University of Cambridge

Part II of the Mathematical Tripos

Applied Probability

Lectured by Sourav Sarkar, Lent 2024–25

Notes by Avish Kumar
ak2461@cam.ac.uk
https://ak1089.github.io/maths/notes

Version 1.3

Contents

1 Continuous-Time Markov Chains

1.1 The Markov Property

A *stochastic process*, or sometimes just a *process*, is a sequence of random variables $X = (X_n)_{n \in \mathbb{N}_0}$.

Definition 1.1 (Markov Chain)

A stochastic process is said to have the *Markov* property if the future and the past are independent, given the present. That is, for any $k > n$, the conditional distribution of X_k given $X_1 \dots X_n$ (the future given the past and present) is the same as that given X_n alone (only the present).

The process X is called a discrete-time Markov Chain with state space I if:

$$\forall x_0, x_1, \dots x_n \in I,\ \mathbb{P}[X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_0 = x_0] = \mathbb{P}[X_n = x_n \mid X_{n-1} = x_{n-1}].$$

If these probabilities are independent of n, then the chain is called time-homogeneous. We then write $P = (P_{xy})_{x,y \in I}$ for the transition matrix with $P_{xy} = \mathbb{P}[X_n = y \mid X_{n-1} = x]$. This is a stochastic matrix, which means all entries are nonnegative and all rows sum to 1.

A time-homogeneous discrete-time Markov Chain has an initial distribution $\mu_0 : I \to [0, 1]$, where $\mu_0(x) = \mathbb{P}[X_0 = x]$ for all $x \in I$. It also has a transition matrix P, as described. From now on, we take the state space I to be some countable (possibly finite) set.

Note: There is some space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to which probabilities are defined, but we do not consider it in detail in this course.

Definition 1.2 (Continuous-Time Process)

We now consider a *continuous-time* process: if $X_t = (X(t) : t \geqslant 0)$ is a right-continuous function taking values in I, then we call $X = X_t$: a continuous-time random process if:

1. for all $t \geqslant 0$, $X(t)$ is a random variable with values in I.
2. for all $\omega \in \Omega$, $t \mapsto X_t(\omega)$ has right-continuous sample paths: for all t. That is, there is some $\varepsilon > 0$ such that $X_s(\omega) = X_t(\omega)$ for all $s \in [t, t + \varepsilon]$.

A right-continuous random process is determined by its finite-dimensional distributions:

$$\mathbb{P}[X_{t_0} = i_0, \dots, X_{t_n} = i_n] : n \in \mathbb{N}, t_k \geqslant 0, i_k \in I.$$

That is, a random process is determined by all finite experiments on it. (We do not prove this).

For every $\omega \in \Omega$, the path $t \mapsto X_t(\omega)$ of a right-continuous process remains constant for some time. There are thus three possibilities for such a path:

1. The path makes infinitely many jumps, but only finitely many within any finite interval. For example, if X_t makes a jump at each integer $t \in \mathbb{N}$, then this case holds.
2. The path makes finitely many jumps in its entire lifetime, before being absorbed in a state. This means that $\exists\, T$ with $X_t = X_T$ for all $t > T$.
3. There is a single finite interval within which the path makes infinitely many jumps. For example, if X_t makes a jump at times $1 - 2^n$ for $n \in \mathbb{N}$, then it will make infinitely many jumps in the first second. Here, $T = 1$ is called the explosion time, and the process starts again after it. We usually ignore everything after the first explosion time.

Definition 1.3 (Jump Times)

The *jump times* J_i for $i = 0, 1, 2 \dots$ (when the process changes value) and the *holding times* S_i for $i = 1, 2, 3 \dots$ (the duration between successive jumps) are formally given by:

$$J_0 = 0, \qquad J_{n+1} = \inf\{t > J_n : J_t \neq J_n\}, \qquad \text{and } S_n = J_{n+1} - J_n.$$

If there are only finitely many jumps, then we take $\inf \varnothing = \infty$, and $\infty - \infty = \infty$, so that these are all well-defined. In this case, we define $X_\infty = \lim_{t\to\infty} X_t$, which is then well-defined.

The first explosion time ζ is then defined as the supremum of the jump times $\sup J_n$. We thus set $X_t = \infty$ for all $t \geqslant \zeta$, adjoining a new state ∞ to I if needed.

With these definitions in mind, we may define a Markov process, which is the continuous-time equivalent of a Markov chain.

Definition 1.4 (Markov Process)

A continuous-time right-continuous random process $X = (X_t)_{t\geqslant 0}$ is called a Markov process if for all $i_1 \dots i_n \in I$ and all $0 < t_1 < \cdots < t_n$, we have

$$\mathbb{P}[X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1}, \dots, X_{t_0} = i_0] = \mathbb{P}[X_n = x_n \mid X_{t_{n-1}} = i_{n-1}].$$

For all $h > 0$, note that $Z_n = X_{hn}$ is a discrete-time Markov chain.

The transition probabilities are now a function $P_{ij}(s,t) = \mathbb{P}[X_t = j \mid X_s = i]$ for $s \leqslant t$ and $i, j \in I$. We call this process time-homogeneous if $P_{ij}(s,t)$ can be expressed as a function of $t - s$, the time difference, independently of s: $P_{ij}(s,t) = P_{ij}(0, t-s)$ for all $t > 0$.

Much like the discrete-time case, continuous-time Markov processes on some state space I are characterised by two things: an initial distribution $\lambda_i = \mathbb{P}[X_0 = i]$ for $i \in I$, and an uncountably infinite family of transition matrices $P(t)_{t\geqslant 0}$.

This family is called the *transition semigroup* of the Markov process. $P(0) = 0$, and for all $t \geqslant 0$, $P(t)$ is a stochatic matrix.

Proposition 1.5 (Transition Semigroup)

The transition semigroup of any Markov Process also satisfies the semigroup property:

$$P(t+s) = P(t)P(s) \quad \forall s, t \geqslant 0.$$

Proof: This can be shown algebraically using the Markov property:

$$\begin{aligned} P_{xz}(t+s) &= \mathbb{P}[X_{t+s} = z \mid X_0 = x] \\ &= \sum_{y\in I} \mathbb{P}[X_{t+s} = z \mid X_t = y, X_0 = x] \times \mathbb{P}[X_t = y \mid X_0 = x] \\ &= \sum_{y\in I} \mathbb{P}[X_{t+s} = z \mid X_t = y] \times \mathbb{P}[X_t = y \mid X_0 = x] \\ &= \sum_{y\in I} P_{xy}(s) \times P_{yz}(t) \end{aligned}$$

which is precisely the definition of $P(t)P(s)$. □

1.2 Holding Times and the Poisson Process

Note: From now on, we suppose that all Markov chains are right-continuous, time-homogeneous, continuous-time, and take values within a countable state space I.

Let's say X is such a Markov chain which starts at x. How long does it stay there? We call S_x the "holding time at x". How long is this holding time? Since X is right-continuous, $S_x > 0$. Now suppose that $s, t \geqslant 0$. We have

$$\begin{aligned}
\mathbb{P}[S_x > t+s \mid S_x > s] &= \mathbb{P}\Big[X_u = x \; \forall u \in [0, t+s] \mid X_u = x \; \forall u \in [0, s]\Big] \\
&= \mathbb{P}\Big[X_u = x \; \forall u \in [s, t+s] \mid X_s = x\Big] \text{ by the Markov property} \\
&= \mathbb{P}\Big[X_u = x \; \forall u \in [0, t] \mid X_0 = x\Big] \text{ by time-homogeneity} \\
&= \mathbb{P}\left[S_x > t\right]
\end{aligned}$$

Therefore S_x has the *memoryless property* for any state $x \in I$.

Theorem 1.6 (Memoryless Property)

Let X be a positive random variable. Then X has the memoryless property

$$\mathbb{P}\left[X > t+s \mid X > s\right] \implies \mathbb{P}\left[X > t\right] \qquad \forall s, t \geqslant 0$$

if and only if X is exponentially distributed.

Proof: ($\Rightarrow$) We have $\mathbb{P}[X > s+t \mid X > s] = e^{-\lambda(s+t)}/e^{-\lambda s} = e^{-\lambda t} = \mathbb{P}[X > t]$.

($\Leftarrow$) Set $G(t) = \mathbb{P}[X > t]$: the probability X exceeds a given value. Then the memoryless property gives us the condition on G:

$$G(t+s) = \mathbb{P}[X > t+s] = \mathbb{P}\left[X > t+s \mid X > s\right] \times \mathbb{P}[X > s] = \mathbb{P}[X > t] \times \mathbb{P}[X > s] = G(t)G(s).$$

Since $X > 0$ almost surely, there is some n such that $G(1/n) = \mathbb{P}[X > 1/n] > 0$. Then $G(1)$ can be written as $G(1) = G(1/n + \cdots + 1/n) = G(1/n)^n > 0$. Therefore we can set $G(1) = e^{-\lambda}$ for some $\lambda \geqslant 0$ (since $G(1) \leqslant 1$).

Therefore for all $k \in \mathbb{N}$, we have $G(k) = G(1 + \cdots + 1) = F(1)^k = e^{-k\lambda}$. This means that for all $p, q \in \mathbb{N}$, we must have $G(p/q) = G(1/q)^p$, but $G(1/q)^q = G(1)$, so this is $e^{-(p/q)\lambda}$.

Thus for any $t \geqslant 0$, we have $r, s \in \mathbb{Q}$ with $r \leqslant t \leqslant s$. Since G is decreasing, $G(s) \leqslant G(t) \leqslant G(r)$. Thus if $s \downarrow t$ and $r \uparrow t$, we have $G(t) = e^{-\lambda t}$ for all $t \geqslant 0$.

But then $\mathbb{P}[X \leqslant t] = 1 - e^{-\lambda t}$ for all t, which is the definition of $X \sim \text{Exponential}(\lambda)$. □

1.3 Poisson Process

Now we consider the canonical continuous-time Markov chain: the Poisson process on $\mathbb{R}^+$.

Definition 1.7 (Basic Poisson Process)

Suppose that $S_1, S_2, \ldots$ is a sequence of iid. random variables, with $S \sim \text{Exp}(\lambda)$. Define the jump times $J_0 = 0$, $J_1 = S_1$, $J_n = S_1 + \cdots + S_n$. Then set $X_t = i$ if $J_i \leqslant t < J_{i+1}$ for $i \in \mathbb{N}_0$.

Then X is called a Poisson process on $\mathbb{R}^+$ with parameter λ. Note that X is right-continuous and non-decreasing.

Note: We sometimes refer to the jumps J_i as the *points* of a Poisson process, so that X_t is the number of points in the interval $[0, t]$.

Theorem 1.8 (Markov Property of Poisson Processes)

Let $(X_t)_{t\geqslant 0}$ be a Poisson process with parameter λ, written $\text{PP}(\lambda)$. Then for any $s \geqslant 0$, the process $(X_{s+t} - X_s)_{t\geqslant 0}$ is also $\text{PP}(\lambda)$ and is independent of $(X_r)_{r<s}$.

Proof: Set $Y_t = X_{s+t} - X_s$ for $t \geqslant 0$. Then let $i \in \mathbb{N}_0$ and condition on $X_s = i$. Then the jump times for the process Y are given by $J_{i+1} - s$, $J_{i+2} - s, \dots$ since $J_i < s$.

The holding times are given by $T_1 = J_{i+1} - s = J_i + S_{i+1} - s = S_{i+1} - (s - J_i)$, and future times are $T_j = S_{i+j}$, where J and S are the jump times and holding times of X.

Since $\{X_s = i\} = \{J_i \leqslant s\} \cup \{S_{i+1} > s - J_i\}$, the distribution of T_1 conditioned on $X_s = i$ is:

$$\mathbb{P}[T_1 > t \mid X_s = i] = \mathbb{P}[S_{i+1} > s - J_i + t \mid J_i \leqslant s, s_{i+1} > s - J_i]$$

Using the independence of S_{i+1} and J_i, this is simply $P[S_{i+1} > t]$ by memorylessness. But the S_i are independent, so we just see that $T_1 \sim \text{Exp}(\lambda)$.

Moreover, the times T_j for $j \geqslant 2$ are independent of S_k for $k \leqslant i + 1$, and hence independent of $(X_r)_{r<s}$, and so they are distributed in the same way. □

Definition 1.9 (Stopping Time)

A random variable T with values in $[0, \infty]$ is called a *stopping time* if for all $t \in \mathbb{R}$, the event $\{T \leqslant t\}$ depends only on $(X_s)_{s\leqslant t}$. That is, a stopping time is an event whereby "you know when you have hit it".

The random variable $T = \inf\{t : X_t \geqslant 2\}$ is thus a stopping time, since we can stop when we first hit 2, but the event $T = \inf\{t : X_t = \sup(X_i)\}$ is not, since we don't know if we will later hit a larger value.

Theorem 1.10 (Strong Markov Property)

Let $(X_t)_{t\geqslant 0} \sim \text{PP}(\lambda)$, with T a stopping time. Then conditioning on $T < \infty$, the process $(X_{s+T} - X_T)_{s\geqslant 0}$ is also $\text{PP}(\lambda)$ and independent of $(X_r)_{r\leqslant T}$.

The following theorem gives three equivalent characterisations of Poisson Processes. Any of these can be used to define the process,

Theorem 1.11 (Poisson Process Formulations)

Let (X_t) be an increasing right-continuous process taking values in $\mathbb{N}_0$ with $X_0 = 0$. Also, let $\lambda > 0$ be a constant. Then the following definitions of a Poisson Process are equivalent:

1. The holding times $S_1, S_2 \dots$ are iid. exponential random variables with parameter λ and the jump chain is $Y_n = n$. This is the traditional definition of a Poisson process.
2. X has independent increments and as $h \downarrow 0$, we have the two relations (uniformly in t) $\mathbb{P}[X_{t+h} - X_t = 1] = \lambda h + \mathcal{O}(h)$, and $\mathbb{P}[X_{t+h} - X_t = 0] = 1 - \lambda h + \mathcal{O}(h)$.
3. X has independent and *stationary* increments, and $X_t \sim \text{Poisson}(\lambda t)$ for all $t \geqslant 0$. That is, the increment process is shift-invariant: the process $(X_{t+s} - X_s)_{t\geqslant 0}$ has the same distribution as $(X_t)_{t\geqslant 0}$, which is therefore time-homogeneous.

All of these uniquely define the Poisson Process on $\mathbb{R}^+$.

Proof: (1 ⇒ 2) If the holding times are iid. exponential random variables, then the increments are independent and stationary. Thus uniformly in t, as $h \downarrow 0$, we have

$$\mathbb{P}[X_{t+h} - X_t = 0] = \mathbb{P}[X_h = 0] = \mathbb{P}[S_1 > h] = e^{-\lambda h} = 1 - \lambda h + \mathcal{O}(h),$$

where equalities arise from stationarity, definition of holding times, the exponential distribution, and the Taylor expansion of the exponential function. Also,

$$\begin{aligned}
\mathbb{P}[X_{t+h} - X_t \geqslant 1] &= \mathbb{P}[X_h \geqslant 1] = \mathbb{P}[S_1 \leqslant h] = 1 - e^{-\lambda h} = \lambda h + \mathcal{O}(h) \\
\mathbb{P}[X_{t+h} - X_t \geqslant 2] &= \mathbb{P}[X_h \geqslant 2] = \mathbb{P}[S_1 + S_2 \leqslant h] \\
&\leqslant \mathbb{P}[S_1 \leqslant h, S_2 \leqslant h] = \mathbb{P}[S_1 \leqslant h] \times \mathbb{P}[S_2 \leqslant h] \text{ by independence} \\
&= (1 - e^{-\lambda h})^2 = \mathcal{O}(\lambda h)
\end{aligned}$$

Therefore the probability $\mathbb{P}[X_{t+h} - X_t = 1]$ is $\lambda h + \mathcal{O}(h)$ uniformly in t. □

Proof: (2 ⇒ 3) If X has independent increments and the probabilities converge uniformly in the way described, X must have stationary increments. Thus we may simply prove $X_t \sim \text{Poisson}(\lambda t)$.

Since the increments of X are independent and X is increasing, we have

$$p_j(t+h) = \mathbb{P}[X_{t+h} = j] = \sum_{i=0}^{j} \mathbb{P}[X_t = j - i] \times \mathbb{P}[(X_{t+h} - X_t) = i]$$

For small h, we can assume $i = 0, 1$: the contribution from other terms is at most $\mathcal{O}(h)$. Thus

$$\begin{aligned}
p_j(t+h) = \mathbb{P}[X_t = j] &\times \mathbb{P}[(X_{t+h} - X_t) = 0] \\
&+ \mathbb{P}[X_t = j - 1] \times \mathbb{P}[(X_{t+h} - X_t) = 1] + \mathcal{O}(h)
\end{aligned}$$

Using the properties given by (2), we can simplify this to

$$p_j(t+h) = p_j(t)(1 - \lambda h) + p_{j-1}(t)\lambda h + \mathcal{O}(h)$$

which means as $h \downarrow 0$, we get $p_j'(t) = -\lambda p_j(t) + \lambda p_{j-1} t$. Differentiating $e^{\lambda t} p_j(t)$ yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{\lambda t} p_j(t)\right) = e^{\lambda t} p_j'(t) + \lambda e^{\lambda t} p_j(t) = \lambda e^{\lambda t} p_{j-1}(t)$$

For $j = 0$ we get $p_0(t+h) = p_0(t)(1 - \lambda h + \mathcal{O}(h))$, so $p_0'(t) = -\lambda p_0(t)$. This gives $p_0(t) = ce^{-\lambda x}$, which combined with the condition $p_0(0) = 0$ gives $p_0(t) = e^{-\lambda x}$.

Now by induction, we get $p_j(t) = e^{-\lambda t} \times (\lambda t)^j / j!$, so $X_t \sim \text{Poisson}(\lambda t)$ for all t as required. □

Proof: (3 ⇒ 1) Observe that (3) determines the finite-dimensional distributions of X. If the increments are independent and stationary, with $X_t \sim \text{Poisson}(\lambda t)$ for all t, then for $t_1 < \cdots < t_n$ and $k_1, \ldots, k_n$ we have

$$\begin{aligned}
\mathbb{P}[X_{t_1} = k_1, \ldots, X_{t_n} = k_n] &= \mathbb{P}[X_{t_1} = k_1, X_{t_2} - X_{t_1} = k_2 - k_1, \ldots, X_{t_n} X_{t_{n-1}} = k_n - k_{n-1}] \\
&= \mathbb{P}[X_{t_1} = k_1] \times \cdots \times \mathbb{P}[X_{t_n} - X_{t_{n-1}} = k_n - k_{n-1}] \\
&= \underbrace{\mathbb{P}[X_{t_1} = k_1]}_{\sim \text{Poisson}(\lambda t_1)} \times \cdots \times \underbrace{\mathbb{P}[X_{t_n - t_{n-1}} = k_n - k_{n-1}]}_{\sim \text{Poisson}(\lambda(t_n - t_{n-1}))}
\end{aligned}$$

which is therefore determinate. Since the finite-dimensional distributions of a right-continuous process define it uniquely, and we already know a process with these distributions, it must be the correct one, and therefore X must be a Poisson process. □

Proof: (overall) Therefore (1), (2), and (3) are equivalent definitions of the Poisson process, and so we may use properties from any of them in future. □

Theorem 1.12 (The Superposition Principle)

Suppose that $X \sim \text{PP}(\lambda)$ and $Y \sim \text{PP}(\mu)$ are independent. Then we may *superpose* the two to obtain a new process $Z = X + Y$ (which this jumps whenever either X or Y jump), which is also a Poisson Process with rate $\lambda + \mu$.

Proof: We use the fact that the increments $X_t \sim \text{Poisson}(\lambda t)$ and $Y_t \sim \text{Poisson}(\mu t)$ of the two processes are all independent. Therefore the increments $Z_t \sim (\text{Poisson}(\lambda t) + \text{Poisson}(\mu t))$, which we know is also a Poisson random variable with rate $\lambda + \mu$. Thus $Z \sim \text{PP}(\lambda + \mu)$. □

Note: We can also prove this by expanding the second definition of the Poisson Process.