}$'s?
- $T_{\{0\}}^{\leftarrow}$ = 1st time to hit the origin
- $T_n^{\leftarrow}$ = 1st return time to get total population = n
- $T_{\{x\}}^{\leftarrow}$ = 1st hitting time to x
- Admissible paths satisfy

$$T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}, T_{\{x\}}^{\leftarrow} < T_n^{\leftarrow} \wedge T_{\{0\}}^{\leftarrow}$$

Towards a useful identity

- What are the admissible paths in terms of the $x_k^{\leftarrow}$'s?
- $T_{\{0\}}^{\leftarrow}$ = 1st time to hit the origin
- $T_n^{\leftarrow}$ = 1st return time to get total population = n
- $T_{\{x\}}^{\leftarrow}$ = 1st hitting time to x
- Admissible paths satisfy

$$T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}, T_{\{x\}}^{\leftarrow} < T_n^{\leftarrow} \wedge T_{\{0\}}^{\leftarrow}$$

- Admissible starting position

$$y \in D_n := \{y \in population = n \text{ \& } P_y[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}] > 0\}$$

A useful identity

Lemma

$$
\begin{aligned}
& P_0\left(T_{\{x\}} < T_n \wedge T_{\{0\}}\right) P_x\left(T_n < T_0\right) \\
= \; & P_0\left(T_n < T_{\{0\}}, T_{\{x\}} < T_n \wedge T_{\{0\}}\right) \\
= \; & \frac{1}{\pi(0)} E_\pi[P_{X_0^{\leftarrow}}[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}, T_{\{x\}}^{\leftarrow} < T_n^{\leftarrow} \wedge T_{\{0\}}^{\leftarrow}] I\left(X_0^{\leftarrow} \in D_n\right)]
\end{aligned}
$$

A useful identity

Conclusion:

$$P_x\left(T_n < T_0\right) = \frac{E_\pi[P_{X_0^{\leftarrow}}[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}, T_{\{x\}}^{\leftarrow} < T_n^{\leftarrow} \wedge T_{\{0\}}^{\leftarrow}]I\left(X_0^{\leftarrow} \in D_n\right)]}{\pi\left(0\right) P_0\left(T_{\{x\}} < T_n \wedge T_{\{0\}}\right)}$$

- Note $P_0\left(T_{\{x\}} < T_n \wedge T_{\{0\}}\right) \geq P_0\left(T_{\{x\}} < T_{\{0\}}\right) > 0$ if $x = O\left(1\right)$.

A useful identity

Conclusion:

$$P_x\left(T_n < T_0\right) = \frac{E_\pi[P_{X_0^{\leftarrow}}[T^{\leftarrow}_{\{0\}} < T^{\leftarrow}_n, T^{\leftarrow}_{\{x\}} < T^{\leftarrow}_n \wedge T^{\leftarrow}_{\{0\}}] I\left(X_0^{\leftarrow} \in D_n\right)]}{\pi\left(0\right) P_0\left(T_{\{x\}} < T_n \wedge T_{\{0\}}\right)}$$

- Note $P_0\left(T_{\{x\}} < T_n \wedge T_{\{0\}}\right) \geq P_0\left(T_{\{x\}} < T_{\{0\}}\right) > 0$ if $x = O\left(1\right)$.
- Focus on $x = 0$ for simplicity, then $P_0\left(T_{\{x\}} < T_n \wedge T_{\{0\}}\right) = 1$.

Implications of the identity for simulation

$$P_0\left(T_n < T_0\right) = \frac{E_\pi\left(P_{X_0^{\leftarrow}}\left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} | X_0^{\leftarrow} \in D_n\right]\right) P_\pi\left(X_0^{\leftarrow} \in D_n\right)}{\pi\left(0\right)}$$

- Suppose:

Implications of the identity for simulation

$$P_0\left(T_n < T_0\right) = \frac{E_\pi\left(P_{X_0^{\leftarrow}}\left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} \middle| X_0^{\leftarrow} \in D_n\right]\right) P_\pi\left(X_0^{\leftarrow} \in D_n\right)}{\pi\left(0\right)}$$

- Suppose:

1. Can estimate $P_\pi\left(X_0^{\leftarrow} \in D_n\right)$ in $O\left(1\right)$ time

Implications of the identity for simulation

$$P_0\left(T_n < T_0\right) = \frac{E_\pi\left(P_{X_0^{\leftarrow}}\left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} \middle| X_0^{\leftarrow} \in D_n\right]\right) P_\pi\left(X_0^{\leftarrow} \in D_n\right)}{\pi\left(0\right)}$$

- Suppose:

1. Can estimate $P_\pi\left(X_0^{\leftarrow} \in D_n\right)$ in $O\left(1\right)$ time
2. Can sample $X_0^{\leftarrow} | X_0^{\leftarrow} \in D_n$ under $\pi\left(\cdot\right)$ in $O\left(1\right)$ time

Implications of the identity for simulation

$$P_0\left(T_n < T_0\right) = \frac{E_\pi\left(P_{X_0^{\leftarrow}}\left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} | X_0^{\leftarrow} \in D_n\right]\right) P_\pi\left(X_0^{\leftarrow} \in D_n\right)}{\pi\left(0\right)}$$

- Suppose:

1. Can estimate $P_\pi\left(X_0^{\leftarrow} \in D_n\right)$ in $O\left(1\right)$ time
2. Can sample $X_0^{\leftarrow} | X_0^{\leftarrow} \in D_n$ under $\pi\left(\cdot\right)$ in $O\left(1\right)$ time
3. Can simulate $\{X_k^{\leftarrow} : k \leq T_0^{-}\}$ on $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}$ given $X_0^{\leftarrow} \in D_n$ in $O\left(n\right)$ time

Implications of the identity for simulation

$$P_0\left(T_n < T_0\right) = \frac{E_\pi(P_{X_0^\leftarrow}[T_{\{0\}}^\leftarrow < T_n^\leftarrow | X_0^\leftarrow \in D_n])P_\pi\left(X_0^\leftarrow \in D_n\right)}{\pi\left(0\right)}$$

- Suppose:

1. Can estimate $P_\pi\left(X_0^\leftarrow \in D_n\right)$ in $O\left(1\right)$ time
2. Can sample $X_0^\leftarrow | X_0^\leftarrow \in D_n$ under $\pi\left(\cdot\right)$ in $O\left(1\right)$ time
3. Can simulate $\{X_k^\leftarrow : k \leq T_0^-\}$ on $T_{\{0\}}^\leftarrow < T_n^\leftarrow$ given $X_0^\leftarrow \in D_n$ in $O\left(n\right)$ time

- **Conclusion: Can estimate $P_0\left(T_n < T_0\right)$ in $O\left(n\right)$ time**

General identity for Jackson networks

- How to simulate $\{X_k^{\leftarrow} : k \leq T_0^{-}\}$ on $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}$ given $X_0^{\leftarrow} \in D_n$ in $O(n)$ time?

Lemma

$$\inf_{y \in D_n} P_y \left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} \right] > 0$$

General identity for Jackson networks

- How to simulate $\{X_k^{\leftarrow} : k \leq T_0^{-}\}$ on $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}$ given $X_0^{\leftarrow} \in D_n$ in $O(n)$ time?

Lemma

$$\inf_{y \in D_n} P_y\left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}\right] > 0$$

- Key steps in the proof:

General identity for Jackson networks

- How to simulate $\{X_k^{\leftarrow} : k \leq T_0^{-}\}$ on $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}$ given $X_0^{\leftarrow} \in D_n$ in $O(n)$ time?

Lemma

$$\inf_{y \in D_n} P_y \left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} \right] > 0$$

- Key steps in the proof:

1. Intersect with event that K people live before any arrival so $||y||_1 = n - K$

General identity for Jackson networks

- How to simulate $\{X_k^{\leftarrow} : k \leq T_0^{-}\}$ on $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}$ given $X_0^{\leftarrow} \in D_n$ in $O(n)$ time?

Lemma

$$\inf_{y \in D_n} P_y[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}] > 0$$

- Key steps in the proof:

1. Intersect with event that K people live before any arrival so $||y||_1 = n - K$
2. Exists compact set C & $m, a > 0$ such that $\sup_{y \notin C}[E_y[||X_m^{\leftarrow}||_1 - ||y||_1] \leq -a < 0$

General identity for Jackson networks

- How to simulate $\{X_k^{\leftarrow} : k \leq T_0^{-}\}$ on $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}$ given $X_0^{\leftarrow} \in D_n$ in $O(n)$ time?

Lemma

$$\inf_{y \in D_n} P_y\left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}\right] > 0$$

- Key steps in the proof:

1. Intersect with event that K people live before any arrival so $||y||_1 = n - K$
2. Exists compact set C & $m, a > 0$ such that $\sup_{y \notin C}\left[E_y\left[||X_m^{\leftarrow}||_1 - ||y||_1\right] \leq -a < 0\right.$
3. Use Chernoff's bound and a union bound argument to show that if $||y||_1 = n - K$ for K large enough

$$P_y\left[T_n^{\leftarrow} < T_{\{0\}}^{\leftarrow}\right] \leq \varepsilon$$

General identity for Jackson networks

- How to simulate $\{X_k^{\leftarrow} : k \leq T_0^{-}\}$ on $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}$ given $X_0^{\leftarrow} \in D_n$ in $O(n)$ time?

Lemma

$$\inf_{y \in D_n} P_y \left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} \right] > 0$$

Conclusion:

General identity for Jackson networks

- How to simulate $\{X_k^{\leftarrow} : k \leq T_0^{-}\}$ on $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}$ given $X_0^{\leftarrow} \in D_n$ in $O(n)$ time?

Lemma

$$\inf_{y \in D_n} P_y[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow}] > 0$$

Conclusion:

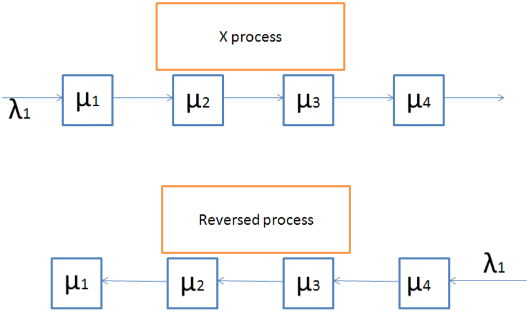
- $E_\pi(P_{X_0^{\leftarrow}}[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} | X_0^{\leftarrow} \in D_n])$ is easily estimated if we know how to sample $X_0^{\leftarrow} | X_0^{\leftarrow} \in D_n$ under π

Conditional steady-state distribution

- How to estimate $P_{\pi}\left(X_0^{\leftarrow} \in D_n\right)$ and sample $X_0^{\leftarrow} | X_0^{\leftarrow} \in D_n$ under $\pi\left(\cdot\right)$ in $O\left(1\right)$ time?

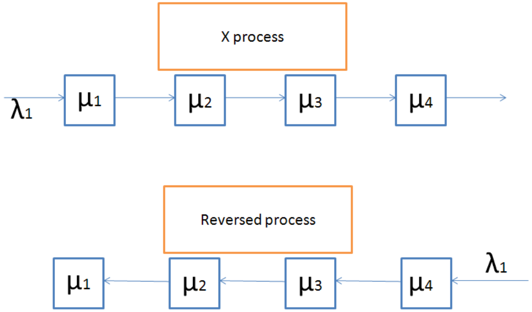
Conditional steady-state distribution

- How to estimate $P_{\pi}\left(X_0^{\leftarrow} \in D_n\right)$ and sample $X_0^{\leftarrow} | X_0^{\leftarrow} \in D_n$ under $\pi(\cdot)$ in $O(1)$ time?
- Observation $y \in D_n$ IF AND ONLY IF $||y|| = n$ AND at least one "receiving stations" is NON-empty



Conditional steady-state distribution

- How to estimate $P_\pi\left(X_0^{\leftarrow} \in D_n\right)$ and sample $X_0^{\leftarrow} | X_0^{\leftarrow} \in D_n$ under $\pi\left(\cdot\right)$ in $O\left(1\right)$ time?
- Observation $y \in D_n$ IF AND ONLY IF $||y|| = n$ AND at least one "receiving stations" is NON-empty



- So concentrate on $X_0^{-}\left(1\right) + \ldots + X_0^{-}\left(d\right) = n -$ *receiving stations*

Conditional steady-state distribution

- How to estimate $P\left(X_0^-(1) + ... + X_0^-(d) = n\right)$, $X_0^-(j)$ independent Geometrics (maybe different parameters)?

Conditional steady-state distribution

- How to estimate $P\left(X_0^-(1)+\ldots+X_0^-(d)=n\right)$, $X_0^-(j)$ independent Geometrics (maybe different parameters)?
- First try: Exponential tilting!

Conditional steady-state distribution

- How to estimate $P\left(X_0^-(1)+\ldots+X_0^-(d)=n\right)$, $X_0^-(j)$ independent Geometrics (maybe different parameters)?
- First try: Exponential tilting!
- Well... good (subexponential complexity) but NOT $O(1)$

Conditional steady-state distribution

Lemma

It is possible to design a sequential importance sampling estimator for

$$P\left(M_1 + \ldots + M_l = n\right)$$

with likelihood ratio $O\left(P\left(M_1 + \ldots + M_l = n\right)\right)$, *where* M_i*'s are independent negative binomial. Thus, one obtains both a strongly efficient estimator and a conditional sampler both with* $O\left(1\right)$ *complexity.*

- Family of samplers: Sort from heavier to lighter tail & use mixtures

$$p\left(i, n-s\right) P\left(M_i = k \middle| M_i \geq n - s_i\right)$$
$$+q\left(i, n-s\right)\right) P\left(M_i = n - s_i - k \middle| M_i \leq n - s_i\right)$$

Conditional steady-state distribution

Lemma

It is possible to design a sequential importance sampling estimator for

$$P\left(M_1 + \ldots + M_l = n\right)$$

with likelihood ratio $O\left(P\left(M_1 + \ldots + M_l = n\right)\right)$*, where* M_i*'s are independent negative binomial. Thus, one obtains both a strongly efficient estimator and a conditional sampler both with* $O\left(1\right)$ *complexity.*

- Family of samplers: Sort from heavier to lighter tail & use mixtures

$$p\left(i, n-s\right) P\left(M_i = k \mid M_i \geq n - s_i\right)$$
$$+q\left(i, n-s\right)) P\left(M_i = n - s_i - k \mid M_i \leq n - s_i\right)$$

- Select probabilities $p\left(i, n-s\right)$ sequentially where $s = M_1 + \ldots + M_{i-1}$

Main result

Theorem (B. 2010)

Given a Jackson network one can estimate

$$P_0\left(T_n < T_0\right)$$

and simulate conditional paths given $T_n < T_0$ *with optimal running time (*$O\left(n\right)$ *complexity).*

Identity

$$\pi\left(0\right)P_0\left(T_n < T_0\right) = E_\pi\left(P_{X_0^{\leftarrow}}\left[T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} | X_0^{\leftarrow} \in D_n\right]\right)P_\pi\left(X_0^{\leftarrow} \in D_n\right)$$

Event $T_{\{0\}}^{\leftarrow} < T_n^{\leftarrow} | X_0^{\leftarrow} \in D_n$ **is not rare for backward process**
Whole problem is on $X_0^{-}\left(1\right) + \ldots + X_0^{-}\left(d\right) = n$... **plenty of tools!**

Conclusions and remarks (what did I say?)

- **Large deviations helps computers BUT much more than direct interpretation is needed!**

Conclusions and remarks (what did I say?)

- **Large deviations helps computers BUT much more than direct interpretation is needed!**
- **Lyapunov adaptation of subsolutions yields complexity $O(n^{2(d-\beta+1)})$ for (tandem)**

Conclusions and remarks (what did I say?)

- **Large deviations helps computers BUT much more than direct interpretation is needed!**
- **Lyapunov adaptation of subsolutions yields complexity $O(n^{2(d-\beta+1)})$ for (tandem)**
- **Guaranteed cost of splitting (general) $O(n^{2\beta+1})$**

Conclusions and remarks (what did I say?)

- **Large deviations helps computers BUT much more than direct interpretation is needed!**
- **Lyapunov adaptation of subsolutions yields complexity** $O(n^{2(d-\beta+1)})$ **for (tandem)**
- **Guaranteed cost of splitting (general)** $O(n^{2\beta+1})$
- **New algorithm: time-reversed sampling yields optimal** $O\left(n\right)$ **complexity (general)**

Conclusions and remarks (what did I say?)

- **Large deviations helps computers BUT much more than direct interpretation is needed!**
- **Lyapunov adaptation of subsolutions yields complexity** $O(n^{2(d-\beta+1)})$ **for (tandem)**
- **Guaranteed cost of splitting (general)** $O(n^{2\beta+1})$
- **New algorithm: time-reversed sampling yields optimal** $O(n)$ **complexity (general)**
- **Main features of new algorithm appear common to most quasi-reversible queues**

References

1. Anantharam, Heidelberger and Tsoucas (1990) Tech. Report IBM.
2. **Blanchet (2010). http://www.columbia.edu/~jb2814/**
3. **Blanchet, Glynn and Leder (2010). http://www.columbia.edu/~jb2814/**
4. **Blanchet, Leder and Shi (2010). http://www.columbia.edu/~jb2814/**
5. Dupuis, Sezer and Wang (2007) Annals of Applied Probability,17.
6. Dupuis and Wang (2009). QUESTA.
7. Dupuis and Dean (2010) Stochastic Processes and Their Applications.
8. Juneja and Nicola (2005) TOMACS
9. Villen-Altamirano (2010) European Journal of Operational Research.