Queueing Theory and Simulation, lecture 11

Nicky van Foreest

June 22, 2021

Contents

1 Overview

1.1 Past

- Batch queues $M^X/M/1$.

1.2 Now

- $\mathsf{E}[W]$ for the $M/G/1$ queue, the so-called Pollaczek-Khinchine equation
- Stationary distribution of $p(n)$ ($= \pi(n)$ by PASTA) of the $M^X/M/1$ queue

1.3 Next

- Numerical method to compute stationary distribution of $p(n)$ ($= \pi(n)$) of the $M^X/M/1$ queue.
- Stationary distribution of $p(n)$ ($= \pi(n)$) of the $M/G/1$ queue.

2 Pollaczek-Khinchine formula derivation

2.1 M/G/1 model

- Jobs arrive as a Poisson process with rate λ
- $\{S_i\}$ are iid, $S_i \sim S$ with distribution F,
- $\mathsf{E}[S]$, $C_s^2 = \mathsf{V}[S]/(\mathsf{E}[S])^2$.
- $\rho = \lambda \mathsf{E}[S]$.
- number of servers is $c = 1$.

- FIFO scheduling

2.2 Derivation of $\mathsf{E}[W]$

- A job sees $\mathsf{E}[Q]$ jobs in queue upon arrival. By PASTA time-average and sample-average Q are the same.
- The waiting time is $\mathsf{E}[W]$: first get rid of the job in service (if any), then the rest of the queue:

$$\mathsf{E}[W] = \mathsf{E}[S_r] + \mathsf{E}[Q]\,\mathsf{E}[S] = \mathsf{E}[S_r] + \lambda\,\mathsf{E}[W]\,\mathsf{E}[S] \tag{1}$$
$$\mathsf{E}[W] = \mathsf{E}[S_r]/(1-\rho). \tag{2}$$

- Why is $\mathsf{E}[S_r] \neq \mathsf{E}[S]$ in general? Job service times are no longer exponentially distributed, hence not memoryless.
- Task ahead: find a formula for $\mathsf{E}[S_r]$.

2.3 Derivation of $\mathsf{E}[S_r]$, job payments

- Use renewal reward $Y = \lambda X$. Find suitable $Y(t)$, λ and $\{X_k\}$.
- We need to let jobs pay in proportion to the remaining service time.
- Suppose a job starts service at time 0 and needs S service.
- Let the job pay $(S-s)\,\mathrm{d}s$ at moment $s \in [0, S]$.
- Then $\int_0^S (S-s)\,\mathrm{d}s = S^2/2$ is the total amount paid for the remaining service time.

2.4 Derivation of $\mathsf{E}[S_r]$, server earnings

- $Y(t)$ is total amount of money earned by the server up to time t.
- At time s, what is $D(s)$? the number of jobs departed
- At time s, what is $D(s)+1$? the id of the next job to depart
- At time s, what is the meaning of $I_{L(s)>0} = 1$? There is a job in service.
- The server has earned up to time t: $Y(t) = \int_0^t (D_{D(s)+1} - s)\, I_{L(s)>0}\,\mathrm{d}s$.

2.5 Derivation of $\mathsf{E}[S_r]$, step 3

- What is the total amount paid by job k: $X_k = Y(D_k) - Y(D_{k-1})$.
- $X_k = S_k \cdot S_k/2 \implies \mathsf{E}[X] = \mathsf{E}\left[S^2\right]/2$.
- $\mathsf{E}[S_r] = Y = \delta\,\mathsf{E}[X] = \lambda\,\mathsf{E}\left[S^2\right]/2$, as $\delta = \lambda$ (rate-stability)
- Why is $\mathsf{E}[S_r] \neq \mathsf{E}[X]$? Sometimes the server is idle, and then makes no money.
- Recall, by PASTA, arriving jobs see the *time average* remaining service time at the server, hence Y!

2.6 Pollaczek-Khinchine formula, final step

- Merging the above

$$\mathsf{E}\left[W\right] = \frac{\mathsf{E}\left[S_r\right]}{1-\rho} = \frac{1}{2}\frac{\mathsf{E}\left[S^2\right]}{1-\rho} \tag{3}$$

$$= \frac{1+C_s^2}{2}\frac{\rho}{1-\rho}\mathsf{E}\left[S\right]. \tag{4}$$

- In an exercise you show the second equation.
- PK is an exact result!

2.7 Relation to Sakasegawa's formula

- PK $\mathsf{E}\left[W(M/G/1)\right] = \frac{1+C_s^2}{2}\frac{\rho}{1-\rho}\mathsf{E}\left[S\right]$.
- $S \sim \text{Exp}(\mu) \implies C_s^2 = 1 \implies (1+C_s^2)/2 = 1$. We get $\mathsf{E}\left[W\right] = \rho/(1-\rho)\,\mathsf{E}\left[S\right]$, i.e., the result for the $M/M/1$ queue.
- So, from $M/G/1$ to $M/M/1$, replace C_s^2 by 1
- Simple approximation, from $M/G/1$ to $G/G/1$, replace 1 by C_a^2. Hence,

$\mathsf{E}\left[W(G/G/1)\right] \approx \frac{C_a^2+C_s^2}{2}\frac{\rho}{1-\rho}\mathsf{E}\left[S\right]$

- I don't know how Sakasegawa's guessed that $\rho^{\sqrt{2(c+1)}-1}$ would work quite well.

2.8 Applications

See Chapter 4 of the book.

3 $M^X/M/1$: distribution of $p(n)$ derivation

3.1 Model (recall)

- Jobs arrive as a Poisson process with rate λ.
- Job batch sizes are iid $\sim B$, $f(k) = \mathsf{P}\left[B=k\right], G(k) = \mathsf{P}\left[B>k\right]$.
- Items are served individually
- Item service times iid $S \sim \text{Exp}(\mu)$.
- $c = 1$
- $\rho = \lambda\,\mathsf{E}\left[B\right]\,\mathsf{E}\left[S\right]$.

3.2 Up crossings

- Jobs arrive as a Poisson process with rate λ.
- Jobs with batch size $> l$ arrive as a Poisson process with rate $\lambda \,\mathsf{P}[B > l] = \lambda G(l)$.
- If $L(A_{k-}) = m$, then $L_{A_k} > n$ if job size $B_k > n - m$. Hence, level n is crossed from state m.
- Probability to see m jobs in the system upon arrival: $\pi(m)$.
- Hence, level n is upcrossed with rate $\lambda \sum_{m=0}^{n} \pi(m)G(n-m)$.

3.3 Level crossing

- Downcrossing rate $\pi(n+1)\mu$.
- Upcrossings and downcrossings rate of level n must be equal.
- In the long run therefore:

$$\lambda \sum_{m=0}^{n} \pi(m)G(n-m) = \mu\pi(n+1) \tag{5}$$

- We can solve for $\pi(n)$ with recursion, and normalize at the end.