Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

LECTURE NOTES

# Fundamentals of Probability and Statistics
# Chapter 1: Lecture 3

Spring 2018

**Written by m**arco **b**roccardo
IBK, ETHZ

# 7 Probability Models for Random Processes

A random process is a probabilistic model that describes the evolution of a random phenomenon over time. Roughly speaking, a random process can be viewed as a sequence of random variables. We start this chapter by assuming discrete events; then, we extend the results for continuous time.

## 7.1 The Bernoulli process

The Bernoulli process can be thought of as a sequence of independent experiments, $X_1, X_2, ..., X_n, ...$, which having two possible outcomes (e.g. fail, not fail). Each experiment is named Bernoulli trial. The probability of success of an experiment is $p$ (with $0 < p < 1$), (e.g. $P(X_n = 1) = p$ and $P(X_n = 0) = 1 - p$), and its mean and expectation are given by

$$\begin{aligned} E[X_n] &= p \times 1 + (1-p) \times 0 = p, &(95)\\ \mathrm{Var}[X_n] &= E[X_n^2] - E[X_n]^2 = p - p^2 = p(1-p). &(96) \end{aligned}$$

As mentioned above, a random process can be described as a collection of random variables. In Section 4, we have seen that in the presence of multiple random variables the complete probabilistic description of the problem is given in terms of the joint distribution. In the Bernoulli process, since each random variable is independent, the joint distribution can be written as $P_{\mathbf{X}}(x_1, x_2, ...x_n...) = p_{X_1}(x_1)p_{X_2}(x_2)...p_{X_n}(x_n)...$. An alternative way to think about the Bernoulli process is to imagine that every infinite sequence of Bernoulli trials is a sample point of a sample space $S$. In this view, every single infinite sequence of Bernoulli trials has probability 0 (since $0 < p < 1$). It follows that the sample space must be continuous.

**Problem** Can you think what might be a good representation of the sample space $S$?

Note that because of the assumption of independence, knowledge of the outcome of an experiment or of a sequence of experiments does not affect the outcome of future experiments. Oftentimes, this property is called memoryless property.

When facing with Bernoulli process there are three main probabilistic questions we might ask:

i. For a given amount of experiments, how many will succeed?

ii. How many experiments should we perform before observing the first success?

iii. How many experiments should we perform before observing the second (or third, or fourth...) success?

i. The first question is analogous to the Example II. In particular given $n$ experiments, the probability that $x$ out of $n$ will succeed are given by the binomial distribution with parameters $p$ and $n$, ($bin(p, n)$), i.e.

$$p_X(x; p, n) = \binom{n}{x} p^x (1-p)^{(n-x)}, \text{ for } x = 0, 1, 2, ..., n. \tag{97}$$

The mean and the variance of the binomial distribution are given by

$$\begin{aligned} E[X] &= np, &(98)\\ \mathrm{Var}[X] &= np(1-p). &(99) \end{aligned}$$

ii. To answer the second question, we define the number of performed experiments before we observe a success as the inter-arrival time, $T_1$. Given this definition, a sequence of outcomes up to $T_1$ is of the following type $0, 0, ..., 0, 1$. Then, given the assumption of independence, the probability of such sequence is

$$p_{T_1}(t;p) = P(T_1 = t) = (1-p)^{t-1}p, \text{ for } t = 1, 2, ... \tag{100}$$

Eq.(100) is known as the geometric distribution, $geom(p)$. The mean and variance of the geometric distribution are given as

$$E[T_1] = \frac{1}{p}, \tag{101}$$

$$\text{Var}[T_1] = \frac{1-p}{p^2}. \tag{102}$$

iii. To answer the third question, we observe that once the first success occurres, the next string of $0, 0, ..., 0, 1$ is statistically independent from the past. It follows that the inter-arrival time $T_{1,2}$ is again a geometrical random variable with parameter $p$. More generally, any inter-arrival time between two sequential successful events is a geometric distribution with parameter $p$. Then, the arrival of the $k$th event is simply the sum of $k$ statistically independent geometrical random variables, i.e. $T_k = T_1 + T_{1,2} + ... + T_{(k-1),k}$. Alternatively, the distribution of $T_k$ can be derived as follows

$$p_{T_k}(t;p) = P(T_k = t) = P(k-1 \text{ experiments succeeded before } t) \times P(\text{the } t\text{th experiment is successful}) \tag{103}$$

which leads to

$$\begin{aligned} p_{T_n}(t;p,k) &= \binom{t-1}{k-1} p^{k-1}(1-p)^{(t-k)} \times p, \\ &= \binom{t-1}{k-1} p^{k}(1-p)^{(t-k)} \text{ for } t = k, k+1, ... \end{aligned} \tag{104}$$

The (104) is known as negative binomial distribution of order $k$, $negbin(p, k)$. Expectation and variance are given as

$$E[T_k] = \frac{k}{p}, \tag{105}$$

$$\text{Var}[T_k] = k\frac{1-p}{p^2}. \tag{106}$$

### 7.1.1 Bernoulli process with random selection

Suppose that you have a Bernoulli process composed of a series of experiments with probability $p$ of success. Whenever there is a success, then you perform an additional experiment. This additional experiment has probability of success $q$, and it is statistically independent from the previous experiments. If the the second experiment succeed, then you keep the sample otherwise you discard it, Figure 14. If you observe the derived process, you might recognize that it is a Bernoulli process with probability $pq$. On the other hand, the process of discarded experiments is also a Bernoulli process, where each sample has a probability $p(1-q)$.

#### 7.1.2 Combining Bernoulli processes

Suppose you have two independent Bernoulli processes with parameters $p$ and $q$ and you would like to combine the two processes into one single process, Figure 15. A success on the merged process is recorded if there is a success in at least one of the two original processes. The probability that there is at least a success for every single experiment is $p + q - pq$. Notice that we are not accounting for the number of success at each sequential experiment, but only if there was at least one success.

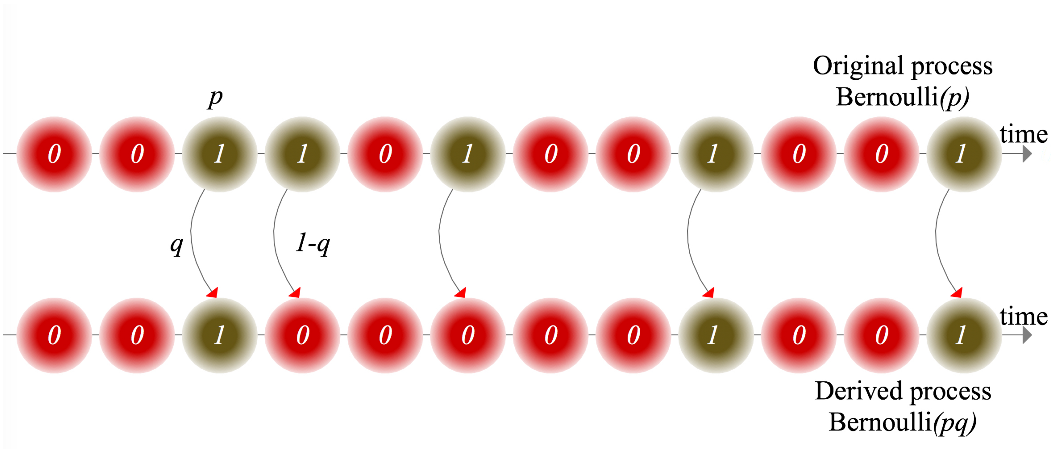


Figure 14: Bernoulli process with random selection

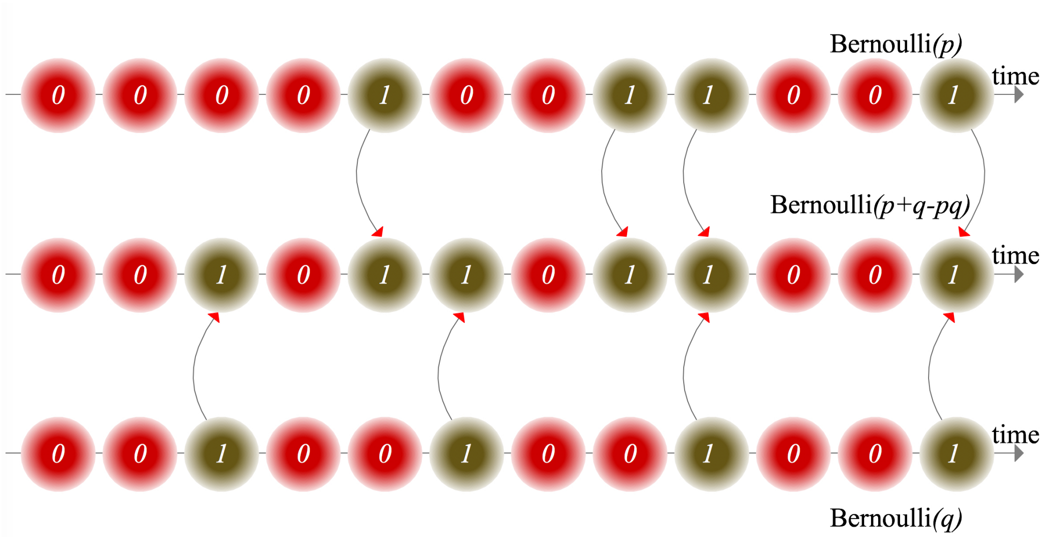


Figure 15: Combining Bernoulli processes

## 7.2 The Homogeneous Poisson Process, HPP

The Poisson[7] process is a continuous time version of the Bernoulli process. Before examining the process in details, let's focus on the following problem.

Suppose that for a given experimental campaign $n$ is large and $p$ is small so that the mean $np$ has a finite reasonable value. In this case, every string of sequential experiments has a large number of 0s and few 1s. If we make these experiments so fast that to the limit they become instantaneous experiments; then, we can record the exact time when an experiment succeed. Hence, we would like to use a continuous variable for the time, rather than discrete. For a given interval, we can address this problem by letting $n$ growing to infinity, while $p$ decreases so that for a fixed interval $t$ , $np = \Lambda$. Thus, given a time interval $t$ we can obtain the number of instantaneous success by finding the limit of the binomial distribution, i.e.

$$\begin{aligned}
p_X(x;\Lambda) &= \lim_{n\to\infty} \binom{n}{x} \left(\frac{\Lambda}{n}\right)^x \left(1-\frac{\Lambda}{n}\right)^{n-x}, \\
&= \lim_{n\to\infty} \frac{n(n-1)...(n-x+1)}{x!} \left(\frac{\Lambda}{n}\right)^x \left(1-\frac{\Lambda}{n}\right)^{(n-x)}, \\
&= \lim_{n\to\infty} \underbrace{\frac{n(n-1)...(n-x+1)}{n^x}}_{1 \text{ for } n\to\infty} \frac{\Lambda^x}{x!} \underbrace{\left(1-\frac{\Lambda}{n}\right)^{-x}}_{1 \text{ for } n\to\infty} \underbrace{\left(1-\frac{\Lambda}{n}\right)^{n}}_{e^{-\Lambda} \text{ for } n\to\infty}, \\
&= \frac{\Lambda^x}{x!} e^{-\Lambda}
\end{aligned} \tag{107}$$

The (107) is known as the Poisson distribution. Observe that $\Lambda$ was fixed for a given duration of the experimental campaign. Then, we can define the mean *rate* of events as $\lambda = \Lambda/t$, where $t$ is the duration of the experimental campaign.

A Poisson process builds on the following three properties:

**Independence:** the number of events for a given interval is independent from any other disjoint interval.

**Stationarity (Homogeneity):** the probability of $x$ success for a given $t$, is the same for any interval of length $t$. This is equivalent of assuming constant $\lambda$. We later relax this condition by defining the non-homogeneous Poisson process.

**Single event occurrence:** given an infinitesimal time $dt$ the probability of 0 event is $P(0, dt) = e^{-\lambda dt} = 1 - \lambda dt + \mathcal{O}(dt^2)$, the probability of one event is equal to $P(1, dt) = \lambda dt e^{-\lambda dt} = \lambda dt + \mathcal{O}(dt^2)$, where $\mathcal{O}(dt^2)$ are high order terms of $dt$. It follows that every single instant of time is a random variable which corresponds to a Bernoulli trial with probability of success (and expectation) $p = \lambda dt$ . Then, $\lambda$ is the expected number of success per unit time.

Like we did for the Bernoulli process, we might want to answer to the following questions:

[7] Siméon Denis Poisson (*1781 †1840) was a French mathematician and physicist. He accomplished many great results including the correction of the Laplace's second order partial differential equation for potential. It is also remembered as strenuous opponent of the wave theory of light; however, he was proven to be wrong by the civil engineer Augustin-Jean Fresnel (*1788 †1827). The famous experiment that brought shame upon Poisson and confirmed the Fresnel's waves theory of light is known as the Arago (*1786 †1853) spot.

i. For a given time interval $t$, how many successful instantaneous events can we observe (equivalently, how many arrivals can we observe)?

ii. How much time should we wait to observe the first success (equivalently, what is the first arrival time)?

iii. How much time should we wait to observe the second (or third, or fourth...) success (equivalently, what is the $k$th arrival time)?

i. Provided the introduction of this subsection, it should be easy to recognize that given a time $t$, the probability of $x$ successes is the Poisson distribution, $Poi(\lambda, t)$, here rewritten as

$$p_X(x; \lambda, t) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \tag{108}$$

with mean and variance given as

$$\begin{aligned} E[X] &= \lambda t, && (109) \\ \mathrm{Var}[X] &= \lambda t. && (110) \end{aligned}$$

**Common pitfalls:** oftentimes students are confused whether the Poisson distribution is a discrete or continuous distribution. It is absolutely a DISCRETE distribution. Observe that the time $t$ is a parameter of the distribution.

ii. The second question can be answered in two way. The first one is to take the limit of the geometric distribution as $n$ goes to infinity and $p = \Lambda/n$. The second approach is similar to the one we used to derive the geometric distribution. We follow the second approach. By definition

$$f_{T_1}(t) = \frac{P(0 \text{ events in } [0,t] \cap 1 \text{ event in } t + dt)}{dt}; \tag{111}$$

then,

$$\begin{aligned} f_{T_1}(t; \lambda) &= \frac{(e^{-\lambda t})(\lambda dt)}{dt}, \\ &= \lambda e^{-\lambda t}, \text{ for } t \geq 0. && (112) \end{aligned}$$

The (112) is known as the exponential distribution, $exp(\lambda)$. Mean and variance are given as follows

$$\begin{aligned} E[T_1] &= \frac{1}{\lambda}, && (113) \\ \mathrm{Var}[T_1] &= \frac{1}{\lambda^2}. && (114) \end{aligned}$$

iii. The third question can be answered in three possible ways. The first one is by taking the limit of the negative binomial as $n$ goes to infinity, and $p = \Lambda/n$. The second approach follows the same reasoning of the previous example. The third one is based on sum of statistically independent exponential random variables. In this notes, we follow the second method. By definition

$$f_{T_k}(t) = \frac{P(k-1 \text{ events in } [0,t] \cap 1 \text{ event in } t + dt)}{dt}; \tag{115}$$

then,

$$
\begin{aligned}
f_{T_k}(t;\lambda,k) &= \frac{1}{dt}\frac{(\lambda t)^{(k-1)}e^{-\lambda t}}{(k-1)!}(\lambda dt), \\
&= \frac{t^{k-1}\lambda^k e^{-\lambda t}}{(k-1)!}. \qquad (116)
\end{aligned}
$$

The (116) is known as Erlang[8] distribution which is simply a special case of the gamma distribution, $\Gamma(\lambda, k)$ with $k = 1, 2, ....$ Mean and variance are given as

$$
\begin{aligned}
E[T_k] &= \frac{k}{\lambda}, \\
\mathrm{Var}[T_k] &= \frac{k}{\lambda^2}. \qquad (117)
\end{aligned}
$$

We have seen that the Poisson process can be seen as the continuous counterpart of the discrete Bernoulli process. In particular, we have seen that the geometric distribution is memoryless, i.e. the first arrivals does not affect the the next sequence of experiments. The exponential distribution has a similar property. It follows that the inter-arrival time $T_{(k-1)-k}$ is exponential. The fact that at every instant of time the process renews itself is oftentimes counterintuitive. In particular, if for a given time $t$ nothing has happen, it must $T_1 > t$. Then, the remaining time $T - t$ is again another exponential distribution, with the same parameter $\lambda$. This can be proved mathematically as follows

$$
\begin{aligned}
P(T > t + \tau | T > t) &= \frac{P(T > t + \tau, T > t)}{P(T > t)}, \\
&= \frac{P(T > t + \tau)}{e^{-\lambda t}}, \\
&= \frac{e^{-\lambda(t+\tau)}}{e^{-\lambda t}} = e^{-\lambda\tau}. \qquad (118)
\end{aligned}
$$

It follows that the sum of statistically independent exponential distributions is the Erlang distribution.

### 7.2.1 Poisson process with random selection

Similar to the Bernoulli process, we start from a Poisson process with rate $\lambda$ and select events as follows: at each event of the original process we perform an additional experiment and we keep the event with probability $p$ and we discard it with probability $1 - p$, Figure 16. The resulting process is a Poisson process with rate $\lambda p$. To confirm this statement, we should verify the three assumptions of the Poisson process. The independence property is satisfied since each additional experiment with probability $p$ is statistically independent. The time stationarity holds because $\lambda$ and $p$ are constant in time. Finally, for a given $dt$ we have for the Poisson process with random selection $P(0) = 1 - (\lambda dt)p$ and $P(1) = (\lambda dt)p$; hence, the third property is satisfied.

**Example VI**

Suppose that the occurrence of earthquakes in the Hayward fault is modeled with a Poisson process with rate of $M \geq 5.5$ equal to $\lambda_{hf} = 0.15$. Suppose that given an earthquake of $M \geq 5.5$

[8] Agner Krarup Erlang (*1878 †1929) was a Danish statistician, mathematician, and engineer. He is known as the father of transportation engineering.

occurred, the probability of observing a peak ground acceleration ($PGA$) greater than $0.1[g]$ is 0.1, i.e. $p = P(PGA \geq 0.1[g]|M \geq 5.5) = 0.1$.

### Problem V

i What is the probability that at least one earthquake of $PGA \geq 0.1[g]$ will hit the site in Oakland in the next 20 years?

Suppose that the probability of exceeding the $PGA > 0.1[g]$ is given conditional to $R$ (epicentral distance) and $M$ magnitude, i.e. $P(PGA \geq 0.1[g]|M = m, R = r)$.

### Problem VI

i Write the joint probability distribution of $PGA$, $R$, and $M$, i.e. $f_{PGA,M,R}(pga, m, r)$.

ii Find the marginal probability of the $PGA$.

iii Describe the Poisson process of the earthquakes with $PGA \geq 0.1[g]$.

### 7.2.2 Combining Poisson processes

Suppose you have two independent Poisson processes with rate $\lambda_1$ and $\lambda_2$. The merged process is another Poisson process with rate $\lambda_1 + \lambda_2$, Figure 17. To verify this statement we should verify that the three assumptions of the Poisson process are satisfied. The independence property is satisfied since different intervals are independent within the two processes (and the two processes are independent) so it must be for the combined. The time stationarity holds because the rate $\lambda_1 + \lambda_2$ is stationary on time. Finally, for a given $dt$ we have for the combined process

$$P(0 \text{ events in } dt) = (1 - \lambda_1 dt)(1 - \lambda_2 dt) = 1 - (\lambda_1 + \lambda_2)dt + \mathcal{O}(dt^2), \quad (119)$$
$$P(1 \text{ events in } dt) = \lambda_1 dt + \lambda_2 dt - \lambda_1\lambda_2 dt^2 = (\lambda_1 + \lambda_2)dt + \mathcal{O}(dt^2); \quad (120)$$

hence, the third property is satisfied.

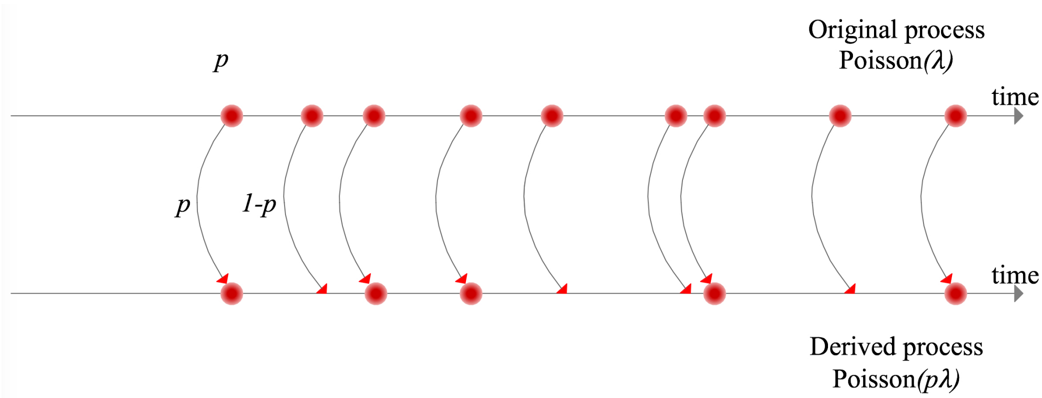


Figure 16: Poisson with random selection

**Example VII**

Suppose that the occurrence of earthquakes in the Hayward fault and St Andrea fault can be modeled as two statistically independent Poisson processes with rate of occurrence of $M \geq 6.7$ equal respectively to $\lambda_{hf} = 0.011$ and $\lambda_{af} = 0.019$.

**Problem VII**

i What is the probability of having at least one earthquake of Magnitude 6.7 or greater in our location in South Oakland?

ii Given that an earthquake occurred what is the probability that it is from St Andrea fault?

Suppose that an earthquake $M = 7$ occurred, and 4 out of the 6 buildings of Example II are heavily damaged. Assume (somewhat not very realistically) that the recovery time follow an exponential distribution, and the recovery of the 4 buildings are statistically independent.

**Problem VIII**

i What is the expected time until the last building is restored?

Table 3 summarize the distributions for the Bernoulli and Poisson processes

| | Bernoulli Process | Poisson Process |
|---|---|---|
| PMF of successes or arrivals | $bin(p, n)$, Equation (97) | $Poi(\lambda, t)$, Equation (108) |
| Interarrival time PDF | $geom(p)$, Equation (100) | $exp(\lambda)$, Equation (112) |
| $k$th arrival time | $negbin(p, k)$, Equation (104) | $\Gamma(\lambda, k)$, Equation (116) |

Table 3: Summary Bernoulli and Poisson processes

# 8 Non-Homogeneous Poisson Process, NHPP

In this section, we derive a generalized form of the Poisson process where the mean rate of events are changing with time. A non-homogeneous Poisson process is defined by the following three assumptions

**Independence:** the number of events for a given interval is independent from any other disjoint interval.

**Infinitesimal Poisson assumption:** given a small interval $dt$, the probability of events in the interval $[t, t + dt]$ can be assumed as a Poisson distribution with mean number of events $d\Lambda(t) = \lambda(t)dt$.

**Single event occurrence:** given a time $t$ and an infinitesimal time $dt$ the probability of 0 events in the time interval $[t, t + dt]$ is $P(0, [t, t + dt]) = e^{-\lambda(t)dt} = 1 - \lambda(t)dt + \mathcal{O}(dt^2)$, the probability of one event is equal to $P(1, [t + dt]) = \lambda(t)te^{-\lambda(t)dt} = \lambda(t)dt + \mathcal{O}(dt^2)$, where $\mathcal{O}(dt^2)$ are high order terms of $dt$. It follows that every single instant of time is a random variable which corresponds to a Bernoulli trial with probability of success (and expectation) $p = \lambda(t)dt$. It follows that $\lambda(t)$ is the instantaneous expected number of success for $[t, t + dt]$.

Denote with $p(x, t)$ the probability of $x$ events in the interval $[0, t]$, then given the above assumption we have

$$\begin{aligned} p(x, t + dt) &= p(x, t)p(0, [t, t + dt]) + p(x - 1, t)p(1, [t, t + dt]), && (121) \\ &= p(x, t)(1 - \lambda(t)dt) + p(x - 1, t)\lambda(t)dt, && (122) \end{aligned}$$

which leads to the following differential equation

$$\begin{aligned} \frac{p(x, t + dt) - p(x, t)}{dt} &= -\lambda(t)[p(x, t) - p(x - 1, t)], \\ \frac{dp(x, t)}{dt} &= -\lambda(t)[p(x, t) - p(x - 1, t)]. && (123) \end{aligned}$$

with initial conditions $p(0, 0) = 1$ and $p(x, 0) = 0$. The (123) can be solved recursively, for a fixed $x$. In particular for $x = 0$ we have $p(-1, 0) = 0$

$$\begin{aligned} \frac{dp(0, t)}{p(0, t)} &= -\lambda(t)dt, \\ \ln(p(0, t)) &= -\int_0^t \lambda(t)dt, \\ p(0, t) &= \exp\left(-\Lambda(t)\right), && (124) \end{aligned}$$

where

$$\Lambda(t) = \int_0^t \lambda(t)dt. \tag{125}$$

Finally we can verify by substitution that

$$p_X(x; \lambda(t), t) = \frac{\Lambda(t)^x}{x!} e^{-\Lambda(t)}. \tag{126}$$

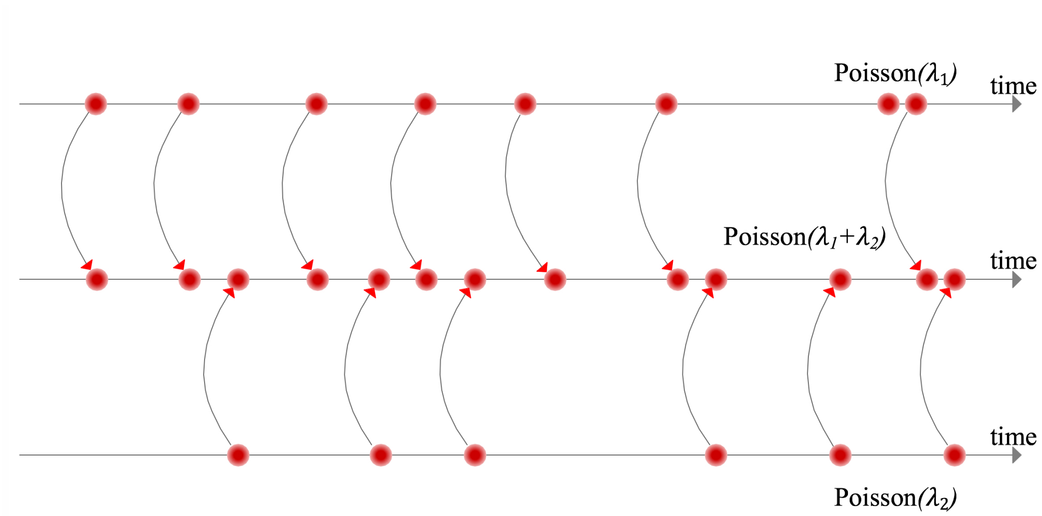


Figure 17: Combining two Poisson processes

Following the same reasoning of Section 7.2, we can derive the distribution of the first arrival as follow

$$f_{T_1}(t) = \frac{P(0 \text{ events in } [0,t] \cap 1 \text{ event in } t+dt)}{dt}; \tag{127}$$

then,

$$\begin{aligned} f_{T_1}(t;\lambda) &= \frac{(e^{-\Lambda(t)})(\lambda(t)dt)}{dt}, \\ &= \lambda(t)e^{-\Lambda(t)} \text{ for } t \geq 0. \end{aligned} \tag{128}$$

The probability distribution of the $k$th event is then

$$f_{T_k}(t) = \frac{P(k-1 \text{ events in } [0,t] \cap 1 \text{ event in } t+dt)}{dt}; \tag{129}$$

then,

$$\begin{aligned} f_{T_k}(t;\lambda,k) &= \frac{1}{dt}\frac{\Lambda(t)^{(k-1)}e^{-\Lambda(t)}}{(k-1)!}(\lambda(t)dt), \\ &= \frac{\Lambda(t)^{k-1}\lambda(t)e^{-\Lambda(t)}}{(k-1)!}. \end{aligned} \tag{130}$$

**Problem IX**

i Find the Probability distribution of the inter-arrival time. Hint: you should use the total probability theorem

ii Assume a rate function of the following type

$$\lambda(t;\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} \tag{131}$$

Derive the distribution of $T_1$.