

Lecture Note 9: Introduction to Stochastic Processes

These notes are based on S. Ross, *Introduction to Probability Models*, Academic Press, and J. Hamilton, *Time Series Analysis*, Princeton University Press.

Definition 1 A stochastic process $\{X(t), t \in T\}$ is a collection of random variables: for each $t \in T$, $X(t)$ is a random variable.

The set T is called an index set, and $t \in T$ is often interpreted as *time*. So $X(t)$ would be the (random) state of a process at time t . Sometimes for simplicity we write $X_t = X(t)$.

Note that each X_t is a random variable, so really we should be writing $X_t(\omega)$. Each X_t is a function from the sample space Ω to a subset of \mathbb{R} . In applications we often are interested in modelling the evolution of some variable over time, so it is reasonable that the range of X_t is the same across time. In that case we call the range of X_t the state space.

If the index set T is a countable set, we call the stochastic process a discrete-time process. If the index set is an interval of the real line, we call the process a continuous-time process.

Although t is often used to indicate time, it can be used in other ways as well. For example, when modelling *spatial* phenomena (e.g. geographical concentrations of pollution), we might use a two-dimensional index t corresponding to longitude and latitude.

Example 1: Consider flipping a coin forever (or for a very long time...). The sample space Ω would contain every possible infinite sequence of H s and T s. We could define the index set as $T = \{1, 2, 3, \dots\}$ and $X_1 = 1$ if the first toss is heads and 0 otherwise, $X_2 = 1$ if the second toss is heads and zero otherwise, and so on. This defines a stochastic process $\{X_t, t \in T\}$, where X_s is independent of X_r for $s \neq r$.

Next, we could define a new stochastic process $\{Y_t, t \in T\}$, where Y_t is the total number of heads up to that point in time: $Y_t = \sum_{i=1}^t X_i$. Now there is a very distinct dependence between, say, Y_s and Y_{s+1} . We could also consider another stochastic process $\{Z_t, t \in T\}$ where

$$Z_t = \frac{Y_t}{t} = \frac{1}{t} \sum_{i=1}^t X_i.$$

Since Z_t is the average of X_t up to that point in time, we might think that Z_t would converge to $1/2$ as t increases. We will come back to this idea in later lecture notes.

Markov Chains

Suppose that $\{X_t, t = 1, 2, 3, \dots\}$ is a discrete-time stochastic process with a finite or countable state space. That is, X_t takes on a finite or countable number of possible values, and for simplicity let us say that the range is actually the nonnegative integers $0, 1, 2, 3, \dots$

We can fully specify the joint probability distribution of these random variables by starting with the marginal probabilities

$$Pr(X_1 = i) = f_{X_1}(i), \quad i = 0, 1, 2, \dots$$

and then defining various conditional probabilities recursively:

$$Pr(X_2 = j | X_1 = i)$$

$$Pr(X_3 = j | X_2 = i, X_1 = k)$$

and so on. Suppose that the conditional probabilities have a simple form, depending only on the most recent past random variable:

$$Pr(X_{t+1} = j | X_t = i, X_{t-1} = k_{t-1}, \dots, X_1 = k_1) = Pr(X_{t+1} = j | X_t = i) = P_{ij}, \quad \forall t = 1, 2, 3, \dots$$

We call such a stochastic process a Markov chain.

The numbers P_{ij} represent the transition probability, the probability of going from state i to state j . They must be nonnegative: $P_{ij} \geq 0$ for all i and j , and for each i , we must have $\sum_{j=0}^{\infty} P_{ij} = 1$. It's handy to collect these into a matrix:

$$P = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & & \end{bmatrix}.$$

Example 1 continued For independent coin flips, the possible values are 0 and 1, and the transition matrix would be

$$P = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}.$$

Next, consider Y_t , the cumulative sum of heads. For any time t , Y_{t+1} can either be equal to Y_t (with probability 1/2) or $Y_t + 1$ (with probability 1/2). Thus $P_{ij} = .5$ for $j = i, i + 1$ and 0 otherwise. \square

Example 2: The paper “Intrafirm Mobility and Sex Differences in Pay,” by Michael Ransom and Ronald Oaxaca, studies employment records for a major grocery store. They construct transition probabilities of transitions between different job categories, separately for males and females. For example, male produce clerks have a .17 probability of being terminated the following year, a .65 probability of remaining in their position, a .04 probability of being promoted to produce manager, and various other probabilities of shifting position within the firm. A female produce clerk has a .25 probability of being terminated, a .375 probability of remaining as a produce clerk, and so on. \square

Next, we can calculate the transition probabilities more than one step ahead: let the m -step transition probabilities be:

$$P_{ij}^m = Pr(X_{t+m} = j | X_t = i), \quad t \geq 0, i, j \geq 0.$$

Result 1 For a Markov chain, and for any $n, m \geq 0$, the $n+m$ step transition probabilities are related to the n and the m step transition probabilities by:

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m. \quad (1)$$

Proof: The intuition is that the right hand side represents the probability of starting at state i , passing through state k at the n th transition, and then going to j after another m

transitions. Formally:

$$\begin{aligned}
P_{ij}^{m+n} &= Pr(X_{m+n} = j | X_0 = i) \\
&= \sum_{k=0}^{\infty} Pr(X_{m+n} = j, X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} Pr(X_{m+n} = j | X_n = k, X_0 = i) Pr(X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} Pr(X_{m+n} = j | X_n = k) Pr(X_n = k | X_0 = i) \quad (\text{by the Markov property}) \\
&= \sum_{k=0}^{\infty} P_{kj}^m P_{ik}^n
\end{aligned}$$

□

The equations (1) are often referred to as the Chapman-Kolmogorov equations. They are particularly easy to write in matrix form. If $P^{(n)}$ denotes the matrix of n step transition probabilities, then (1) can be written as

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)}.$$

where the multiplication is the usual matrix multiplication. Thus,

$$P^{(2)} = P \cdot P, \quad P^{(3)} = P \cdot P \cdot P,$$

and in general:

$$P^{(n)} = P^n.$$

In some cases, as $n \rightarrow \infty$, the P^n converge to a constant matrix. The following theorem is one of the key results in Markov chain theory:

Theorem 1 *Suppose that a Markov chain with transition matrix P satisfies:*

1. *Aperiodicity: for all i , $P_{ii} > 0$.*
2. *Positive recurrence: starting in any state i , the expected time to return to state i is finite.*
3. *Irreducibility: for all states i and j , there is some n such that $P_{ij}^n > 0$. Thus, state j is “accessible” from state i .*

Then, $\lim_{n \rightarrow \infty} P_{ij}^n$ exists and is independent of i . Furthermore, letting

$$\pi_j = \lim_{n \rightarrow \infty} P_{ij}^n,$$

the limiting probabilities are the unique nonnegative solution to

$$\begin{aligned}
\pi_j &= \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad j \geq 0, \\
\sum_{j=0}^{\infty} \pi_j &= 1.
\end{aligned}$$

Notice that if we start with probabilities π_i of being in state i , then at the next step we have the same probabilities of being in any given state. Thus, the vector π could be thought of as an “invariant” or “steady-state” distribution over the states.

Also: π_j is interpreted as the limiting probability that the process is in state j at time n . It can be shown that π_j also equal the long-run proportion of times that the process is in state j .

Poisson Processes

Next, we will consider a simple but important example of a continuous-time process that is used to model the total number of events up to different points in time. In a sense, we have already discussed counting processes, especially in connection with the exponential and Poisson distribution, so some of the following will be familiar.

Definition 2 A counting process is a stochastic process $\{N(t), t \geq 0\}$ with index set $[0, \infty)$ such that:

1. $N(t) \geq 0$.
2. $N(t)$ is integer-valued.
3. if $s < t$, then $N(s) \leq N(t)$.
4. for $s < t$, $N(t) - N(s)$ equals the number of events that have occurred in the interval (s, t) .

The Poisson process is a special kind of counting process:

Definition 3 A counting process $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda > 0$ if

1. $N(0)=0$.
2. If (s, t) and (a, b) are disjoint intervals of time, then $N(t) - N(s)$ is independent of $N(b) - N(a)$.
3. The number of events in any interval of length t is Poisson with mean λt : for any $s, t \geq 0$,

$$Pr(N(s+t) - N(s) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad \text{for } n = 0, 1, \dots$$

This should look very familiar. In Lecture Note 5, we considered such a process, and showed that the time to the next event is an exponentially distributed random variable with parameter λ .

Autoregressive and Moving Average Processes

The Poisson process is a continuous-time, discrete state space stochastic process. Next, let us consider discrete-time, continuous state space processes. Here is a particularly simple one:

Definition 4 A discrete-time stochastic process $\{\epsilon_t, t = 1, 2, \dots\}$ is a white noise process if:

1. for all t , $E[\epsilon_t] = 0$.
2. for all t , $V[\epsilon_t] = \sigma^2$.
3. for all t, s such that $t \neq s$, $Cov(\epsilon_t, \epsilon_s) = 0$.

A white noise process has constant mean 0 and constant variance, and zero correlation across time periods. The terminology comes from signal processing and Fourier analysis, where this type of process was used to describe excess noise that attached to a signal at all frequencies.

Often we work with a particularly convenient type of white noise process, the Gaussian white noise, where

$$\epsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2).$$

The notation “iid” stands for “independent and identically distributed,” and $\stackrel{\text{iid}}{\sim}$ means that the random variables are independent with the same $N(0, \sigma^2)$ distribution. (The term “Gaussian” is another name for the normal distribution.)

The Gaussian white noise process exhibits no dependence across observations. However, we can use this process as a building block to construct many interesting processes with various types of dependence.

1. First-order autoregression: let Y_1 have some distribution F_1 , and let

$$Y_t = \alpha + \rho Y_{t-1} + \epsilon_t.$$

The term “autoregression” deserves some explanation. In general, a “regression” is a model describing how the mean of a variable depends on another variable.¹ An “autoregression” means that we are explaining the mean of Y_t in terms of its own past value. Since the white noise term has mean zero,

$$E[Y_t | Y_{t-1}, Y_{t-2}, \dots] = \alpha + \rho Y_{t-1}.$$

In the special case that $\alpha = 0, \rho = 1$, we get

$$Y_t = Y_{t-1} + \epsilon_t.$$

This is a type of random walk, where the current value of the process is the best (squared error) predictor of next period’s value.

2. p th-order autoregression ($AR(p)$):

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t.$$

¹Regression analysis is the main focus of Economics 522A.

3. q th-order Moving average ($MA(q)$):

$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}.$$

Thus, the Y_t is a weighted average of $\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-q}$.

4. $ARMA(p, q)$ process:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.$$

A key issue with these processes is whether they are in some sense “stable.” One notion of stability is covariance stationarity:

Definition 5 A discrete time stochastic process $\{Y_t, t = 1, 2, \dots\}$ is said to be covariance stationary or weakly stationary if

1. $E[Y_t] = \mu$ for all t .
2. $E[(Y_t - \mu)(Y_{t-j} - \mu)] = \gamma_j$ for all t and all j .

This says that the mean is constant, the variance is constant, and the covariances only depend on the distance between the two points in time.

It's easy to see that any white noise process is covariance stationary: the mean is constant at 0, and

$$E[(Y_t - 0)(Y_{t-j} - 0)] = \begin{cases} \sigma^2 & j = 0 \\ 0 & j \neq 0 \end{cases}$$

What about the AR(1) process $Y_t = \alpha + \rho Y_{t-1} + \epsilon_t$? It turns out that if

$$-1 < \rho < 1,$$

then the process is covariance stationary. The random walk, where $\rho = 1$, is not stationary.