Notes for Lecture 23

klin@math.arizona.edu

April 14, 2020
Revised April 16, 2020

In today's lecture, I continued to discuss the Poisson process. Recall the defining properties of the Poisson process are:

(i) $N(0)=0$;

(ii) $N(t+s)-N(s)\sim \text{Poisson}(\lambda t)$ for all $s,t\geqslant 0$; and

(iii) for all n and $0\leqslant t_0\leqslant \cdots t_n$, the *increments* $N(t_1)-N(t_0),\cdots,N(t_n)-N(t_{n-1})$ are independent random variables.

Last time, we introduced two different ways of thinking about the Poisson process:

- *Model A:* let $\tau_1,\tau_2,\cdots$ be a sequence of independent $\text{Exp}(\lambda)$ random variables, which we view as waiting times (or "gaps") between events. Then $T_n=\tau_1+\cdots+\tau_n$ are the times of actual events, and $N(t)=\max\{n\mid T_n\leqslant t\}$ is a Poisson process.
- *Model B:* fix n large, and subdivide $[0,t]$ into n bins. Put a ball in the ith bin with probability $p_n=\lambda t/n$. Do this independently for each bin. Then the total number N_n of balls is a binomial random variable with distribution $\text{Bin}(n,p_n)$, and N_n converges to a Poisson process "in distribution" as $n\to\infty$.

The phrase "in distribution" means, roughly speaking, that the probability distribution of N_n and functions of N_n converge to those of the Poisson process as $n\to\infty$. For example, for nonnegative integers k, $P(N_n=k)\to e^{-\lambda t}(\lambda t)^k/k!$ as $n\to\infty$.

Note: there is a subtlety in the definition of Model A that is used below, namely that we use "$\leqslant t$" in defining $N(t)$, instead of "$< t$". This implies that if $\tau_1=7$, then $N(t)=0$ for all $t<7$ but $N(7)=1$. It also tells us that for all $0\leqslant s\leqslant t$, $N(t)-N(s)$ counts the number of events in $(s,t]$.

From Model B to Model A. First, as the text pointed out, the Poisson Approximation Theorem (Theorem 2.5) tells us that $P(N_n=k)$ tends to the Poisson distribution as $n\to\infty$. So this suggests that Model B approximates a Poisson process for large n, and that this approximation becomes exact as $n\to\infty$. These things are true, but I won't prove them here.[1]

Now let $\gamma_1,\cdots$ denote the gaps between events, i.e., these are the number of empty bins between filled ones or, if you prefer flipping coins, the lengths of runs of tails. So $\gamma_i\geqslant 0$ for each

[1] If you are curious how this works, let me know and I'll try to find you a good reference.

i. The γ_i are analogous to the waiting timse τ_i in Model A. As explained in class, however, γ_i count time in terms of bins, while τ_i count time as a real number. To relate them, let $\delta_n = t/n$ be the bin width. Then $\gamma_i \delta_n$ measures time in the correct "units," i.e., as a real number.

To show that Model B is equivalent to Model A in the limit $n \to \infty$, we need to show

(a) the random variables $\gamma_i \delta_n$ have the exponential distribution as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} P(\gamma_i \delta_n > s) = e^{-\lambda s} \tag{1}$$

for all $s > 0$.

(b) $\gamma_1 \delta_n, \cdots$ become independent as $n \to \infty$.

Condition (b) is actually false for fixed n: the gaps cannot be independent because for a finite number of bins, a long gap in one place makes a long gap elsewhere less likely (since there are only so many bins). To show (b) means to show that this effect goes away as $n \to \infty$, which I won't do here.

Condition (a) is easier: first, for any integer $k \geqslant 0$,

$$P(\gamma_i \geqslant k) = (1-p)^k. \tag{2}$$

(This is just the probability of flipping at least k tails in a row.) Now, for a fixed $s > 0$, let k_n be the nearest integer to s/δ_n, so that $s \approx k_n \delta_n$, and keep in mind the approximation improves as $n \to \infty$. Then

$$\begin{aligned} P(\gamma_i \delta_n \geqslant s) &\approx P(\gamma_i \delta_n \geqslant k_n \delta_n) && \text{(3a)} \\ &= P(\gamma_i \geqslant k_n) && \text{(3b)} \\ &= (1-p_n)^{k_n} && \text{(3c)} \end{aligned}$$

where $p_n = \lambda t/n$ as before. Observe

$$\begin{aligned} (1-p_n)^{k_n} &\approx \Big(1 - \frac{\lambda t}{n}\Big)^{s/\delta_n} && \text{(4a)} \\ &= \Big(1 - \frac{\lambda t}{n}\Big)^{ns/t}. && \text{(4b)} \end{aligned}$$

By an application of l'Hôpital's rule, this $\to e^{-\lambda s}$ as $n \to \infty$. So the $\gamma_i \delta_n$ approach exponential random variables as $n \to \infty$, as claimed.

Model A has the properties of a Poisson process. The text gives an explanation of Property (ii) using the gamma distribution (see Lemma 2.6). Here is the alternate inductive proof I gave in class, with a little cleaning up. For simplicity, let $s = 0$. The goal is then to show that the $N(t)$ given by Model A has the property

$$P(N(t) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}. \tag{5}$$

For the base case $n = 0$, note that

$$P(N(t) = 0) = P(\tau_1 > t) \tag{6a}$$
$$= e^{-\lambda t}, \tag{6b}$$

which is exactly Eq. (5) with $n = 0$. For the inductive step, assume $n > 0$ and that Eq. (5) holds up to $n - 1$. Then using the continuous law of alternatives to condition on τ_1, we get

$$P(N(t) = n) = \int_0^t P(N(t) = n \mid \tau_1 = s)\, f_{\tau_1}(s)\, ds \qquad \text{(Law of Alt.)} \tag{7a}$$
$$= \int_0^t P(N(t) = n \mid \tau_1 = s)\, \lambda\, e^{-\lambda s}\, ds \qquad (\tau_1 \sim \text{Exp}(\lambda)). \tag{7b}$$

Now observe that

$$P(N(t) = n \mid \tau_1 = s) = P(N(t) - N(s) = n - 1). \tag{8}$$

In words: if the first event occurred at time s, then the probability of having n events in $[0, t]$ equals the probability of having $n - 1$ events in $(s, t]$. (The half-open interval is important because we don't want to count the event at time s again. It is also what's consistent with the definition of $N(t)$.) Now by Lemma 2.7, $N(t) - N(s)$ is again Poisson, so by our induction hypothesis, we have

$$P(N(t) - N(s) = n - 1) = \frac{e^{-\lambda(t-s)}(\lambda(t-s))^{n-1}}{(n-1)!}. \tag{9}$$

Plugging this into Eq. (7) gives

$$P(N(t) = n) = \int_0^t \frac{e^{-\lambda(t-s)}(\lambda(t-s))^{n-1}}{(n-1)!}\, \lambda\, e^{-\lambda s}\, ds \tag{10a}$$
$$= \frac{\lambda^n}{(n-1)!} \int_0^t e^{-\lambda(t-s)}(t-s)^{n-1}\, e^{-\lambda s}\, ds \tag{10b}$$
$$= \frac{e^{-\lambda t}\lambda^n}{(n-1)!} \int_0^t (t-s)^{n-1}\, ds. \tag{10c}$$

(From the 1st to 2nd line, we just moved some factors out of the integral. From 2nd to 3rd, we used $e^{\lambda s} \cdot e^{-\lambda s} = 1$.) The remaining integral is easy to evaluate, and is just t^n/n. Combining it all gives

$$P(N(t) = n) = \frac{e^{-\lambda t}\lambda^n t^n}{n!}, \tag{11}$$

which was what we wanted.

As for Property (iii), this is the content of Lemmas 2.7 and 2.8. Lemma 2.7, in particular, is the key. My explanation of it in lecture was a bit garbled. I don't currently have a better explanation than what's in the text, so take a look at that and let me know if you have questions!

Here, I'll just point out one more thing: recall that a sum of independent Poisson random variables is again Poisson. This is *compatible* with the independent increments property, as illustrated by the following example: let $N_1 = N(2) - N(1)$ and $N_2 = N(4) - N(2)$, i.e., N_1 and N_2 count the number of events in $(1,2]$ and $(2,4]$, respectively. By Property (ii), they are Poisson with rates $\lambda_1 = \lambda$ and $\lambda_2 = 2\lambda$. Now, $N_1 + N_2$ is just the total numbr of events in $(1,4]$. By Property (ii), they are independent, so $N_1 + N_2$ is again Poisson, with rate $\lambda_1 + \lambda_1 = 3\lambda$. This is exactly what one would expect from applying Property (ii) directly to $N(4) - N(1)$. Note that this reasoning extends to any number of disjoint intervals.

From Properties (i)–(iii) to Model A. While I don't have anything better to say about why Model A has the Poisson properties (i)–(iii), it is relatively straightforward to go in the *reverse* direction, i.e., derive Model A from Properties (i)–(iii). I record the ideas here.

More precisely, suppose we have a Poisson process $N(t)$, and that all we know about it is Properties (i)–(iii). Our first task is to try to define the waiting times $\tau_1, \tau_2, \cdots$ from just $N(t)$. This can be done as follows: we know $N(t)$ is integer valued, and Property (ii) tells us that $N(t) \geqslant N(s)$ if $t \geqslant s$, i.e., $N(t)$ is nondecreasing. So, $N(t)$ can only increas by discrete jumps of size $+1$. The *jump times* T_i can thus be defined as the times at which $N(t)$ has jumps, i.e.,

$$
\begin{aligned}
T_1 &= \min\{t \mid N(t) = 1\} \\
T_2 &= \min\{t \mid N(t) = 2\} \\
&\vdots
\end{aligned}
$$

The waiting times are defined by $\tau_n = T_n - T_{n-1}$, with $T_0 = 0$.

We now need to show that the τ_i are independent $\text{Exp}(\lambda)$ random variables. To do this, note that for any $t \geqslant 0$,

$$
\begin{aligned}
P(\tau_1 > t) &= P(N(t) = 0) && \text{(12a)} \\
&= e^{-\lambda t}. && \text{(12b)}
\end{aligned}
$$

(We used these equalitiess just a little earlier, but in a different way.) This shows that $\tau_1 \sim \text{Exp}(\lambda)$.

Next, to show that $\tau_2 \sim \text{Exp}(\lambda)$ and that it is independent of τ_1, let $s, t \geqslant 0$. Then

$$
\begin{aligned}
P(\tau_1 > s, \tau_2 > t) &= \int_0^\infty P\big(\tau_1 > s, \tau_2 > t \,\big|\, \tau_1 = u\big)\, f_{\tau_1}(u)\, du \qquad \text{(Law of Alt.)} && \text{(13a)} \\
&= \int_s^\infty P\big(\tau_2 > t \,\big|\, \tau_1 = u\big)\, f_{\tau_1}(u)\, du && \text{(13b)} \\
&= \int_s^\infty P\big(\tau_2 > t \,\big|\, \tau_1 = u\big)\, \lambda e^{-\lambda u}\, du. && \text{(13c)}
\end{aligned}
$$

Observe

$$
P\big(\tau_2 > t \,\big|\, \tau_1 = u\big) = P\big(N(u+t) - N(u) = 0 \,\big|\, N(u) = 1\big). \tag{14}
$$

(Draw a time line!) But by Property (iii), $N(t) - N(u)$ and $N(u)$ are independent, so

$$
P\big(N(u+t) - N(u) = 0 \,\big|\, N(u) = 1\big) = P(N(u+t) - N(u) = 0). \tag{15}
$$

By Property (ii), we have

$$P(N(u+t)-N(u)=0)=e^{-\lambda t}. \tag{16}$$

Thus

$$\begin{aligned}
P(\tau_1 > s, \tau_2 > t) &= \int_s^\infty e^{-\lambda t}\ \lambda e^{-\lambda u}\ du && \text{(17a)}\\
&= \lambda e^{-\lambda t}\int_s^\infty e^{-\lambda u}\ du && \text{(17b)}\\
&= e^{-\lambda t}e^{-\lambda s} && \text{(17c)}
\end{aligned}$$

and

$$P(\tau_1 \leqslant s, \tau_2 \leqslant t)=(1-e^{-\lambda t})(1-e^{-\lambda s}). \tag{18}$$

The above is a bivariate analog of the CDF. Taking partial derivatives $\frac{\partial^2}{\partial s \partial t}$ tells us the joint PDF of τ_1 and τ_2 is

$$\lambda^2 e^{-\lambda t}e^{-\lambda s}, \tag{19}$$

i.e., τ_1 and τ_2 are independent.

The argument to show show τ_n is exponential for $n > 2$, and independent of τ_m for all $m < n$, is a little more complicated, but can be constructed along similar lines.