Poisson processes

Stochastic processes

A stochastic process is just a collection of random variables $\{X_t : t \in T\}$ indexed by some set T. The set T is often $\mathbb{R}^+$ and interpreted as time, though there are many other possibilities. What distinguishes the study of stochastic processes is the objective of studying the random function $t \mapsto X_t$. Mostly we'll study these next semester, but the Poisson process is so fundamental and useful in modeling that some study is in order, so I am taking a little time to set it in the context of stochastic processes.

Some more examples of stochastic processes are:

1. Let $\{X_n\}$ be IID. Then $\{S_n\}$ is a stochastic process with $T = \{1, 2, 3, \ldots\}$. So is $\{X_n\}$ for that matter, but we are less likely to ask questions about the function $n \mapsto X_n$ than about the function $n \mapsto X_n$.

2. A markov chain is also a stochastic process on $1, 2, 3, \ldots$.

3. Brownian motion is a stochastic process on $T = \mathbb{R}^+$.

4. A **Gaussian process** on the set T is a collection $\{X_t : t \in T\}$ such that for each finite $A \subseteq T$, the collection $\{X_t : t \in A\}$ is multivariate normal. A (centered) Gaussian process is completely determined by the covariances $V(s, t) := \mathbb{E}X_s X_t$.

A **Poisson process** with rate $\lambda > 0$ is designed to model the following situation. I would like a series of alarms to ring; independently of what has happened in the past, at time t I would like an alarm to go off in the next dt units of time with probability $\lambda\, dt$. Such a process would be a very useful building block for more complicated processes.

We will construct such a process, but before doing so, let's pause to see what properties it would have. Let $N(I)$ denote the number of alarms going off at times in the half-open interval $I := (s, t]$. Let us extend the Poisson convergence theorem as in (6.7) and then apply it to the number of alarms that go intervals obtained by partitioning I into m equal sub-intevals. By the Poisson convergence theorem, $N(I) \to \mathcal{P}(\lambda(t - s))$. But also, if $I_1, \ldots, I_k$ are disjoint subintervals then $\{N(I_j) : 1 \leq j \leq k\}$ converges to a collection of independent

Poissons with means $|I_j|$. Note: $N(s,t] = N[0,t] - N[0,s]$, so we may define such a process in terms of a collection $\{N_t := N[0,t]\}$.

Accordingly, we say that $\{N_t : t \geq 0\}$ is a Poisson process with rate λ on $\mathbb{R}^+$ if it satisfies the following axioms.

1. for $s < t$, the law of $N(s,t] := N_t - N_s$ is $\mathcal{P}(t-s)$.
2. for a finite collection $\{I_1, \ldots, I_k\}$ of finite, disjoint intervals, $\{N(I_j) : 1 \leq j \leq k\}$ are independent.

Durrett shows that these properties follow from (i)–(iv) in Theorem (6.7). We now concern ourselves with constructing such a process. Since it is important, I would like to discuss two very different constructions.

One quick preliminary observation: we think intuitively about events such as the first timer going off (or second or third, etc.) but all we will have constructed are the cumulative random variables: N_t is the number of timers that have gone off by time t. If we let $\tau_k = \inf\{t : N_t = k\}$, then we recover the successive alarm times.

Durrett's exponential construction

Let's go a little farther down the what-if path. If such a process exists, then what is the distribution of τ_1? Well, $\mathbb{P}(\tau_1 > x) = P(N_x = 0) = e^{-\lambda x}$. Thus $\tau_1 \sim \exp(\lambda)$, an exponential of rate λ (mean $1/\lambda$). Next, let's see that τ_1 and $\tau_2 - \tau_1$ are IID. We have

$$\mathbb{P}(\tau_1 > x, \tau_2 - \tau_1 > y) \leq \sum_{m \geq 0} \mathbb{P}(x + m\epsilon < \tau_1 \leq x + (m+1)\epsilon, \tau_2 > y + x + (m+1)\epsilon)\,,$$

which implies that

$$\mathbb{P}(\tau_1 > x, \tau_2 - \tau_1 > y) \leq \sum_{m \geq 0} e^{-\lambda(x+m\epsilon)} \lambda \epsilon e^{-\lambda y} = e^{-\lambda(x+y)} + o(\epsilon)$$

as $m \to \infty$. Similarly,

$$\mathbb{P}(\tau_1 > x, \tau_2 - \tau_1 > y) \geq \sum_{m \geq 0} \mathbb{P}(x + m\epsilon < \tau_1 \leq x + (m+1)\epsilon, \tau_2 > y + x + m\epsilon)\,,$$

whence

$$\mathbb{P}(\tau_1 > x, \tau_2 - \tau_1 > y) \geq \sum_{m \geq 0} e^{-\lambda(x+m\epsilon)} \lambda(\epsilon + o(1)) e^{-\lambda(y-\epsilon)} = e^{-\lambda(x+y)} + o(\epsilon)$$

as $m \to \infty$. Thus

$$\mathbb{P}(\tau_1 > x, \tau_2 - \tau_1 > y) = e^{-\lambda(x+y)}$$

for all $x, y > 0$, implying that τ_1 and $\tau_2 - \tau_1$ are IID exponentials of rate λ. A similar computation shows that $\{\tau_{k+1} - \tau_k\}$ are IID.

The construction is now clear. Let $\{\xi_n\}$ be IID exponentials of rate λ and let $\{\tau_n\}$ be their partial sums. Then $N_t := \sup\{k : \tau_k \leq t\}$ should be a Poisson process with rate λ. This is a theorem, whose proof may be found in Durrett, pages 144–145.

Thinning

You might wonder why we go through so much trouble to construct a "process" that is just the partial sums of IID exponentials. The answer lies in the second construction, which generalizes to a construction of much greater use.

For our second construction, we expand upon Durrett's last paragraph in Section 2.6. Given a measure space $(S, \mathcal{S})$ and a σ-finite measure μ, a **Poisson process with mean measure** μ is a collection $\{N_A : A \in \mathcal{S}\}$ of non-negative integer random variables satisfying

1. $N_A \sim \mathcal{P}(\mu(A))$;
2. for a finite collection $\{A_1, \ldots, A_k\}$ of disjoint sets of finite μ-measure, $\{N(A_j) : 1 \leq j \leq k\}$ are independent.
3. $N_{\bigcup_j A_j} = \sum_n N_{A_j}$ for disjoint $\{A_n\}$.

Clearly, when $S = \mathbb{R}^+$ and μ is λ times Lebesgue measure, the definition boils down to our previous definition of a Poisson process with rate λ [note that (3) is satisfied because of the construction of $N(s, t]$ as $N_t - N_s$].

Durrett's construction of this is based on an exercise, but since I prefer this proof to the proof that the exponential construction works (notice that I referred to the book for that and did not prove it in class), I will state and prove a version of this exercise. (Also, the result is worthwhile in its own right.)

Recall that if X and Y are independent Poissons with means μ and ν, then $Z := X + Y$ is a Poisson with mean $\mu + \nu$, that is, $\mathcal{P}(\mu) * \mathcal{P}(\nu) = \mathcal{P}(\mu + \nu)$. Given Z, a Poisson with mean $\mu + \nu$, is there a natural way to reverse engineer a variable X on the same probability space such that $X \sim \mathcal{P}(\mu)$ and X is independent of $Z - X$ (in which case, necessarily $Z - X \sim \mathcal{P}(\nu)$)?

Lemma 1 (thinning) *Let Z be Poisson λ, and let $\{\xi_n\}$ be independent Bernoullis with mean p. Let $X = \sum_{n=1}^{Z} \xi_n$ be the "result of retaining each point in Z independently with probability p". Then $X \sim \mathcal{P}(p\lambda)$.*

Proof: there are three arguments. One is that you computed the characterstic function of this compound Poisson on your homework and it is equal to the characteristic function of a $\mathcal{P}(p\lambda)$. A second, similar argument is the the generating function for a $\mathcal{P}(\lambda)$ is $\exp(\lambda(z-1))$, so the generating function for the thinned Poisson is $\exp(\lambda(pz+(1-p)-1)) = \exp(\lambda p(z-1))$.

These arguments are short but the third argument will lead to the reverse engineering we wanted. Let us see how to generate a Poisson with mean λ and then thin it by p. Given λ and p, let $\mu := p\lambda$ and $\nu := (1-p)\lambda$. Consider N bins, into each of which is put $\mathcal{P}(\mu/N)$ many red balls and $\mathcal{P}(\nu/N)$ many green balls. We may also generate this as follows. Place $\mathcal{P}(\lambda/N)$ many balls in each bin. For every bin containing exactly one ball, flip a p'-coin to see whether to color it red or green, where $p'/(1-p') = \mu e^{-\mu/N}/(\nu e^{\nu/N})$. If a bin contains more than one ball, consult an oracle as to the right way to color the balls. As $N \to \infty$, this procedure converges to choosing colors with IID p-coins, since $p' \to p$ and the probability of more than one ball in some bin goes to zero. Thus we see that the joint distribution of Z thinned by p and Z minus this is the original pair count: (red balls , green balls). These are independent. By induction, we derive Durrett's Exercise 6.12: If $N \sim \mathcal{P}(\lambda)$ and $\{\xi_n\}$ are IID with $\mathbb{P}(\xi_1 = j) = p_j$ for $1 \leq j \leq k$, then $\{N_j : 1 \leq j \leq k\}$ are independent Poissons with means λp_j where $N_j = |\{i \leq \lambda : \xi_i = j\}|$.

From here, given a finite measure μ on a space S, one may construct the Poisson process of intensity μ on S as follows. Let $\lambda := ||\mu||$ be the total mass of μ and let $\hat{\mu} = \mu/||\mu||$ be the normalized measure. Let $Z \sim \mathcal{P}(\lambda)$ and let $\{\xi_i : i \in \mathbb{Z}^+\}$ be S-valued random variables independent of Z and IID $\sim \hat{\mu}$. For measurable $A \subseteq S$, define a random variable

$N_A := \#\{j \le Z : \xi_j \in A\}$; this definition makes countable additivity (condition 3) trivial. Using the thinning lemma, we may verify

1. $N_A \sim \mathcal{P}(\mu(A))$.
2. If $A_1, A_2, \ldots$ are disjoint then $\{N_{A_j}\}$ are independent.

We have verified that $\{N_A\}$ is a Poisson process of intensity μ.

Finally, if μ is any σ-finite measure, we may partition space into $S_1, S_2, \ldots$; letting $\mu_j := \mu(\cdot \cap S_j)$, we may construct a Poisson process of intensity μ by summing independent Poisson processes of intensities μ_j.

Poissonization

There is one more application of the Poisson machinery worth mentioning, namely Poissonization. To illustrate, suppose we want probability bounds on the event A that when 20000 balls are distributed IID uniformly into 10000 boxes, no box will have more than 9 balls. The problem is, the numbers of balls in different boxes are dependent. To solve the problem, consider instead the Poissonization, where we distribute $\mathcal{P}(\lambda)$ many balls instead of exactly 20000. Under the Poisson distribution, the numbers of balls in each box will be independent $\mathcal{P}(\lambda/10000)$.

Observe that $\mathbb{P}_\lambda(A)$, the probability of finding no more than 9 balls in each box when $\mathcal{P}(\lambda)$ balls are distributed, is decreasing in λ. Suppose λ satisfies

$$\mathbb{P}(\mathcal{P}(\lambda) > 20000) < \epsilon\,.$$

Then

$$\mathbb{P}_\lambda(A) < \epsilon + \mathbb{P}(A)\,.$$

Let us use this for various values of λ. When $\lambda = 20000$, the probability of at most 9 balls in one box is $0.9989033\ldots$, thus we expect that $\mathbb{P}(A) \approx \mathbb{P}_\lambda(A) \approx 0.6281383\ldots$. We compute

$$\begin{aligned} \mathbb{P}_{19700}(A) &\approx 0.6632327\ldots\,, \\ \mathbb{P}_{20300}(A) &\approx 0.5913313\ldots\,. \end{aligned}$$

When $\lambda = 19700$, we take $\epsilon = .024$:

$$\mathbb{P}(\mathcal{P}(19700) > 20000) < .024\,.$$

Similarly,

$$\mathbb{P}(\mathcal{P}(20300) > 20000) < .024\,.$$

of balls distributed as $\mathcal{P}(19700)$ and $\mathbb{P}(\mathcal{P}(19700) > 20000) \approx \mathbb{P}(N(0,1) \geq 1.58) \approx 0.024$. Therefore,

$$0.567 < \mathbb{P}(A) < 0.688\,.$$

For another way to find this value, you should have gone to Herb Wilf's talk on October 31.

Applications of Poisson processes

The following constructions typify the use of Poisson processes in probability modeling.

Queuing: Let $\{\tau_n\}$ be the arrival times of a Poisson process with rate λ. This will represent customers arriving at a FIFO queue. Let $\{Y_n\}$ be IID random variables representing service times: how long it will take to serve the n^{th} customer. Let X_t denote the number of people in the queue at time t. This process may be defined as the right-continuous process that increases by 1 at each time τ_n and decreases by 1 at each completion time W_n where $W_n = W_{n-1} + Y_n$ if $X_{W_{n-1}} > 0$ (if the n^{th} customer was already in line when the $n-1^{st}$ finished) and $W_n = \tau_n + Y_n$ otherwise. The classical theory shows that $\{X_t\}$ is positive recurrent if $\lambda \mathbb{E} Y_1 < 1$ (the queue is clear a positive fractino of the time), null recurrent if $\lambda \mathbb{E} Y_1 = 1$ and transient otherwise (the queue length goes to infinity).

Mixing: Let $M(ij)$ be a nonnegative $n \times n$ symmetric matrix, and for unordered pairs $e := \{1, j\}$ with $M_{ij} \neq 0$, let $\tau_n^{(e)}$ be independent Poisson processes with rates M_{ij}. Let S_t be the right-continuous process on the symmetric group with $S_0 = id$ and jumps at times $\tau_n^{(e)}$ from σ to $\tau_e \sigma$ where τ_e is the transposition switching the elements of e.

Lilly pad percolation: Let $\{N_A\}$ be the Poisson process on $(\mathbb{R}^2, \mathcal{B})$ with intensity equal to λ times Lebesgue measure. Let S be the random set of points x such that $N_{B(x,1)} > 0$ where $B(x,1)$ is the ball of radius 1 centered at x. We interpret S as the set of points covered by Lilly pads (disks of radius 1) growing at random locations. Of great interest is the event H that S has an infinite connected component: you can go arbitrarily far traveling only on overlapping lilly pads. It is known that there is a positive finite λ_c such that for $\lambda > \lambda_c$ $\mathbb{P}(H) = 1$ and for $\lambda < \lambda_c$, $\mathbb{P}(H) = 0$. The value of λ_c is not known, nor (I believe) is it proved that $\mathbb{P}(H) = 0$ at $\lambda = \lambda_c$.

Random objects: Suppose we wish to model Poisson process of objects with random attributes. For example, suppose we model the US by an infinite plane, and we want a Poisson process of cell towers with random heights. Let F denote the height distribution (we wish the heights to be IID and independent of the locations). When working in $\mathbb{R}^1$, we have a natural ordering of the points, so we could construct the locations as $\{\tau_n\}$ and the heights then as $\{X_n\}$. In higher dimension, there is no natural ordering but a convenient way to construct the model is to construct a Poisson process on $(\mathbb{R}^d \times S, \mu \times F)$ where μ is the location intensity and (S, F) is the attribute space and distribution (e.g., here $S = \mathbb{R}^1$ and F is the height distribution).