Efficiency of Splitting Algorithm for Estimating Rare Events in Jackson Networks

Kevin Leder
Joint Work with Jose Blanchet and Yixi Shi

February 14, 2012

Outline

Two Widely Used Approaches to Rare Event Simulation

- Importance sampling: Simulate system under alternative dynamics so that event of interest is no longer rare. Keep track of likelihood ratio so that you can renormalize final answer to create unbiased estimator.
- Particle based methods: Simulate lots of correlated copies of system under original dynamics.

Splitting Method

- Will focus on a specific particle method called 'splitting method', which has been in use for over 60 years.
- Dean and Dupuis '08 presented a procedure for construction of efficient and stable splitting schemes. Will follow their notation.
- Model problem: X^n a sequence of stochastic processes on domain $D \subset \mathbb{R}^d$, and two disjoint sets A and B, define the sequence of stopping times $\tau_n = \min\{i : X^n(i) \in A \cup B\}$
- Goal estimate the probabilities

$$p_n(y) = \mathbb{P}\left(X^n(\tau_n) \in B | X^n(0) = y\right).$$

- Assume that there is a non-negative measurable function L such that

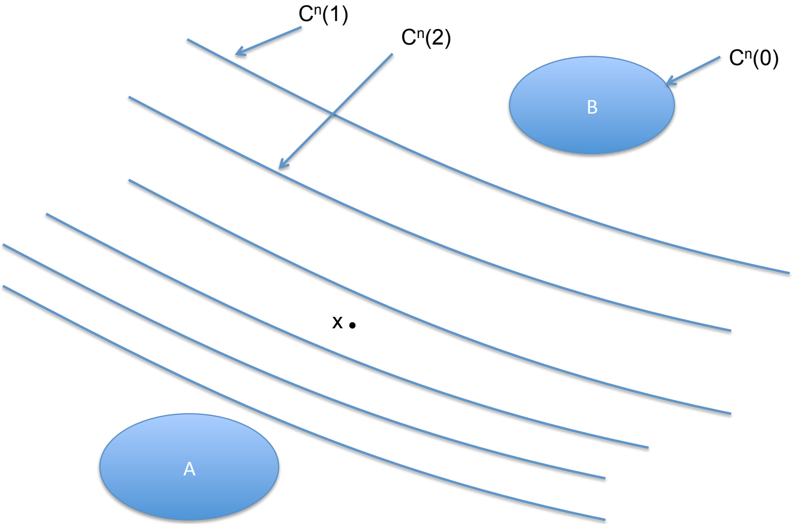
$$\lim_{n\to\infty} -\frac{1}{n} \log p_n(y)$$
$$= \inf\{\int_0^t L\left(\phi(s), \dot{\phi}(s)\right) ds : \phi(0) = y, \phi(t) \in B, \phi(s) \in A^c \text{ for all } s \le t\}.$$

The Splitting Algorithm

- Consider collection of nested sets $B = C_0^n \subset C_1^n \subset \ldots \subset C_{M_n}^n$

1. Initiate simulation procedure with a single particle starting from position $x \in C_k^n$ for some $k \geq 1$. Let $w_1 = 1$ initial weight associated to particle.
2. Evolve initial particle according to original transition kernel until either it hits A (dies) or level C_{k-1}^n. If it hits C_{k-1}^n it is replaced by r identical particles ($r > 1$). Weight of descendant particles is weight of parent particle $\times 1/r$.
3. Procedure from step 1 is replicated for each descendant particle, carrying over the value of the weights at each level for the surviving particles.
4. Steps 1 to 3 are repeated until all particles have either died or reached level $C_0^n = B$.

Splitting Method



The Splitting Estimator

- Consider collection of nested sets $B = C_0^n \subset C_1^n \subset \ldots \subset C_{M_n}^n$ (Note will want $M_n = c * n$ for some $c > 0$).
- Nested sets are based on level sets of an 'importance function', U. Specifically define $L_z = \{y \in D : U(y) \leq z\}$ then

$$C_j^n = L_{(j-1)/n}.$$

- An important function is the level function

$$\ell^n(y) = \min\{j \geq 0 : y \in C_j^n\},$$

- Estimator for $p_n(y)$ is

$$R_n(y) = N_n(y)/r^{\ell_n(y)},$$

where $N_n(y)$ is number of particles that made it to B.

Potential Issues with Splitting Estimators

- Instability or too much splitting: Computational cost of algorithm has potential to grow exponentially in n, i.e. total number of particles created can be exponential in number of levels.
- High relative variance or too little splitting: run a simulation with no particles arriving at target set.

Fully Branching Algorithm

- Dean and Dupuis introduce alternative algorithm for studying distributional properties of splitting estimator.
- Suppose that we continue to split particles that end up in set A, then total number of particles generated starting from y is given by $r^{\ell n(y)}$.
- This gives the 'fully branching' representation of $R_n(y)$,

$$R_n(y) = r^{-\ell n(y)} \sum_{j=1}^{r^{\ell n(y)}} 1_{\{X_j \in B\}}.$$

- This gives a simple way to check for stability

$$r^{-\ell n(y)} \mathbb{E}[N_n(y)] = \mathbb{E}_y \left[\sum_{j=1}^{r^{\ell n(y)}} 1_{\{X_j \in B\}} r^{-\ell n(y)} \right] = p_n(y)$$

that is to maintain subexponential complexity we need at least $\mathbb{E}[N_n(y)] = r^{\ell n(y)} p_n(y) = \exp(o(n))$.

Second Moment of Estimator

- To prevent high relative variance want

$$\mathbb{E}_y[R_n(y)^2] = p_n(y)^2 \exp(o(n)).$$

- From fully branching representation we get

$$\begin{aligned}\mathbb{E}_y[R_n(y)^2] &= \mathbb{E}_y \left[\sum_{j=1}^{r^{\ell n(y)}} 1_{\{X_j \in B\}} r^{-\ell n(y)} \right]^2 \\ &\geq \mathbb{E}_y \left[\sum_{j=1}^{r^{\ell n(y)}} 1_{\{X_j \in B\}} r^{-2\ell n(y)} \right] = r^{-\ell n(y)} p_n(y)\end{aligned}$$

- Thus need $r^{-\ell n(y)} = p_n(y) exp(o(n))$, which matches criteria needed for logarithmic stability.

Analysis of Splitting Estimators

- In order for stability and logarithmic optimality need at least $r^{\ell_n(y)} p_n(y) = \exp(o(n))$.
- Suppose we have a function $W(y)$ such that

$$p_n(y) \le \exp\left(-nW(y) + o(n)\right),$$

then it suffices to establish that

$$\ell_n(y) \log r - nW(y) = o(n)$$

- It's easy to see that $\ell_n(y) = \lceil nU(y) \rceil$ therefore we choose our importance function as $U(y) = W(y)/\log(r)$.
- See Dean and Dupuis (08) for more details

First Entrance Time of Finite State Space Markov Chains

- Well designed splitting schemes show subexponential growth in complexity for estimating first entrance probabilities.
- This allows for favorable comparison with naive Monte Carlo, which has exponential growth in complexity.
- If X^n lives in a countable state space $\mathcal{S}$ with transition kernel K then we have the following
$$p_n(x) = \sum_{y \in \mathcal{S}} K(x,y) p_n(y)$$
subject to the boundary condition
$$p_n(x) = \begin{cases} 1, & x \in A \\ 0, & x \in B. \end{cases}$$
- If the size of our state space is n^d then the total complexity for finding $p_n(x)$ is $O(n^{3d})$, i.e., it is subexponential in n.

Jackson Networks

Jackson Networks and Overflow Probabilities

Open Jackson Networks

- Consider a network of d stations. Customers arrive to the network with arrival rate $\lambda = (\lambda_1, \ldots, \lambda_d)^T$, and the service rate of the d stations is encoded by $\mu = (\mu_1, \ldots, \mu_d)^T$.
- A job that leaves station i joins station j with probability $P_{i,j}$, and leaves the system with probability

$$P_{i,0} = 1 - \sum_{j=1}^{d} P_{i,j},$$

this is called the routing matrix.

- We are interested in stable open Jackson networks, that is
  i) $\forall i$, either $\lambda_i > 0$ or $\lambda_{j_1} P_{j_1 j_2} \ldots P_{j_k i} > 0$ for some $j_1, \ldots, j_k$.
  ii) $\forall i$, either $P_{i0} > 0$ or $P_{ij_1} P_{j_1 j_2} \ldots P_{j_k 0} > 0$ for some $j_1, \ldots, j_k$.
  iii) The network is stable (i.e. a stationary distribution exists).

Basic Properties of Jackson Networks

- Assume without loss of generality that $\sum_{j=1}^{d}(\lambda_j + \mu_j) = 1$.
- Under the stability assumption the following

$$\phi_i = \lambda_i + \sum_{j=1}^{d} \phi_j P_{ji}, \qquad \forall i = 1, 2, ..., d$$

  has a unique solution $\phi^T = \lambda^T (I - P)^{-1}$.
- The traffic intensity at station i is in equilibrium is given by $\rho_i = \phi_i / \mu_i \in (0, 1)$.
- Define $\rho_* = \max_{1 \leq i \leq d} \rho_i$, and then set $\beta = |\{i : \rho_i = \rho_*\}|$.
- Study system through embedded discrete time Markov chain $Q = \{Q(k) : k \geq 0\}$ where $Q(k) = (Q_1(k), \ldots, Q_d(k))$, and $Q_i(k)$ represents number of customers in station i immediately after kth transition.
- Notice that we will mostly work with the queue length and NOT the scaled queue length

Overflow Probabilities in Jackson Networks

- Consider a subset of stations encoded by the vector v, denote the total population in this subset by $N_v(x) = \langle x, v \rangle$.
- Will be interested in the following probability:

$$p_n^v = \mathbb{P}\{\text{ total population in stations encoded by } v \text{ reaches } \\ n \text{ before returning to 0, starting from 0}\}.$$

- Can also define p_n^v via stopping times

$$T_{\{x\}} \triangleq \inf\{k \geq 1 : Q(k) = x\},$$
$$T_n^V \triangleq \inf\{k \geq 1 : N_v(Q(k)) \geq n\}.$$

- If we define $\mathbb{P}(\cdot) \doteq \mathbb{P}(\cdot | Q(0) = x)$ then

$$p_n^V = \mathbb{P}_0(T_n^V \leq T_{\{0\}}).$$

or more generally

$$p_n^V(x) = \mathbb{P}_x(T_n^V \leq T_{\{0\}}).$$

Dynamics of Q

- Queue length process is just a state-dependent random walk

$$Q(k+1) = Q(k) + \zeta\left(Q(k), Y(k+1)\right),$$

- ζ is a reflection function that prevents the queue-length process from taking negative values.
- The noise term $Y(k)$ represents outcome of next transition and has following pdf

$$\mathbb{P}\left(Y(k) = w\right) = \begin{cases} \lambda_i & \text{arrival at station } i, \\ \mu_i P_{ij} & \text{dep. at station } i \text{ goes to station } j, \\ \mu_i P_{i0} & \text{dep. at station } i \text{ leaves sys.} \end{cases}.$$

Logarithmic Asymptotics of Overflow Probabilities in Jackson Networks I

- Set $y = x/n$ and assume the existence of a function W_V

$$p_n^V(y) = \exp\left(-nW_V(y) + o(n)\right).$$

- By looking at Q/n we have the following via formal Taylor expansion

$$\begin{aligned}
1 &= \frac{1}{p_n^V(y)} \mathbb{E}\left[p_n^V(y + \frac{1}{n}\zeta(y, Y(1)))\right] \\
&\approx \mathbb{E}\exp\{-nW_V[y + \frac{1}{n}\zeta(y, Y(1))] + nW_V(y)\} \\
&= \mathbb{E}\exp\{-\partial W_V(y)^T \zeta(y, Y(1)) + o(1)\} \\
&= \exp\left(\psi(y, -\partial W_V(y)) + o(1)\right),
\end{aligned}$$

where $\psi(y, \theta) = \log \mathbb{E}\left[\exp\left(\theta^T \zeta(y, Y(k))\right)\right]$.

Logarithmic Asymptotics of Overflow Probabilities in Jackson Networks II

- In order to characterize logarithmic asymptotics of $p_n(y)$ need to find function W_V that satisfies

$$\psi(y, -\partial W_V(y)) = 0, \quad W_V(y) = 0, \text{ if } V(y) \geq 1.$$

- Only need an upper bound, so search for a subsolution, i.e., $\bar{W}_V$ that satisfies

$$\psi(y, -\partial \bar{W}_V(y)) \leq 0, \quad \bar{W}_V(y) \leq 0, \text{ if } V(y) \geq 1.$$

- Subsolution property guarantees that $\bar{W}_v(y) \leq W_V(y)$, and hence gives logarithmic upper bound for $p_n(y)$.
- A function that satisfies this condition is

$$\bar{W}_V(y) = \langle \varrho, y \rangle - \log \rho_*^V,$$

where $\varrho_i = \log \rho_i$ and $\rho_*^V = \max\{\rho_i : v_i = 1\}$.

- Can build splitting scheme out of this function, i.e. the importance function is given by $U(y) = \bar{W}_V(y)/\log(r)$.

Stationary Distributions of Jackson Networks

- Stationary distribution of Jackson networks:

$$\pi(m_1, ..., m_d) = \prod_{j=1}^{d} \mathbb{P}(Q_j(\infty) = m_j)$$
$$= \prod_{j=1}^{d} (1-\rho_j)\rho_j^{m_j}, \quad j = 1, ..., d, \text{ and } m_j \geq 0.$$

- Suppose that transition kernel of Q is given by $K(\cdot,\cdot)$, then define new transition kernel via

$$\tilde{K}(x,y) = K(x,y)\pi(x)/\pi(y),$$

this is transition kernel of time reversed Markov chain $\tilde{Q}$.

- $\tilde{Q}$ is queue length process of open stable Jackson network with stationary distribution π.
- Denote the probability measure on path space associated with the reversed chain by $\tilde{\mathbb{P}}$.

Asymptotics of Overflow Probabilities in Jackson Networks

- Can establish following representation based on stationary distribution and reversed process

$$p_n^V(x) = \frac{\tilde{\mathbb{P}}_\pi\left(\tilde{Q}(0) \in nC_0^n, \tilde{T}_{\{x\}} \le \tilde{T}_{\{0\}}, \tilde{T}_{\{x\}} < \tilde{T}_n^V\right)}{\pi(x)P_x(T_{\{x\}} \ge T_n^V \wedge T_{\{0\}})}$$
$$= \frac{\tilde{\mathbb{P}}_\pi\left(\tilde{Q}(0) \in nC_0^n, \tilde{\sigma}_{\{x\}} < \tilde{T}_{\{0\}} < \tilde{T}_n^V\right)}{\pi(0)P_0\left(\sigma_{\{x\}} < T_n^V \wedge T_{\{0\}}\right)}$$

where $\tilde{T}_n^V = \inf\{k \ge 1 : V\left(\tilde{Q}(k)\right) \ge n\}$, $\tilde{T}_{\{x\}} = \inf\{k \ge 1 : \tilde{Q}(k) = x\}$, and $\sigma_{\{x\}} \triangleq \inf\{k \ge 0 : Q(k) = x\}$.

- Can establish that there exists a $\delta > 0$ such that for all $x \in \mathbb{Z}_d^+$

$$P_x(T_{\{x\}} \ge T_n^V \wedge T_{\{0\}}) \ge \delta.$$

and therefore

$$p_n^V(x) = O(\pi(nC_0^n)/\pi(x)).$$

Asymptotics of Overflow Probabilities in Jackson Networks

- Using results from Blanchet ('11) it can be established that for x in a compact set
$$p_n^V(x) = \Omega(\pi(nC_0^n)).$$
- Therefore it suffices to find asymptotics of $\pi(nC_0^n)$, however this is found easily by using properties of negative binomial distribution
$$\pi(nC_0^n) = \Theta(e^{-n\gamma_V} n^{\beta_V - 1}),$$
where $\gamma_V = -\log \rho_*^V$, $\rho_*^V = \max\{\rho_i : v_i = 1\}$, and
$$\beta_V = \sum_{i=1}^{d} I(\rho_i = \rho_*^V, v_i = 1).$$

Splitting for Rare Events in Jackson Networks

Computational Cost of Splitting in for Jackson Networks

Computational Effort for Single Run of Splitting Algorithm

- In Blanchet, Leder, and Shi (11) we looked at the computational effort necessary to use a well designed splitting algorithm.
- Define $C = \frac{-\log \rho_*^V}{\log r}$, then rewrite importance function and level function as

$$U(x/n) = C(1 - \frac{\varrho^T x}{n \log \rho_*^V})$$

$$\ell_n(x) = \lceil C(n - \frac{\varrho^T x}{\log \rho_*^V}) \rceil .$$

- Consider the total number of particles that make it to overflow set

$$\mathbb{E}[N_n(x)] = r^{\ell_n(x)} p_n^V(x) = O\left(e^{-\gamma_V n} n^{\beta_V - 1} r^{\ell_n(x)} / \pi(x)\right) .$$

- Notice that $e^{-\gamma_V} = e^{\log \rho_*^V} = e^{-C \log r} = r^{-C}$, and $\pi(x) = c_1 r^{-C \varrho^T x / \log \rho_*^V}$ so that

$$\mathbb{E}[N_n(x)] = O\left(n^{\beta_V - 1} r^{\ell_n(x) - nC} / \pi(x)\right) = O\left(n^{\beta_V - 1}\right)$$

Computational Effort for Single Run of Splitting Algorithm

- From previous slide we saw that the number of particles to survive is $\approx n^{\beta_V - 1}$, which serves as a proxy for computational effort.
- Let N_m^n be total number of particles that survive to level m, and $Q_{m,j}$ be the position of particle j in generation m.
- Claim that there is a positive constant c_0 such that

$$\mathbb{E}[N_m^n] = r^m \mathbb{P}_x(Q_{m,1} \in nC^n_{\ell_n(x)-m}) \leq c_0(m-1)^{\beta_V - 1}.$$

- Define $\eta_{m,j}$ as the computational effort by jth particle starting at level m until it dies or reaches next level, and $\bar{\eta}_{m,j}(x_j)$ the expected effort given that it starts level m at x_j. Notice that $\bar{\eta}_{m,j}(x_j) \leq c_1 m$ for a positive constant c_1.
- We can write the total expected effort as

$$\mathbb{E} \sum_{m=0}^{\ell_n(x)-1} \sum_{j=1}^{N_m^n} \eta_{m,j} = \sum_{m=0}^{\ell_n(x)-1} \mathbb{E} \sum_{j=1}^{N_m^n} \bar{\eta}_{m,j}(x_j) \leq \sum_{m=0}^{\ell_n(x)-1} c_1 \mathbb{E}[N_m^n] m$$
$$= O\left(n^{\beta_V + 1}\right)$$

Refined Performance of Splitting

- Want to discuss the claim that

$$\mathbb{E}[N_m^n] = r^m \mathbb{P}_x(Q_{m,1} \in nC^n_{\ell_n(x)-m}) \leq c_0(m-1)^{\beta_V - 1}.$$

- Recalling that $\left(\rho_*^V\right)^{(m-1)/C} = r^{-m+1}$ the claim follows by establishing

$$P\left(Q_{m,1} \in nC^n_{\ell_n(x)-m}\right) = O\left(\left(\frac{m-1}{C}\right)^{\beta_V - 1}\left(\rho_*^V\right)^{(m-1)/C}\right).$$

- This result follows by first defining the hitting time

$$\hat{T}_{\frac{m}{C}} = \inf\{k \geq 1 : \varrho^T Q(k) \leq \varrho^T x - (m-1)\log r\},$$

  and then noticing that $Q_{m,1} \in nC^n_{\ell_n(x)-m} \Leftrightarrow \hat{T}_{\frac{m}{C}} < T_0$.

- We can bound the probability $\mathbb{P}_x\left(\hat{T}_{\frac{m}{C}} < T_0\right)$ using the reversed process, and results on the negative binomial distribution.

Second Moment of Splitting Estimator I

- Based on the 'fully branching' representation the second moment of the estimator is given by

$$\mathbb{E}_x[R_n(x)^2] = \mathbb{E}_x\left[\sum_{j=1}^{r^{\ell_n(x)}} 1_{\{X_j \in B\}} r^{-\ell_n(x)}\right]^2$$

- By looking at the most recent common ancestor of cells that make it to C_0^n, Dean and Dupuis (08) showed that

$$\mathbb{E}_x[R_n(x)^2] = \frac{r-1}{r}\sum_{m=0}^{\ell_n(x)-1} r^{-m}\mathbb{E}_x\left[I(V(Q_{m,1})>0)p_n^V(Q_{m,1})^2\right] + r^{-\ell_n(x)}p_n^V(x)$$

Second Moment of Splitting Estimator II

- Consider the summands from the off-diagonal term

$$\mathbb{E}_x\left[I(V(Q_{m,1})>0)p_n^V(Q_{m,1})^2\right]$$
$$=\mathbb{E}_x\left[p_n^V(Q_{m,1})^2\,|\varrho^T Q_{m,1}\le\varrho^T x-(m-1)\log r\right]\mathbb{P}_x\left(\hat{T}_{\frac{m}{C}}<T_0\right)$$

- Can use time reversed process and stationary distribution to show that there exists a $K>0$ such that for all $x\in\mathbb{Z}_+^d$, $p_n^V(x)\le K\pi(nC_0^n)/\pi(x)$,

$$\mathbb{E}_x\left[I(V(Q_{m,1})>0)p_n^V(Q_{m,1})^2\right]$$
$$\le K\pi(nC_0^n)\mathbb{E}\left[\pi(Q_{m,1})^{-2}|\varrho^T Q_{m,1}\le\varrho^T x-(m-1)\log r\right]\mathbb{P}_x\left(\hat{T}_{\frac{m}{C}}<T_0\right)$$

- Q has bounded increments, so there exists $\eta>0$ s.t.

$$\mathbb{E}\left[\pi(Q_{m,1})^{-2}|\varrho^T Q_{m,1}\le\varrho^T x-(m-1)\log r\right]=$$
$$\mathbb{E}\left[\pi(Q_{m,1})^{-2}|\varrho^T x-(m-1)\log r-\eta\le\varrho^T Q_{m,1}\le\varrho^T x-(m-1)\log r\right]$$

- Based on π there exists k_1 s.t.

$$\mathbb{E}\left[\pi(Q_{m,1})^{-2}|\varrho^T x-(m-1)\log r-\eta\le\varrho^T Q_{m,1}\le\varrho^T x-(m-1)\log r\right]$$
$$\le k_1 r^{2m-1}.$$

Second Moment of Splitting Estimator III

- Combine results from previous slide with upper bound on $\mathbb{P}_x\left(\hat{T}_{\frac{m}{C}} < T_0\right)$ to see that

$$\mathbb{E}_x\left[I(V(Q_{m,1}) > 0)p_n^V(Q_{m,1})^2\right] = p_n^V(x)^2 O\left(r^{m-1}m^{\beta_V - 1}\right).$$

- Using results from previous slide we see that

$$\begin{aligned}\mathbb{E}[R_n(x)^2] = & \frac{r-1}{r}\sum_{m=0}^{\ell_n(x)-1} r^{-m} E_x\left[I(V(Q_{m,1}) > 0)p_n^V(Q_{m,1})^2\right] \\ & + r^{-\ell_n(x)}p_n^V(x) \\ = & \, p_n^V(x)^2 O(n^{\beta_V}) + r^{-\ell_n(x)}p_n^V(x)\end{aligned}$$

- Since $r^{-\ell_n(x)}p_n^V(x) = O\left(n^{\beta_V - 1}\left(\rho_*^V\right)^{2n}\right)$ is dominated by $p_n^V(x)^2 O(n^{\beta_V})$ we see that

$$\mathbb{E}[R_n(x)^2] = p_n^V(x)^2 O(n^{\beta_V}).$$

Performance of Well Designed Splitting Schemes

- The computational effort required to achieve a fixed level of relative error is given by

$$C_n \frac{\mathbb{E}[R_n(x)^2]}{p_n^V(x)^2},$$

  where C_n is the computational cost per replication of the estimator i.e. roughly n^{β_V+1}.
- Therefore computational effort for a fixed level of relative error is $O(n^{2\beta_V+1})$
- If one were to solve via Gaussian elimination the computational cost is $O(n^{3d})$.
- Can take advantage of sparse structure of linear system, in particular get a banded matrix with bandwidth $O(n^{d-1})$, so computational complexity can be reduced to $O(n^{3d-2})$.
- Assuming that $d > 3$ well designed splitting algorithm significantly outperforms direct solution of linear system.

Comparison with other algorithms

- For tandem network Blanchet, Leder and Glynn (12) show that well designed importance sampling algorithm requires computational effort $O(n^{2(d-\beta+1)})$ to achieve a fixed level of relative error.
- Thus prefer importance sampling if more than half the stations are bottlenecks, and splitting otherwise.
- Conjecture: This property holds for all Jackson networks.
- Future work: Generalize analysis to DPR algorithm from Dean and Dupuis (09).