STATISTICAL COMPUTATIONAL METHODS

Queuing Systems

Parameters and random variables of a queuing system

- λ_A: **arrival rate**
- λ_S: **service rate**
- $\mu_A = 1/\lambda_A$: **mean interarrival time**
- $\mu_S = 1/\lambda_S$: **mean service time**
- $r = \lambda_A/\lambda_S = \mu_S/\mu_A$: **utilization**
- $X_s(t)$: **nr. of jobs receiving service at time** t
- $X_w(t)$: **nr. of jobs waiting in a queue at time** t
- $X(t)$: **total nr. of jobs in the system at time** t $(X(t) = X_s(t) + X_w(t))$
- S: **service time for a job**
- W: **waiting time for a job**
- R: **response time for a job (total time a job spends in the system from arrival to departure)** $(R = S + W)$.

1. Little's Law

$$E(X) = \lambda_A E(R).$$

2. Bernoulli single-server queuing system: a discrete-time Markov queuing process with
- one server
- unlimited capacity
- arrivals occur according to a Binomial process, with prob. of a new arrival $p_A = \lambda_A\Delta$
- prob. of service completion (i.e. departure) during each frame is $p_S = \lambda_S\Delta$, provided there is at least one job in the system at the beginning of a frame
- service times and interarrival times are independent.

Properties:
- $r < 1$, in order for the system to be functional;
- there is a SGeo(p_A) nr. of frames between successive arrivals; the interarrival time is $\Delta \cdot SGeo(p_A)$;
- each service takes a SGeo(p_S) nr. of frames;
- service of any job takes at least one frame;
- Markov property, transition probability matrix

$$\begin{aligned}
p_{00} &= 1 - p_A, \\
p_{01} &= p_A, \\
p_{i,i-1} &= (1-p_A)p_S, \quad i \geq 1, \\
p_{i,i} &= (1-p_A)(1-p_S) + p_A p_S, \quad i \geq 1, \\
p_{i,i+1} &= p_A(1-p_S), \quad i \geq 1.
\end{aligned}$$

- irregular Markov chain, no steady-state distribution;
- if limited capacity C,

$$p_{C,C} = 1 - (1-p_A)p_S.$$

2. **M/M/1 queuing system** : a continuous-time Markov queuing process with
- one server
- unlimited capacity
- Exponential interarrival times with arrival rate λ_A
- Exponential service times with service rate λ_S
- service times and interarrival times are independent.

Properties:
- $r < 1$, in order for the system to be functional;
- steady-state distribution of the number of jobs, X

$$\begin{aligned}
\pi_x &= P(X = x) = (1-r)r^x, \quad x = 0, 1, \dots \ (\text{Geo}(r)), \\
E(X) &= \frac{r}{1-r}, \\
V(X) &= \frac{r}{(1-r)^2}.
\end{aligned}$$

Main performance characteristics
- **expected response time**: $E(R) = \dfrac{\mu_S}{1-r} = \dfrac{1}{\lambda_S(1-r)}$,
- **expected waiting time**: $E(W) = \dfrac{\mu_S r}{1-r} = \dfrac{r}{\lambda_S(1-r)}$,
- **expected service time**: $E(S) = \mu_S = \dfrac{1}{\lambda_S}$,
- **expected nr. of jobs in the system**: $E(X) = \dfrac{r}{1-r}$,
- **expected queue length**: $E(X_w) = \dfrac{r^2}{1-r}$,
- **expected nr. of jobs being serviced**: $E(X_s) = r$,
- $r = P(X > 0) = 1 - \pi_0 = P(\text{system is busy})$,
- $1 - r = P(X = 0) = \pi_0 = P(\text{system is idle})$.

3. **Bernoulli k-server queuing system**: a discrete-time Markov queuing process with
- k servers
- unlimited capacity
- arrivals occur according to a Binomial process, with prob. of a new arrival during each frame $p_A = \lambda_A \Delta$
- during each frame, each busy server completes its job with probability $p_S = \lambda_S \Delta$ independently of the other servers and independently of the process of arrivals.

Properties:
- $r < k$, in order for the system to be functional;
- all interarrival times and all service times are independent Shifted Geometric random variables (multiplied by Δ) with parameters p_A and p_S, respectively;
- if there are j jobs in the system, then the nr. of busy servers is $n = \min\{j, k\}$;
- nr of departures during a frame, X_d is Binomial(n, p_s);
- Markov property, transition probability matrix

$$\begin{aligned}
p_{i,i+1} &= p_A(1-p_S)^n, \\
p_{i,i} &= p_A C_n^1 p_S (1-p_S)^{n-1} + (1-p_A)(1-p_S)^n, \\
p_{i,i-1} &= p_A C_n^2 p_S^2 (1-p_S)^{n-2} + (1-p_A) C_n^1 p_S (1-p_S)^{n-1}, \\
p_{i,i-2} &= p_A C_n^3 p_S^3 (1-p_S)^{n-3} + (1-p_A) C_n^2 p_S^2 (1-p_S)^{n-2}, \\
&\dots \\
p_{i,i-n} &= (1-p_A) p_S^n.
\end{aligned}$$

- if limited capacity C,

$$\begin{aligned} p_{C,C} &= p_A C_n^1 p_S (1-p_S)^{n-1} + (1-p_A)(1-p_S)^n + p_A(1-p_S)^n \\ &= n p_A p_S (1-p_S)^{n-1} + (1-p_S)^n. \end{aligned}$$

4. <u>M/M/k queuing system</u> : a continuous-time Markov queuing process with
- k servers
- unlimited capacity
- Exponential interarrival times with arrival rate λ_A
- Exponential service time for each server with service rate λ_S, independent of all arrival times and the other servers.

Properties:
- $r < k$, in order for the system to be functional;
- steady-state distribution of the number of jobs, X

$$\pi_x = P(X=x) = \begin{cases} \dfrac{r^x}{x!}\pi_0, & \text{for } x \leq k \\ \dfrac{r^k}{k!}\pi_0 \left(\dfrac{r}{k}\right)^{x-k}, & \text{for } x > k \end{cases}$$

where

$$\pi_0 = P(X=0) = \frac{1}{\displaystyle\sum_{i=0}^{k-1} \frac{r^i}{i!} + \frac{r^k}{k!(1-r/k)}}.$$

5. <u>M/M/∞ queuing system</u> : an M/M/k queuing system with $k=\infty$.

Properties:
- $X = X_s$;
- $R = S$;
- $X_w = W = 0$;
- steady-state distribution of the number of jobs, X

$$\begin{aligned} \pi_x &= P(X=x) = \frac{r^x}{x!}e^{-r}, \quad x = 0, 1, \ldots \ (\text{Poisson}(r)), \\ E(X) &= V(X) = r. \end{aligned}$$