Simulation Information

June 15, 2018

Definitions:

- Let μ be the *mean service rate* (i.e., the average number of jobs per second the server can complete), so that the mean service time is $1/\mu$.
- Let λ be the *mean arrival rate* (i.e., the average number of jobs per second that arrive to the system), so that the mean interarrival time is $1/\lambda$.
- A *probability distribution* tells you the probability of each possible outcome. For example, when you flip a coin there are two possible outcomes: heads and tails. If your coin is fair, you have a 1/2 probability of landing on heads, and a 1/2 probability of landing on tails. This is a *distribution* on the possible outcomes.
- A *random variable* is a function on the outcome of an experiment. For example, if we flip a coin, whether the coin comes up heads or tails is a random variable. It takes on value 0 or 1 with probabilities according to the associated probability distribution.

Distributions and experiments to try:

When you run your experiments, record the response time for each job and then compute the mean over all jobs. You will want to run your simulation for *many* arrivals—think on the order of 100,000 arrivals. This will ensure two things: (1) that you are averaging together enough jobs' response times to reduce noise, thereby giving you a more accurate result, and (2) that you have run the simulation long enough to reach what is called *steady state*. The idea here is that the first few arrivals in a simulation have an "unusual" experience because the system is always empty at the beginning of the simulation; later on in the simulation jobs might be more likely to have to wait in the queue, and we don't want our results to be skewed towards the initial "unusual" behavior. So we need to run the simulation long enough to wash out these effects.

Below are some experiments that I recommend running to give yourself practice with: (1) writing a queueing simulator, (2) generating random variables, and (3) making graphs of your results and interpreting them. I've given you several questions to think about, and my hope is that thinking about these questions will help you develop some intuition about the factors that affect mean response time in queueing systems.

- **Deterministic:** there is only one possible outcome.
  Recommendation:
  - Set the service time always equal to 1.
  - Set the arrival rate λ equal to, in turn, each of the values in $\{0.1, 0.2, \ldots, 0.9\}$ (so the interarrival times will be $\{1/0.1, 1/0.2, \ldots, 1/0.9\}$).
  - Make a graph with λ on the x-axis and mean response time on the y-axis.
  - **Questions:** What does your graph show? Why do the results look like this?

- **Uniform:** all outcomes in the range $[x_{min}, x_{max}]$ are equally likely (this is a *continuous* distribution)
  Recommendation:
  - Set the range for service times to $[1-\delta, 1+\delta]$, where δ is some constant between 0 and 1.
  - Set the range for interarrival times to $[\frac{1-\epsilon}{\lambda}, \frac{1+\epsilon}{\lambda}]$, where ϵ is some constant between 0 and 1. Set λ equal to, in turn, each of the values in $\{0.1, 0.2, \ldots, 0.9\}$.
  - Make a graph with λ on the x-axis and mean response time on the y-axis. Include different lines for different combinations of δ and ϵ
  - **Questions:** For a fixed δ, how does changing ϵ affect mean response time? For a fixed ϵ, how does changing δ affect mean response time? If the goal is to achieve low mean response time, is it better to have small or large δ and ϵ? Why?
- **2-point:** the outcome is x_1 with probability p and x_2 with probability $1-p$
  Recommendation:
  - Set the service time equal to $x_1 < 1$ with probability p and $1/x_1$ with probability $1-p$. Vary x_1 and then solve for p so that $px_1 + \frac{1-p}{x_1} = 1$.
  - Set the interarrival time equal to $x_2 < 1$ with probability q and $1/x_2$ with probability $1-q$. Vary x_2 and then solve for q so that $qx_2 + \frac{1-q}{x_2} = \frac{1}{\lambda}$. Set λ equal to, in turn, each of the values in $\{0.1, 0.2, \ldots, 0.9\}$.
  - Make a graph with λ on the x-axis and mean response time on the y-axis. Include different lines for different combinations of x_1 and x_2.
  - **Questions:** For a fixed x_1, how does changing x_2 affect mean response time? For a fixed x_2, how does changing x_1 affect mean response time? If the goal is to achieve low mean response time, is it better to have x_1 (respectively, x_2) relatively small (so x_1 and $1/x_1$ are far apart) or relatively large (so x_1 and $1/x_1$ are close together)? Why?
- **Geometric:** the outcome is i with probability $(1-p)^{i-1}p$. This is the number of times you need to flip a coin before getting the first heads, where p is the probability that the coin comes up heads on a single flip. (Note: the mean of a geometric is $1/p$.)
  Recommendation:
  - To compute a service time, generate a geometric random variable with probability p, then scale the result by multiplying the value you generated by p (this will ensure that the mean service time is 1).
  - To compute an interarrival time, generate a geometric random variable with probability p, then scale the result by multiplying the value you generated by p/λ (this will ensure that the mean interarrival time is $1/\lambda$). Set the arrival rate λ equal to, in turn, each of the values in $\{0.1, 0.2, \ldots, 0.9\}$.
  - Make a graph with λ on the x-axis and mean response time on the y-axis. Include lines for different choices of p between 0 and 1.
  - **Questions:** How does your choice of p affect what the graph looks like?