Queueing

Christian Berrig

May 31, 2024

0 Queueing of exponential distributions

Suppose that two exponentially distributed processes are queued. Then the distribution of the total waiting time distribution can be calculated as:

$$E_{\lambda_1 \lambda_2}(t) = \int_0^t d\tau \lambda_1 \exp\{-\lambda_1 \tau\} \lambda_2 \exp\{-\lambda_2(t-\tau)\}$$

$$= \lambda_1 \lambda_2 \exp\{-\lambda_2 t\} \int_0^t d\tau \exp\{-(\lambda_1 - \lambda_2)\tau\}$$

$$= \frac{-\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} \exp\{-\lambda_2 t\}[\exp\{-(\lambda_1 - \lambda_2)t\} - 1]$$

$$= \frac{-\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} [\exp\{-\lambda_1 t\} - \exp\{-\lambda_2 t\}]$$

Note that in the special case of $\lambda_1 = \lambda_2$, the apparent singularity is removable. This is actually a generalization of the Erlang distribution, with gives the distribution of the waiting times of k (identically) queued exponential distributions.

In the special case, however where $\lambda_1 = \lambda_2 = \lambda$, we have the special case of an Erlang-process; each subprocess is distributed exponentially, but the total waiting time is no longer exponential but follows an Erlang distribution.

The form of this distribution can be found pr induction to be:

$$P(t|\lambda, k)$$

since:

$$E_\lambda(t|\lambda, 1) = \exp\{-\lambda t\} E_\lambda(t|\lambda, 2) = \qquad \int_0^t d\tau \lambda \exp\{-\lambda \tau\} \lambda \exp\{-\lambda(t-\tau)\}$$

$$= \lambda \lambda \exp\{-\lambda t\} \int_0^t d\tau \exp\{-(\lambda - \lambda)\tau\}$$

Note the two obvious ways in which the Erlang distributed can be used:

- in terms of queued processes, in which each of the sub-processes has a well defined waiting time, that do not scale with number of subprocesses. This is the archetypical example of the supermarket queue, where the assumption is that a single customers expedition time ex exponentially distributed, and the assembler works with constant rate, that is the rate of expedition is constant no matter the length of the queue.

- in the particular case in which the rate scales with the queue length, such that the average waiting time for the queued process do not change with the length of the queue, but the mean waiting time becomes more well defined. This is the way in which the typical rate of the entire process is well known, and the shape-parameter k adjust how well-defined the reciprocal process rate adheres to the mean waiting time. The larger the shape parameter, the more this distribution approaches a Dirac-delta distribution.