Chapter 8

8.1 Let in the $M/M/1$ queue the arrival rate be λ and the departure rate μ. Let A be the (random) number of arrivals during a service period and let T be the (random) time this service period takes (which is exponential with rate μ). Then, $\mathbb{E}[A] = \mathbb{E}[\mathbb{E}[A|T]] = \mathbb{E}[\lambda T]$, because up to time T arrivals occur according to a homogeneous Poisson Process with intensity λ. So, $\mathbb{E}[A] = \mathbb{E}[\lambda T] = \lambda/\mu$.

The probability that no customer arrives during a service period can be computed in a similar fashion. Note that if $T = t$, the probability of no arrivals during the service period is $e^{-\lambda t}$, So,

$$\mathbb{P}(A = 0) = \int_0^\infty \mu e^{-\mu t} e^{-\lambda t} dt = \frac{\mu}{\lambda + \mu}.$$

Note that this latter result can also be obtained by observing that the time until the next arrival and the next departure are (as long as there is at least one customer in the system) independent exponential random variables. And we can use the theory on the minimum of independent exponential random variables from Session 1.

8.6 Using page 490 the time in the system of a typical customer for an $M/M/1$ queue with arrival rate λ and departure rate 2μ (with $2\mu > \lambda$) is given by $W_1 = \frac{1}{2\mu - \lambda}$.

To compute properties of the $M/M/2$ queue with arrival rate λ and service rate μ per server we use Example 8.6 (page 502). We obtain that

$$P_0 = \frac{1}{1 + \lambda/\mu + (\lambda/\mu)^2/2 + 2(\sum_{n=0}^\infty (\lambda/2\mu)^n - 1 - \lambda/2\mu - (\lambda/2\mu)^2)}$$
$$= \frac{1}{2\sum_{n=0}^\infty (\lambda/2\mu)^n - 1} = \frac{1}{\frac{2}{1-\lambda/2\mu} - 1} = \frac{1 - \lambda/(2\mu)}{2 - 1 + \lambda/(2\mu)} = \frac{2\mu - \lambda}{2\mu + \lambda}.$$

Furthermore, $P_1 = 2P_0(\lambda/2\mu)$, $P_2 = 2P_0(\lambda/2\mu)^2$ and $P_n = 2P_0(\lambda/2\mu)^n$. So, the average system size in this model is

$$\sum_{n=1}^\infty nP_n = 2P_0 \sum_{n=1}^\infty n(\lambda/2\mu)^n = 2P_0 \sum_{n=1}^\infty \sum_{k=1}^n (\lambda/2\mu)^n = 2P_0 \sum_{k=1}^\infty \sum_{n=k}^\infty (\lambda/2\mu)^n$$
$$= 2P_0 \sum_{k=1}^\infty (\lambda/2\mu)^k \frac{1}{1 - \lambda/2\mu} = 2P_0 \left(\frac{1}{1 - \lambda/2\mu} - 1\right) \frac{1}{1 - \lambda/2\mu} = 2P_0 \frac{2\mu\lambda}{(2\mu - \lambda)^2}$$
$$= \frac{4\mu\lambda}{(2\mu - \lambda)(2\mu + \lambda)}.$$

Then we use that the average time a customer is in the system the average system size divided by the arrival rate, which in this case is

$$W_2 = \frac{4\mu}{(2\mu - \lambda)(2\mu + \lambda)} = \frac{4\mu}{2\mu + \lambda} \frac{1}{2\mu - \lambda} > \frac{1}{2\mu - \lambda},$$

where the inequality is because $2\mu > \lambda$ and thus $\frac{4\mu}{2\mu + \lambda} > 1$. So, the average time in the system is larger for the double server queue. This is because the customers arrive and depart from the two systems according to the same rate, appart from when there is only one customer in the system, then the rate of departure in the single server queue is faster.

The W_Q for the single server queue is $W_1 - 1/(2\mu) = \frac{\lambda}{2\mu(2\mu-\lambda)}$. While for the double server queue it is

$$W_2 - 1/\mu = \frac{4\mu}{(2\mu-\lambda)(2\mu+\lambda)} - \frac{1}{\mu} = \frac{\lambda^2}{\mu(2\mu-\lambda)(2\mu+\lambda)} = \frac{\lambda}{2\mu(2\mu-\lambda)}\frac{2\lambda}{2\mu+\lambda},$$

which is less than $\frac{\lambda}{2\mu(2\mu-\lambda)}$. This implies that the queue length is usually longer in the 1 sever queue.

8.8 Note that in this problem items arrive (as long as there are less than k items already) on the shelf according to a Poisson process with rate λ and leave (as long as there is at least 1 item on the shelf) at rate μ. So, the number of items on the shelf can be described as an $M/M/1$ system with finite capacity k (see page 500, part (b)). We can use the resuts on page 501 to compute $P_0, P_1, \cdots, P_k$, where $P_j = (\lambda/\mu)^j P_0$ for $0 \leq j \leq k$ and $P_j = 0$ otherwise. So, because the probabilities have to sum up to 1 we obtain

$$1 = \sum_{j=0}^{k} P_0(\lambda/\mu)^j = P_0\frac{1-(\lambda/\mu)^{k+1}}{1-(\lambda/\mu)} \Rightarrow P_0 = \frac{1-(\lambda/\mu)}{1-(\lambda/\mu)^{k+1}}.$$

Only customers that arrive when there are 0 items on the shelf go away empty handed. So the fraction of customers leaving empty handed is P_0.

To find the average time an item is on the shelf we compute first the average number of items on the shelf (which is L in the notation of this chapter), which can be computed (among other ways)

$$\begin{aligned}\sum_{j=0}^{k} jP_j &= P_0\sum_{j=1}^{k} j(\lambda/\mu)^j = P_0\sum_{j=1}^{k}\sum_{i=1}^{j}(\lambda/\mu)^j = P_0\sum_{i=1}^{k}\sum_{j=i}^{k}(\lambda/\mu)^j \\ &= P_0\sum_{i=1}^{k}(\lambda/\mu)^i\sum_{j=i}^{k}(\lambda/\mu)^{j-i} = P_0\sum_{i=1}^{k}(\lambda/\mu)^i\frac{1-(\lambda/\mu)^{k-i}}{1-(\lambda/\mu)} = P_0\sum_{i=1}^{k}\frac{(\lambda/\mu)^i-(\lambda/\mu)^k}{1-(\lambda/\mu)} \\ &= \frac{P_0}{1-(\lambda/\mu)}\left((\lambda/\mu)\frac{1-(\lambda/\mu)^k}{1-(\lambda/\mu)} - k(\lambda/\mu)^k\right) = \frac{1}{1-(\lambda/\mu)^{k+1}}\left(\lambda\frac{1-(\lambda/\mu)^k}{\mu-\lambda} - k(\lambda/\mu)^k\right).\end{aligned}$$

To compute the average time an item is on the shelf (W in the notation of the chapter we use $W = L/\lambda_a$). Where we note that λ_a is $\lambda(1-P_k)$.

8.12 The model described here is just an $M/M/2$ queue (what a server does if he is not serving is really irrelevant). So, we can use Example 8.6 on page 502 with $k = 2$ to answer part a. The rate of going from 0 to 1 is $P_0\lambda$, while the rate of going from 2 to 1 is $P_2 \times 2\mu$.

8.28 Let A be the event that the first departure we consider leaves the system empty, A^c be its complement and D be the time between two departures. Then $\mathbb{P}(D > t|A^c)$ is the probabitily that the service time of the customer after the one just served is larger than t, which is $e^{-\mu t}$. $\mathbb{P}(D > t|A)$ is the probability that the time until arrival of the next customer plut its service rate exceeds t, which is the probability that the sum of two independent exponential random variables with expectations $1/\lambda$ and $1/\mu$ exceeds t. Because $a_n = d_n = P_n$ (section 8.2.2) we know that the probability that the system is empty after the arrival under consideration is $P_0 = 1 - \lambda/\mu$.

Recall that the moment generating function $\psi(r)$ of an exponentially distributed random variable with mean γ is for $r > \gamma$ given by

$$\int_0^\infty \gamma e^{-\gamma t} e^{rt} dt = \frac{\gamma}{\gamma - r}$$

and that the moment generating function of the sum of two independent random variables is the product of the moment generating functions of those random variables. Computing now the moment generating function of D (say $\psi_D(r)$) we obtain for $r < \lambda < \mu$.

$$\begin{aligned}\psi_D(r) = \mathbb{E}[e^{-rD}] &= (1 - \mathbb{P}(A))\frac{\mu}{\mu - r} + \mathbb{P}(A)\frac{\mu}{\mu - r} \times \frac{\lambda}{\lambda - r} \\ &= \frac{\lambda}{\mu}\frac{\mu}{\mu - r} + (1 - \lambda/\mu)\frac{\mu\lambda}{(\mu - r)(\lambda - r)} = \frac{\lambda(\lambda - r) + \lambda(\mu - \lambda)}{(\mu - r)(\lambda - r)} = \frac{\lambda}{\lambda - r},\end{aligned}$$

which is the moment generation function of an exponential distributed random variable with expectation $1/\lambda$.

8.36 Replacing First-Come-First-Served by Last-Come-First-Served has no influence on the system size and the queue lengths, since in both cases, at the end of a service a customer from the queue enters (if queue is not empty) and the workloads of different customers. in the queue are independent. The busy period distribution does not change either, because still you have to clear up all workload in the system, and the time this takes does not depend on the order you deal with them. The time spend in the system changes in distribution though. Intuitively speaking this can be seen by that in LCLS if no customers arrive between your arrival and the departure of the person in service during your arrival, then you don't stay long in the queue, independently of the number of customers in the queue at the time of your arrival. The same arguments hold for chosing a random customer from the queue.

The expectations are still the same for all mentioned service disciplines because the equations deduce section 8.2 (in particular $W = L/\lambda$) are independent of the service discipline.

8.37 In an M/G/1 queue customers arrive and depart one at a time, so the number of times the system enters state 0 is equal or 1 different from the number of times it departs state 0. By PASTA the fraction of arriving customers in state 0 is P_0. Now note that for time T by the strong law of large numbers the ammount of work brought in by customers divided by T converges as $T \to \infty$ to $\lambda\mathbb{E}[S]$, which is also the assymptotic fraction of time that the server is busy (which is $1 - P_0$).

The average work a customer sees in the system at departure is $\mathbb{E}[S]$ times the average number of customers a departure sees. However, by the same argument as above, this is equal to $\mathbb{E}[S]$ times the number of other customers a new arrival sees, which by PASTA is $\mathbb{E}[S]$ times L, where L is given by L in (8.34).

8.40 a(i) The change in the number of customers in the system follows the same law for a first come first serve discipline and a last come first serve discipline. In the latter discipline you only start touching the $n-1$ other initial customers in the system once the number of customers in the system is n. What happens with the other customers up to this time can be exactly described by what happens with the busy period of an $M/G/1$ queue with Last Come First Serve discipline. We know from Section 8.5.3 that the expected busy period is given by

$$\mathbb{E}[B] = \frac{\mathbb{E}[S]}{1-\lambda\mathbb{E}[S]}.$$

a(ii) would then be $(n-1)\mathbb{E}[B]$.

b(i) $\mathbb{E}[T|N] = N\mathbb{E}[B]$

b(ii) $\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T|N]] = \mathbb{E}[N]\mathbb{E}[B] = \lambda A\mathbb{E}[B]$.

8.23 a) appropriate states might be B for broken down and states $0, 1, 2, \cdots$ representing the number of customers in the queue if the stat is not broken down.
b) The rate at which customers enter state B is given by $\alpha \times \sum_{k=1}^{\infty} P_k = \alpha(1-P_B-P_0)$ and the rate at which customers leave state B is given by βP_B. The rate at which customers enter state 0 is $\beta P_B + \mu P_1$ and the rate at which they leave is λP_0 , while for state $k > 0$, the rate of entering the state is $\lambda P_{k-1} + \mu P_{k+1}$ and the rate of leaving the state is $(\lambda + \mu + \alpha)P_k$. It is possible to solve these equations but not needed for this question.
c) The expected system size is $L = \sum_{k=1}^{\infty} kP_k$, the rate at which customers enter $\lambda_a = \lambda(1-P_B)$. So the avarage amount of time a customer spends in the system is

$$L/\lambda_a = \sum_{k=1}^{\infty} kP_k/[\lambda(1-P_B)].$$

d) The rate at which customers leave through breakdowns is $\alpha\sum_{k=1}^{\infty} kP_k = \alpha L$, while the rate at which customers leave should be equal to the rate at which customers enter, which is $\lambda_a = \lambda(1-P_B)$. So, the proportion of entering customers leaving at the end of their service is given by $1 - \frac{\alpha L}{\lambda(1-P_B)}$.
e) Customers arrive according to a Poisson process, so by PASTA the answer to the question is P_B.