Chapter 19

Poisson Processes

19.1 Poisson distribution

The discrete random variable $X \in \{0, 1, 2, ...\}$ has a Poisson distribution if the probability mass function has the form:

$$P(X = x) = f_X(x) = \begin{cases} e^{-\mu} \frac{\mu}{x!} & x = 0, 1, 2, \ldots \\ 0 & \text{all other } x \end{cases}$$

where $\mu > 0$ is the single parameter of the distribution.

It was shown in the earlier lectures that the mean and variance are given by

$$\mathbb{E}(X) = \mu \quad \text{and} \quad \text{var}(X) = \mu \qquad .$$

The Poisson distribution can be viewed as the limiting form of the Binomial distribution, which in itself is used to represent a series of Bernoulli trials. This limit, or rather approximation, occurs when the probability of an event is very unlikely but many trials (experiments) are involved.

So consider a sequence of n Bernoulli trials, for example the probability of two cars colliding on a busy road during rush hour in a 100 minute period.

Let p be the probability of a collision in any 1 minute period, and let us assume that $p \ll 1$ (the event is very unlikely). Assume that this probability is the same for each of the 100 minutes. Then the period of 100 minutes can be considered as $n = 100$ Bernoulli trials.

Let X be the number of accidents in n trials. Then the number of accidents will have a binomial distribution.

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad , \qquad x = 0, 1, 2, 3, \ldots, n \qquad . \tag{19.1}$$

where $q = 1 - p$.

Suppose n is very large, and we are only interested is small values of x. Then, $n \gg x$, and it follows that

$$n - 1 \approx n \quad , \quad n - 2 \approx n \quad \cdots \quad n - x \approx n \qquad . \tag{19.2}$$

Then it follows that:

$$P(X=x) = \frac{n(n-1)(n-2)...(n-x+1)}{x!}p^x q^{n-x} \approx \frac{n^x}{x!}p^x q^n \qquad . \tag{19.3}$$

Let us assume that, although, $p \to 0$, np has a finite value,

$$np = \mu \qquad . \tag{19.4}$$

Then, under these conditions, to a good approximation:

$$P(X=x) \approx \frac{n^x}{x!}\left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^n \qquad . \tag{19.5}$$

Since

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = \mathrm{e}^x \tag{19.6}$$

it follows that:

$$\lim_{n\to\infty} P(X=x) \approx \frac{n^x}{x!}\frac{\mu^x}{n^x}\mathrm{e}^{-\mu} \qquad . \tag{19.7}$$

That is:

$$P(X=x) \approx \mathrm{e}^{-\mu}\frac{\mu^x}{x!} \qquad (x=0,1,2,\ldots) \qquad . \tag{19.8}$$

So the Binomial distribution tends to a Poisson distribution when $n \to \infty$ and $p \to 0$, with $np \to \mu$. The accuracy of the approximation is very good when $p \leq 0.1$, as indicated in the figure (19.1). It is not too surprising that the shapes of the probability masses have similarities when we recall that the mean and variance of the binomial distribution are, np and npq, respectively. In the approximation $\mu = np$ the means are in exact agreement, while the variance of the Poisson distribution $\mu = np \approx npq$ (since $q = 1 - p \approx 1$).

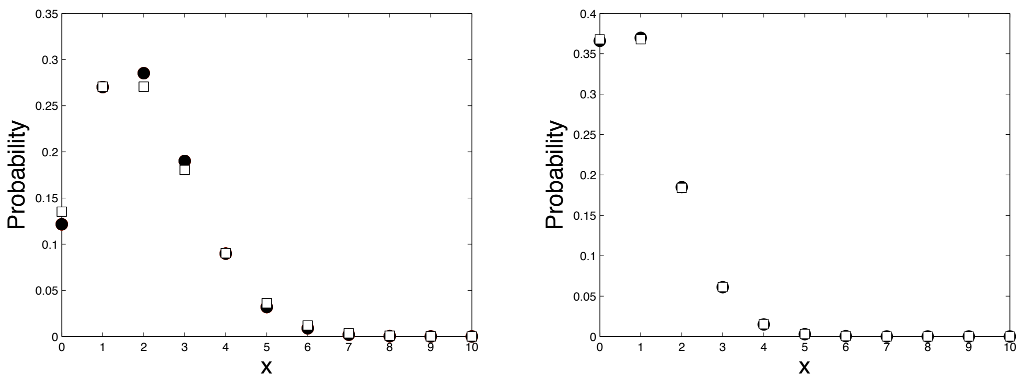


Figure 19.1: Comparison of the binomial distribution (black circles) and the equivalent Poisson approximation (white squares). Left $p = 0.1$, $n = 20$, giving $\mu = 2$. RIght $p = 0.01$, $n = 100$. For these parameters, the approximation is very good.

EXAMPLE

Suppose each car on the road, each day, has the same probability $p \approx 0.001$ of having an accident. If there are $n = 3000$ cars on the road each day. Then we have $n = 3000$ Bernoulli trials for having an accident.

This is an example of a rare event, but with many trials.

The probability of *no* accidents in a day is:

$$P(X=0) = \mathrm{e}^{-\mu}\frac{\mu^0}{0!} \qquad .$$

Here

$$\mu = np = 3000 \times 0.001 = 3.0 \qquad . \tag{19.9}$$

So that:

$$P(X=0) = \mathrm{e}^{-3} \approx 0.0498 \qquad . \tag{19.10}$$

Therefore, the probability of one or more accidents is 95%.

The probability of no accidents is given exactly by the binomial distribution:

$$P(X=0) = \binom{3000}{0} q^{3000} p^0 = (0.999)^{3000} \approx 0.0497 \qquad . \tag{19.11}$$

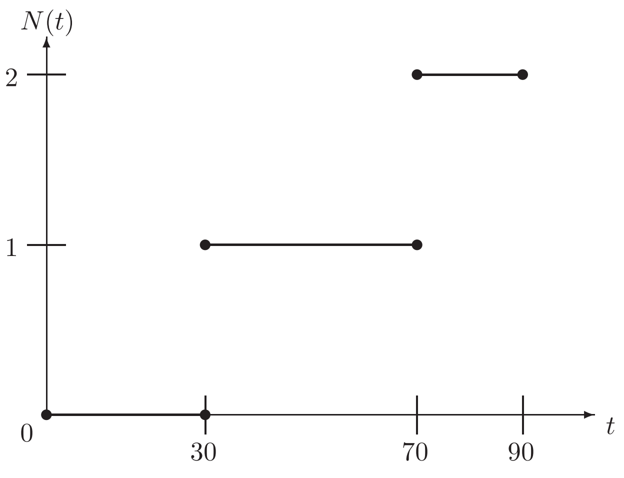
The following properties of the Poisson distribution are certainly familiar to you, but are well worth memorising:

$$\boxed{\mathbb{E}\,(X) = \mu} \qquad , \qquad \boxed{\mathrm{var}\,(X) = \mu} \qquad . \tag{19.12}$$

19.2 Poisson process

Consider random events in time. A 'counting process' $\{N(t), t \geqslant 0\}$ is the total number of events that have occurred by time t. $N(t)$ is a discrete random variable that depends on a continuous parameter - time. For example, $N(t)$ = number of goals scored by time t in a football match.

Arsenal vs Newcastle
(1 - 1)



(Dyer 30 mins, Henry 70 mins)

Any 'counting process' $N(t)$ must satisfy the conditions:

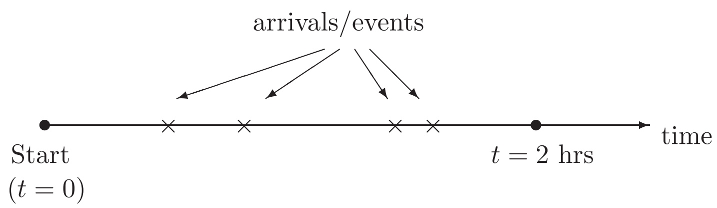
(i) $N(t) \geqslant 0$

(ii) $N(t) \in \{0, 1, 2, \ldots\}$

(iii) if $s < t$, then $N(s) \leqslant N(t)$

(iv) if $s < t$, then $N(t) - N(s)$ = number of events that occurred in the interval $s <$ time $\leqslant t$

Another example of a counting process could be the number of text messages received on a phone during a two-hour period.



A 'counting process' has '*independent increments*' if the number of events occurring in *disjoint* time intervals are *independent*. For example, the number of text messages received between 10.15 and 10.25 being independent of the number received between 16.10 and 16.40. Perhaps the number of goals scored in the first half is independent of the number of goals scored in the second half of a football match.

A 'counting process' has 'stationary increments' if the probability distribution of the number of events in a time interval depends *only* on the duration (length of interval) and *not* on the timing (time of measurement).

That is:

$$\begin{array}{c}\text{Probability of number of events}\\ \text{in interval } (t_1, t_2)\end{array} = \begin{array}{c}\text{Probability of number of events}\\ \text{in interval } (t_1 + s, t_2 + s)\end{array}$$

for all s. Clearly this is not true for text messages for daytime and night-time. The most important 'counting process' is the *Poisson process* which we define as follows.

The 'counting process' $\{N(t), t \geqslant 0\}$ is a *Poisson process* with a *rate* (or *intensity*) λ if:

(i) $N(t = 0) = 0$.

(ii) the process has independent increments.

(iii) the number of events in *any* interval of duration t is Poisson distributed with mean λt.

That is, for all $s, t \geqslant 0$:

$$P\left\{N(t+s) - N(s) = n\right\} = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!} \tag{19.13}$$

From (19.13) it is clear that, if this holds for all s, then the increments are stationary. Then, $\mathbb{E}(N(t))$ (that is, the expected number of events between $t = 0$ and $t = t$) is:

$$\begin{aligned}\mathbb{E}(N(t)) &= \sum_{n=0}^{\infty} n \cdot P(N(t) = n)\\ &= \sum_{n=0}^{\infty} n \cdot \mathrm{e}^{-(\lambda t)}\frac{(\lambda t)^n}{n!}\\ &\boxed{\mathbb{E}(N(t)) = \lambda \cdot t}\end{aligned}$$

That the number of expected events is linearly proportional to the time of observation, and the proportionality constant is called the *rate* of the process.

19.2.1 $O(h)$ and $o(h)$

Now for a bit of calculus. We say $f(h) = O(h)$ if:

$$\lim_{h\to 0} \frac{f(h)}{h} = \text{finite constant}$$

We say $f(h) = o(h)$ if:

$$\lim_{h\to 0} \frac{f(h)}{h} = 0$$

That is, $o(h)$ means 'goes to zero faster than h'.

$$\begin{array}{lll} \text{eg} & f(x) = x^2 & \text{is} \quad O(h) \to \text{goes to zero faster.} \\ \text{but} & f(x) = x & \text{is} \quad O(h) \to \text{goes to zero as } h. \end{array}$$

This notation leads to some slightly bizarre arithmetic:

$$o(h) + o(h) = o(h) \quad , \quad \text{and} \qquad o(h) = -o(h)$$

which is just our way of saying 'something tiny plus something tiny gives something else tiny', and subtracting or adding a very very small number makes a tiny difference. These are not 'real equations'. In fact these equations are really only true in the limit $h \to 0$. So whatever we do, we must make sure to take this limit at the end to make our mathematics consistent.

19.3 Alternative definition of a Poisson process

There is an equivalent way of defining a Poisson process (let's call it definition 2 to avoid confusion), and it is often of more practical use.

The 'counting process' $\{N(t), t \geqslant 0\}$ is Poisson if:

(i) $N(t = 0) = 0$.

(ii) the increments are *independent* and *stationary*.

(iii) $P\{N(h) = 1\} = \lambda h + o(h)$.

(iv) $P\{N(h) > 1\} = o(h)$.

The meaning of (iii) and (iv) is that, over a *very* short interval of time (h), the probability of more than one event happening is negligible. That is, one will observe, at most, one event in this short time, and (according to (iii)) the probability of the single event is proportional to the interval of time (duration of observation).

Note that we can infer $P(N(h) = 0)$ from this definition. Since either 0,1,2,... events must occur in the time interval h (law of total probability - that is something happens or nothing happens) we can express this as:

$$\sum_{n=0}^{\infty} P(N(h) = n) = 1 \qquad . \tag{19.14}$$

This in turn means:

$$P(N(h)=0)+P(N(h)=1)+P(N(h)=2)+P(N(h)=3)+\cdots=1 \quad . \tag{19.15}$$

That is

$$P(N(h)=0)+\lambda h+o(h)+o(h)+o(h)+\cdots=1 \tag{19.16}$$

which gives:

$$P(N(h)=0)=1-\lambda h+o(h) \quad . \tag{19.17}$$

Let us verify that *definition* 1 is consistent with *definition* 2, by considering the probability of no events occurring in a very small interval h. According to definition 1:

$$P(N(h)=0)=e^{\lambda h}\times\frac{(\lambda h)^0}{0!}=e^{-\lambda h} \quad . \tag{19.18}$$

Now since h is very small, we can make a series approximation:

$$P(N(h)=0)=e^{-\lambda h}=1+\frac{(-\lambda h)}{1!}+\frac{(-\lambda h)^2}{2!}+\cdots \quad . \tag{19.19}$$

That is:

$$P(N(h)=0)=1-\lambda h+o(h) \quad . \tag{19.20}$$

This relation is consistent with what we derived from definition 2. Similarly, again using definition 1, and employing a series expansion for the exponential function, we find:

$$P(N(h)=1)=e^{\lambda h}\times\frac{(\lambda h)^1}{1!}=\lambda h+o(h) \quad . \tag{19.21}$$

Later we will show that the converse is also true: definition 2 will imply definition 1. For the present let's assume that they are entirely equivalent.

19.4 Examples of Poisson processes

19.4.1 Poissonian football

Let us consider some real data for a change. A statistical study was carried out by Chu in 2003 [1] of football matches in the four World Cup tournaments 1990, 1994, 1997 and 2002. The probability distribution of goals per mach and the waiting times between goals were analysed. A total of 232 matches were taken as the data set.To ensure that each match was of identical duration, The number of goals scored in regulation time (90 minutes each match) was recorded.

A total of 575 goals was recorded giving a sample average of $575/232 \approx 2.4784$ goals per game. The variance of the number of goals per game was found to be 2.4584. This being close to the mean is suggestive of a Poisson distribution, and this is confirmed by the data shown in figure (19.2). It is clear that, for this data, there is a good fit to the Poisson distribution. However, this does not, of itself, prove that goal scoring is a Poisson process. This requires a closer examination of the scoring process, and this is considered in the next chapter.

[1] Chu, S. (2003), "Motivating the Poisson Process Using Goals in Soccer." INFORMS Transactions on Education, Vol. 3, No. 2,

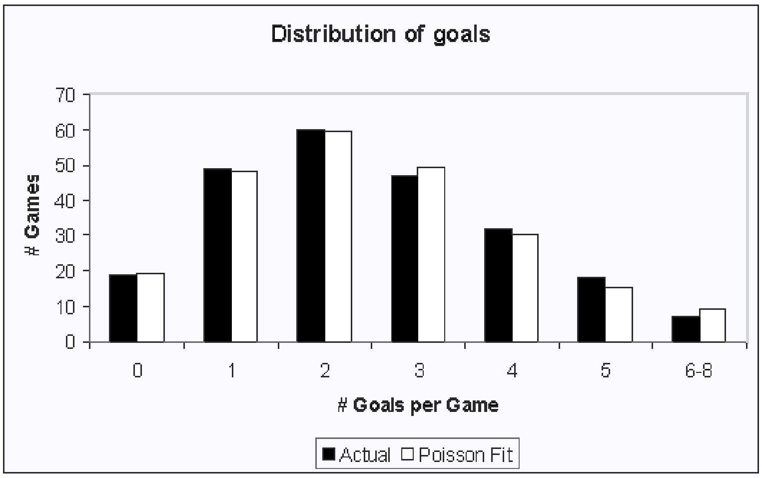


Figure 19.2: The sample distribution of goals per match compared with a theoretical Poisson distribution, $\lambda = 2.478$ goals per game. The data match the Poisson model very well. Note that the range of events, 6-8 goals, is binned into a single value

19.4.2 Poissonian flu

Flu epidemics are unpredictable in their timing, but devastating in their effect on human populations. The most recent pandemics occurred in 1918, 1957, and 1968/69 prior to H1N1 this year 2009. So one question is what precautions should be taken to prepare for a pandemic. In particular, is it cost effective to stock pile anti-viral medicine. SInce anti-viral drugs have a limited shelf-life (roughly 5 years) long-term maintenance of stock is a significant cost. THe UK government spent (wisely it appears) £500 million on antiviral doses.

The arrival of an epidemic is treated as a Poisson process in calculating the risk associated with buying such drugs,

19.4.3 Credit risk

One of the major challenges facing the financial markets is the pricing of credit risk, For example, in extending a loan, what is the likelihood that the loan will not be repaid. A bank extending a large number of loans can view the default process (failure to pay) as a Poisson process. That is, the timing of defaults is to some extent predictable if it is Poissonian.

19.4.4 Data traffic

The transmission of data over a network is not a smooth continuous flow. Requests for data, for example visit to web pages, tend to be made in a random manner, by a group of unpredictable individuals. These requests for data tend to be in bursts that mimic a classic Poisson process. That is the interarrival times tend, to a good approximation, towards an exponential distribution. Thus, in designing servers and routers this randomness needs to be taken into account.

The 3G network technology uses time-sharing. The base station transmits to a mobile in bursts lasting 2 milliseconds, and this is shared evenly in what is called a round-robin protocol. Thus the available bandwidth depends on the capacity of the hand-set and the number of users. The calls to the base station arise at random times that are well describe by a Poisson process model.

19.5 Poisson bunching

Since a random Poisson process is completely unpredictable, and the probability of an event happening in any 5-minute interval is the same, one might expect that such a process leads to an even distribution of events. However, a closer analysis reveals that the are often long time gaps with no arrivals and occasionally multiple events within a short period. This explains why buses often come in pairs even when they depart by equally spaced intervals.

The *grouping* or *bunching* of events in time can be illustrated by using an example. Suppose that a football team scores goals, in the manner of a Poisson process, at a rate $\lambda = 2.5$ per hour. Recall that a football match lasts 90 mins (1.5 hours) in total.

QUESTION

Calculate the probability that the team scores no goals in a match.

Calculate the probability that the team scores exactly 1 goal in the second half.

If we are told that exactly 3 goals were scored. Calculate the probability that the goals were evenly separated. one goal in the first 30 minutes, one in the second 30 minutes, and one in the third 30 minutes?

SOLUTION

We are told this is a Poisson process with rate (intensity) $\lambda = 2.5$ per hour. That means in any interval of time, t, the probability of $N(t) = n$ goals is:

$$P(N(t) = n) = e^{-\lambda t} \times \frac{(\lambda t)^n}{n!} \qquad , \qquad n = 0, 1, 2, \ldots \qquad . \tag{19.22}$$

So the probability of $n = 0$ goals in $t = 1.5$ hours is:

$$P(N(1.5) = 0) = e^{-2.5\times 1.5} \times \frac{(2.5 \times 1.5)^0}{0!} = e^{-3.75} = 0.0235 \qquad . \tag{19.23}$$

That is, this is a very unlikely event.

The probability of exactly one goal in the second half is:

$$P(N(1.5) - N(0.75) = 1) = P(N(0.75) - N(0) = 1)$$

where we have used the homogeneity property, that is we have no memory of the first half. Indeed the probability of scoring one goal (exactly) in the first half would be the same.

$$P(N(0.75) = 1) = e^{-2.5\times 0.75} \times 2.5 \times 0.75 = 0.2875 \qquad .$$

We seek the value of:

$$P\left[N(1.5) - N(1) = 1, N(1.0) - N(0.5) = 1, N(0.5) - N(0) = 1 | N(1.5) = 3\right] \qquad .$$

The three thirds are disjoint intervals of time, and hence independent. Then we can write:

$$\frac{P\left[N(1.5) - N(1) = 1\right] P\left[N(1.0) - N(0.5) = 1\right] P\left[N(0.5) - N(0) = 1\right]}{P\left[N(1.5) = 3\right]} \qquad .$$

$$\frac{(P[N(0.5)=1])^3}{P[N(1.5)=3]} = \frac{3!}{27} = \frac{2}{9} \approx 0.2222 \qquad .$$

So, the probability that the scores were evenly spread throughout the match is also small. Conversely the probability that that two goals were scored in one of the thirds is higher. There are 6 ways that this can happen, if we consider the goals scored in each third we find.

$$\{0,1,2\},\{0,2,1\},\{1,0,2\},\{1,2,0\},\{2,0,1\},\{2,1,0\} \qquad .$$

So we should include all 6 possibilities, that is:

$$6 \times \frac{P[N(1.5)-N(1)=0]\,P[N(1.0)-N(0.5)=1]\,P[N(0.5)-N(0)=2]}{P[N(1.5)=3]} = \frac{6}{9} \approx 0.6667 \qquad .$$

That is, it is three times more likely that the goals are 'bunched' together than spread out. Finally, we can calculate the probability that all 3 goals occurred in the same third of the game. The possible scenarios are

$$\{0,0,3\},\{0,3,0\},\{3,0,0\} \qquad .$$

and each of these has a probability 1/27, so the chance that all 3 goals were scored in a third of the game is: 1/9.

19.6 Mixing and thinning

19.6.1 Mixing or superposition

Suppose we have two independent Poisson processes occurring simultaneously. For example the arrival of customers at a pizza restaurant, $N_1(t)$, and the number of telephone orders for take-away pizza, $N_2(t)$. We denote the rates of each process are λ_1 and λ_2, respectively.

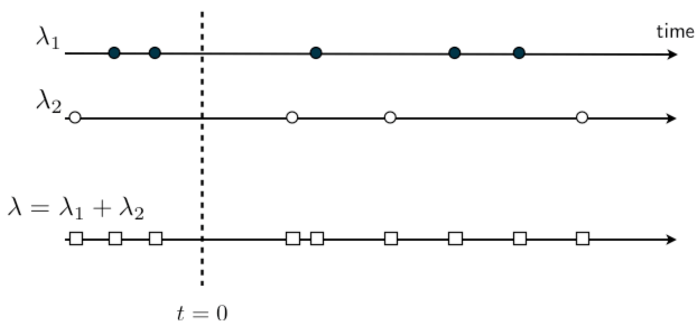


Figure 19.3: Mixing: A Poisson process (top line) of events (of type I) with rate λ_1 when added to a second (independent) Poisson process (with events of type II) with rate λ_2, creates a mixed Poisson process with rate $\lambda = \lambda_1 + \lambda_2$.

Then the combined process, that is mixing or superposing the two processes:

$$N(t) \equiv N_1(t) + N_2(t) \tag{19.24}$$

which counts all customers, is a Poisson process with rate

$$\boxed{\lambda = \lambda_1 + \lambda_2} \quad . \tag{19.25}$$

Although the result seems fairly obvious, particularly if we inspect figure (19.3) and hardly worth proving. Nonetheless let's show this is true mathematically.

Firstly we note that:

$$N(0) = N_1(0) + N_2(0) = 0 + 0 = 0$$

which is one of the four defining equations satisfied. Next, consider a short time interval $[0, h]$, and the probability that one event, of either kind, would be observed:

$$P(N(h) = 1) = P(N_1(h) = 0, N_2(h) = 1 \quad \text{or} \quad N_1(h) = 1, N_2(h) = 0)$$

Now these are disjoint (mutually exclusive events), $N_1 = 0$ or $N_= 1$, but not both, we can write:

$$P(N_1(h) = 0, N_2(h) = 1) + P(N_1(h) = 1, N_2(h) = 0) \quad .$$

Since, we are told that N_1 and N_2 are independent - there is no correlation between the two types of event, then

$$P(N_1(h) = 0, N_2(h) = 1) = P(N_1(h) = 0)P(N_2(h) = 1) \quad .$$

Using definition 2 we write:

$$P(N(h) = 1) = (1 - \lambda_1 h + o(h))(\lambda_2 h + o(h)) + (\lambda_1 h + o(h))(1 - \lambda_2 h + o(h)) = (\lambda_1 + \lambda_2)h + o(h) \quad .$$

Therefore $N(t)$ is a Poisson process with a rate $\lambda = \lambda_1 + \lambda_2$. To complete the proof, we should show that:

$$P(N(h) = 0) = 1 - (\lambda_1 + \lambda_2)h + o(h)$$

and we leave this as an exercise.

19.6.2 Thinning

Suppose that the Poisson process $\{N(t), t \geq 0\}$ with rate λ, consists of two types of event - type I and type II. For example the Poisson process could be the arrival of customers at a shop, and the customers are type I (male) or type II (female).

If the type of event is also random, say given an event has arrived, the probability that it is type I is $0 \leq p \leq 1$. Furthermore, let us assume that this is true for each event (that is the type is independent and identically distributed for all events). Of course the event must be either of two types, then it follows that type II will have a probability $q = 1 - p$. If we count type I with the variable, $N_1(t)$ and type II with the variable $N_2(t)$, then clearly:

$$N(t) = N_1(t) + N_2(t)$$

The we can show that:
$N_1(t)$ is a Poisson process with rate $\lambda_1 = \lambda p$, and
$N_1(t)$ is a Poisson process with rate $\lambda_2 = \lambda q$.

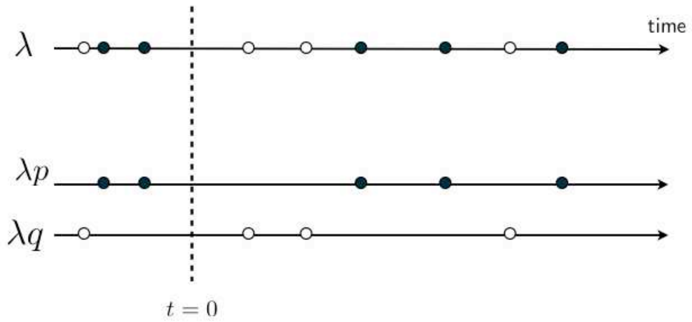


Figure 19.4: Thinning: A Poisson process (top line) with rate λ has events of two types. The two types occur randomly, type I having a probability $0 \leq p \leq 1$, type II having a probability $1-p$, and every event is independent in this regard. Suppose we *thin* this sequence of event, by listing the sequences of each type separately. Then these subsequences are also both Poisson processes. Type I is a Poisson process with rate $\lambda_1 = \lambda p$ and Type II is Poisson with rate $\lambda_2 = \lambda(1-p)$.

We use the same idea of a short time interval $[0, h]$ and consider the probability that we observe a type I event in this interval, that is that an event occurs, and that it is of type I, namely:

$$P(N_1(h) = 1) = P(N(h) = 1 \text{ and type I}) = P(N(h) = 1) \times P(\text{type I}) = (\lambda h + o(h))p = (\lambda p)h + o(h) \quad .$$

Thus, $N_1(t)$, is Poissonian with a rate λp. The same argument can be used for type 2. This is called thinning as the Poisson process is separated into its component types.

19.7 Differential equations for the Poisson process

Consider the Poisson process defined by the occurrence of events in small intervals of time. This leads us to differential equations for the event arrivals. Recall that, for a Poisson process, $N(0) = 0$, and the events (increments) are independent and stationary.

Let us use the shorthand:

$$P[N(t) = n] \equiv p_n(t) \quad . \tag{19.26}$$

Then for a Poisson process, let us consider the value of $p_0(t + h)$, the probability that no events have occurred on/before a time $t + h$. This means that no event occur in the time intervals $[0, t]$ and $[t, t + h]$, as for the frog crossing the road. Then we have:

$$P[N(t + h) = 0] = P[N(t) = 0, \text{and}, N(t + h) - N(t) = 0] \tag{19.27}$$

Since these time intervals are disjoint, then they are independent, by our axiom of independent increments.

$$P[N(t + h) = 0] = P[N(t) = 0] \times P[N(t + h) - N(t) = 0] \quad , \tag{19.28}$$

Then, we have the stationary (homogeneous) property:

$$P[N(t + h) - N(t) = 0] = P[N(h) - N(0) = 0] = P[N(h) = 0] = 1 - \lambda h + o(h) \quad . \tag{19.29}$$

Then (19.28) can be written as:

$$p_0(t+h) = p_0(t) \times \{1 - \lambda h + o(h)\} \tag{19.30}$$

Then we can write,

$$\frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t) + \frac{o(h)}{h}\} \tag{19.31}$$

Now taking the limit $h \to 0$, we get the first-order differential equation:

$$\boxed{\frac{dp_0(t)}{dt} = -\lambda p_0(t)} \quad . \tag{19.32}$$

Recall that our initial condition are: $N(0) = 0$, therefore it is certain that no events will occur the moment we start counting:

$$p_0(0) = 1 \quad . \tag{19.33}$$

This provides the initial value condition for the differential equation, which can be integrate by separating the variables:

$$\int \frac{dp_0(t)}{p_0(t)} = -\int \lambda dt \quad , \tag{19.34}$$

$$\ln p_0(t) = -\lambda t + C \quad , \tag{19.35}$$

where C is the constant of integration. Using the initial condition (19.33):

$$\ln 1 = C \quad \textsf{giving} \qquad C = 0 \quad . \tag{19.36}$$

Then:

$$p_0(t) = e^{-\lambda t} \quad , \tag{19.37}$$

which agrees with definition 1.

We expressions for $p_n(t)$, for $n = 1, 2, 3, \ldots$ extending this idea. Conditioning on the number of events at time t, we have:

$$P[N(t+h) = n] = \sum_{m=0}^{\infty} P[N(t+h) = n | N(t) = m] P[N(t) = m] \tag{19.38}$$

Now consider the term $P[N(t+h) = n|N(t) = m]$. Clearly, since we have a counting process, it is impossible for the count to decrease as time increases:

$$P[N(t+h) = n|N(t) = m] = 0 \quad m > n \quad . \tag{19.39}$$

So, in the sum (19.38) we can ignore all terms $m > n$. But we can also ignore most of the terms $m \leq n$, for the following reason. If $m = n - 2$ for example:

$$P[N(t+h) = n|N(t) = n-2] = P[N(t+h) - N(t) = 2] = o(h) \quad . \tag{19.40}$$

That is, in a tiny time interval h the chance of two events occurring is negligible. At most we would have one event, according to our definition 2. So all terms in the sum, for $m \leq n - 2$ can be ignored. Then

(19.38) collapses to:

$$P[N(t+h)=n] = P[N(t+h)=n|N(t)=n]P[N(t)=n]+P[N(t+h)=n|N(t)=n-1]P[N(t)=n-1] \tag{19.41}$$

Then, using our definition, this can be simplified to:

$$p_n(t+h) = P[N(t+h)=n|N(t)=n]p_n(t) + P[N(t+h)=n|N(t)=n-1]p_{n-1}(t) \quad . \tag{19.42}$$

Note that:

$$P[N(t+h)=n|N(t)=n] = P[N(t+h)-N(t)=0] = P[N(h)=0] = 1-\lambda h + o(h) \tag{19.43}$$

while,

$$P[N(t+h)=n|N(t)=n-1] = P[N(t+h)-N(t)=1] = P[N(h)=1] = \lambda h + o(h) \tag{19.44}$$

This gives us:

$$p_n(t+h) = \{1-\lambda h + o(h)\}p_n(t) + \{\lambda h + o(h)\}p_{n-1}(t) \quad , \tag{19.45}$$

and then,

$$\frac{p_n(t+h)-p_n(t)}{h} = \left\{-\lambda + \frac{o(h)}{h}\right\}p_n(t) + \left\{\lambda + \frac{o(h)}{h}\right\}p_{n-1}(t) \quad , \tag{19.46}$$

Taking the limit $h \to 0$ we get:

$$\boxed{\frac{dp_n(t)}{dt} = -\lambda p_n(t) + \lambda p_{n-1}(t) \quad n = 1, 2, 3, \ldots} \quad . \tag{19.47}$$

The boundary (initial) conditions for this set of differential equations is:

$$p_n(0) = 0 \quad , \quad n = 1, 2, 3, \ldots \quad , \tag{19.48}$$

since no events have occurred before we start counting.

We can show, by substitution, that the solutions of this (infinite) set of differential equations is:

$$\boxed{p_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!} \quad , \quad n = 1, 2, 3, \ldots} \tag{19.49}$$

This is the definition 1 that we started with, showing that the two definitions are equivalent.

One can prove this solution by direct integration of the differential equations (19.47) as follows. Firstly, rewriting,

$$\frac{dp_n(t)}{dt} + \lambda p_n(t) = \lambda p_{n-1}(t) \quad n = 1, 2, 3, \ldots \quad . \tag{19.50}$$

First we use an integrating factor to simplify, that is we multiply across by $e^{\lambda t}$.

$$e^{\lambda t}\frac{dp_n(t)}{dt} + \lambda e^{\lambda t}p_n(t) = \lambda e^{\lambda t}p_{n-1}(t) \quad n = 1, 2, 3, \ldots \quad . \tag{19.51}$$

This can be written as

$$\frac{d}{dt}\left\{e^{\lambda t}p_n(t)\right\} = \lambda\left\{e^{\lambda t}p_{n-1}(t)\right\} \quad n = 1, 2, 3, \ldots \quad . \tag{19.52}$$

We can solve this iteratively, since we know $p_0(t)$. For $n = 1$,

$$\frac{d}{dt}\left\{e^{\lambda t} p_1(t)\right\} = \lambda \left\{e^{\lambda t} p_0(t)\right\} = \lambda \quad . \tag{19.53}$$

So, integrating and rearranging.

$$p_1(t) = e^{-\lambda t} \lambda t \quad . \tag{19.54}$$

Then, for $n = 2$:

$$\frac{d}{dt}\left\{e^{\lambda t} p_2(t)\right\} = \lambda \left\{e^{\lambda t} p_1(t)\right\} = \lambda(\lambda t) \quad . \tag{19.55}$$

Now integrating gives:

$$p_2(t) = e^{-\lambda t} \frac{(\lambda t)^2}{2} \quad . \tag{19.56}$$

And this process can be repeated (recursively) to arrive at the general case:

$$p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad , \quad n = 0, 1, 2, 3 \ldots \quad . \tag{19.57}$$