Applied Probability

February 1, 2018

Contents

0   Miscellaneous

Some speech

Google lecture's name to find his homepage and example sheets or probably some notice of a change of room

1   Poisson process

Suppose we have a Geiger counter. We model the "click process" as a family $\{N(t) : t \geq 0\}$, where $N(t)$ denotes the total number of ticks up to time t. Now note that $N(t) \in \{0, 1, ...\}$, $N(s) \leq N(t)$ if $s \leq t$, N increases by unit jumps, and $N(0) = 0$. We also assert that N is right-continuous, i.e. $\lim_{x \to t^+} N(x) = N(t)$.

Definition. (infinitesimal definition)
A *Poisson process* with intensity λ is a process $N = (N(t) : t \geq 0)$ which takes values in $S = \{0, 1, 2, ...\}$, s.t.:
(a) $N(0) = 0$, $N(s) \leq N(t)$ if $s \leq t$;
(b)

$$\mathbb{P}(N(t+h) = n+m | N(t) = n) = \begin{cases} \lambda h + o(h) & m = 1 \\ o(h) & m > 1 \\ 1 - \lambda h & m = 0 \end{cases}$$

Recall that $g(h) = o(h)$ means that $\frac{g(h)}{h} \to 0$ as $h \to 0$;
(c) if $s < t$, then $N(t) - N(s)$ is independent of all arrivals prior to s.

Theorem. $N(t)$ has the Poisson distribution with parameter λt.

Proof. Study $N(t+h)$ given $N(t)$. We have

$$\mathbb{P}(N(t+h) = j) = \sum_{i \leq j} \mathbb{P}(N(t+h)$$
$$= j | N(t) = i) \mathbb{P}(N(t) = i)$$
$$= (1 - \lambda h) \mathbb{P}(N(t) = j) + \lambda h \mathbb{P}(N(t) = h - 1) + o(h)$$

So

$$\frac{\mathbb{P}(N(t+h) = j) - \mathbb{P}(N(t) = j)}{h} = -\lambda \mathbb{P}(N(t) = j) + \lambda \mathbb{P}(N(t) = j - 1) + \frac{o(h)}{h}$$

write $p_n(t) = \mathbb{P}(N(t) = n)$, then let $h \to 0^+$ we get

$$p_j'(t) = -\lambda p_j(t) + \lambda p_{j-1}(t) \quad j \geq 1$$
$$p_0'(t) = -\lambda p_0(t)$$

with boundary condition $p_0(0) = 1$.
We solve p_0 to get $p_0(t) = e^{-\lambda(t)}$. Then we can use this to inductively solve $p_1, p_2, ...$ to get the desired result. $\square$

An alternative derivation from the differential equations:
Let $G(s, t) = \sum_j s^j p_j(t)$. Now we take the set of differential equation, multiplying each one by s^j, then we get
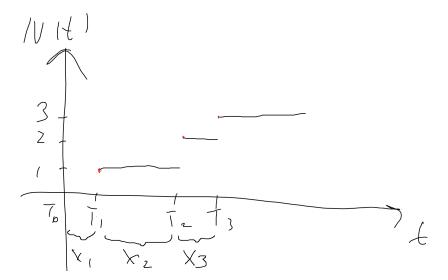
$$\frac{\partial G}{\partial t} = \lambda(s - 1)G$$

Then we have

$$G(s, t) = A(s)e^{\lambda(s-1)t}$$

We also have $G(s, 0) = 1$ so we should be able to plug in a suitable value of s to get the desired result (I probably missed that).

Definition. (Holding/interarrival times) In a poisson process (pp) with parameter λ, let $N(t)$ denote the total number of "clicks". Define the arrival times $T_0 = 0$, $T_n = \inf\{t \geq 0 : N(t) = n\}$, i.e. the first time t that N reaches n (note right continuity of N). We also define the interarrival times $X_n = T_n - T_{n-1}$.

Theorem. Suppose $X_1, X_2, ...$ are known. Let $T_n = \sum_1^n x_i$, note $N(t) = \max\{n : T_n \leq t\}$. Then the random variables $X_1, X_2, ...$ are independent and they have the exponential distribution with parameter λ ($Exp(\lambda)$).

Proof.
$$\mathbb{P}(X_1 > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t}$$

So X_1 has $Exp(\lambda)$ distribution. Now consider $\mathbb{P}(X_2 > t|X_1 = t_1)$. This doesn't look to make much sense as X_1 has a continuous distribution so $\mathbb{P}(X_1 = t_1) = 0$; however we could could consider the conditional densiy as $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$. Then $\mathbb{P}(X_2 > t|X_1 = t_1) = \mathbb{P}(\text{ no arrivals in } (t_1, t_1 + t)|X_1 = t_1) = \mathbb{P}(\text{ no arrivals in } (t_1, t_1 + t)$ by independence. This is then equal to $\mathbb{P}(\text{ no arrivals in } (0, t)) = \mathbb{P}(N(t) = 0) = e^{-\lambda t}$. Then continue by induction. $\square$

Proposition. (properties of a poisson process N)
(a) N has stationary independent increments, i.e.:
(i) If $0 < t_1 < ... < t_n$, then $N(t_1), N(t_2) - N(t_1), ..., N(t_n) - N(t_{n-1})$ are independent;
(ii) $N(s + t) - N(s) \xrightarrow{d} N(t) - N(0)$.
Amongst processes which are right continuous, non-decreasing, has only jump discontinuities of size 1, (i) and (ii) are characteristics of the Poisson process, meaning that Poisson process is the only process that has those two properties.
(b) Thinning:
Suppose insects arrive as a poisson process with parameter λ. Each insect is a mosquito with probability α, or a skeet with probability $1 - \alpha$, and the occurences of the two insects are independent. Then

(i) the mosquito-arrival process F is a $PP(\alpha\lambda)$, (ii) the skeet-arrival process is S a $PP((1-\alpha)\lambda)$, and (iii) these processes are independent.

Proof. (i) and (ii) are immediate by infinitestimal definition of a poisson process. For (iii), by independence we mean that $\mathbb{P}(F(t_1) = f_1, S(t_1) = s_1, ..., F(t_n) = f_n, S(t_n) = s_n) = \mathbb{P}(F(t_1) = f_1, ..., F(t_n) = f_n)\mathbb{P}(S(t_1) = s_1, ..., S(t_n) = s_n)$ $\forall t_1, ..., t_n, f_1, ..., f_n, s_1, ..., s_n)$.

The simple case is

$$\mathbb{P}(F(t) = f, S(t) = s) = \frac{(\lambda t)^{f+s}e^{-\lambda t}}{(f+s)!}\binom{f+s}{f}\alpha^f(1-\alpha)^s$$

$$= \frac{(\alpha\lambda t)^f}{f!}e^{-\alpha\lambda t}\frac{((1-\alpha)\lambda t)^s}{s!}e^{-(1-\alpha)\lambda t}$$

$$= \mathbb{P}(F(t) = f)\mathbb{P}(S(t) = s)$$

$\square$

(c) Superposition:
F: Flies arrive as $PP(\lambda_1)$;
S: Skeets arrive as $PP(\lambda_2)$, and these processes are independent. Then $N = F + S$ is a $PP(\lambda_1 + \lambda_2)$. This follows by infinitesimal construction of PP.
(d) Given $N(t) = n$, write $\mathbf{T} = (T_1, ..., T_n)$, $\mathbf{t} = (t_1, ..., t_n)$, we have $f_{\mathbf{T}}(\mathbf{t}|N(t) = n) = \left(\frac{1}{t}\right)^n n!L(\mathbf{t})$, where $L(\mathbf{t}) = 1$ iff $t_1 < t_2 < ... < t_n$.

Proof. Next time. $\square$

Let's complete the proof left last lecture.

Theorem. Conditional on $\{N(t) = n\}$, the times $T_1, ..., T_n$ have joint pdf

$$f_{\mathbf{T}|N(t)=n}(\mathbf{t}) = \frac{n!}{t^n}L(\mathbf{t})1_{\{t_n \leq t\}}$$

where $L(\mathbf{t}) = 1_{\{t_1 \leq t_2 \leq ... \leq t_n\}}$.

Proof. The interarrival times $X_1, X_2, ..., X_n$ have joint pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \lambda^n \exp(-\lambda\sum_i^n x_i)$$

by change of variables, we now have (noting $T_i = X_1 + ... + X_i$)

$$f_{\mathbf{T}}(\mathbf{t}) = \lambda^n e^{-\lambda t_n} L(\mathbf{t})$$

Now for $C \subseteq \mathbb{R}^n$, we have

$$
\begin{aligned}
\mathbb{P}(T \in C | N(t) = n) &= \frac{\mathbb{P}(T \subseteq C, N(t) = n)}{\mathbb{P}(N(t) = n)} \\
&= \frac{1}{\mathbb{P}(N(t) = n)} \int_C \mathbb{P}(N(t) \\
&= n | \mathbf{T} = \mathbf{t}) f_{\mathbf{T}}(\mathbf{t}) d\mathbf{t} \\
&= \frac{1}{\mathbb{P}(N(t) = n)} \int_{t_n \leq t} e^{-\lambda(t - t_n)} \lambda^n e^{-\lambda t_n} L(\mathbf{t}) d\mathbf{t}
\end{aligned}
$$

the last equation is because we need ther to be no arrival between t and t_n. Now the conditional pdf of $\mathbf{T}$ given $N(t) = n$ is

$$
\frac{1}{(\lambda t)^n e^{-\lambda t}/n!} e^{-\lambda(t - t_n)} \lambda^n e^{-\lambda t_n} L(\mathbf{t}) = \frac{n! L(\mathbf{t})}{t^n} 1_{\{t_n \leq t\}}
$$

I think somewhere in this proof we used $\mathbb{P}(X \in C) = \int_C g(u) du \iff f_X(u) = g(u)$, otherwise the lecture wouldn't have written this down on a separate board. $\qquad \square$

2 Continuous-time Markov chains

This is actually quite a complicated topic, so we are going to make a lot of assumptions to simplify it.

Assume state space S is countable, and we often take $S \subseteq \mathbb{Z} = \{..., -1, 0, 1, ...\}$ (sometimes useful to assume $|S| < \infty$).

Definition. A process $X = \{X(t) : t \geq 0\}$ taking values in S satisfies the *Markov property* if:

$$\mathbb{P}(X(t_n) = j | X(t_1) = i_1, ..., X(t_{n-1}) = i_{n-1})$$
$$= \mathbb{P}(X(t_n) = j | X(t_{n-1}) = i_{n-1})$$

for all $i_1, i_2, ..., i_{n-1}, j \in S, t_1 < t_2 < ... < t_n$.

We have the transition probabilities $p_{i,j}(s,t) = \mathbb{P}(X(t) = j | X(s) = i)$. We, however, assume the process is homogeneous, i.e.

$$p_{i,j}(s,t) = p_{i,j}(0, t-s) := p_{i,j}(t-s) \forall s, t, i, j$$

so the transition probabilities only depend on the duration of time passed instead of the absolute time. We can then write this as a transition matrix $(p_{i,j}(t))_{i,j \in S} = P_t$.

Proposition. The family $\{P_t : t \geq 0\}$ satisfies
(a) $P_0 = I$;
(b) P_t is a *stochastic* matrix, i.e. a non-negative matrix with row sum 1;
(c) $P_{s+t} = P_s P_t$ for $s, t \geq 0$.

Proof. (of (c))
$p_{i,j}(s+t) = \sum_{k \in S} p_{i,k}(s) p_{j,k}(t)$ by Markov Property which is just the component form of $P_{s+t} = P_s P_t$.
$P_{s+t} = P_s P_t$ is sometimes called the *semigroup property* $(s, t \geq 0)$.
$(P_t : t \geq 0)$ is called a *stochastic semigroup*. $\qquad\square$

General theory involves conditions of regularity.

We assume X is a right-continuous jump process.

Holding times for general chains:

Assume $X(t_0) = i$.
Let $H = \inf\{t > t_0 : X(t) \neq i\}$. We have

$$\mathbb{P}(H > u + v | H > u) = \mathbb{P}(H > v) \ (*)$$

by Markov Property $(u, v \geq 0)$.
Let $G(u) = \mathbb{P}(H > u)$. By (*), we get $\frac{G(u+v)}{G(u)} = G(v)$, so $G(u+v) = G(u)G(v)$.
We know $G(0) = 1$, and G is non-increasing.
Solution: $G(n) = G(1)G(n-1) = G(1)^n \ \forall n \in \mathbb{N}$. Also $G(p/q)...G(p/q) =$

$G(p) = G(1)^p$ so $G(p/q) = G(1)^{p/q}$, hence $G(u) = G(1)^u$ for $u \geq 0$. We deduce that $G(u) = e^{-\alpha u}$ for some $\alpha > 0$.

Lemma. A random variable $X > 0$ has an exponential distribution iff it has the *lack of memory property*: $\mathbb{P}(X > u + v | X > u) = \mathbb{P}(X > v) \ \forall u, v > 0$.

A MC s a combination of exponential-distribution holding times, and a transition matrix for the *jump chain* $Y = (Y_n)$ given by $Y_0 = X(0)$, $Y_1 = X(T_1)$, where $T_1 = \inf\{t : X(t) \neq X(0)\}$, and $Y_n = X(T_n)$ where $T_{n+1} = \inf\{t > T_n : X(t) \neq X(T_n)\}$. Y is a *discrete-time Markov chain*.

If in state i, want H, we jump to state $j \neq i$ with probability $\frac{g_{ij}}{\alpha_i}$. Intensity of a jump is α_i, and intensity of a jump to state j is g_{ij}.
Note a transition from i to itself is not deemed to be a transition.

We have $p_{ij}(h) = g_{ij}h + o(j)$ $(j \neq i)$, $p_{ij}(h) = 1 - \sum_{j \neq i} p_{ij}(h) = 1 - h\sum_{j \neq i} g_{ij} + o(h) = 1 - \alpha_i h + o(h) = 1 + g_{ii}h + o(h)$, where we let $g_{ii} = -\alpha_i$. Now we let G be the matrix (g_{ij}), with the off-diagonal terms the previous g_{ij}'s, but the diagonal terms g_{ii} as defined just now (so as to make row sums 0). Now the off-diagonal terms are non-negative, and diagonal terms are non-positive. We call G the *generator* of the chain (otherwise known as the Q-matrix).

Conclusion: $\frac{P_t - I}{t} \xrightarrow{t \to 0^+} G$.
Questions of regularity: OK if $|S| < \infty$.

(?)
$$p_{ij}(t + h) = \sum_k p_{ik}(t)p_{kj}(h)$$
$$= \sum_{k \neq j} p_{ik}(t)[g_{kj}h + o(h) + p_{ij}(t)(1 + g_{jj}h + o(h))$$
$$= \sum_k p_{ik}(t)g_{kj}$$
$$= P_t \cdot G$$

this is the (Kolmogov) Forward Equation.

$p_{ij}(t + h) = \sum_k p_{ik}(h)p_{kj}(t)$, so $P't = GP_t$, called the K-Backward equation.

Interchange of limits requires justification – it's OK if $|S| < \infty$.

Now $P'_t = P_t G$, we can rewrite this as $f' = fg$, which gives $f(t) = Ae^{gt}$. So the solution should be $P_t = P_0(= I)e^{tG}$ (i.e. $= \sum_{k=0}^{\infty} \frac{t^k}{k!}G^k$).

In many cases, the solution to the forward and/or backward equation is the function $P = e^{tG}$.

A mistake has been made! The definition for holding time is wrong. It should be $H = \inf\{t - t_0 : X(t) \neq X(t_0), t > t_0\}$ (the length rather than the absolute time).

Let's look at an example now.

Example. Let $S = \{1, 2\}$, $G = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$, where $\alpha\beta > 0$. What are the $p_{ij}(t)$? There are only two states, so let's find

$$p'_{11}(t) = -\alpha p_{11} + \beta p_{12},$$
$$p'_{12}(t) = \alpha p_{11}(t) - \beta p_{12}(t)$$

$$\cdots$$

So we get a bunch of differential equations, with boundary conditions $p_{11}(0) = p_{22}(0) = 1$. This has a unique solution (check). We get $P_t = e^{tG} = \sum_n \frac{t^n}{n!} A\Lambda^n A^{-1}$ wher we diagonalize $G = A\Lambda A^{-1}$. But $e^{t\Lambda}$ is equal to $Diag(e^{\lambda_1 t}, e^{\lambda_2 t}...)$. Each $p_{ij}(t)$ has the form $\sum_k e^{\lambda_k t} c_K(i, j)$, where we need to find the constants $c_k(i, j)$.

Lemma. Let $i, s \in S$. Then either $p_{ij}(t) = 0 \; \forall t > 0$, or $p_{ij}(t) > 0 \; \forall t > 0$ (this is because our time here is continuous; we can fit in any number of jumps in any time length with positive probability).

Proof. Assume $p_{ij}(T) > 0$ for some $T > 0$. Then for $t > T$, $p_{ij}(t) \geq p_{ij}(T)\mathbb{P}_j(X_j > t - T) > 0$ (start in j and holding time $> t - T$, which is positive).
For $t < T$: there exists finite sequence of jumps from i to j in time T, so there exists $i_1, i_2, ..., i_n \in S$ with $g_{i,i_1} g_{i_1,i_2}...g_{i_n,j} > 0$, where $g_{i,j}$ is as defined previously. Then we can just divide the time $(0, t)$ into $n + 1$ intervals. Then $p_{i,j}(t) \geq p_{i,i_1}(t/(n+1))...p_{t_n,j}(t/(n+1)) > 0$ (this is an applied course, so we don't care that much). $\square$

Definition. The chain X on state space is *irreducible* if $\forall i, j \in S$, $\forall t > 0, p_{i,j}(t) > 0$ ($\iff \exists t > 0, p_{i,j}(t) > 0$).

A distribution π on S is invariant (or stationary, or equilibrium distribution) if $\pi = \pi P_t$ for all $t \geq 0$.

Note: if $X(0)$ has distribution μ_0, then $X(t)$ has distribution $\mu_t = \mu_0 P_t$ ($\mu_t(j) = \sum_i \mu_0(i) p_{i,j}(t)$).

Note: (a) Differentiate $\pi = \pi P_t$ to get $0 = \pi G$;
(b) If $P_t = e^{tG}$ then $\pi G = 0$ iff $\pi G^n = 0$ for $n \geq 1$ iff $\pi \sum_n \frac{t^n}{n!} G^n = \pi$ iff $\pi P_t = \pi$.

Theorem. Let X be irreducible. If there exist an invariant distribution π, then it is unique (what a surpise), and $p_{ij}(t) \to \pi_j$ as $t \to \infty$. Can we prove this in the remaining 11 minutes? Let's try:

Proof. Let $h > 0$. Then $Y_n = X(nh)$. So Y is a 'skeleton' of X. Y is a markov chain. Since X is irreducible, so is Y. Y has invariant distribution π, hence π is unique. (?) Since X is irreducible, Y is aperiodic(?). Hence $p_{ij}(nh) \to \pi_{ij}$ as $n \to \infty$. Since this holds for all $h \in \mathbb{Q}^+$, we deduce that $p_{ij}(t) \to \pi_j$ as $t \to \infty$ through the rationals. The conclusion follows by continuity of $p_{ij}(\cdot)$. $\square$

Lemma. The functions $p_{ij}(\cdot)$ are continuous.

Last time we claimed that if $\forall h > 0$, $p_{ij}(nh)\pi_j$ as $n \to \infty$, then $p_{ij}(t) \to \pi_j$ as $t \to \infty$. However this is wrong as the rate of convergence might depend on h.

From some theorem in Linear Analysis it is known that if p is continuous then the above is actually true. But this is an applied course so let's not assume Linear Analysis. Let's now prove it with the following lemma:

Lemma. $p_{ij}(\cdot)$ is uniformly continuous in t.

Proof.

$$|p_{ij}(t+h) - p_{ij}(t)| = |[\sum_k p_{ik}(h)p_{kj}(t)] - p_{ij}(t)|$$
$$\leq |\sum_{k\neq i} p_{ik}(h)p_{kj}(t)| + |p_{ij}(t)[1 - p_{ii}(h)]|$$
$$\leq [1 - p_{ii}(h)]$$
$$\leq 2(1 - e^{-g_i h}) \to 0$$

as $h \to 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Back to the theorem: let $t > 0$, $\exists h$ s.t. $|p_{ij}(t+h) - p_{ij}(t)| < \frac{1}{2}\varepsilon \forall t$. Then $|p_{ij}(t) - p_{ij}(\lfloor t/h \rfloor h)| < \frac{1}{2}\varepsilon$. Pick N s.t. $|p_{ij}(nh) - \pi_j| < \varepsilon/2$ for $n \geq N$. For $t > (n+1)h$ we have $|p_{ij}(t) - \pi_j| \leq |p_{ij}(t) - p_{ij}(\lfloor t/h \rfloor h)| + |p_{ij}(\lfloor t/h \rfloor h) - \pi_j) \leq \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon$ so done.

Explosion:
Let S be countable. $H = (h_{i,j} : i, j \in S)$ is the transition matrix of a discrete time markov chain $Z = (Z_n : n \geq 0)$ on S. Assume $h_{i,i} = 0 \forall i \in S$. Let $(g_i : i \in S)$ be non-negative reals. Inifinitesimal definition of X: $g_{ij} = g_i h_{ij}$ if $i \neq j$, and $-g_i$ if $i = j$. Holding time definition: $X(0) = Z_0$. Given Z, let $U_0, U_1, ...$ be independent exponential random variables, where U_n has parameter g_{Z_n}.

We define $T_n = U_0 + ... + U_{n-1}$, the time of the nth jump. $X(t) = Z_n$ if $T_n \leq t < T_{n+1}$.

$T_n \to T_\infty = \sum_0^\infty U_n$. Say the process [explodes if $T_\infty < \infty$], or explodes if $\mathbb{P}(T_\infty < \infty) > 0$.

We augment the state space S to $S' = S \cup \{\infty\}$ (cemetery state, means ∞ is absorbing). Assume: at T_∞, the chain enters the cemetery state labelled ∞. Such a process is called minimal.

Theorem. The process X, constructed via holding times as above, does not explode if any of the following occurs:
(a) $|S| < \infty$;
(b) $\sup_i g_i < \infty$;
(c) $X(0) = i$, where i is recurrent for the jump chain Z.

Lemma. Let $X_1, ...$ be independent random variales, and X_i has distribution $Exp(\lambda_{i-1})$. Let $T_\infty = \sum_1^\infty X_i$. $\mathbb{P}(T_\infty < \infty = 0$ if $\sum \lambda_i^{-1} = \infty$, and 1 otherwise.

Proof.

$$\mathbb{E}(T_\infty) = \mathbb{E}(\sum_1^\infty X_i)$$

$$= \mathbb{E}(\lim_{N \to \infty} \sum_1^N X_i)$$

$$= \lim_{N \to \infty} \mathbb{E}(\sum_1^N X_i)$$

$$= \lim_{N \to \infty} \sum_1^N \frac{1}{\lambda_{i-1}}$$

$$= \sum_1^\infty \frac{1}{\lambda_{i-1}}$$

Now If $\sum \frac{1}{\lambda_{i-1}} < \infty$, then $\mathbb{E}(T_\infty) < \infty$. So $\mathbb{P}(T_\infty = \infty) = 0$.
The other part will be proven in next lecture.
Problem sheet 2 is online! □