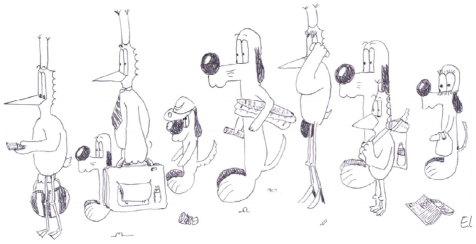
PERFORMANCE EVALUATION EXERCISES

QUEUING

With Solutions Jean-Yves Le Boudec, Spring 2021



1. An information server can be modelled as an M/GI/1 queue. Doubling the capacity of the server would...

   (a) □ Reduce the mean queuing time by a factor 2

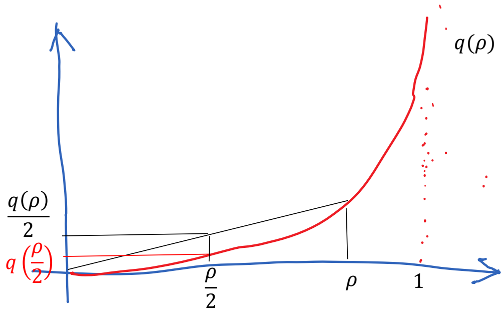
   (b) ☒ Reduce the mean queuing time by a factor larger than 2

   (c) □ Reduce the mean queuing time by a factor smaller than 2

   (d) □ It depends on the utilization factor

   Solution. Doubling the capacity $\Rightarrow$ ρ is divided by 2. The mean queuing delay $q(\rho)$ is convex in ρ and $q(0) = 0$. Thus

$$q\left(\frac{\rho}{2}\right) \leq \frac{1}{2}\left(q(0) + q(\rho)\right)$$
$$q\left(\frac{\rho}{2}\right) \leq \frac{1}{2}\, q(\rho)$$

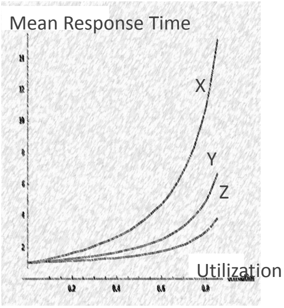


2. The 3 curves are for an M/GI/1 queue with different distributions of the service time S. Say which curve is for which distribution.

   B Rescaled Bernoulli with $p = 0.2$, i.e. $S = s$ with probability $p = 0.2$ and $S = 0$ with probability $1 - p$.

   C Constant

   E Exponential



	X	Y	Z
(a) □	B	C	E
(b) ☒	B	E	C
(c) □	C	B	E

(d) ☐ E B C

(e) ☐ C E B

(f) ☐ E C B

Solution. See the mean response time of the M/GI/1 queue. The larger the coefficient of variation, the larger the response time. The curve X is for the largest coefficient of variation and Z for the smallest.

Bernoulli: it is a random variable that takes values 0 with probability $1-p$ and s with probability p. The mean is $ps = 0.2$ and the variance is $p(1-p)s^2 = 0.16s^2$. The coefficient of variation is $0.4/0.2 = 2$.

For a constant random variable, the coefficient of variation is 0.

For an exponential random variable it is 1.

3. Which sentences are true ? λ = arrival rate, $\bar{S}$ = mean service time.

A For a single server queue, if $\lambda < \frac{1}{\bar{S}}$ the queue has a stationary regime.

B For an M/GI/1 queue, if $\lambda < \frac{1}{\bar{S}}$ the queue has a stationary regime.

(a) ☐ A

(b) ☐ B

(c) ☒ Both

(d) ☐ None

Solution. See Loyne's theorem

4. A train with 200 tourists arrive at the skilift. A queue builds up. Doubling the capacity of the skilift would...

(a) ☒ Reduce the queuing time by a factor 2

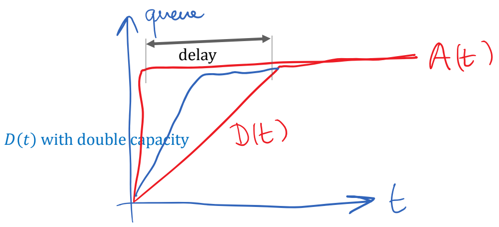
(b) ☐ Reduce the queuing time by a factor larger than 2

(c) ☐ Reduce the queuing time by a factor smaller than 2

(d) ☐ It depends on the utilization factor

Solution.

A deterministic analysis consists in plotting the cumulative arrival and departure curves. The worst case and the mean delay are given by horizontal distances. Doubling the capacity doubles the rate of departure and halves the worst-case and mean queuing times.



5. The average number of customers present in an M/GI/∞ queue is ($\bar{S}$ is the mean service time) ...

(a) ☒ $\bar{N} = \lambda \bar{S}$

(b) ☐ $\bar{N} = \frac{\rho}{1-\rho}$ with $\rho = \lambda \bar{S}$

(c) ☐ None of the above, the result depends on the distribution

(d) □ There is no closed form formula

Solution. There is an infinite number of server, therefore there is no queuing. The average response time is the average service time $\bar{S}$. By Little's formula $\bar{N} = \lambda \bar{S}$.

6. At a GI/GI/1 FIFO queue, the expected waiting time for a job, given that its service time is s, is...

(a) ☒ Independent of s

(b) □ Proportional to s

(c) □ Dependent on s but not proportional (in general)

Solution. The waiting time depends only on the amount of unfinished work found upon arrival.

7. At a M/GI/1 Processor Sharing queue, the expected response time for a job, given that its service time is s, is...

(a) □ Independent of s

(b) ☒ Proportional to s

(c) □ Dependent on s but not proportional (in general)

Solution. See formula 8.19

8. The European Commission has 50 anti-dumping cases on file and opens 20 cases a year. What is the average case duration ?

Solution. By Little's formula $N = \lambda T$, it is

$$T = \frac{N}{\lambda} = \frac{50}{20 \text{ year }^{-1}} = 2.5 \text{ years}$$

9. The autolib company has a fleet of electric cars and one single charging station. Every car visits the charging station to charge its batteries. Only one car can be charged at a time, other cars wait in a queue. The charging time is 30 mn. Every car spends in average 2 hours when it is charged before returning to the charging station. There are N cars in total. Can you approximately plot (1) the average waiting time at the charging station (2) the intensity of visits to the charging station as a function of N ? (3) Can you estimate the worst case waiting time ?

Solution. We can model this system as an interactive user system, where cars iterate between charging station and being out in the field (i.e. in "think time"). Let W be the average waiting time and λ the system throughput, i.e. number of car visits to charging stations per hour. By Little's law (the time unit is 1 hour):

$$\begin{aligned} N &= \lambda \cdot (W + \text{ mean charging time } + \text{ mean think time }) \\ N &= \lambda (W + 2.5) \end{aligned}$$

Bottleneck analysis gives:

$$W \geq 0 \tag{1}$$

and

$$\lambda \cdot \text{ mean charging time=resource utilization at charger } \leq 1$$

i.e.

$$0.5\lambda \leq 1 \tag{2}$$

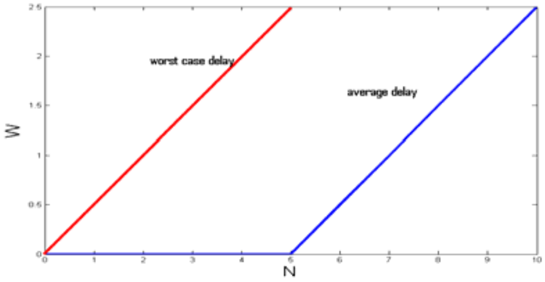
We expect (1) to be accurate at low load, i.e. for small N, and (2) to be accurate at high load, i.e. for large N. We derive the bounds for W:

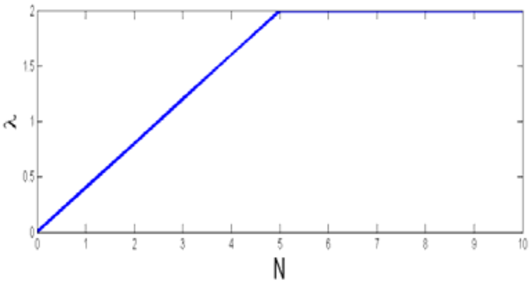
$$\begin{cases} W \geq 0 \\ W = \frac{N}{\lambda} - 2.5 \geq \frac{N}{2} - 2.5 \end{cases}$$

and for λ:

$$\begin{cases} \lambda \leq 2 \\ \lambda = \frac{N}{W+2.5} \leq \frac{N}{2.5} \end{cases}$$

We obtain the following plots (in blue):





These curves suggest that $N = 5$ is a critical fleet size, beyond which average queuing delays start to become significant, and below which the system works smoothly.

For the worst case waiting time we do a deterministic analysis. The cumulative arrival curve when all cars arrive at the same time at the charging station is a step function with a jump of N, which we assume to be the worst case. The service is at a rate of $1/0.5 = 2$ cars per hour so the departure curve is $D(t) = 2t$. The worst case delay is the horizontal deviation and is equal to $N/2$ (hours) (red line on plot).