2.2 Filtrations

Let $(\Omega, \mathcal{F})$ be a measureable space. A filtration in discrete time is a sequence of σ–algebras $\{\mathcal{F}_t\}$ such that

$$\mathcal{F}_t \subset \mathcal{F}$$

and

$$\mathcal{F}_t \subset \mathcal{F}_{t+1}$$

for all $t = 0, 1, \ldots$.

In continuous time, the second condition is replaced by

$$\mathcal{F}_s \subset \mathcal{F}_t$$

for all $s \leq t$.

3 Markov processes

The idea of a Markov process is to capture the idea of a short-memory stochastic process: once its current state is known, past history is irrelevant from the point of view of predicting its future.

Definition. Let $(\Omega, \mathcal{F})$ be a measurable space and let $(P, \underline{\mathcal{F}})$ be, respectively, a probability measure on and a filtration of this space. Let X be a stochastic process in discrete time on $(\Omega, \mathcal{F})$. Then X is called a $(P, \underline{\mathcal{F}})$-*Markov process* if

1. X is $\underline{\mathcal{F}}$–adapted, and

2. For each $t \in \mathbb{Z}_+$ and each Borel set $B \subset \mathcal{B}(\mathbb{R})$

$$P\left(X_{t+1} \in B | \mathcal{F}_t\right) = P\left(X_{t+1} \in B | \sigma\left(X_t\right)\right). \tag{9}$$

Sometimes when the probability measure and filtration are understood, we will talk simply of a Markov process.

Remark. Often the filtration $\underline{\mathcal{F}}$ is taken to be that generated by the process X itself.

Proposition. Let $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ be a filtered probability space and let X be a $(P, \underline{\mathcal{F}})$-Markov process. Let $t, k \in \mathbb{Z}_+$. Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel function such that $f\left(X_{t+k}\right)$ is integrable. Then,

$$E\left[f\left(X_{t+k}\right) | \mathcal{F}_t\right] = E\left[f\left(X_{t+k}\right) | \sigma\left(X_t\right)\right] \tag{10}$$

and hence there is a Borel function $g : \mathbb{Z}_+ \times \mathbb{Z}_+ \times \mathbb{R} \to \mathbb{R}$ such that, for each t,

$$E\left[f\left(X_{t+k}\right) | \mathcal{F}_t\right] = g\left(t, k, X_t\right). \tag{11}$$

Proof. To show it for $k = 1$, use that $f\left(X_{t+1}\right)$ is a $\sigma\left(X_{t+1}\right)$ −measurable random variable and hence is the limit of a monotone increasing sequence of $\sigma\left(X_{t+1}\right)$ −simple random variables. But such random variables are linear combinations of indicator functions of sets $X_{t+1}^{-1}(B)$ with B a Borel set. This completes the proof for $k = 1$. To prove it for arbitrary positive k, use induction. To prove it for $k + 1$ assuming it true for k, use the law of iterated expectations. ■

The vector case is a simple extension of the scalar case. However, it is important that the definition of a vector Markov process is *not* that each component is Markov.

Instead, we require that all the relevant (for the future of X) bits of information in $\mathcal{F}_t$ are in the $\sigma-$algebra generated by *all* the stochastic variables in X_t, i.e. $\sigma(X_t)$ is defined as the *single* $\sigma-$algebra $\sigma(\{X_{1,t}, X_{2,t}, ..., X_{n,t}\})$. This means that, for each Borel function $f : \mathbb{R}^n \to \mathbb{R}^m$,

$$E[f(X_{t+k})|\mathcal{F}_t] = E[f(X_{t+k})|\sigma(X_t)] \tag{12}$$

and hence that there is a Borel function $g : \mathbb{Z}_+ \times \mathbb{R}^n \to \mathbb{R}^m$ such that, for each t,

$$E[f(X_{t+1})|\mathcal{F}_t] = g(t, X_t). \tag{13}$$

(The case $k = 1$ is so important that we stress it here by ignoring greater values of k.)

3.0.1 Probability transition functions and time homogeneity

Definition. Let $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ be a filtered probability space and let X be a $(P, \underline{\mathcal{F}})$-Markov process. Then, for each $t = 0, 1, 2...$ its probability transition function $Q_t : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0, 1]$ is defined via

$$Q_t(x, B) = P(X_t \in B | X_{t-1} = x). \tag{14}$$

Note that any Markov process has a sequence of probability transition functions. Note also that for each fixed t and x, $Q_{t+1}(x, \cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R})$. Meanwhile, if we fix B, $Q_{t+1}(X_t(\cdot), B)$ is a random variable. Indeed, it is the conditional probability of $X_{t+1} \in B$ given X_t, i.e. $Q_{t+1}(X_t, B) = E\left[I_{X_{t+1}^{-1}(B)} \middle| \sigma(X_t)\right]$. Moreover, the conditional expectation of any $\sigma(X_{t+1})$-measurable random variable (given X_t) is an integral with respect to the measure Q_{t+1} in the following sense.

Proposition. Let $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ be a filtered probability space and let X be a $(P, \underline{\mathcal{F}})$-Markov process. Let $\langle Q_t \rangle$ be its probability transition functions and let $Z \in \mathcal{L}^1 (\Omega, \sigma (X_{t+1}), P)$. Then, for each $t = 0, 1, ...$

$$E[Z|X_t] = \int_{\mathbb{R}} f(y) Q_{t+1}(X_t, dy) \tag{15}$$

or, put differently, we have for each $t = 0, 1, ...$ and each x,

$$E[Z|X_t = x] = \int_{\mathbb{R}} f(y) Q_{t+1}(x, dy). \tag{16}$$

Proof.

We will show it first for an indicator variable $Z = I_{X_{t+1}^{-1}(A)}$ where $A \in \mathcal{B}(\mathbb{R})$. Then $f(y) = I_A(y)$. We now need to show that the random variable $\int_{\mathbb{R}} f(y) Q_{t+1}(X_t, dy)$ qualifies as the conditional expectation $E[Z|X_t]$. Clearly it is $\sigma(X_t)-$measurable. But does it integrate to the right thing? Well, let $G \in \sigma(X_t)$ and recall that, by definition, $Q_{t+1}(X_t, A) = E\left(I_{X_{t+1}^{-1}(A)}|\sigma(X_t)\right)$. Hence

$$\int_G \int_{\mathbb{R}} f(y) Q_{t+1}(X_t, dy) P(d\omega) = \int_G \int_{\mathbb{R}} I_A(y) Q_{t+1}(X_t, dy) P(d\omega) =$$

$$= \int_G Q_{t+1}(X_t, A) P(d\omega) = \int_G E\left(I_{X_{t+1}^{-1}(A)}|\sigma(X_t)\right) P(d\omega) = \int_G I_{X_{t+1}^{-1}(A)} P(d\omega). \tag{17}$$

Meanwhile, since $Z = I_{X_{t+1}^{-1}(A)}$ we obviously have

$$\int_G Z P(d\omega) = \int_G I_{X_{t+1}^{-1}(A)} P(d\omega). \tag{18}$$

To show the theorem for an arbitrary $Z \in \mathcal{L}^1(\Omega, \sigma(X_{t+1}), P)$, use the Monotone Convergence Theorem. ■

We now use the probability transition function to define a *time homogeneous* Markov process.

Definition. Let $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ be a filtered probability space and let X be a $(P, \underline{\mathcal{F}})$-Markov process. Let $\langle Q_t \rangle_{t=1}^{\infty}$ be its probability transition functions. If there is a Q such that $Q_t = Q$ for all $t = 1, 2, ...$ then X is called a *time homogeneous* Markov process.

Proposition. Let $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ be a filtered probability space and let X be a time homogeneous $(P, \underline{\mathcal{F}})$-Markov process. For any nonnegative integers k, t, let $Y_{t+k} \in \mathcal{L}^1(\Omega, \sigma(X_{t+k}), P)$. Then for each $k = 0, 1, ...$ there is a Borel function $g_k : \mathbb{R} \to \mathbb{R}$ such that, for each $t = 0, 1, ...$

$$E[Y_{t+k} | \mathcal{F}_t] = g_k(X_t). \tag{19}$$

In particular, there is a Borel function h such that, for each $t = 0, 1, ...$

$$E[Y_{t+1} | \mathcal{F}_t] = h(X_t). \tag{20}$$

3.1 Finite–state Markov chains in discrete time

This is perhaps the simplest class of Markov processes. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ be a finite set. $X : \mathbb{Z}_+ \to \mathcal{X}$ be a stochastic process. Denote by μ_t the vector of probabilities that $X_t = x_i$ and suppose there is a sequence of matrices Γ_t such that

$$\mu_{t+1} = \Gamma_t \mu_t. \tag{21}$$

Then X is said to be a finite–state Markov chain in discrete time. Call Γ_t the probability transition matrix.

The interpretation of (21) is the following.

$$P(X_{t+1} = x_i | X_t = x_j) = \Gamma_t(i, j).$$

If $\Gamma_t = \Gamma$ we call the process and **time–homogenenous** or **stationary**.

If Γ is sufficiently well–behaved, then X has a unique stationary distribution.

Definition. Let X be a stationary finite–state Markov chain and let T be the time of the first visit to state j after $t = 0$. Then state j is called **recurrent** (opposite: **transient**) if

$$P(\{T < \infty\} | X_0 = x_j) = 1.$$

Definition. The j is called **periodic** with period $\delta > 1$ if δ is the largest integer for which

$$P(\{T = n\delta \text{ for some } n \geq 1\} | X_0 = j) = 1.$$

If there is no such $\delta > 1$, then j is called **aperiodic**.

Definition. A Markov chain is said to be aperiodic if all its states are aperiodic.

Definition. A state j can be reached from i if there exists an integer $n \geq 0$ such that

$$\Gamma^n(i, j) > 0.$$

where by $\Gamma^n(i,j)$ we mean the (i,j) element of the matrix Γ^n.

Definition. A set of states is said to be **closed** if no state outside it can be reached from any state in it.

Definition. A set of states is said to be **ergodic** if it is closed and no proper subset is closed.

Definition. A Markov chain is called **irreducible** if its only closed set is the set of all states.

Theorem. Let X be a finite state stationary Markov chain with transition matrix Γ and suppose X irreducible and aperiodic. Then

$$\begin{cases} \mu = \Gamma\mu \\ \mu \cdot \mathbf{1} = 1 \end{cases}$$

has a unique solution μ^* and this μ^* has the property that

$$\lim_{t\to\infty} \Gamma^t \mu_0 = \mu^*$$

for all μ_0 such that $\mu \cdot \mathbf{1} = 1$.

Proof. See Cinlar (1975). ■

3.2 Finite–state Markov chains in continuous time

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ be a finite set. Let $X : \mathbb{R}_+ \to \mathcal{X}$ be a stochastic process. Denote by $\mu(t)$ the vector of probabilities that $X(t) = x_i$ and suppose there is a matrix-valued function $\Gamma(t)$ such that $\mu(t)$ satisfies

$$\dot{\mu}(t) = \Gamma(t)\mu(t).$$

Then we call X a finite-state continuous time Markov chain and If $\Gamma(t) = \Gamma$ we call the process and **time–homogenenous** or **stationary**.

3.3 Poisson processes

3.3.1 The Poisson distribution

Intuitively, the Poisson distribution comes from taking the limit of a sum of Bernoulli random variables.

Definition. A random variable X is said to be **Bernoulli** if there is a real number $0 \leq p \leq 1$ such that

$$P(\{X = 1\}) = p$$

and

$$P(\{X = 0\}) = 1 - p.$$

Definition. A random variable Y is said to be **binomial** distribution if there is an integer n and a real number $0 \leq p \leq 1$ such that

$$P(\{Y = k\}) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

where the **binomial coefficient** is defined as follows.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Proposition. If $\{X_i,\, i = 1, \ldots, n\}$ is a collection of independent Bernoulli random variables, then Y defined via

$$Y = \sum_{i=1}^{n} X_i$$

is binomially distributed.

Proposition. Let X be binomial with parameters n, p. Then

$$E[X] = np.$$

Proof. Exercise. ■

Now imagine that we are on a fishing expedition. For some reason we dip the fishing pole into the water 10 times for one minute at a time. The probability of catching a fish during any one dipping is p, independently of whether I caught a fish in any previous dipping. Then the total number of fish caught is a binomial variable.

But what if I dip 20 times for half a minute, or 40 times for 15 seconds etc. Assume that the probability of catching a fish during a half–minute dip is $p/2$ and similarly for shorter dips. What happens in the limit? As we take limits, let the expected total number of fish caught $\lambda = np$ be constant and let $n \to \infty$. (It follows that $p \to 0$.) Then it turns out that the distribution of the total number of fish caught tends to the Poisson distribution with parameter λ.

Definition. A random variable X is said to be **Poisson** distributed if there is a real number $\lambda \geq 0$ such that

$$P(\{X = k\}) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Proposition Suppose two independent random variables X and Y are Poisson distributed with parameters λ and μ, respectively. Then $Z = X + Y$ is Poisson distributed with parameter $\lambda + \mu$. **Proof.** Use the characteristic function. ■

Definition. A continuous time stochastic process is a function $X(t)$ where for each fixed $t \geq 0$, $X(t)$ is a random variable.

Definition. Let $(\Omega, \mathcal{F}, P, \underline{\mathcal{F}})$ be a filtered probability space. A stochastic process $\{N(t, \omega); t \geq 0\}$ is said to be a $(P, \underline{\mathcal{F}})$–Poisson process with intensity λ if

- N is $\underline{\mathcal{F}}$–adapted.
- The trajectories of N are (with probability one) right continuous and piecewise continuous.
- $N(0) = 0$.
- $\Delta N(t) = 0$ or 1 (with probability one) where

$$\Delta N(t) = N(t) - N(t-).$$

- For all $s \leq t$, $N(t) - N(s)$ is independent of $\mathcal{F}_s$.
- $N(t) - N(s)$ is Poisson distributed with parameter $\lambda(t - s)$, that is

$$P\left(N(t) - N(s) = k \,|\, \mathcal{F}_s\right) = P\left(N(t) - N(s) = k\right) = e^{-\lambda(t-s)} \frac{\lambda^k (t-s)^k}{k!}.$$

Proposition. The time between jumps is is exponentially distributed. More precisely, let τ be defined via

$$\tau = \inf_{t \geq 0} \{N(t) > 0\}.$$

Then

$$P(\{\tau < t\}) = F(t) = 1 - e^{-\lambda t}$$

and the corresponding probability density function is

$$f(t) = F'(t) = \lambda e^{-\lambda t}.$$

Proof. Exercise. ■

References

Chung, K. L. (2001). *A Course in Probability Theory, Third Edition.* Academic Press.

Cinlar, E. (1975). *Introduction to Stochastic Processes.* Prentice–Hall.

Williams, D. (1991). *Probability with Martingales.* Cambridge University Press.