7 The $M/G/1$ system

In the previous chapters we considered a production system with Poisson arrivals, which is in many cases a fairly realistic model for the arrival process, and with exponentially distributed production times. In practice exponentially distributed processing times are not very common. In most production systems the coefficient of variation of the processing times will be smaller than 1. Therefore it is essential to extend the theory to the case of more generally distributed processing times. In this chapter we will treat the case of Poisson arrivals and generally distributed, though independent, processing times. So we will consider the $M/G/1$ system. Jobs are again processed in order of arrival.

7.1 The mean value approach

Performance measures like the mean waiting time and the mean number of jobs in the queue can be, similar to the $M/M/1$ queue, obtained by the mean value approach. First, let us derive the arrival relation. A newly arriving job first has to wait for the *residual processing time* of the job in production (if there is one) and then continues to wait for the processing of all jobs which are already waiting in the queue on arrival. By PASTA we know that with probability ρ the machine is working on arrival. Let the random variable B denote the processing time, R the residual processing time and let L^q denote the number of jobs waiting in the queue. Hence,

$$E(W) = E(L^q)E(B) + \rho E(R).$$

Furthermore, we get by Little's law (applied to the queue of waiting jobs),

$$E(L^q) = \lambda E(W).$$

Combining these two relations, we find for the mean waiting time

$$E(W) = \frac{\rho E(R)}{1-\rho}\,. \tag{1}$$

Formula (1) is commonly referred to as the *Pollaczek-Khinchin mean value formula.* It remains to calculate the mean residual processing time. In the following section we will show that

$$E(R) = \frac{E(B^2)}{2E(B)}\,, \tag{2}$$

which may also be written in the form

$$E(R) = \frac{E(B^2)}{2E(B)} = \frac{\sigma_B^2 + E(B)^2}{2E(B)} = \frac{1}{2}(c_B^2 + 1)E(B), \tag{3}$$

where c_B^2 denotes the squared coefficient of variation of the processing time. An important observation is that, clearly, the mean waiting time only depends upon the first two moments of the processing time (and not upon its distribution). So in practice it is sufficient to know the mean and standard deviation of the processing time in order to estimate the mean waiting time.

Finally, expressions for $E(S)$ and $E(L)$ easily follow from the relations $E(S) = E(W) + E(B)$ and $E(L) = E(L^q) + \rho$.

Example 7.1 *(Exponential processing times)*
For exponential processing times we have $c_B^2 = 1$ and hence $E(R) = E(B)$ (memoryless property!). So, in this case the expressions for the mean performance measures simplify to

$$E(W) = \frac{\rho}{1-\rho}E(B), \qquad E(L^q) = \frac{\rho^2}{1-\rho}, \qquad E(S) = \frac{1}{1-\rho}E(B), \qquad E(L) = \frac{\rho}{1-\rho}.$$

Example 7.2 *(Deterministic processing times)*
For deterministic processing times we have $c_B^2 = 0$ and hence $E(R) = E(B)/2$. In this case we have

$$\begin{aligned} E(W) &= \frac{\rho}{1-\rho}\frac{E(B)}{2}, \qquad E(L^q) = \frac{\rho^2}{2(1-\rho)}, \\ E(S) &= \frac{\rho}{1-\rho}\frac{E(B)}{2} + E(B), \qquad E(L) = \rho + \frac{\rho^2}{2(1-\rho)}. \end{aligned}$$

7.2 Residual processing time

Suppose that a job arrives when the machine is working and denote the total processing time of the job in production by X. Further let $f_X(\cdot)$ denote the density of X. The basic observation to find $f_X(\cdot)$ is that it is more likely that a job arrives in a long processing time than in a short one. So the probability that X is of length x should be proportional to the length x as well as the frequency of such processing times, which is denoted by $f_B(x)dx$. Thus we may write

$$P(x \le X \le x + dx) = f_X(x)dx = Cxf_B(x)dx,$$

where C is a constant to normalize this density. So

$$C^{-1} = \int_0^\infty xf_B(x)dx = E(B).$$

Hence

$$f_X(x) = \frac{xf_B(x)}{E(B)}.$$

We conclude that

$$E(X) = \int_0^\infty xf_X(x)dx = \frac{1}{E(B)}\int_0^\infty x^2 f_B(x)dx = \frac{E(B^2)}{E(B)}.$$

Because the time of arrival of the job will be a random point somewhere in this processing time X (so on average in the middle of X), we conclude that the mean residual processing time is given by

$$E(R) = \frac{E(X)}{2} = \frac{E(B^2)}{2E(B)}\,.$$

Example 7.3 *(Erlang processing times)*
For an Erlang-r processing time with mean r/μ we have

$$E(B) = \frac{r}{\mu}\,, \qquad \sigma^2(B) = \frac{r}{\mu^2}\,,$$

so

$$E(B^2) = \sigma^2(B) + (E(B))^2 = \frac{r(1+r)}{\mu^2}\,.$$

Hence

$$E(R) = \frac{1+r}{2\mu}.$$

7.3 Batch arrivals

In this section we consider a production system where jobs do not arrive one by one, but in batches. These batches arrive according to a Poisson process with rate λ. The batch size is denoted by the random variable G with probability distribution

$$g_k = P(G = k), \qquad k = 0, 1, 2, \ldots$$

Note that we also admit zero-size groups to arrive. Our interest lies in the mean waiting time of a (single) job, for which we can write down the following equation.

$$E(W) = E(L^q)E(B) + \rho E(R) + \sum_{k=1}^{\infty} r_k(k-1)E(B)\,, \tag{4}$$

where ρ is the utilization of the machine, so

$$\rho = \lambda E(G)E(B),$$

and r_k is the probability that our job is the kth job processed in his group. The first two terms at the right-hand side of (4) correspond to the mean waiting time of the whole batch. The last one indicates the mean waiting time due to the processing of members in his own batch.

To find r_k we first determine the probability h_n that our job is a member of a batch of size n (cf. section 7.2). Since it is more likely that our job belongs to a large batch than to a small one, it follows that h_n is proportional to the batch size n as well as the frequency of such batches. Thus we can write

$$h_n = Cng_n,$$

where C is a constant to normalize this distribution. So

$$C^{-1} = \sum_{n=1}^{\infty} n g_n = E(G).$$

Given that our job is a member of a batch of size n, he will be with probability $1/n$ the kth job in his batch going into processing (of course, $n \geq k$). So we obtain

$$r_k = \sum_{n=k}^{\infty} h_n \cdot \frac{1}{n} = \frac{1}{E(G)} \sum_{n=k}^{\infty} g_n,$$

and for the last term in (4) it immediately follows that

$$\begin{aligned}
\sum_{k=1}^{\infty} r_k (k-1) E(B) &= \frac{1}{E(G)} \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} g_n (k-1) E(B) \\
&= \frac{1}{E(G)} \sum_{n=1}^{\infty} \sum_{k=1}^{n} g_n (k-1) E(B) \\
&= \frac{1}{E(G)} \sum_{n=1}^{\infty} \frac{1}{2} n(n-1) g_n E(B) \\
&= \frac{E(G^2) - E(G)}{2E(G)} E(B). \qquad (5)
\end{aligned}$$

From (4) and (5) and Little's law stating that

$$E(L^q) = \lambda E(G) E(W),$$

we finally obtain (cf. formula (1))

$$E(W) = \frac{\rho E(R)}{1-\rho} + \frac{(E(G^2) - E(G))E(B)}{2E(G)(1-\rho)}.$$

The first term at the right-hand side is equal to the mean waiting time in the system where jobs arrive one by one according to a Poisson process with rate $\lambda E(G)$. Clearly, the second term indicates the extra mean delay due to batching of arrivals.

Example 7.4 *(Uniform batch sizes)*
In case the batch size is uniformly distributed over $1, 2, \ldots, n$, so

$$g_k = \frac{1}{n}, \qquad k = 1, \ldots, n,$$

we find

$$E(G) = \sum_{k=1}^{n} \frac{k}{n} = \frac{n+1}{2}, \qquad E(G^2 - G) = \sum_{k=1}^{n} \frac{k^2 - k}{n} = \frac{(n-1)(n+1)}{3}.$$

Hence

$$E(W) = \frac{\rho E(R)}{1-\rho} + \frac{(n-1)E(B)}{3(1-\rho)}.$$