21

The Exponential Distribution

From Discrete-Time to Continuous-Time:

In Chapter 6 of the text we will be considering Markov processes in continuous time. In a sense, we already have a very good understanding of continuous-time Markov chains based on our theory for discrete-time Markov chains. For example, one way to describe a continuous-time Markov chain is to say that it is a discrete-time Markov chain, except that we explicitly model the times between transitions with continuous, positive-valued random variables and we explicity consider the process at any time t, not just at transition times.

The single most important continuous distribution for building and understanding continuous-time Markov chains is the exponential distribution, for reasons which we shall explore in this lecture.

The Exponential Distribution:

A continuous random variable X is said to have an Exponential(λ) distribution if it has probability density function

$$f_X(x|\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases},$$

where $\lambda > 0$ is called the *rate* of the distribution.

In the study of continuous-time stochastic processes, the exponential distribution is usually used to model the *time until something happens in the process*. The mean of the Exponential(λ) distribution is calculated using integration by parts as

$$\begin{aligned} E[X] &= \int_0^\infty x\lambda e^{-\lambda x} dx \\ &= \lambda \left[\frac{-xe^{-\lambda x}}{\lambda}\Big|_0^\infty + \frac{1}{\lambda}\int_0^\infty e^{-\lambda x} dx \right] \\ &= \lambda \left[0 + \frac{1}{\lambda} \frac{-e^{-\lambda x}}{\lambda}\Big|_0^\infty \right] \\ &= \lambda \frac{1}{\lambda^2} = \frac{1}{\lambda}. \end{aligned}$$

So one can see that as λ gets larger, the thing in the process we're waiting for to happen tends to happen more quickly, hence we think of λ as a rate.

As an exercise, you may wish to verify that by applying integration by parts twice, the second moment of the Exponential(λ) distribution is given by

$$E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} = \ldots = \frac{2}{\lambda^2}.$$

From the first and second moments we can compute the variance as

$$\mathsf{Var}(X) = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

The Memoryless Property:

The following plot illustrates a key property of the exponential distribution. The graph after the point s is an exact copy of the original function. The important consequence of this is that the distribution of X conditioned on $\{X > s\}$ is *again exponential*.

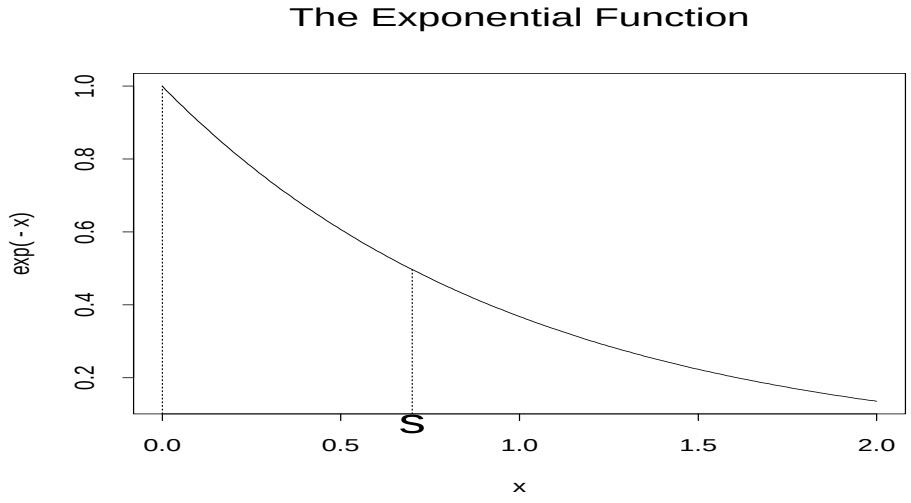

Figure 21.1: The Exponential Function e^{-x}

To see how this works, imagine that at time 0 we start an alarm clock which will ring after a time X that is exponentially distributed with rate λ. Let us call X the *lifetime* of the clock. For any $t > 0$, we have that

$$P(X > t) = \int_t^\infty \lambda e^{-\lambda x} dx = \lambda \left. \frac{-e^{-\lambda x}}{\lambda} \right|_t^\infty = e^{-\lambda t}.$$

Now we go away and come back at time s to discover that the alarm has not yet gone off. That is, we have observed the event $\{X > s\}$. If we let Y denote the *remaining* lifetime of the clock given that $\{X > s\}$, then

$$\begin{aligned} P(Y > t | X > s) &= P(X > s + t | X > s) \\ &= \frac{P(X > s + t, X > s)}{P(X > s)} \\ &= \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t}. \end{aligned}$$

But this implies that the remaining lifetime after we observe the alarm has not yet gone off at time s has the same distribution as the original lifetime X. The really important thing to note, though, is that this implies that the distribution of the remaining lifetime *does not depend on* s. In fact, if you try setting X to have *any other* continuous distribution, then ask what would be the distribution of the remaining lifetime after you observe $\{X > s\}$, the distribution will depend on s.

This property is called the *memoryless* property of the exponential distribution because I don't need to remember when I started the clock. If the distribution of the lifetime X is Exponential(λ), then if I come back to the clock at any time and observe that the clock has not yet gone off, regardless of when the clock started I can assert that the distribution of the time till it goes off, starting at the time I start observing it again, is Exponential(λ). Put another way, given that the clock has currently not yet gone off, I can forget the past and still know the distribution of the time from my current time to the time the alarm will go off. The resemblance of this property to the Markov property should not be lost on you.

It is a rather amazing, and perhaps unfortunate, fact that the exponential distribution is the only one for which this works. The memoryless property is like enabling technology for the construction of continuous-time Markov chains. We will see this more clearly in Chapter 6. But the exponential distribution is even more special than just the memoryless property because it has a second enabling type of property.

Another Important Property of the Exponential:

Let $X_1, \ldots, X_n$ be independent random variables, with X_i having an Exponential(λ_i) distribution. Then the distribution of $\min(X_1, \ldots, X_n)$ is Exponential($\lambda_1 + \ldots + \lambda_n$), and the probability that the minimum is X_i is $\lambda_i/(\lambda_1 + \ldots + \lambda_n)$.

Proof:

$$\begin{aligned} P(\min(X_1, \ldots, X_n) > t) &= P(X_1 > t, \ldots, X_n > t) \\ &= P(X_1 > t) \ldots P(X_n > t) \\ &= e^{-\lambda_1 t} \ldots e^{-\lambda_n t} \\ &= e^{-(\lambda_1 + \ldots + \lambda_n)t}. \end{aligned}$$

The preceding shows that the CDF of $\min(X_1, \ldots, X_n)$ is that of an Exponential($\lambda_1 + \ldots + \lambda_n$) distribution. The probability that X_i is the minimum can be obtained by conditioning:

$$\begin{aligned}
&P(X_i \text{ is the minimum}) \\
&= P(X_i < X_j \text{ for } j \neq i) \\
&= \int_0^\infty P(X_i < X_j \text{ for } j \neq i | X_i = t) \lambda_i e^{-\lambda_i t} dt \\
&= \int_0^\infty P(t < X_j \text{ for } j \neq i) \lambda_i e^{-\lambda_i t} dt \\
&= \int_0^\infty \lambda_i e^{-\lambda_i t} \prod_{j \neq i} P(X_j > t) dt \\
&= \int_0^\infty \lambda_i e^{-\lambda_i t} \prod_{j \neq i} e^{-\lambda_j t} dt \\
&= \lambda_i \int_0^\infty e^{-(\lambda_1 + \ldots + \lambda_n)t} dt \\
&= \lambda_i \left. \frac{-e^{-(\lambda_1 + \ldots + \lambda_n)t}}{\lambda_1 + \ldots + \lambda_n} \right|_0^\infty \\
&= \frac{\lambda_i}{\lambda_1 + \ldots + \lambda_n},
\end{aligned}$$

as required. □

To see how this works together with the the memoryless property, consider the following examples.

Example: (Ross, p.332 #20). Consider a two-server system in which a customer is served first by server 1, then by server 2, and then departs. The service times at server i are exponential random variables with rates μ_i, $i = 1, 2$. When you arrive, you find server 1 free and two customers at server 2 — customer A in service and customer B waiting in line.

(a) Find P_A, the probability that A is still in service when you move over to server 2.

(b) Find P_B, the probability that B is still in the system when you move over to 2.

(c) Find $E[T]$, where T is the time that you spend in the system.

Solution:

(a) A will still be in service when you move to server 2 if your service at server 1 ends before A's remaining service at server 2 ends. Now A is currently in service at server 2 when you arrive, but because of memorylessness, A's remaining service is Exponential(μ_2), and you start service at server 1 that is Exponential(μ_1). Therefore, P_A is the probability that an Exponential(μ_1) random variable is less than an Exponential(μ_2) random variable, which is

$$P_A = \frac{\mu_1}{\mu_1 + \mu_2}.$$

(b) B will still be in the system when you move over to server 2 if your service time is less than the sum of A's remaining service time and B's service time. Let us condition on the first thing to happen, either A finishes service or you finish service:

$$P(B \text{ in system}) = P(B \text{ in system}|A \text{ finishes before you})\frac{\mu_2}{\mu_1+\mu_2} + P(B \text{ in system}|\text{you finish before } A)\frac{\mu_1}{\mu_1+\mu_2}$$

Now $P(B \text{ in system}|\text{you finish before } A) = 1$ since B will still be waiting in line when you move to server 2. On the other hand, if the first thing to happen is that A finishes service, then at that point, by memorylessness, your remaining service at server 1 is Exponential(μ_1), and B will still be in the system if your remaining service at server 1 is less than B's service at server 2, and the probability of this is $\mu_1/(\mu_1+\mu_2)$. That is,

$$P(B \text{ in system}|A \text{ finishes before you}) = \frac{\mu_1}{\mu_1+\mu_2}.$$

Therefore,

$$P(B \text{ in system}) = \frac{\mu_1\mu_2}{(\mu_1+\mu_2)^2} + \frac{\mu_1}{\mu_1+\mu_2}.$$

(c) To compute the expected time you are in the system, we first divide up your time in the system into

$$T = T_1 + R,$$

where T_1 is the time until the first thing that happens, and R is the rest of the time. The time until the first thing happens is Exponential($\mu_1+\mu_2$), so that

$$E[T_1] = \frac{1}{\mu_1+\mu_2}.$$

To compute $E[R]$, we condition on what was the first thing to happen, either A finished service at server 2 or you finished service

at server 1. If the first thing to happen was that you finished service at server 1, which occurs with probability $\mu_1/(\mu_1+\mu_2)$, then at that point you moved to server 2, and your remaining time in the system is the remaining time of A at server 2, the service time of B at server 2, and your service time at server 2. A's remaining time at server 2 is again Exponential(μ_2) by memorylessness, and so your expected remaining time in service will be $3/\mu_2$. That is,

$$E[R|\text{first thing to happen is you finish service at server 1}] = \frac{3}{\mu_2},$$

and so

$$E[R] = \frac{3}{\mu_2}\frac{\mu_1}{\mu_1+\mu_2} + E[R|\text{first thing is } A \text{ finishes}]\frac{\mu_2}{\mu_1+\mu_2}.$$

Now if the first thing to happen is that A finishes service at server 2, we can again compute your expected remaining time in the system as the expected time until the next thing to happen (either you or B finishes service) plus the expected remaining time after that. To compute the latter we can again condition on what was that next thing to happen. We will obtain

$$\begin{aligned} E[R|\text{first thing is } A \text{ finishes}] &= \frac{1}{\mu_1+\mu_2} + \frac{2}{\mu_2}\frac{\mu_1}{\mu_1+\mu_2} \\ &\quad + \left(\frac{1}{\mu_1}+\frac{1}{\mu_2}\right)\frac{\mu_2}{\mu_1+\mu_2} \end{aligned}$$

Plugging everything back gives $E[T]$. □

As an exercise you should consider how you might do the preceding problem assuming a different service time distribution, such as a Uniform distribution on $[0,1]$ or a deterministic service time such as 1 time unit.

22

The Poisson Process: Introduction

We now begin studying our first continuous-time process – the Poisson Process. Its relative simplicity and significant practical usefulness make it a good introduction to more general continuous time processes. Today we will look at several equivalent definitions of the Poisson Process that, each in their own way, give some insight into the structure and properties of the Poisson process.

Stationary and Independent Increments:

We first define the notions of *stationary increments* and *independent increments*. For a continuous-time stochastic process $\{X(t) : t \geq 0\}$, an *increment* is the difference in the process at two times, say s and t. For $s < t$, the increment from time s to time t is the difference $X(t) - X(s)$.

A process is said to have *stationary* increments if the distribution of the increment $X(t) - X(s)$ depends on s and t only through the difference $t - s$, for all $s < t$. So the distribution of $X(t_1) - X(s_1)$ is the same as the distribution of $X(t_2) - X(s_2)$ if $t_1 - s_1 = t_2 - s_2$. Note that the intervals $[s_1, t_1]$ and $[s_2, t_2]$ may overlap.

A process is said to have *independent* increments if any two increments involving disjoint intervals are independent. That is, if $s_1 < t_1 < s_2 < t_2$, then the two increments $X(t_1) - X(s_1)$ and $X(t_2) - X(s_2)$ are independent.

Not many processes we will encounter will have both stationary and independent increments. In general they will have neither stationary increments nor independent increments. An exception to this we have already seen is the simple random walk. If $\xi_1, \xi_2, \ldots$ is a sequence of independent and identically distributed random variables with $P(\xi_i = 1) = p$ and $P(\xi_i = -1) = q = 1 - p$, the the simple random walk $\{X_n : n \geq 0\}$ starting at 0 can be defined as $X_0 = 0$ and

$$X_n = \sum_{i=1}^{n} \xi_i.$$

From this representation it is not difficult to see that the simple random walk has stationary and independent increments.

Definition 1 of a Poisson Process:

A continuous-time stochastic process $\{N(t) : t \geq 0\}$ is a Poisson process with rate $\lambda > 0$ if

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

(ii) It has stationary and independent increments.

(iii) The distribution of $N(t)$ is Poisson with mean λt, i.e.,

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \text{for } k = 0, 1, 2, \ldots.$$

This definition tells us some of the structure of a Poisson process immediately:

- By stationary increments the distribution of $N(t) - N(s)$, for $s < t$ is the same as the distribution of $N(t-s) - N(0) = N(t-s)$, which is a Poisson distribution with mean $\lambda(t-s)$.
- The process is *nondecreasing*, for $N(t) - N(s) \geq 0$ with probability 1 for any $s < t$ since $N(t) - N(s)$ has a Poisson distribution.
- The state space of the process is clearly $S = \{0, 1, 2, \ldots\}$.

We can think of the Poisson process as counting events as it progresses: $N(t)$ is the number of events that have occurred up to time t and at time $t+s$, $N(t+s) - N(t)$ more events will have been counted, with $N(t+s) - N(t)$ being Poisson distributed with mean λs.

For this reason the Poisson process is called a *counting* process. Counting processes are a more general class of processes of which the Poisson process is a special case. One common modeling use of the Poisson process is to interpret $N(t)$ as the number of arrivals of tasks/jobs/customers to a system by time t.

Note that $N(t) \to \infty$ as $t \to \infty$, so that $N(t)$ itself is by no means stationary, even though it has stationary increments. Also note that, in the customer arrival interpetation, as λ increases customers will tend to arrive faster, giving one justification for calling λ the rate of the process.

We can see where this definition comes from, and in the process try to see some more low level structure in a Poisson process, by considering a discrete-time analogue of the Poisson process, called a *Bernoulli* process, described as follows.

The Bernoulli Process: A Discrete-Time "Poisson Process":

Suppose we divide up the positive half-line $[0, \infty)$ into disjoint intervals, each of length h, where h is small. Thus we have the intervals $[0, h)$, $[h, 2h)$, $[2h, 3h)$, and so on. Suppose further that each interval corresponds to an independent Bernoulli trial, such that in each interval, independently of every other interval, there is a successful event (such as an arrival) with probability λh. Define the Bernoulli process to be $\{B(t) : t = 0, h, 2h, 3h, \ldots\}$, where $B(t)$ is the number of successful trials up to time t.

The above definition of the Bernoulli process clearly corresponds to the notion of a process in which events occur randomly in time, with an intensity, or rate, that increases as λ increases, so we can think of the Poisson process in this way too, assuming the Bernoulli process is a close approximation to the Poisson process. The way we have defined it, the Bernoulli process $\{B(t)\}$ clearly has stationary and independent increments. As well, $B(0) = 0$. Thus the Bernoulli process is a discrete-time approximation to the Poisson process with rate λ if the distribution of $B(t)$ is approximately Poisson(λt).

For a given t of the form nh, we know the exact distribution of $B(t)$. Up to time t there are n independent trials, each with probability λh of success, so $B(t)$ has a Binomial distribution with parameters n and λh. Therefore, the mean number of successes up to time t is $n\lambda h = \lambda t$. So $E[B(t)]$ is correct. The fact that the distribution of $B(t)$ is approximately Poisson(λt) follows from the Poisson approximation to the Binomial distribution (p.32 of the text), which we can re-derive here. We have, for k a nonnegative integer and $t > 0$, (and keeping in mind that $t = nh$ for some positive integer n),

$$\begin{aligned} P(B(t) = k) &= \binom{n}{k}(\lambda h)^k(1-\lambda h)^{n-k} \\ &= \frac{n!}{(n-k)!k!}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k} \\ &= \frac{n!}{(n-k)!n^k}\left(1-\frac{\lambda t}{n}\right)^{-k}\frac{(\lambda t)^k}{k!}\left(1-\frac{\lambda t}{n}\right)^{n} \\ &\approx \frac{n!}{(n-k)!n^k}\left(1-\frac{\lambda t}{n}\right)^{-k}\frac{(\lambda t)^k}{k!}e^{-\lambda t}, \end{aligned}$$

for n very large (or h very small). But also, for n large

$$\left(1-\frac{\lambda t}{n}\right)^{-k} \approx 1$$

and

$$\frac{n!}{(n-k)!n^k} = \frac{n(n-1)\ldots(n-k+1)}{n^k} \approx 1.$$

Therefore, $P(B(t) = k) \approx (\lambda t)^k/k!e^{-\lambda t}$ (this approximation gets exact as $h \to 0$).

Thinking intuitively about how the Poisson process can be expected to behave can be done by thinking about the conceptually simpler Bernoulli process. For example, given that there are n events in the interval $[0, t)$ (i.e. $N(t) = n$), the times of those n events should be uniformly distributed in the interval $[0, t)$ because that is what we would expect in the Bernoulli process. This intuition is true, and we'll prove it more carefully later.

Thinking in terms of the Bernoulli process also leads to a more low-level (in some sense better) way to define the Poisson process. This way of thinking about the Poisson process will also be useful later when we consider continuous-time Markov chains. In the Bernoulli process the probability of a success in any given interval is λh and the probability of two or more successes is 0 (that is, $P(B(h) = 1) = \lambda h$ and $P(B(h) \geq 2) = 0$). Therefore, in the Poisson process we have the approximation that $P(N(h) = 1) \approx \lambda h$ and $P(N(h) \geq 2) \approx 0$.

We write this approximation in a more precise way by saying that

$$P(N(n) = 1) = \lambda h + o(h) \text{ and } P(N(h) \geq 2) = o(h).$$

The notation "$o(h)$" is called Landau's $o(h)$ notation, read "little o of h", and it means any function of h that is of *smaller order* than h. This means that if $f(h)$ is $o(h)$ then $f(h)/h \to 0$ as $h \to 0$ ($f(h)$ goes to 0 faster that h goes to 0). Notationally, $o(h)$ is a very clever and useful quantity because it lets us avoid writing out long, complicated, or simply unknown expressions when the only crucial property of the expression that we care about is how fast it goes to 0. We will make extensive use of this notation in this and the next chapter, so it is worthwhile to pause and make sure you understand the properties of $o(h)$.

Landau's "Little o of h" Notation:

Note that $o(h)$ doesn't refer to any specific function. It denotes any quantity that goes to 0 at a faster rate than h, as $h \to 0$:

$$\frac{o(h)}{h} \to 0 \text{ as } h \to 0.$$

Since the sum of two such quantities retains this rate property, we get the potentially disconcerting property that

$$o(h) + o(h) = o(h)$$

as well as

$$\begin{aligned} o(h)o(h) &= o(h) \\ c \times o(h) &= o(h), \end{aligned}$$

where c is any constant (note that c can be a function of other variables as long as it remains constant as h varies).

Example: The function h^k is $o(h)$ for any $k > 1$ since

$$\frac{h^k}{h} = h^{k-1} \to 0 \text{ as } h \to 0.$$

h however is not $o(h)$. The infinite series $\sum_{k=2}^{\infty} c_k h^k$, where $|c_k| < 1$, is $o(h)$ since

$$\begin{aligned} \lim_{h\to 0} \frac{\sum_{k=2}^{\infty} c_k h^k}{h} &= \lim_{h\to 0} \sum_{k=2}^{\infty} c_k h^{k-1} \\ &= \sum_{k=2}^{\infty} c_k \lim_{h\to 0} h^{k-1} = 0, \end{aligned}$$

where taking the limit inside the summation is justified because the sum is bounded by $1/(1-h)$ for $h < 1$. □

Definition 2 of a Poisson Process:

A continuous-time stochastic process $\{N(t) : t \geq 0\}$ is a Poisson process with rate $\lambda > 0$ if

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

(ii) It has stationary and independent increments.

(iii) $P(N(h) = 1) = \lambda h + o(h)$,
$P(N(h) \geq 2) = o(h)$, and
$P(N(h) = 0) = 1 - \lambda h + o(h)$.

This definition can be more useful than Definition 1 because its conditions are more "primitive" and correspond more directly with the Bernoulli process, which is more intuitive to imagine as a process evolving in time.

We need to check that Definitions 1 and 2 are equivalent (that is, they define the same process). We will show that Definition 1 implies Definition 2. The proof that Definition 2 implies Definition 1 is shown in the text in Theorem 5.1 on p.292 (p.260 in the 7th Edition), which you are required to read.

Proof that Definition 1 $\Rightarrow$ Definition 2: (Problem #35, p.335)

We just need to show part(iii) of Definition 2. By Definition 1, $N(h)$ has a Poisson distribution with mean λh. Therefore,

$$P(N(h) = 0) = e^{-\lambda h}.$$

If we expand out the exponential in a Taylor series, we have that

$$\begin{aligned} P(N(h) = 0) &= 1 - \lambda h + \frac{(\lambda h)^2}{2!} - \frac{(\lambda h)^3}{3!} + \dots \\ &= 1 - \lambda h + o(h). \end{aligned}$$

Similarly,

$$\begin{aligned} P(N(h)=1) &= \lambda h e^{-\lambda h} \\ &= \lambda h \left[1 - \lambda h + \frac{(\lambda h)^2}{2!} - \frac{(\lambda h)^3}{3!} + \ldots\right] \\ &= \lambda h - \lambda^2 h^2 + \frac{(\lambda h)^3}{2!} - \frac{(\lambda h)^4}{3!} + \ldots \\ &= \lambda h + o(h). \end{aligned}$$

Finally,

$$\begin{aligned} P(N(h) \geq 2) &= 1 - P(N(h)=1) - P(N(h)=0) \\ &= 1 - (\lambda h + o(h)) - (1 - \lambda h + o(h)) \\ &= -o(h) - o(h) = o(h). \end{aligned}$$

Thus Definition 1 implies Definition 2. □

A third way to define the Poisson process is to define the distribution of the time between events. We will see in the next lecture that the times between events are independent and identically distributed Exponential(λ) random variables. For now we can gain some insight into this fact by once again considering the Bernoulli process.

Imagine that you start observing the Bernoulli process at some arbitrary trial, such that you don't know how many trials have gone before and you don't know when the last successful trial was. Still you would know that the distribution of the time until the next successful trial was h times a Geometric random variable with parameter λh. In other words, you don't need to know anything about the past of the process to know the distribution of the time to the next success, and in fact this is the same as the distribution until the first success. That is, the distribution of the time between successes in the Bernoulli process is *memoryless*.

When you pass to the limit as $h \to 0$ you get the Poisson process with rate λ, and you should expect that you will retain this memoryless property in the limit. Indeed you do, and since the only continuous distribution on $[0, \infty)$ with the memoryless property is the Exponential distribution, you may deduce that this is the distribution of the time between events in a Poisson process. Moreover, you should also inherit from the Bernoulli process that the times between successive events are independent and identically distributed.

As a final aside, we remark that this discussion also suggests that the Exponential distribution is a limiting form of the Geometric distribution, as the probability of success λh in each trial goes to 0. This is indeed the case. As we mentioned above, the time between successful trials in the Bernoulli process is distributed as $Y = hX$, where X is a Geometric random variable with parameter λh. One can verify that for any $t > 0$, we have $P(Y > t) \to e^{-\lambda t}$ as $h \to 0$:

$$\begin{aligned} P(Y > t) &= P(hX > t) \\ &= P(X > t/h) \\ &= (1 - \lambda h)^{\lceil t/h \rceil} \\ &= (1 - \lambda h)^{t/h} (1 - \lambda h)^{\lceil t/h \rceil - t/h} \\ &= \left(1 - \frac{\lambda t}{t/h}\right)^{t/h} (1 - \lambda h)^{\lceil t/h \rceil - t/h} \\ &\to e^{-\lambda t} \quad \text{as } h \to 0, \end{aligned}$$

where $\lceil t/h \rceil$ is the smallest integer greater than or equal to t/h. In other words, the distribution of Y converges to the Exponential(λ) distribution as $h \to 0$.

Note that the above discussion also illustrates that the Geometric distribution is a discrete distribution with the memoryless property.

23

Properties of the Poisson Process

Today we will consider the distribution of the times between events in a Poisson process, called the *interarrival times* of the process. We will see that the interarrival times are independent and identically distributed Exponential(λ) random variables, where λ is the rate of the Poisson process. This leads to our third definition of the Poisson process.

Using this definition, as well as our previous definitions, we can deduce some further properties of the Poisson process. Today we will see that the time until the nth event occurs has a Gamma(n,λ) distribution. Later we will consider the sum, called the *superposition*, of two independent Poisson processes, as well as the *thinned* Poisson process obtained by independently marking, with some fixed probability p, each event in a Poisson process, thereby identifying the events in the thinned process.

Interarrival Times of the Poisson Process:

We can think of the Poisson process as a counting process with a given interarrival distribution That is, $N(t)$ is the number of events that have occurred up to time t, where the times between events, called the *interarrival* times, are independent and identically distributed random variables.

Comment: We will see that the interarrival distribution for a Poisson process with rate λ is Exponential(λ), which is expected based on the discussion at the end of the last lecture. In general, we can replace the Exponential interarrival time distribution with any distribution on $[0, \infty)$, to obtain a large class of counting processes. Such processes (when the interarrival time distribution is general) are called *Renewal Processes*, and the area of their study is called *Renewal Theory*. We will not study this topic in this course, but for those interested this topic is covered in Chapter 7 of the text. However, we make the comment here that if the interarrival time is not Exponential, then the process will not have stationary and independent increments. That is, the Poisson process is the only Renewal process with stationary and independent increments.

Proof that the Interarrival Distribution is Exponential(λ):

We can prove that the interarrival time distribution in the Poisson process is Exponential directly from Definition 1. First, consider the time until the first event, say T_1. Then for any $t > 0$, the event $\{T_1 > t\}$ is equivalent to the event $\{N(t) = 0\}$. Therefore,

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}.$$

This shows immediately that T_1 has an Exponential distribution with rate λ.

In general let T_i denote the time between the $(i-1)$st and the ith event. We can use an induction argument in which the nth proposition is that $T_1, \ldots, T_n$ are independent and identically distributed Exponential(λ) random variables:

$$\text{Proposition } n : T_1, \ldots, T_n \text{ are i.i.d. Exponential}(\lambda).$$

We have shown that Proposition 1 is true. Now assume that Proposition n is true (the induction hypothesis). Then we show this implies Proposition $n+1$ is true. To do this fix $t, t_1, \ldots, t_n > 0$. Proposition $n+1$ will be true if we show that the distribution of T_{n+1} conditioned on $T_1 = t_1, \ldots, T_n = t_n$ does not depend on $t_1, \ldots, t_n$ (which shows that T_{n+1} is independent of $T_1, \ldots, T_n$), and $P(T_n > t) = e^{-\lambda t}$. So we wish to consider the conditional probability

$$P(T_{n+1} > t | T_n = t_n, \ldots, T_1 = t_1).$$

First, we will re-express the event $\{T_n = t_n, \ldots, T_1 = t_1\}$ which involves the first n interarrival times into an equivalent event which involves the first n *arrival* times. Let $S_k = T_1 + \ldots + T_k$ be the kth

arrival time (the time of the kth event) and let $s_k = t_1 + \ldots + t_k$, for $k = 1, \ldots, n$. Then

$$\{T_n = t_n, \ldots, T_1 = t_1\} = \{S_n = s_n, \ldots, S_1 = s_1\},$$

and we can rewrite our conditional probability as

$$P(T_{n+1} > t | T_n = t_n, \ldots, T_1 = t_1) = P(T_{n+1} > t | S_n = s_n, \ldots, S_1 = s_1)$$

The fact that the event $\{T_{n+1} > t\}$ is independent of the event $\{S_n = s_n, \ldots, S_1 = s_1\}$ is because of independent increments, though it may not be immediately obvious. We'll try to see this in some detail.

If the event $\{S_n = s_n, \ldots, S_1 = s_1\}$ occurs then the event $\{T_{n+1} > t\}$ occurs if and only if there are no arrivals in the interval $(s_n, s_n + t]$, so we can write

$$\begin{aligned} &P(T_{n+1} > t | S_n = s_n, \ldots, S_1 = s_1) \\ &\quad = \; P(N(s_n + t) - N(s_n) = 0 | S_n = s_n, \ldots, S_1 = s_1). \end{aligned}$$

Therefore, we wish to express the event $\{S_n = s_n, \ldots, S_1 = s_1\}$ in terms of increments disjoint from the increment $N(s_n + t) - N(s_n)$. At the cost of some messy notation we'll do this, just to see how it might be done at least once. Define the increments

$$\begin{aligned} I_1^{(k)} &= N(s_1 - 1/k) - N(0) \\ I_i^{(k)} &= N(s_i - 1/k) - N(s_{i-1} + 1/k) \quad \text{for } i = 2, \ldots, n, \end{aligned}$$

for $k > M$, where M is chosen so that $1/k$ is smaller than the smallest interarrival time, and also define the increments

$$\begin{aligned} B_i^{(k)} &= N(s_i + 1/k) - N(s_i - 1/k) \quad \text{for } i = 1, \ldots, n-1 \\ B_n^{(k)} &= N(s_n) - N(s_n - 1/k), \end{aligned}$$

for $k > M$. The increments $I_1^{(k)}, B_1^{(k)}, \ldots, I_n^{(k)}, B_n^{(k)}$ are all disjoint and account for the entire interval $[0, s_n]$. Now define the event

$$A_k = \{I_1 = 0\} \bigcap \ldots \bigcap \{I_n = 0\} \bigcap \{B_1 = 1\} \bigcap \ldots \bigcap \{B_n = 1\}.$$

Then A_k implies A_{k-1} (that is, A_k is contained in A_{k-1}) so that the sequence $\{A_k\}_{k=M}^{\infty}$ is a decreasing sequence of sets, and in fact

$$\{S_n = s_n, \ldots, S_1 = s_1\} = \bigcap_{k=M}^{\infty} A_k,$$

because one can check that each event implies the other.

However (and this is why we constructed the events A_k), for any k the event A_k is independent of the event $\{N(s_n+t)-N(s_n) = 0\}$ because the increment $N(s_n + t) - N(s_n)$ is independent of all the increments $I_1^{(k)}, \ldots, I_n^{(k)}, B_1^{(k)}, \ldots, B_n^{(k)}$, as they are all disjoint increments. But if the event $\{N(s_n + t) - N(s_n) = 0\}$ is independent of A_k for every k, it is independent of the intersection of the A_k. Thus, we have

$$\begin{aligned}
&P(T_{n+1} > t | S_n = s_n, \ldots, S_1 = s_1) \\
&= P(N(s_n + t) - N(s_n) = 0 | S_n = s_n, \ldots, S_1 = s_1) \\
&= P\Big(N(s_n + t) - N(s_n) = 0 \;\Big|\; \bigcap_{k=M}^{\infty} A_k\Big) \\
&= P((N(s_n + t) - N(s_n) = 0) \\
&= P(N(t) = 0) = e^{-\lambda t},
\end{aligned}$$

and we have shown that T_{n+1} has an Exponential(λ) distribution and is independent of $T_1, \ldots, T_n$. We conclude from the induction argument that the sequence of interarrival times $T_1, T_2, \ldots$ are all independent and identically distributed Exponential(λ) random variables. $\square$

Definition 3 of a Poisson Process:

A continuous-time stochastic process $\{N(t) : t \geq 0\}$ is a Poisson process with rate $\lambda > 0$ if

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

(ii) $N(t)$ counts the number of events that have occurred up to time t (i.e. it is a counting process).

(iii) The times between events are independent and identically distributed with an Exponential(λ) distribution.

We have seen how Definition 1 implies (i), (ii) and (iii) in Definition 3. One can show that Exponential(λ) interarrival times implies part(iii) of Definition 2 by expanding out the exponential function as a Taylor series, much as we did in showing that Definition 1 implies Definition 2. One can also show that Exponential interarrival times implies stationary and independent increments by using the memoryless property. As an exercise, you may wish to prove this. However, we will not do so here. That Definition 3 actually is a definition of the Poisson process is nice, but not necessary. It suffices to take either Definition 1 or Definition 2 as the definition of the Poisson process, and to see that either definition implies that the times between events are i.i.d. Exponential(λ) random variables.

Distribution of the Time to the nth Arrival:

If we let S_n denote the time of the nth arrival in a Poisson process, then $S_n = T_1 + \ldots + T_n$, the sum of the first n interarrival times. The distribution of S_n is *Gamma* with parameters n and λ. Before showing this, let us briefly review the Gamma distribution.

The Gamma(α, λ) Distribution:

A random variable X on $[0, \infty)$ is said to have a Gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$ if its probability density function is given by

$$f_X(x|\alpha, \lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases},$$

where $\Gamma(\alpha)$, called the *Gamma* function, is defined by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy.$$

We can verify that the density of the Gamma(α,λ) distribution integrates to 1, by writing down the integral

$$\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

and making the substitution $y = \lambda x$. This gives $dy = \lambda dx$ or $dx = (1/\lambda)dy$, and $x = y/\lambda$, and so

$$\begin{aligned} \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx &= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{y}{\lambda}\right)^{\alpha-1} e^{-\lambda y/\lambda} \frac{1}{\lambda} dy \\ &= \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} dy = 1, \end{aligned}$$

by looking again at the definition of $\Gamma(\alpha)$.

The $\Gamma(\alpha)$ function has a useful recursive property. For $\alpha > 1$ we can start to evaluate the integral defining the Gamma function using integration by parts:

$$\int_a^b u dv = uv\Big|_a^b - \int_a^b v du.$$

We let

$$u = y^{\alpha-1} \quad \text{and} \quad dv = e^{-y}dy,$$

giving

$$du = (\alpha-1)y^{\alpha-2}dy \quad \text{and} \quad v = -e^{-y},$$

so that

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty y^{\alpha-1}e^{-y}dy \\ &= -y^{\alpha-1}e^{-y}\Big|_0^\infty + (\alpha-1)\int_0^\infty y^{\alpha-2}e^{-y}dy \\ &= 0 + (\alpha-1)\Gamma(\alpha-1).\end{aligned}$$

That is, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. In particular, if $\alpha = n$, a positive integer greater than or equal to 2, then we recursively get

$$\begin{aligned}\Gamma(n) = (n-1)\Gamma(n-1) = \ldots &= (n-1)(n-2)\ldots(2)(1)\Gamma(1) \\ &= (n-1)!\Gamma(1).\end{aligned}$$

However,

$$\Gamma(1) = \int_0^\infty e^{-y}dy = -e^{-y}\Big|_0^\infty = 1,$$

is just the area under the curve of the Exponential(1) density. Therefore, $\Gamma(n) = (n-1)!$ for $n \geq 2$. However, since $\Gamma(1) = 1 = 0!$ we have in fact that $\Gamma(n) = (n-1)!$ for any positive integer n.

So for n a positive integer, the Gamma(n,λ) density can be written as

$$f_X(x|n,\lambda) = \begin{cases} \frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}.$$

An important special case of the Gamma(α, λ) distribution is the Exponential(λ) distribution, which is obtained by setting $\alpha = 1$. Getting back to the Poisson process, we are trying to show that the sum of n independent Exponential(λ) random variables has a Gamma(n,λ) distribution, and for $n = 1$, the result is immediate. For $n > 1$, the simplest way to get our result is to observe that the time of the nth arrival is less than or equal to t if and only if the number of arrivals in the interval $[0, t]$ is greater than or equal to n. That is, the two events

$$\{S_n \leq t\} \quad \text{and} \quad \{N(t) \geq n\}$$

are equivalent. However, the probability of the first event $\{S_n \leq t\}$ gives the CDF of S_n, and so we have a means to calculate the CDF:

$$F_{S_n}(t) \equiv P(S_n \leq t) = P(N(t) \geq n) = \sum_{j=n}^{\infty} \frac{(\lambda t)^n}{n!}e^{-\lambda t}.$$

To get the density of S_n, we differentiate the above with respect to t, giving

$$\begin{aligned} f_{S_n}(t) &= -\sum_{j=n}^{\infty} \lambda \frac{(\lambda t)^j}{j!}e^{-\lambda t} + \sum_{j=n}^{\infty} \lambda \frac{(\lambda t)^{j-1}}{(j-1)!}e^{-\lambda t} \\ &= \lambda \frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t} = \frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}. \end{aligned}$$

Comparing with the Gamma(n,λ) density above we have our result.

24

Further Properties of the Poisson Process

Today we will consider two further properties of the Poisson process that both have to do with deriving new processes from a given Poisson process. Specifically, we will see that

(1) The sum of two independent Poisson processes (called the *superposition* of the processes), is again a Poisson process but with rate $\lambda_1 + \lambda_2$, where λ_1 and λ_2 are the rates of the constituent Poisson processes.

(2) If each event in a Poisson process is *marked* with probability p, independently from event to event, then the marked process $\{N_1(t) : t \geq 0\}$, where $N_1(t)$ is the number of marked events up to time t, is a Poisson process with rate λp, where λ is the rate of the original Poisson process. This is called *thinning* a Poisson process.

The operations of taking the sum of two or more independent Poisson processes and of thinning a Poisson process can be of great practical use in modeling many systems where the Poisson process(es) represent arrival streams to the system and we wish to classify different types of arrivals because the system will treat each arrival differently based on its type.

Superposition of Poisson Processes:

Suppose that $\{N_1(t) : t \geq 0\}$ and $\{N_2(t) : t \geq 0\}$ are two independent Poisson processes with rates λ_1 and λ_2, respectively. The sum of $N_1(t)$ and $N_2(t)$,

$$\{N(t) = N_1(t) + N_2(t) : t \geq 0\},$$

is called the *superposition* of the two processes $N_1(t)$ and $N_2(t)$. Since $N_1(t)$ and $N_2(t)$ are independent and $N_1(t)$ is Poisson($\lambda_1 t$) and $N_2(t)$ is Poisson($\lambda_2 t$), their sum has a Poisson distribution with mean $(\lambda_1 + \lambda_2)t$. Also, it is clear that $N(0) = N_1(0) + N_2(0) = 0$. That is, properties (i) and (iii) of Definition 1 of a Poisson process are satisfied by the process $N(t)$ if we take the rate to be $\lambda_1 + \lambda_2$. Thus, to show that $N(t)$ is indeed a Poisson process with rate $\lambda_1 + \lambda_2$ it just remains to show that $N(t)$ has stationary and independent increments.

First, consider any increment $I(t_1, t_2) = N(t_2) - N(t_1)$, with $t_1 < t_2$. Then

$$\begin{aligned} I(t_1, t_2) &= N(t_2) - N(t_1) \\ &= N_1(t_2) + N_2(t_2) - (N_1(t_1) + N_2(t_1)) \\ &= (N_1(t_2) - N_1(t_1)) + (N_2(t_2) - N_2(t_1)) \\ &\equiv I_1(t_1, t_2) + I_2(t_1, t_2), \end{aligned}$$

where $I_1(t_1, t_2)$ and $I_2(t_1, t_2)$ are the corresponding increments in the $N_1(t)$ and $N_2(t)$ processes, respectively. But the increment $I_1(t_1, t_2)$ has a Poisson($\lambda_1(t_2 - t_1)$) distribution and the increment $I_2(t_1, t_2)$ has a Poisson($\lambda_2(t_2 - t_1)$) distribution. Furthermore, $I_1(t_1, t_2)$ and $I_2(t_1, t_2)$ are independent. Therefore, as before, their sum has a Poisson distribution with mean $(\lambda_1 + \lambda_2)(t_2 - t_1)$. That is, the distribution of the increment $I(t_1, t_2)$ depends on t_1 and t_2 only through the difference $t_2 - t_1$, which says that $N(t)$ has stationary increments.

Second, for $t_1 < t_2$ and $t_3 < t_4$, let $I(t_1, t_2) = N(t_2) - N(t_1)$ and $I(t_3, t_4) = N(t_4) - N(t_3)$ be any two disjoint increments (i.e. the intervals $(t_1, t_2]$ and $(t_3, t_4]$ are disjoint). Then

$$I(t_1, t_2) = I_1(t_1, t_2) + I_2(t_1, t_2)$$

and

$$I(t_3, t_4) = I_1(t_3, t_4) + I_2(t_3, t_4).$$

But $I_1(t_1, t_2)$ is independent of $I_1(t_3, t_4)$ because the $N_1(t)$ process has independent increments, and $I_1(t_1, t_2)$ is independent of $I_2(t_3, t_4)$ because the processes $N_1(t)$ and $N_2(t)$ are independent. Similarly, we can see that $I_2(t_1, t_2)$ is independent of both $I_1(t_3, t_4)$ and $I_2(t_3, t_4)$. From this it is clear that the increment $I(t_1, t_2)$ is independent of the increment $I(t_3, t_4)$. Therefore, the process $N(t)$ also has independent increments.

Thus, we have shown that the process $\{N(t) : t \geq 0\}$ satisfies the conditions in Definition 1 for it to be a Poisson process with rate $\lambda_1 + \lambda_2$.

Remark 1: By repeated application of the above arguments we can see that the superposition of k independent Poisson processes with rates $\lambda_1, \ldots, \lambda_k$ is again a Poisson process with rate $\lambda_1 + \ldots + \lambda_k$.

Remark 2: There is a useful result in probability theory which says that if we take N independent counting processes and sum them up, then the resulting superposition process is approximately a Poisson process. Here N must be "large enough" and the rates of the individual processes must be "small" relative to N (this can be made mathematically precise, but here in this remark our interest is in just

the practical implications of the result), but the individual processes that go into the superposition can otherwise be *arbitrary*.

This can sometimes be used as a justification for using a Poisson process model. For example, in the classical voice telephone system, each individual produces a stream of connection requests to a given telephone exchange, perhaps in a way that does not look at all like a Poisson process. But the stream of requests coming from any given individual typically makes up a very small part of the total aggregate stream of connection requests to the exchange. It is also reasonable that individuals make telephone calls largely independently of one another. Such arguments provide a theoretical justification for modeling the aggregate stream of connection requests to a telephone exchange as a Poisson process. Indeed, empirical observation also supports such a model.

In contrast to this, researchers in recent years have found that arrivals of *packets* to gateway computers in the internet can exhibit some behaviour that is not very well modeled by a Poisson process. The packet traffic exhibits large "spikes", called *bursts*, that do not suggest that they are arriving uniformly in time. Even though many users may make up the aggregate packet traffic to a gateway or router, the number of such users is likely still not as many as the number of users that will make requests to a telephone exchange. More importantly, the aggregate traffic at an internet gateway tends to be dominated by just a few individual users at any given time. The connection to our remark here is that as the bandwidth in the internet increases and the number of users grows, a Poisson process model should theoretically become more and more reasonable.

Thinning a Poisson Process:

Let $\{N(t) : t \geq 0\}$ be a Poisson process with rate λ. Suppose we mark each event with probability p, independently from event to event, and let $\{N_1(t) : t \geq 0\}$ be the process which counts the marked events. We can use Definition 2 of a Poisson process to show that the *thinned* process $N_1(t)$ is a Poisson process with rate λp. To see this, first note that $N_1(0) = N(0) = 0$. Next, the probability that there is one marked event in the interval $[0, h]$ is

$$\begin{aligned} P(N_1(h) = 1) &= P(N(h) = 1)p + \sum_{k=2}^{\infty} P(N(h) = k)\binom{k}{1}p(1-p)^{k-1} \\ &= (\lambda h + o(h))p + \sum_{k=2}^{\infty} o(h)kp(1-p)^{k-1} \\ &= \lambda ph + o(h). \end{aligned}$$

Similarly,

$$\begin{aligned} P(N_1(h) = 0) &= P(N(h) = 0) + P(N(h) = 1)(1-p) \\ &\quad + \sum_{k=2}^{\infty} P(N(h) = k)(1-p)^k \\ &= 1 - \lambda h + o(h) + (\lambda h + o(h))(1-p) \\ &\quad + \sum_{k=2}^{\infty} o(h)(1-p)^k \\ &= 1 - \lambda ph + o(h). \end{aligned}$$

Finally, $P(N_1(h) \geq 2)$ can be obtained by subtraction:

$$\begin{aligned} P(N_1(h) \geq 2) &= 1 - P(N_1(h) = 0) - P(N_1(h) = 1) \\ &= 1 - (1 - \lambda ph + o(h)) - (\lambda ph + o(h)) = o(h). \end{aligned}$$

We can show that the increments in the thinned process are stationary by computing $P(I_1(t_1,t_2)=k)$, where $I_1(t_1,t_2) \equiv N_1(t_2)-N_1(t_1)$ is the increment from t_1 to t_2 in the thinned process, by conditioning on the increment $I(t_1,t_2) \equiv N(t_2)-N(t_1)$ in the original process:

$$\begin{aligned}
P(I_1(t_1,t_2)=k) &= \sum_{n=0}^{\infty} P(I_1(t_1,t_2)=k|I(t_1,t_2)=n)P(I(t_1,t_2)=n) \\
&= \sum_{n=k}^{\infty} P(I_1(t_1,t_2)=k|I(t_1,t_2)=n)P(I(t_1,t_2)=n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{[\lambda(t_2-t_1)]^n}{n!} e^{-\lambda(t_2-t_1)} \\
&= \frac{[\lambda p(t_2-t_1)]^k}{k!} e^{-\lambda p(t_2-t_1)} \\
&\quad \times \sum_{n=k}^{\infty} \frac{[\lambda(1-p)(t_2-t_1)]^{n-k}}{(n-k)!} e^{-\lambda(1-p)(t_2-t_1)} \\
&= \frac{[\lambda p(t_2-t_1)]^k}{k!} e^{-\lambda p(t_2-t_1)}.
\end{aligned}$$

This shows that the distribution of the increment $I_1(t_1,t_2)$ depends on t_1 and t_2 only through the difference t_2-t_1, and so the increments are stationary. Finally, the fact that the increments in the thinned process are independent is directly inherited from the independence of the increments in the original Poisson process $N(t)$.

Remark: The process consisting of the *unmarked* events, call it $N_2(t)$, is also a Poisson process, this time with rate $\lambda(1-p)$. The text shows that the two processes $N_1(t)$ and $N_2(t)$ are independent. Please read this section of the text (Sec.5.3.4).

The main practical advantage that the Poisson process model has over other counting process models is the fact that many of its properties are explicitly known. For example, it is in general difficult or impossible to obtain explicitly the distribution of $N(t)$ for any t if $N(t)$ were a counting process other than a Poisson process. The memoryless property of the exponential interarrival times is also extremely convenient when doing calculations that involve the Poisson process.

(Please read the class notes for material on the filtered Poisson process and Proposition 5.3 of the text).