COMP312/DATA304/DATA474
Simulation & Stochastic Models

More about the Poisson Process

Alejandro C. Frery

T1 2023

VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA

School of Mathematics and Statistics
New Zealand

What is it about?

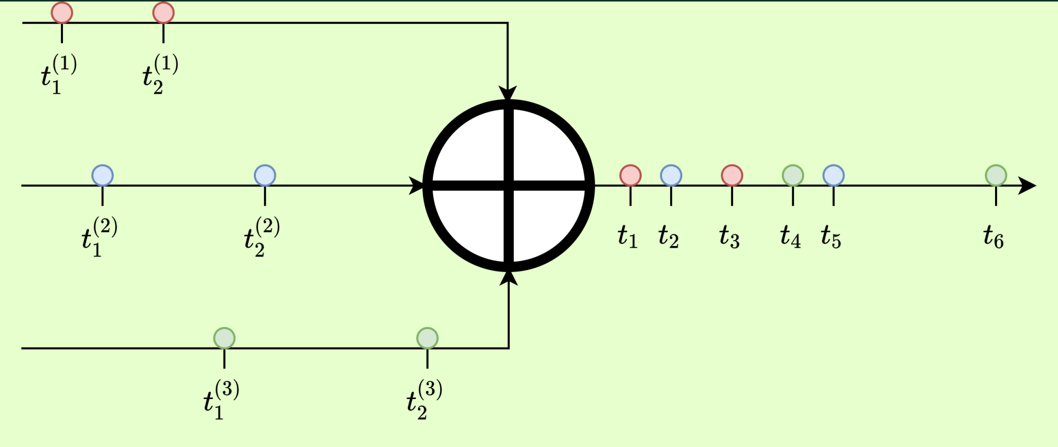
We will see some of the properties that arise when we assume that a queueing system follows a Poisson process.

We will see order statistics, and the Erlang distribution.

Invariance

Merging processes

We often see two or more queues converging into a single queue. We call this process *merging*.

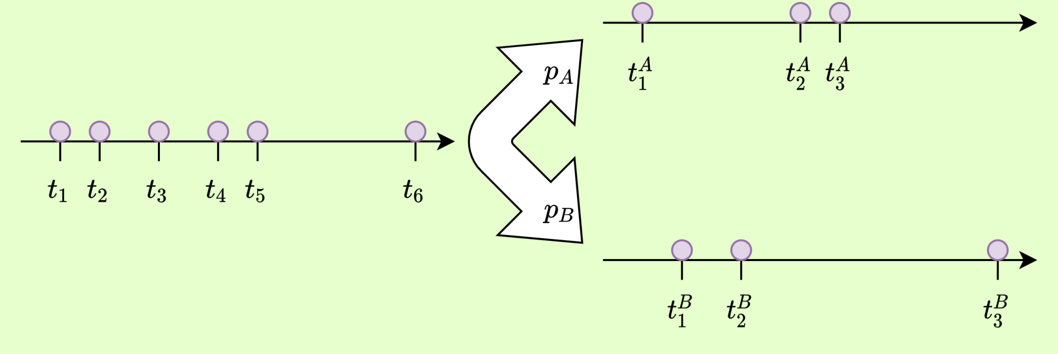


Merging processes

Assume there are k independent Poisson processes $\mathrm{PP}(\theta_1), \mathrm{PP}(\theta_2), \ldots, \mathrm{PP}(\theta_k)$ with rates $\theta_1, \theta_2, \ldots, \theta_k$. The resulting process of a *merging* is a new Poisson process with rate $\theta = \sum_{i=1}^{k} \theta_i$.

Partitioning a process

Assume we have one queue, and whenever one customer arrives we divert it to stream A or B with fixed probabilities $p_A > 0$ and $p_B > 0$ such that $p_A + p_B = 1$.



Partitioning a process

Recall that the decision is random.

If the original process has rate θ, the new process A has rate θp_A and the new process B has rate θp_B.

We may partition in as many new processes as desired: $p_1, p_2, \ldots, p_k$.

Partitioning and aggregating

What happens if we first partition, and then aggregate in any order?

Invariance
The Poisson process is invariant before aggregation and partitioning.

Long-term properties

Steady-state

We say that a queuing system is in transient state when its state depends on the initial conditions.

When the behaviour of a queuing system becomes stable, i.e., it does not depend on how the system began operating, we say it is in steady state. We refer to the distribution in steady state as stationary distribution.

Most results about queuing system relate to their steady state.

Terminology and notation

We will follow closely the notation used by Hillier & Lieberman (2001, Chapter 17).

State of system = number of customers in queueing system.

Queue length = number of customers waiting for service to begin
= state of system *minus* number of customers being served.

$N(t)$ = number of customers in queueing system at time t ($t \geq 0$).

$P_n(t)$ = probability of exactly n customers in queueing system at time t, given number at time 0.

s = number of servers (parallel service channels) in queueing system.

λ_n = mean arrival rate (expected number of arrivals per unit time) of new customers when n customers are in system.

μ_n = mean service rate for overall system (expected number of customers completing service per unit time) when n customers are in system. *Note:* μ_n represents *combined* rate at which all *busy* servers (those serving customers) achieve service completions.

Terminology and notation

Usually,

- $\lambda_n = \lambda$ (the arrival rate does not depend on the system state), and
- $\mu_n = \mu$ (the service rate does not depend on the system state).

Discuss these hypotheses

Under those conditions,

- the system capacity is $s\mu$ being utilised by,
- λ arrivals per unit time, in mean.

So

$$\rho = \frac{\lambda}{s\mu}$$

is the utilisation factor: the expected fraction of time the individual servers are busy.

More notation

Again, verbatim from Hillier & Lieberman (2001, Chapter 17).

P_n = probability of exactly n customers in queueing system.

$$L = \text{expected number of customers in queueing system} = \sum_{n=0}^{\infty} nP_n.$$

$$L_q = \text{expected queue length (excludes customers being served)} = \sum_{n=s}^{\infty} (n-s)P_n.$$

$\mathcal{W}$ = waiting time in system (includes service time) for each individual customer.

$W = E(\mathcal{W})$.

$\mathcal{W}_q$ = waiting time in queue (excludes service time) for each individual customer.

$W_q = E(\mathcal{W}_q)$.

Little's Formula

Little (1961) provided proofs for the following results, that hold only if the queue is in steady state:

$$
\begin{aligned}
L &= \lambda W, \\
L_q &= \lambda W_q, \\
W &= W_q + \frac{1}{\mu}.
\end{aligned}
$$

Order Statistics

Consider the random sample $\boldsymbol{X} = X_1, X_2, \ldots, X_n$ from independent and identically distributed continuous random variables whose distribution is characterized by the cumulative distribution function $F_X(t)$.

We are interested in the properties of $\underset{\sim}{\boldsymbol{X}} = X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$, the order statistics of the sample $\boldsymbol{X}$. We also encounter the following notation: $\underset{\sim}{\boldsymbol{X}} = X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, which is more compact but omits the fundamental information of the sample size n.

Results

Please follow the derivations.

$$f_{X_1,X_2,\ldots,X_n}(t_1,t_2,\ldots,t_n) = \prod_{\ell=1}^{n} f_X(t_\ell), \tag{1}$$

$$f_{X_{(n:n)}}(t) = nF_X^{n-1}(t)f_X(t), \tag{2}$$

$$f_{X_{(1:n)}}(t) = n\big[1 - F_X(t)\big]^{n-1} f_X(t), \tag{3}$$

$$f_{X_{(1:n)},X_{(n:n)}}(t_1,t_2) = n(n-1)\big[F_X(t_2) - F_X(t_1)\big]^{n-2} f_X(t_1)f_X(t_2)\mathbb{1}_{[t_1<t_2]}. \tag{4}$$

Please do!

Consider the sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ of iid $\mathcal{U}(0,1)$-distributed random variables.

1. Analyse how $f_{X_{1:n}}$ varies with n.
2. Analyse how $f_{X_{n:n}}$ varies with n.
3. Simulate 1000 observations of $(X_{1:n}, X_{n:n})$ for $n = 2, 5, 10$, plot the observations and overlap theoretical and empirical level curves of their joint density.

Consider the sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ of iid $\text{Exp}(1)$-distributed random variables.

1. Analyse how $f_{X_{1:n}}$ varies with n.
2. Analyse how $f_{X_{n:n}}$ varies with n.
3. Simulate 1000 observations of $(X_{1:n}, X_{n:n})$ for $n = 2, 5, 10$, plot the observations and overlap theoretical and empirical level curves of their joint density.

All the n servers are busy. Their service times follow $\mathrm{E}(\theta)$, with rate $\theta > 0$. What is the distribution of the time until the next available server?

The serving times are $T_\ell \sim \mathrm{Exp}(1/\theta)$ distributed, $1 \leq \ell \leq n$, i.e., $f_T(t) = \theta e^{-\theta t} \mathbb{1}_{\mathbb{R}_+}(t)$, and $F_T(t) = (1 - e^{-\theta t}) \mathbb{1}_{\mathbb{R}_+}(t)$.

Using (3):

$$f_{T_{1:n}(t)} = n\theta \exp\{-n\theta t\} \mathbb{1}_{\mathbb{R}_+}(t),$$

so its distribution is $\mathrm{Exp}(1/(n\theta))$.

In mean, the next customer will have to wait n times less than if there was a single server.

Modelling events under the PP

- Whatever happens in $(t_0, t_0 + t)$ is described by the elapsed time t and does not depend on the starting time t_0.
- The probability of no arrival in $(0, t)$ is $\Pr_0(t) = \Pr(X = 0) = e^{-\lambda t}$.
- The probability of one arrival in $(0, t)$ is $\Pr_1(t) = \Pr(X = 1) = \lambda t e^{-\lambda t}$.
- The probability of k arrivals in $(0, t)$ is

$$\Pr_k(t) = \Pr(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Several events

How long will we have to wait until exactly k customers arrive? Each customer follows an $\mathrm{Exp}(\theta)$ distribution with rate $\theta > 0$. The arrival time of the fourth customer is $Z = \sum_{\ell=1}^{4} T_\ell$, in which $T_\ell \sim \mathrm{Exp}(\theta)$ are iid random variables.

Recalling the properties of Exponential random variables, we have that the sum of k independent identically distributed random variables follows a distribution characterized by the density

$$f_Z(z) = \frac{\theta^k}{(k-1)!} z^{k-1} e^{-\theta z} \mathbb{1}_{\mathbb{R}_+}(z).$$

This is known as Erlang distribution with parameters k and θ, which you should have also seen as Gamma distribution with shape k and mean $1/\theta$. They are denoted, respectively, as $\mathrm{Erlang}(k, \theta)$ and $\Gamma(k, 1/\theta)$. Notice that the Gamma law can be (and often is) parametrised in several ways.

Please do!

Express the Erlang(k, θ) distribution as

- A Gamma random variable with the parametrisation employed by the function `dgamma` in R.
- A Gamma random variable with probability density function

$$f_W(w; \alpha, \gamma) = \frac{\gamma^\alpha}{\Gamma(\alpha)} w^{\alpha-1} \exp\{-\gamma w\} \mathbb{1}_{\mathbb{R}_+}(x).$$

More useful results

Consider the positive random variable $X : \Omega \to \mathbb{R}_+$. Its expected value is

$$\mathrm{E}(X) = \int_{\mathbb{R}_+} x f_X(x) dx = \int_{\mathbb{R}_+} \left[1 - F_X(x)\right] dx,$$

where f_X and F_X are, respectively, the density and cumulative distribution function that characterize the behaviour of X.

As a consequence, if $X : \Omega \to \mathbb{N}_0$, then

$$\mathrm{E}(X) = \sum_{\ell=0}^{\infty} \Pr(X > \ell) = \sum_{\ell=1}^{\infty} \Pr(X \geq \ell).$$

References

Hillier, F. S. & Lieberman, G. J. (2001), *Introduction to Operations Research*, 7 edn, McGraw-Hill, New York.

Little, J. D. C. (1961), 'A proof for the queueing formula: $L = \lambda W$', *Operations Research* **9**(3), 383–387.