Stochastic Processes

Week 3 Notes

Recall that a *discrete-time Markov chain* on a (discrete) state space S consists of a collection of random variables $(M_k)_{k\in\mathbb{Z}}$ and a transition matrix P, with entries indexed by $S\times S$. The entries of P correspond to the probabilities of transition from one state to another, i.e.

$$P_{xy} = \mathbb{P}\{M_{k+1} = y | M_k = x\}$$

Note that this process is *time-homogeneous*; for any $k,\ell\in\mathbb{Z}$, $\mathbb{P}\{M_{k+\ell+1} = y|M_{k+\ell} = x\} = \mathbb{P}\{M_{k+1} = y|M_k = x\}$. Moreover, it also *memoryless*; that is, for any finite number of states $x_1, ..., x_n$ and times $i_1 < ... < i_n < k$, we have

$$\mathbb{P}\{M_{k+\ell} = y|M_k = x, M_{i_1} = x_1, ..., M_{i_n} = x_n\} = (P^\ell)_{xy} = \mathbb{P}\{M_{k+\ell} = y|M_k = x\}$$

where $(P^\ell)_{xy}$ is the (x,y)-th entry of P^ℓ. In other words, the probability of ending up at some future state only depends on the current state and not on any previous states.

This motivates the definition of a continuous-time Markov chain, when the transition times are allowed to take values in $\mathbb{R}_{\geq 0}$ instead of $\mathbb{Z}$. We give two such definitions and prove that they are equivalent.

Definition 1: A *continuous-time Markov chain* on a state space S consists of a generator matrix G and a collection of random variables $(X_t)_{t\geq 0}$ such that for any $\epsilon > 0$

$$\mathbb{P}\{X_{t+\epsilon} = y|X_t = x\} = -G_{xy}\epsilon + \mathcal{O}(\epsilon^2),\ \text{for } x \neq y$$

$$\mathbb{P}\{X_{t+\epsilon} = x|X_t = x\} = 1 + \epsilon G_{xx} + \mathcal{O}(\epsilon^2) \tag{1}$$

It follows that the entries of G must satisfy $G_{xy} \geq 0$ when $x\neq y$ and $G_{xx} = -\sum_{y\neq x} G_{xy}$. Note that just as in the discrete case, the X_t are memoryless and time-homogeneous.

Definition 2: A *continuous-time Markov chain* on a state space S consists of a generator matrix G and random variables $(X_t)_{t\geq 0}$ such that if $\tau_t = \inf\{s > 0|X_{t+s} \neq X_t\}$ (the time from t it takes to jump to a state different from X_t), then $(\tau_t|X_t = x)$ is exponentially distributed with mean $\sum_{y\neq x} G_{xy}$ and

$$\mathbb{P}\{X_{\tau_t} = y|X_t = x\} = \frac{G_{xy}}{\sum_{z\neq x} G_{xz}} \tag{2}$$

These conditions say that when at state x, the chain jumps at rate $\sum_{y\neq x} G_{xy}$ to a state y with probability proportional to G_{xy}.

Proposition 1: The two definitions above agree.

Proof: First note that since everything is time-homogeneous, we can assume WLOG that $t = 0$. For the first implication, we need to show that $\mathbb{P}\{\tau_0 > s|X_0 = x\} = e^{-\lambda s}$ where $\lambda = \sum_{y\neq x} G_{xy} = -G_{xx}$. We claim

$$\mathbb{P}\{\tau_0 > s|X_0 = x\} = \lim_{\epsilon\to 0^+} \mathbb{P}\{X_{k\epsilon} = x \forall 1 \leq k \leq s/\epsilon | X_0 = x\}$$

Indeed, this simply says that the probability of the first jump occurring after time s is just the probability that the process remains in the state x for all times up to s. By the memoryless property, we can compute the above as

$$\mathbb{P}\{X_{k\epsilon} = x, \forall 1 \leq k \leq s/\epsilon|X_0 = x\} = \prod_{k=1}^{\text{floor}(s/\epsilon)} \mathbb{P}\{X_{k\epsilon} = x|X_{(k-1)\epsilon} = x\} = (1+\epsilon G_{xx})^{s/\epsilon}$$

This last expression goes to $e^{-\lambda s}$ as desired. The second implication is left as an exercise. $\square$

Any given continuous-time Markov chain $(X_t)_{t\geq 0}$ with matrix G can be produced from a discrete-time Markov chain $(M_k)_{k\in\mathbb{N}}$ through a process called *Poissonization*. The construction is as follows. Let $\lambda = \max_x\{-G_{xx}\}$ be the largest jump rate in G and define the matrix P by

$$P_{xy} = \begin{cases} \frac{1}{\lambda}G_{xy}, & x \neq y \\ 1 + \frac{1}{\lambda}G_{xx}, & x = y \end{cases}$$

In other words, P is the sum of $\frac{1}{\lambda}G$ and a suitably-sized identity matrix. It is easily verified that $P_{xy} \geq 0$ and $\sum_{y\neq x} P_{xy} = 1$. Now let $N(t) = N([0,t])$ be a Poisson process on $\mathbb{R}_{\geq 0}$ with rate λ, and let $(M_k)_{k\in\mathbb{N}}$ be the discrete-time Markov chain with transition matrix P.

Proposition 2: The original continuous-time Markov chain is recovered by setting $X_t = M_{N(t)}$.

Proof: Once again, time-homogeneity implies that it's enough to consider the time 0 case. We first need to show that $\tau = \inf\{t \geq 0 | X_t \neq X_0\}$ is exponentially distributed with rate $-G_{xx}$; i.e. $\mathbb{P}\{\tau > t | X_0 = x\} = e^{tG_{xx}}$. To that end, label the points of N by the discrete states of the process; i.e. the label (t, x) indicates a jump to state x at time t. Let $\tilde{N}(t)$ be the number of points in $[0, t]$ with label not equal to x. Then by the labeling theorem, $\tilde{N}$ is a Poisson point process with mean density $\lambda(1 - P_{xx}) = G_{xx}$ and we have

$$\mathbb{P}\{\tau > t | X_0 = x\} = \mathbb{P}\{\tilde{N}(t) = 0\} = e^{tG_{xx}}$$

Finally, (2) is easily verified using the properties of P $\square$

For any continuous-time Markov chain $(X_t)_{t\geq 0}$, one can calculate transition probabilities using the exponential of the transition matrix G.

Proposition 3: $\mathbb{P}\{X_t = y | X_0 = x\} = \left(e^{tG}\right)_{xy} = \sum_{n\in\mathbb{N}} \frac{t^n}{n!}(G^n)_{xy}$

Proof: Let $P = I + \lambda G$ denote the matrix of the corresponding discrete-time Markov chain. As mentioned above, the entries of P^n give the probability of transitioning from x to y given that there are n jumps, i.e.

$$\mathbb{P}\{X_t = y | X_0 = x, N(t) = n\} = (P^n)_{xy}$$

By averaging over n, we obtain

$$\mathbb{P}\{X_t = y | X_0 = x\} = \sum_{n\in\mathbb{N}} \frac{(\lambda t)^n}{n!} e^{-\lambda t}(P^n)_{xy} = e^{-\lambda t}\left(e^{\lambda P t}\right)_{xy} = \left(e^{\lambda t(P-I)}\right)_{xy} = \left(e^{Gt}\right)_{xy}$$

$\square$