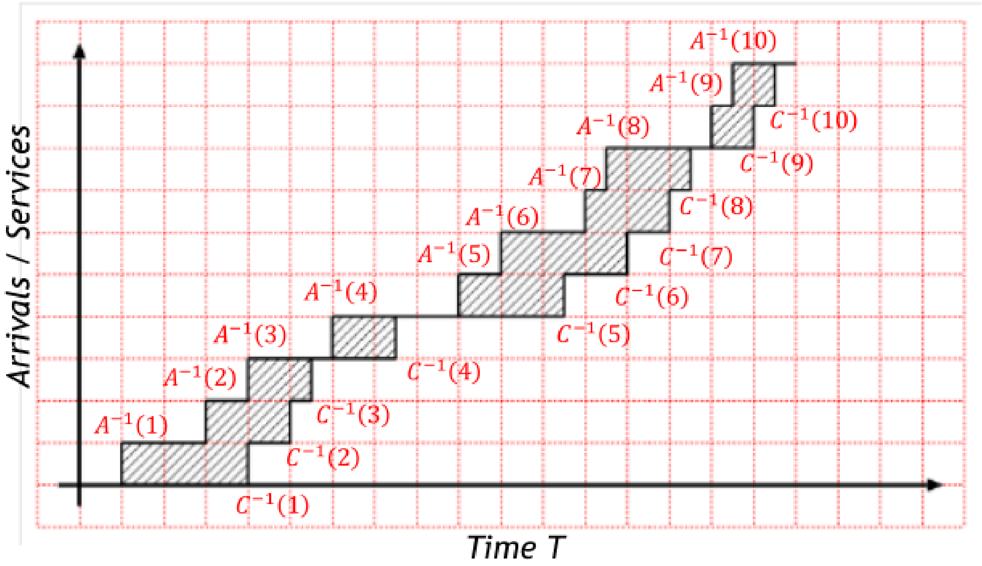
02.Basic Performance Metrics

Arrival and Completion time functions are step functions so we can extract the time instant when there is a jump in the functions. These time instant can be easily obtained.



From them we can derive the interarrivals times, the time between the arrivals of two consecutive jobs, $a_i = A^{-1}(i+1) - A^{-1}(i)$.
From this we can derive the A(T) function. Lets call I(X) the indicator function which returns 1 if proposition X is true 0 otherwise and assume that a_0 accounts for the arrival time of the first job:

$$A(T) = \sum_{K=!} I(\sum_{i=0}^{K-1} a_i \leq T), A^{-1}(i) = \sum_{k=0}^{i-1} a_k$$

For the response time we have an harder time as jobs can be stopped and resumed, but if we have an extra hypothesis as jobs are served one at a time and they are served in order of arrive we will have the departures in order and we can estimate $r_i = C^{-1}(i) - A^{-1}(i)$
Under the same assumptions we can compute the completion time of each job:

$$C(T) = \sum_{K} I(A^{-1}(K) + r_k \leq T)$$

If we have the arrival instant and the service time we can compute the instant completion time as:

$$C^{-1}(i) = max(A^{-1}(i), C^{-1}(i-1)) + s_i$$

From that we can obtain the service time s_i isolating it in the equation.

Basic Relations

If we set the time T as the time between the first arrival and the moment before another arrival in an empty system(one cycle of the system) we can have:

$$A(T) = C(T),\ \sum_{i=1}^{A(T)} a_i = T,\ \sum_{i=1}^{C(T)} s_i = B(T)$$

We can consider the average inter-arrival time:

$$\bar{A} = \lim_{T \to \infty} \frac{\sum_{i=1} A(T) a_i}{A(T)}, \lambda = \frac{1}{\bar{A}}$$

If a system is stable there will always be a point in time where A(T)=C(T). Thus if the system is stable and there are no losses throughput and arrival rates are always equal: $\lambda = X$
If the system is unstable, A(T) and C(T) will diverge, and after a given point in time, the system will never return empty again: $\lambda > X$.

Stability Conditions

By construction, since B(T) is less or equal to T, then the utilization should be less than one:

$$B(T) \leq T \Rightarrow U = \frac{B(T)}{T} \leq 1$$

Although there exists special cases in which the system is stable with U exactly equal to one (U = 1), they are extremely rare. In most of the cases, B(T) = T means that the system never returns
to an empty state, thus it is unstable. For this reason, we usually prefer to check that:

$$B(T) < T \Rightarrow U < 1$$

Stability condition allows to find limiting relations between the arrival rate and the average service:

$$XS = \lambda S = \frac{S}{\bar{A}} \leq 1$$

$$\lambda \leq \frac{1}{S}, S \leq \frac{1}{\lambda}, S \leq \bar{A}$$

$$X \leq \frac{1}{S}, S \leq \frac{1}{X}, \frac{1}{X} = \bar{A}$$

Response Time Distribution

Another thing we can compute is that the response time is less than a threshold:

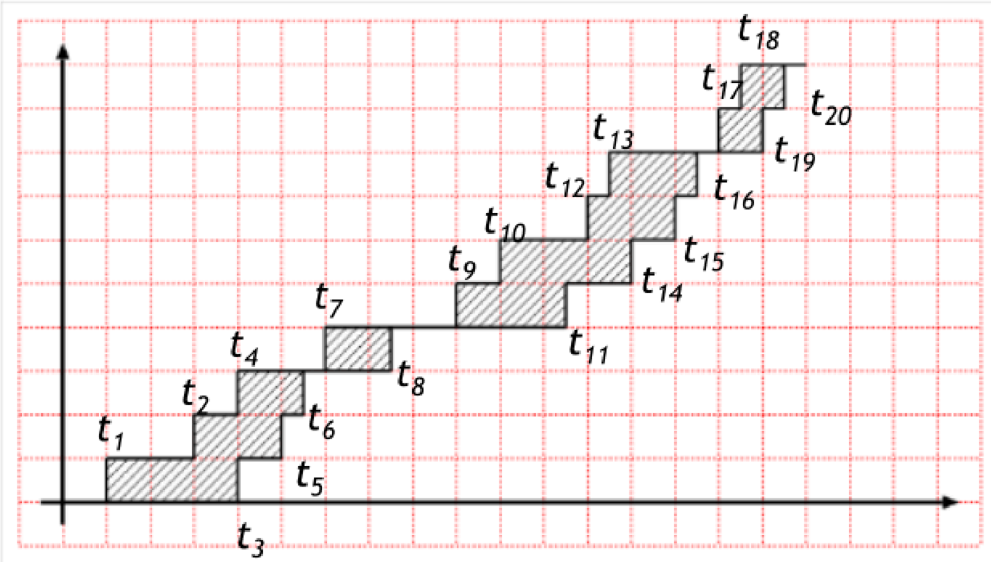
$$p(R < \tau) = \frac{\sum_{i=!}^{C} I(r_i < \tau)}{C}$$

This relation can be extended to any predicate $\Psi(R)$ and it can be used to compute the probability that the response time respects a given property:

$$p(\Psi(R)) = \frac{\sum_{i=1}^{C} I(\Psi(r_i))}{C}$$

Queue Length Distribution

With a slightly more complex procedure, we can determine the probability of having n jobs in the system from A(t) and C(t). First we observe that between arrivals or services, the population in the system remains constant. Let's call t_i the time at which either an arrival, or a departure occur.

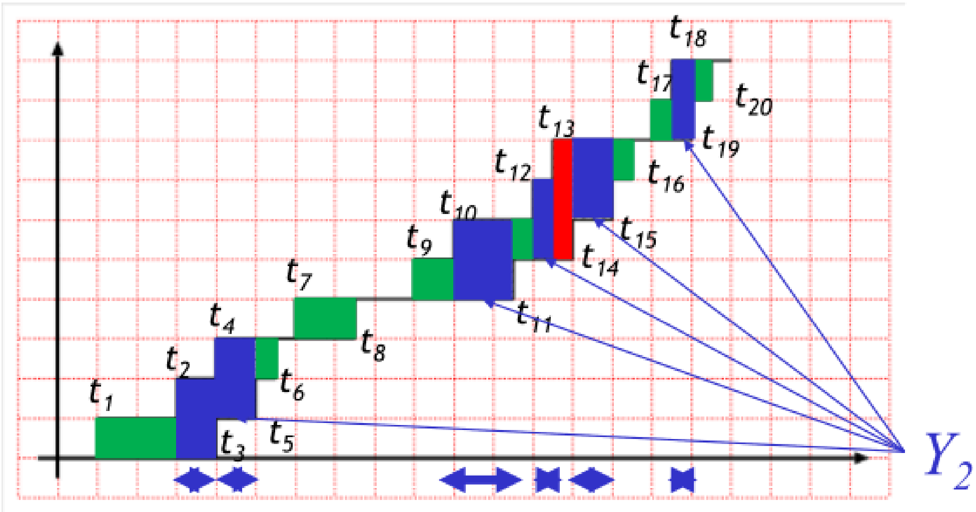


Note that, although very close, we suppose that the departure of the first job t_3 is just slightly before the arrival of the third job t_4 . At a given point in time t between two instants ti and t_{i+1}, the number of jobs in the system n(t) is constant and equal to: $n(t) = A(T) - C(T)$, remember the assumption the system start empty. We can then compute Y_m as the fraction of time the system has m jobs:

$$Y_m = \int_0^T I(n(t) = m)dt$$

This integral can also be computed as a summation of the differences between consecutive time instants where we have the given number of jobs in the system

$$Y_2 = \int_0^T I(n(t) = m)\,dt = (t_3 - t_2) + (t_5 - t_4) + (t_{11} - t_{10}) + (t_{13} - t_{12}) + (t_{15} - t_{14}) + (t_{19} - t_{18})$$



We can then approximate the probability of having n jobs in the system in the following way:

$$p(N = m) = \frac{Y_m}{T}$$

Note that also in this case, the technique can be extended to compute the probability that a given predicate Y(N) on the number of jobs is true. If we call $\Psi(N)$ the time in which the system fullfill such property, we have:

$$p(\Psi(N)) = \frac{Y_{\Psi(N)}}{T}$$

With these relations, we can estimate B, W and N in other ways:

$$B = \sum_{m=1} Y_m = T - T_0, W = \sum_{m=1} m \cdot Y_m, N = \sum_{m=1} m \cdot p(N = m)$$