

16.1: Introduction to Markov Processes

A *Markov process* is a random process indexed by time, and with the property that the future is independent of the past, given the present. Markov processes, named for Andrei Markov, are among the most important of all random processes. In a sense, they are the stochastic analogs of [differential equations and recurrence relations](#), which are of course, among the most important deterministic processes.

The complexity of the theory of Markov processes depends greatly on whether the time space T is \mathbb{N} (discrete time) or $[0, \infty)$ (continuous time) and whether the state space is *discrete* (countable, with all subsets measurable) or a more general topological space. When $T = [0, \infty)$ or when the state space is a general space, continuity assumptions usually need to be imposed in order to rule out various types of weird behavior that would otherwise complicate the theory.

When the state space is discrete, Markov processes are known as *Markov chains*. The general theory of Markov chains is mathematically rich and relatively simple.

- When $T = \mathbb{N}$ and the state space is discrete, Markov processes are known as *discrete-time Markov chains*. The theory of such processes is mathematically elegant and complete, and is understandable with minimal reliance on measure theory. Indeed, the main tools are basic probability and linear algebra. Discrete-time Markov chains are studied in this chapter, along with a number of special models.
- When $T = [0, \infty)$ and the state space is discrete, Markov processes are known as *continuous-time Markov chains*. If we avoid a few technical difficulties (created, as always, by the continuous time space), the theory of these processes is also reasonably simple and mathematically very nice. The Markov property implies that the process, sampled at the random times when the state changes, forms an embedded *discrete-time* Markov chain, so we can apply the theory that we will have already learned. The Markov property also implies that the holding time in a state has the memoryless property and thus must have an exponential distribution, a distribution that we know well. In terms of what you may have already studied, the Poisson process is a simple example of a continuous-time Markov chain.

For a general state space, the theory is more complicated and technical, as noted above. However, we can distinguish a couple of classes of Markov processes, depending again on whether the time space is discrete or continuous.

- When $T = \mathbb{N}$ and $S = \mathbb{R}$, a simple example of a Markov process is the partial sum process associated with a sequence of independent, identically distributed real-valued random variables. Such sequences are studied in the chapter on random samples (but not as Markov processes), and revisited [below](#).
- In the case that $T = [0, \infty)$ and $S = \mathbb{R}$ or more generally $S = \mathbb{R}^k$, the most important Markov processes are the *diffusion processes*. Generally, such processes can be constructed via stochastic differential equations from Brownian motion, which thus serves as the quintessential example of a Markov process in continuous time and space.

The goal of this section is to give a broad sketch of the general theory of Markov processes. Some of the statements are not completely rigorous and some of the proofs are omitted or are sketches, because we want to emphasize the main ideas without getting bogged down in technicalities. If you are a new student of probability you may want to just browse this section, to get the basic ideas and notation, but skipping over the proofs and technical details. Then jump ahead to the study of discrete-time Markov chains. On the other hand, to understand this section in more depth, you will need to review topics in the chapter on foundations and in the chapter on stochastic processes.

Basic Theory

Preliminaries

As usual, our starting point is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, so that Ω is the set of outcomes, \mathcal{F} the σ -algebra of events, and \mathbb{P} the probability measure on (Ω, \mathcal{F}) . The time set T is either \mathbb{N} (discrete time) or $[0, \infty)$ (continuous time). In the first case, T is given the discrete topology and in the second case T is given the usual Euclidean topology. In both cases, T is given the Borel σ -algebra \mathcal{T} , the σ -algebra generated by the open sets. In the discrete case when $T = \mathbb{N}$, this is simply the power set of T so that every subset of T is measurable; every function from T to another measurable space is measurable; and every function from T to another topological space is continuous. The time space (T, \mathcal{T}) has a natural measure; counting measure $\#$ in the discrete case, and Lebesgue in the continuous case.

The set of states S also has a σ -algebra \mathcal{S} of admissible subsets, so that (S, \mathcal{S}) is the *state space*. Usually S has a topology and \mathcal{S} is the Borel σ -algebra generated by the open sets. A typical set of assumptions is that the topology on S is *LCCB*: locally compact, Hausdorff, and with a countable base. These particular assumptions are general enough to capture all of the most important processes that occur in applications and yet are restrictive enough for a nice mathematical theory. Usually, there is a natural positive measure λ on the state space (S, \mathcal{S}) . When S has an LCCB topology and \mathcal{S} is the Borel σ -algebra, the measure λ will usually be a *Borel measure* satisfying $\lambda(C) < \infty$ if $C \subseteq S$ is compact. The term *discrete state space* means that S is countable with $\mathcal{S} = \mathcal{P}(S)$, the collection of all subsets of S . Thus every subset of S is measurable, as is every function from S to another measurable space. This is the Borel σ -algebra for the discrete topology on S , so that every function from S to another topological space is continuous. The compact sets are simply the finite sets, and the reference measure is $\#$, counting measure. If $S = \mathbb{R}^k$ for some $k \in \mathbb{N}$ (another common case), then we usually give S the Euclidean topology (which is LCCB) so that \mathcal{S} is the usual Borel σ -algebra. The compact sets are the closed, bounded sets, and the reference measure λ is k -dimensional Lebesgue measure.

Clearly, the topological and measure structures on T are not really necessary when $T = \mathbb{N}$, and similarly these structures on S are not necessary when S is countable. But the main point is that the assumptions unify the discrete and the common continuous cases. Also, it should be noted that much more general state spaces (and more general time spaces) are possible, but most of the important Markov processes that occur in applications fit the setting we have described here.

Various spaces of real-valued functions on S play an important role. Let \mathcal{B} denote the collection of bounded, measurable functions $f : S \rightarrow \mathbb{R}$. With the usual (pointwise) addition and scalar multiplication, \mathcal{B} is a vector space. We give \mathcal{B} the supremum norm, defined by $\|f\| = \sup\{|f(x)| : x \in S\}$.

Suppose now that $\mathbf{X} = \{X_t : t \in T\}$ is a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ with state space S and time space T . Thus, X_t is a random variable taking values in S for each $t \in T$, and we think of $X_t \in S$ as the *state* of a system at *time* $t \in T$. We also assume that we have a collection $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ of σ -algebras with the properties that X_t is measurable with respect to \mathcal{F}_t for $t \in T$, and the $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s, t \in T$ with $s \leq t$. Intuitively, \mathcal{F}_t is the collection of event up to time $t \in T$. Technically, the assumptions mean that \mathfrak{F} is a filtration and that the process \mathbf{X} is *adapted* to \mathfrak{F} . The most basic (and coarsest) filtration is the *natural filtration* $\mathfrak{F}^0 = \{\mathcal{F}_t^0 : t \in T\}$ where $\mathcal{F}_t^0 = \sigma\{X_s : s \in T, s \leq t\}$, the σ -algebra generated by the process up to time $t \in T$. In continuous time, however, it is often necessary to use slightly finer σ -algebras in order to have a nice mathematical theory. In particular, we often need to assume that the filtration \mathfrak{F} is *right continuous* in the sense that $\mathcal{F}_{t+} = \mathcal{F}_t$ for $t \in T$ where $\mathcal{F}_{t+} = \bigcap\{\mathcal{F}_s : s \in T, s > t\}$. We can accomplish this by taking $\mathfrak{F} = \mathfrak{F}_+^0$ so that $\mathcal{F}_t = \mathcal{F}_{t+}^0$ for $t \in T$, and in this case, \mathfrak{F} is referred to as the *right continuous refinement* of the natural filtration. We also sometimes need to assume that \mathfrak{F} is *complete* with respect to \mathbb{P} in the sense that if $A \in \mathcal{S}$ with $\mathbb{P}(A) = 0$ and $B \subseteq A$ then $B \in \mathcal{F}_0$. That is, \mathcal{F}_0 contains all of the *null events* (and hence also all of the *almost certain* events), and therefore so does \mathcal{F}_t for all $t \in T$.

Definitions

The random process \mathbf{X} is a *Markov process* if

$$\mathbb{P}(X_{s+t} \in A \mid \mathcal{F}_s) = \mathbb{P}(X_{s+t} \in A \mid X_s) \quad (16.1.1)$$

for all $s, t \in T$ and $A \in \mathcal{S}$.

The defining condition, known appropriately enough as the *Markov property*, states that the conditional distribution of X_{s+t} given \mathcal{F}_s is the same as the conditional distribution of X_{s+t} just given X_s . Think of s as the present time, so that $s+t$ is a time in the future. If we know the present state X_s , then any additional knowledge of events in the past is irrelevant in terms of predicting the future state X_{s+t} . Technically, the conditional probabilities in the definition are random variables, and the equality must be interpreted as holding with probability 1. As you may recall, conditional expected value is a more general and useful concept than conditional probability, so the following theorem may come as no surprise.

The random process \mathbf{X} is a Markov process if and only if

$$\mathbb{E}[f(X_{s+t}) \mid \mathcal{F}_s] = \mathbb{E}[f(X_{s+t}) \mid X_s] \quad (16.1.2)$$

for every $s, t \in T$ and every $f \in \mathcal{B}$.

Proof sketch

The condition in this theorem clearly implies the Markov property, by letting $f = \mathbf{1}_A$, the indicator function of $A \in \mathcal{S}$. The converse is a classical bootstrapping argument: the Markov property implies the expected value condition

1. First when $f = \mathbf{1}_A$ for $A \in \mathcal{S}$ (by definition).
2. Next when $f \in \mathcal{B}$ is a simple function, by linearity.
3. Next when $f \in \mathcal{B}$ is nonnegative, by the monotone convergence theorem.
4. Finally for general $f \in \mathcal{B}$ by considering positive and negative parts.

Technically, we should say that \mathbf{X} is a Markov process *relative to the filtration* \mathfrak{F} . If \mathbf{X} satisfies the Markov property relative to a filtration, then it satisfies the Markov property relative to any coarser filtration.

Suppose that the stochastic process $\mathbf{X} = \{X_t : t \in T\}$ is adapted to the filtration $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ and that $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$ is a filtration that is finer than \mathfrak{F} . If \mathbf{X} is a Markov process relative to \mathfrak{G} then \mathbf{X} is a Markov process relative to \mathfrak{F} .

Proof

First recall that \mathbf{X} is adapted to \mathfrak{G} since \mathbf{X} is adapted to \mathfrak{F} . If $s, t \in T$ and $f \in \mathcal{B}$ then

$$\mathbb{E}[f(X_{s+t}) \mid \mathcal{F}_s] = \mathbb{E}(\mathbb{E}[f(X_{s+t}) \mid \mathcal{G}_s] \mid \mathcal{F}_s) = \mathbb{E}(\mathbb{E}[f(X_{s+t}) \mid X_s] \mid \mathcal{F}_s) = \mathbb{E}[f(X_{s+t}) \mid X_s] \quad (16.1.3)$$

The first equality is a basic property of conditional expected value. The second uses the fact that \mathbf{X} is Markov relative to \mathfrak{G} , and the third follows since X_s is measurable with respect to \mathcal{F}_s .

In particular, if \mathbf{X} is a Markov process, then \mathbf{X} satisfies the Markov property relative to the natural filtration \mathfrak{F}^0 . The theory of Markov processes is simplified considerably if we add an additional assumption.

A Markov process \mathbf{X} is *time homogeneous* if

$$\mathbb{P}(X_{s+t} \in A \mid X_s = x) = \mathbb{P}(X_t \in A \mid X_0 = x) \quad (16.1.4)$$

for every $s, t \in T$, $x \in S$ and $A \in \mathcal{S}$.

So if \mathbf{X} is homogeneous (we usually don't bother with the *time* adjective), then the process $\{X_{s+t} : t \in T\}$ given $X_s = x$ is equivalent (in distribution) to the process $\{X_t : t \in T\}$ given $X_0 = x$. For this reason, the initial distribution is often unspecified in the study of Markov processes—if the process

is in state $x \in S$ at a particular time $s \in T$, then it doesn't really matter how the process got to state x ; the process essentially “starts over”, independently of the past. The term *stationary* is sometimes used instead of homogeneous.

From now on, we will usually assume that our Markov processes are homogeneous. This is not as big of a loss of generality as you might think. A non-homogeneous process can be turned into a homogeneous process by enlarging the state space, as shown [below](#). For a homogeneous Markov process, if $s, t \in T$, $x \in S$, and $f \in \mathcal{B}$, then

$$\mathbb{E}[f(X_{s+t}) \mid X_s = x] = \mathbb{E}[f(X_t) \mid X_0 = x] \quad (16.1.5)$$

Feller Processes

In continuous time, or with general state spaces, Markov processes can be very strange without additional continuity assumptions. Suppose (as is usually the case) that S has an LCCB topology and that \mathcal{S} is the Borel σ -algebra. Let \mathcal{C} denote the collection of bounded, continuous functions $f : S \rightarrow \mathbb{R}$. Let \mathcal{C}_0 denote the collection of continuous functions $f : S \rightarrow \mathbb{R}$ that *vanish at ∞* . The last phrase means that for every $\epsilon > 0$, there exists a compact set $C \subseteq S$ such that $|f(x)| < \epsilon$ if $x \notin C$. With the usual (pointwise) operations of addition and scalar multiplication, \mathcal{C}_0 is a vector subspace of \mathcal{C} , which in turn is a vector subspace of \mathcal{B} . Just as with \mathcal{B} , the supremum norm is used for \mathcal{C} and \mathcal{C}_0 .

A Markov process $\mathbf{X} = \{X_t : t \in T\}$ is a *Feller process* if the following conditions are satisfied.

1. *Continuity in space*: For $t \in T$ and $y \in S$, the distribution of X_t given $X_0 = x$ converges to the distribution of X_t given $X_0 = y$ as $x \rightarrow y$.
2. *Continuity in time*: Given $X_0 = x$ for $x \in S$, X_t converges in probability to x as $t \downarrow 0$.

Additional details

1. This means that $\mathbb{E}[f(X_t) \mid X_0 = x] \rightarrow \mathbb{E}[f(X_t) \mid X_0 = y]$ as $x \rightarrow y$ for every $f \in \mathcal{C}$.
2. This means that $\mathbb{P}[X_t \in U \mid X_0 = x] \rightarrow 1$ as $t \downarrow 0$ for every neighborhood U of x .

Feller processes are named for William Feller. Note that if S is discrete, (a) is automatically satisfied and if T is discrete, (b) is automatically satisfied. In particular, every discrete-time Markov chain is a Feller Markov process. There are certainly more general Markov processes, but most of the important processes that occur in applications are Feller processes, and a number of nice properties flow from the assumptions. Here is the first:

If $\mathbf{X} = \{X_t : t \in T\}$ is a Feller process, then there is a version of \mathbf{X} such that $t \mapsto X_t(\omega)$ is continuous from the right and has left limits for every $\omega \in \Omega$.

Again, this result is only interesting in continuous time $T = [0, \infty)$. Recall that for $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is a *sample path* of the process. So we will often assume that a Feller Markov process has sample paths that are right continuous have left limits, since we know there is a version with these properties.

Stopping Times and the Strong Markov Property

For our next discussion, you may need to review again the section on filtrations and stopping times. To give a quick review, suppose again that we start with our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the filtration $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ (so that we have a *filtered probability space*).

Since time (past, present, future) plays such a fundamental role in Markov processes, it should come as no surprise that random times are important. We often need to allow random times to take the value ∞ , so we need to enlarge the set of times to $T_\infty = T \cup \{\infty\}$. The topology on T is extended to T_∞ by the rule that for $s \in T$, the set $\{t \in T_\infty : t > s\}$ is an open neighborhood of ∞ . This is the one-point compactification of T and is used so that the notion of *time converging to infinity* is preserved. The Borel σ -algebra \mathcal{T}_∞ is used on T_∞ , which again is just the power set in the discrete case.

If $\mathbf{X} = \{X_t : t \in T\}$ is a stochastic process on the sample space (Ω, \mathcal{F}) , and if τ is a random time, then naturally we want to consider the state X_τ at the random time. There are two problems. First if τ takes the value ∞ , X_τ is not defined. The usual solution is to add a new “death state” δ to the set of states S , and then to give $S_\delta = S \cup \{\delta\}$ the σ algebra $\mathcal{S}_\delta = \mathcal{S} \cup \{A \cup \{\delta\} : A \in \mathcal{S}\}$. A function $f \in \mathcal{B}$ is extended to S_δ by the rule $f(\delta) = 0$. The second problem is that X_τ may not be a valid random variable (that is, measurable) unless we assume that the stochastic process \mathbf{X} is measurable. Recall that this means that $\mathbf{X} : \Omega \times T \rightarrow S$ is measurable relative to $\mathcal{F} \otimes \mathcal{T}$ and \mathcal{S} . (This is always true in discrete time.)

Recall next that a random time τ is a stopping time (also called a *Markov time* or an *optional time*) relative to \mathfrak{F} if $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \in T$. Intuitively, we can tell whether or not $\tau \leq t$ from the information available to us at time t . In a sense, a stopping time is a random time that does not require that we see into the future. Of course, the concept depends critically on the filtration. Recall that if a random time τ is a stopping time for a filtration $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ then it is also a stopping time for a finer filtration $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$, so that $\mathcal{F}_t \subseteq \mathcal{G}_t$ for $t \in T$. Thus, the finer the filtration, the larger the collection of stopping times. In fact if the filtration is the trivial one where $\mathcal{F}_t = \mathcal{F}$ for all $t \in T$ (so that all information is available to us from the beginning of time), then *any* random time is a stopping time. But of course, this trivial filtration is usually not sensible.

Next, recall that if τ is a stopping time for the filtration \mathfrak{F} , then the σ -algebra \mathcal{F}_τ associated with τ is given by

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \in T\} \quad (16.1.6)$$

Intuitively, \mathcal{F}_τ is the collection of events up to the random time τ , analogous to the \mathcal{F}_t which is the collection of events up to the deterministic time $t \in T$. If $\mathbf{X} = \{X_t : t \in T\}$ is a stochastic process adapted to \mathfrak{F} and if τ is a stopping time relative to \mathfrak{F} , then we would hope that X_τ is measurable with respect to \mathcal{F}_τ just as X_t is measurable with respect to \mathcal{F}_t for deterministic $t \in T$. However, this will generally not be the case unless \mathbf{X} is *progressively measurable* relative to \mathfrak{F} , which means that $\mathbf{X} : \Omega \times T_t \rightarrow S$ is measurable with respect to $\mathcal{F}_t \otimes \mathcal{T}_t$ and \mathcal{S} where $T_t = \{s \in T : s \leq t\}$

and \mathcal{F}_t the corresponding Borel σ -algebra. This is always true in discrete time, of course, and more generally if S has an LCCB topology with \mathcal{S} the Borel σ -algebra, and \mathbf{X} is right continuous. If \mathbf{X} is progressively measurable with respect to \mathfrak{F} then \mathbf{X} is measurable and \mathbf{X} is adapted to \mathfrak{F} .

The *strong Markov property* for our stochastic process $\mathbf{X} = \{X_t : t \in T\}$ states that the future is independent of the past, given the present, when the present time is a stopping time.

The random process \mathbf{X} is a *strong Markov process* if

$$\mathbb{E}[f(X_{\tau+t}) \mid \mathcal{F}_\tau] = \mathbb{E}[f(X_{\tau+t}) \mid X_\tau] \quad (16.1.7)$$

for every $t \in T$, stopping time τ , and $f \in \mathcal{B}$.

As with the regular Markov property, the strong Markov property depends on the underlying filtration \mathfrak{F} . If the property holds with respect to a given filtration, then it holds with respect to a coarser filtration.

Suppose that the stochastic process $\mathbf{X} = \{X_t : t \in T\}$ is progressively measurable relative to the filtration $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ and that the filtration $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$ is finer than \mathfrak{F} . If \mathbf{X} is a strong Markov process relative to \mathfrak{G} then \mathbf{X} is a strong Markov process relative to \mathfrak{F} .

Proof

Recall again that since \mathbf{X} is adapted to \mathfrak{F} , it is also adapted to \mathfrak{G} . Suppose that τ is a finite stopping time for \mathfrak{F} and that $t \in T$ and $f \in \mathcal{B}$. Then τ is also a stopping time for \mathfrak{G} , and $\mathcal{F}_\tau \subseteq \mathcal{G}_\tau$. Hence

$$\mathbb{E}[f(X_{\tau+t}) \mid \mathcal{F}_\tau] = \mathbb{E}(\mathbb{E}[f(X_{\tau+t}) \mid \mathcal{G}_\tau] \mid \mathcal{F}_\tau) = \mathbb{E}(\mathbb{E}[f(X_{\tau+t}) \mid X_\tau] \mid \mathcal{F}_\tau) = \mathbb{E}[f(X_{\tau+t}) \mid X_\tau] \quad (16.1.8)$$

The first equality is a basic property of conditional expected value. The second uses the fact that \mathbf{X} has the strong Markov property relative to \mathfrak{G} , and the third follows since X_τ measurable with respect to \mathcal{F}_τ . In continuous time, it's last step that requires progressive measurability.

So if \mathbf{X} is a strong Markov process, then \mathbf{X} satisfies the strong Markov property relative to its natural filtration. Again there is a tradeoff: finer filtrations allow more stopping times (generally a good thing), but make the strong Markov property harder to satisfy and may not be reasonable (not so good). So we usually don't want filtrations that are too much finer than the natural one.

With the strong Markov and homogeneous properties, the process $\{X_{\tau+t} : t \in T\}$ given $X_\tau = x$ is equivalent in distribution to the process $\{X_t : t \in T\}$ given $X_0 = x$. Clearly, the strong Markov property implies the ordinary Markov property, since a fixed time $t \in T$ is trivially also a stopping time. The converse is true in discrete time.

Suppose that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ is a (homogeneous) Markov process in discrete time. Then \mathbf{X} is a strong Markov process.

As always in continuous time, the situation is more complicated and depends on the continuity of the process \mathbf{X} and the filtration \mathfrak{F} . Here is the standard result for Feller processes.

If $\mathbf{X} = \{X_t : t \in [0, \infty)\}$ is a Feller Markov process, then \mathbf{X} is a strong Markov process relative to filtration \mathfrak{F}_+^0 , the right-continuous refinement of the natural filtration.

Transition Kernels of Markov Processes

For our next discussion, you may need to review the section on kernels and operators in the chapter on expected value. Suppose again that $\mathbf{X} = \{X_t : t \in T\}$ is a (homogeneous) Markov process with state space S and time space T , as described above. The kernels in the following definition are of fundamental importance in the study of \mathbf{X}

For $t \in T$, let

$$P_t(x, A) = \mathbb{P}(X_t \in A \mid X_0 = x), \quad x \in S, A \in \mathcal{S} \quad (16.1.9)$$

Then P_t is a probability kernel on (S, \mathcal{S}) , known as the *transition kernel* of \mathbf{X} for time t .

Proof

Fix $t \in T$. The measurability of $x \mapsto \mathbb{P}(X_t \in A \mid X_0 = x)$ for $A \in \mathcal{S}$ is built into the definition of conditional probability. Also, of course, $A \mapsto \mathbb{P}(X_t \in A \mid X_0 = x)$ is a probability measure on \mathcal{S} for $x \in S$. In general, the conditional distribution of one random variable, conditioned on a value of another random variable defines a probability kernel.

That is, $P_t(x, \cdot)$ is the conditional distribution of X_t given $X_0 = x$ for $t \in T$ and $x \in S$. By the time homogenous property, $P_t(x, \cdot)$ is also the conditional distribution of X_{s+t} given $X_s = x$ for $s \in T$:

$$P_t(x, A) = \mathbb{P}(X_{s+t} \in A \mid X_s = x), \quad s, t \in T, x \in S, A \in \mathcal{S} \quad (16.1.10)$$

Note that $P_0 = I$, the *identity kernel* on (S, \mathcal{S}) defined by $I(x, A) = \mathbf{1}(x \in A)$ for $x \in S$ and $A \in \mathcal{S}$, so that $I(x, A) = 1$ if $x \in A$ and $I(x, A) = 0$ if $x \notin A$. Recall also that usually there is a natural reference measure λ on (S, \mathcal{S}) . In this case, the transition kernel P_t will often have a *transition density* p_t with respect to λ for $t \in T$. That is,

$$P_t(x, A) = \mathbb{P}(X_t \in A \mid X_0 = x) = \int_A p_t(x, y) \lambda(dy), \quad x \in S, A \in \mathcal{S} \quad (16.1.11)$$

The next theorem gives the *Chapman-Kolmogorov* equation, named for Sydney Chapman and Andrei Kolmogorov, the fundamental relationship between the probability kernels, and the reason for the name *transition kernel*.

Suppose again that $\mathbf{X} = \{X_t : t \in T\}$ is a Markov process on S with transition kernels $\mathbf{P} = \{P_t : t \in T\}$. If $s, t \in T$, then $P_s P_t = P_{s+t}$. That is,

$$P_{s+t}(x, A) = \int_S P_s(x, dy) P_t(y, A), \quad x \in S, A \in \mathcal{S} \quad (16.1.12)$$

Proof

The Markov property and a conditioning argument are the fundamental tools. Recall again that $P_s(x, \cdot)$ is the conditional distribution of X_s given $X_0 = x$ for $x \in S$. Let $A \in \mathcal{S}$. Conditioning on X_s gives

$$P_{s+t}(x, A) = \mathbb{P}(X_{s+t} \in A \mid X_0 = x) = \int_S P_s(x, dy) \mathbb{P}(X_{s+t} \in A \mid X_s = y, X_0 = x) \quad (16.1.13)$$

But by the Markov and time-homogeneous properties,

$$\mathbb{P}(X_{s+t} \in A \mid X_s = y, X_0 = x) = \mathbb{P}(X_t \in A \mid X_0 = y) = P_t(y, A) \quad (16.1.14)$$

Substituting we have

$$P_{s+t}(x, A) = \int_S P_s(x, dy) P_t(y, A) = (P_s P_t)(x, A) \quad (16.1.15)$$

In the language of functional analysis, \mathbf{P} is a *semigroup*. Recall that the commutative property generally does not hold for the product operation on kernels. However the property does hold for the transition kernels of a homogeneous Markov process. That is, $P_s P_t = P_t P_s = P_{s+t}$ for $s, t \in T$. As a simple corollary, if S has a reference measure, the same basic relationship holds for the transition densities.

Suppose that λ is the reference measure on (S, \mathcal{S}) and that $\mathbf{X} = \{X_t : t \in T\}$ is a Markov process on S and with transition densities $\{p_t : t \in T\}$. If $s, t \in T$ then $p_s p_t = p_{s+t}$. That is,

$$p_t(x, z) = \int_S p_s(x, y) p_t(y, z) \lambda(dy), \quad x, z \in S \quad (16.1.16)$$

Proof

The transition kernels satisfy $P_s P_t = P_{s+t}$. But P_s has density p_s , P_t has density p_t , and P_{s+t} has density p_{s+t} . From a basic result on kernel functions, $P_s P_t$ has density $p_s p_t$ as defined in the theorem.

If $T = \mathbb{N}$ (discrete time), then the transition kernels of \mathbf{X} are just the powers of the *one-step* transition kernel. That is, if we let $P = P_1$ then $P_n = P^n$ for $n \in \mathbb{N}$.

Recall that a kernel defines two operations: operating on the left with positive measures on (S, \mathcal{S}) and operating on the right with measurable, real-valued functions. For the transition kernels of a Markov process, both of these operators have natural interpretations.

Suppose that $s, t \in T$. If μ_s is the distribution of X_s then X_{s+t} has distribution $\mu_{s+t} = \mu_s P_t$. That is,

$$\mu_{s+t}(A) = \int_S \mu_s(dx) P_t(x, A), \quad A \in \mathcal{S} \quad (16.1.17)$$

Proof

Let $A \in \mathcal{S}$. Conditioning on X_s gives

$$\mathbb{P}(X_{s+t} \in A) = \mathbb{E}[\mathbb{P}(X_{s+t} \in A \mid X_s)] = \int_S \mu_s(dx) \mathbb{P}(X_{s+t} \in A \mid X_s = x) = \int_S \mu_s(dx) P_t(x, A) = \mu_s P_t(A) \quad (16.1.18)$$

So if \mathcal{P} denotes the collection of probability measures on (S, \mathcal{S}) , then the left operator P_t maps \mathcal{P} back into \mathcal{P} . In particular, if X_0 has distribution μ_0 (the *initial distribution*) then X_t has distribution $\mu_t = \mu_0 P_t$ for every $t \in T$.

A positive measure μ on (S, \mathcal{S}) is *invariant* for \mathbf{X} if $\mu P_t = \mu$ for every $t \in T$.

Hence if μ is a probability measure that is invariant for \mathbf{X} , and X_0 has distribution μ , then X_t has distribution μ for every $t \in T$ so that the process \mathbf{X} is identically distributed. In discrete time, note that if μ is a positive measure and $\mu P = \mu$ then $\mu P^n = \mu$ for every $n \in \mathbb{N}$, so μ is invariant for \mathbf{X} . The operator on the right is given next.

Suppose that $f : S \rightarrow \mathbb{R}$. If $t \in T$ then (assuming that the expected value exists),

$$P_t f(x) = \int_S P_t(x, dy) f(y) = \mathbb{E}[f(X_t) | X_0 = x], \quad x \in S \quad (16.1.19)$$

Proof

This follows directly from the definitions:

$$P_t f(x) = \int_S P_t(x, dy) f(y), \quad x \in S \quad (16.1.20)$$

and $P_t(x, \cdot)$ is the conditional distribution of X_t given $X_0 = x$.

In particular, the right operator P_t is defined on \mathcal{B} , the vector space of bounded, linear functions $f : S \rightarrow \mathbb{R}$, and in fact is a linear operator on \mathcal{B} . That is, if $f, g \in \mathcal{B}$ and $c \in \mathbb{R}$, then $P_t(f + g) = P_t f + P_t g$ and $P_t(cf) = cP_t f$. Moreover, P_t is a *contraction operator* on \mathcal{B} , since $\|P_t f\| \leq \|f\|$ for $f \in \mathcal{B}$. It then follows that P_t is a continuous operator on \mathcal{B} for $t \in T$.

For the right operator, there is a concept that is complementary to the invariance of of a positive measure for the left operator.

A measurable function $f : S \rightarrow \mathbb{R}$ is *harmonic* for \mathbf{X} if $P_t f = f$ for all $t \in T$.

Again, in discrete time, if $Pf = f$ then $P^n f = f$ for all $n \in \mathbb{N}$, so f is harmonic for \mathbf{X} .

Combining two results above, if X_0 has distribution μ_0 and $f : S \rightarrow \mathbb{R}$ is measurable, then (again assuming that the expected value exists), $\mu_0 P_t f = \mathbb{E}[f(X_t)]$ for $t \in T$. That is,

$$\mathbb{E}[f(X_t)] = \int_S \mu_0(dx) \int_S P_t(x, dy) f(y) \quad (16.1.21)$$

The result [above](#) shows how to obtain the distribution of X_t from the distribution of X_0 and the transition kernel P_t for $t \in T$. But we can do more. Recall that one basic way to describe a stochastic process is to give its *finite dimensional distributions*, that is, the distribution of $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ for every $n \in \mathbb{N}_+$ and every $(t_1, t_2, \dots, t_n) \in T^n$. For a Markov process, the initial distribution and the transition kernels determine the finite dimensional distributions. It's easiest to state the distributions in differential form.

Suppose $\mathbf{X} = \{X_t : t \in T\}$ is a Markov process with transition operators $\mathbf{P} = \{P_t : t \in T\}$, and that $(t_1, \dots, t_n) \in T^n$ with $0 < t_1 < \dots < t_n$. If X_0 has distribution μ_0 , then in differential form, the distribution of $(X_0, X_{t_1}, \dots, X_{t_n})$ is

$$\mu_0(dx_0) P_{t_1}(x_0, dx_1) P_{t_2-t_1}(x_1, dx_2) \cdots P_{t_n-t_{n-1}}(x_{n-1}, dx_n) \quad (16.1.22)$$

Proof

This follows from induction and repeated use of the Markov property. For example, if $t \in T$ with $t > 0$, then conditioning on X_0 gives

$$\mathbb{P}(X_0 \in A, X_t \in B) = \int_A \mathbb{P}(X_t \in B | X_0 = x) \mu_0(dx) = \int_A P_t(x, B) \mu_0(dx) = \int_A \int_B P_t(x, dy) \mu_0(dx) \quad (16.1.23)$$

for $A, B \in \mathcal{S}$. So in differential form, the distribution of (X_0, X_t) is $\mu(dx) P_t(x, dy)$. If $s, t \in T$ with $0 < s < t$, then conditioning on (X_0, X_s) and using our previous result gives

$$\mathbb{P}(X_0 \in A, X_s \in B, X_t \in C) = \int_{A \times B} \mathbb{P}(X_t \in C | X_0 = x, X_s = y) \mu_0(dx) P_s(x, dy) \quad (16.1.24)$$

for $A, B, C \in \mathcal{S}$. But by the Markov property,

$$\mathbb{P}(X_t \in C | X_0 = x, X_s = y) = \mathbb{P}(X_t \in C | X_s = y) = P_{t-s}(y, C) = \int_C P_{t-s}(y, dz) \quad (16.1.25)$$

Hence in differential form, the distribution of (X_0, X_s, X_t) is $\mu_0(dx) P_s(x, dy) P_{t-s}(y, dz)$. Continuing in this manner gives the general result.

This result is very important for constructing Markov processes. If we know how to define the transition kernels P_t for $t \in T$ (based on modeling considerations, for example), and if we know the initial distribution μ_0 , then the last result gives a consistent set of finite dimensional distributions. From the Kolmogorov construction theorem, we know that there *exists* a stochastic process that has these finite dimensional distributions. In continuous time, however, two serious problems remain. First, it's not clear how we would construct the transition kernels so that the crucial Chapman-Kolmogorov equations [above](#) are satisfied. Second, we usually want our Markov process to have certain properties (such as continuity properties of the sample paths) that go beyond the finite dimensional distributions. The first problem will be addressed in the next section, and fortunately, the second problem can be resolved for a Feller process.

Suppose that $\mathbf{X} = \{X_t : t \in T\}$ is a Markov process on an LCCB state space (S, \mathcal{S}) with transition operators $\mathbf{P} = \{P_t : t \in [0, \infty)\}$. Then \mathbf{X} is a Feller process if and only if the following conditions hold:

1. *Continuity in space:* If $f \in \mathcal{C}_0$ and $t \in [0, \infty)$ then $P_t f \in \mathcal{C}_0$
2. *Continuity in time:* If $f \in \mathcal{C}_0$ and $x \in S$ then $P_t f(x) \rightarrow f(x)$ as $t \downarrow 0$.

A semigroup of probability kernels $\mathbf{P} = \{P_t : t \in T\}$ that satisfies the properties in this theorem is called a *Feller semigroup*. So the theorem states that the Markov process \mathbf{X} is Feller if and only if the transition semigroup of transition \mathbf{P} is Feller. As before, (a) is automatically satisfied if S is discrete, and (b) is automatically satisfied if T is discrete. Condition (a) means that P_t is an operator on the vector space \mathcal{C}_0 , in addition to being an operator on the larger space \mathcal{B} . Condition (b) actually implies a stronger form of continuity in time.

Suppose that $\mathbf{P} = \{P_t : t \in T\}$ is a Feller semigroup of transition operators. Then $t \mapsto P_t f$ is continuous (with respect to the supremum norm) for $f \in \mathcal{C}_0$.

Additional details

This means that for $f \in \mathcal{C}_0$ and $t \in [0, \infty)$,

$$\|P_{t+s}f - P_t f\| = \sup\{|P_{t+s}f(x) - P_t f(x)| : x \in S\} \rightarrow 0 \text{ as } s \rightarrow 0 \quad (16.1.26)$$

So combining this with the remark above, note that if \mathbf{P} is a Feller semigroup of transition operators, then $f \mapsto P_t f$ is continuous on \mathcal{C}_0 for fixed $t \in T$, and $t \mapsto P_t f$ is continuous on T for fixed $f \in \mathcal{C}_0$. Again, the importance of this is that we often start with the collection of probability kernels \mathbf{P} and want to know that there exists a nice Markov process \mathbf{X} that has these transition operators.

Sampling in Time

If we *sample* a Markov process at an increasing sequence of points in time, we get another Markov process in discrete time. But the discrete time process may not be homogeneous even if the original process is homogeneous.

Suppose that $\mathbf{X} = \{X_t : t \in T\}$ is a Markov process with state space (S, \mathcal{S}) and that (t_0, t_1, t_2, \dots) is a sequence in T with $0 = t_0 < t_1 < t_2 < \dots$. Let $Y_n = X_{t_n}$ for $n \in \mathbb{N}$. Then $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ is a Markov process in discrete time.

Proof

For $n \in \mathbb{N}$, let $\mathcal{G}_n = \sigma\{Y_k : k \in \mathbb{N}, k \leq n\}$, so that $\{\mathcal{G}_n : n \in \mathbb{N}\}$ is the natural filtration associated with \mathbf{Y} . Note that $\mathcal{G}_n \subseteq \mathcal{F}_{t_n}$ and $Y_n = X_{t_n}$ is measurable with respect to \mathcal{G}_n for $n \in \mathbb{N}$. Let $k, n \in \mathbb{N}$ and let $A \in \mathcal{S}$. Then

$$\mathbb{P}(Y_{k+n} \in A \mid \mathcal{G}_k) = \mathbb{P}(X_{t_{k+n}} \in A \mid \mathcal{G}_k) = \mathbb{P}(X_{t_{k+n}} \in A \mid X_{t_k}) = \mathbb{P}(Y_{n+k} \in A \mid Y_k) \quad (16.1.27)$$

If we sample a homogeneous Markov process at multiples of a fixed, positive time, we get a homogeneous Markov process in discrete time.

Suppose that $\mathbf{X} = \{X_t : t \in T\}$ is a homogeneous Markov process with state space (S, \mathcal{S}) and transition kernels $\mathbf{P} = \{P_t : t \in T\}$. Fix $r \in T$ with $r > 0$ and define $Y_n = X_{nr}$ for $n \in \mathbb{N}$. Then $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ is a homogeneous Markov process in discrete time, with one-step transition kernel Q given by

$$Q(x, A) = P_r(x, A); \quad x \in S, A \in \mathcal{S} \quad (16.1.28)$$

In some cases, sampling a strong Markov process at an increasing sequence of stopping times yields another Markov process in discrete time. The point of this is that discrete-time Markov processes are often found naturally embedded in continuous-time Markov processes.

Enlarging the State Space

Our first result in this discussion is that a non-homogeneous Markov process can be turned into a homogeneous Markov process, but only at the expense of enlarging the state space.

Suppose that $\mathbf{X} = \{X_t : t \in T\}$ is a non-homogeneous Markov process with state space (S, \mathcal{S}) . Suppose also that τ is a random variable taking values in T , independent of \mathbf{X} . Let $\tau_t = \tau + t$ and let $Y_t = (X_{\tau_t}, \tau_t)$ for $t \in T$. Then $\mathbf{Y} = \{Y_t : t \in T\}$ is a homogeneous Markov process with state space $(S \times T, \mathcal{S} \otimes \mathcal{T})$. For $t \in T$, the transition kernel P_t is given by

$$P_t[(x, r), A \times B] = \mathbb{P}(X_{r+t} \in A \mid X_r = x) \mathbf{1}(r+t \in B), \quad (x, r) \in S \times T, A \times B \in \mathcal{S} \otimes \mathcal{T} \quad (16.1.29)$$

Proof

By definition and the substitution rule,

$$\begin{aligned} \mathbb{P}[Y_{s+t} \in A \times B \mid Y_s = (x, r)] &= \mathbb{P}(X_{\tau_{s+t}} \in A, \tau_{s+t} \in B \mid X_{\tau_s} = x, \tau_s = r) \\ &= \mathbb{P}(X_{\tau+s+t} \in A, \tau+s+t \in B \mid X_{\tau+s} = x, \tau+s = r) \\ &= \mathbb{P}(X_{r+t} \in A, r+t \in B \mid X_r = x, \tau+s = r) \end{aligned}$$

But τ is independent of \mathbf{X} , so the last term is

$$\mathbb{P}(X_{r+t} \in A, r+t \in B \mid X_r = x) = \mathbb{P}(X_{r+t} \in A \mid X_r = x) \mathbf{1}(r+t \in B) \quad (16.1.30)$$

The important point is that the last expression does not depend on s , so \mathbf{Y} is homogeneous.

The trick of enlarging the state space is a common one in the study of stochastic processes. Sometimes a process that has a weaker form of “forgetting the past” can be made into a Markov process by enlarging the state space appropriately. Here is an example in discrete time.

Suppose that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ is a random process with state space (S, \mathcal{S}) in which the future depends stochastically on the last two states. That is, for $n \in \mathbb{N}$

$$\mathbb{P}(X_{n+2} \in A \mid \mathcal{F}_{n+1}) = \mathbb{P}(X_{n+2} \in A \mid X_n, X_{n+1}), \quad A \in \mathcal{S} \quad (16.1.31)$$

where $\{\mathcal{F}_n : n \in \mathbb{N}\}$ is the natural filtration associated with the process \mathbf{X} . Suppose also that the process is time homogeneous in the sense that

$$\mathbb{P}(X_{n+2} \in A \mid X_n = x, X_{n+1} = y) = Q(x, y, A) \quad (16.1.32)$$

independently of $n \in \mathbb{N}$. Let $Y_n = (X_n, X_{n+1})$ for $n \in \mathbb{N}$. Then $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ is a homogeneous Markov process with state space $(S \times S, \mathcal{S} \otimes \mathcal{S})$. The one step transition kernel P is given by

$$P[(x, y), A \times B] = I(y, A)Q(x, y, B); \quad x, y \in S, A, B \in \mathcal{S} \quad (16.1.33)$$

Proof

Note first that for $n \in \mathbb{N}$, $\sigma\{Y_k : k \leq n\} = \sigma\{(X_k, X_{k+1}) : k \leq n\} = \mathcal{F}_{n+1}$ so the natural filtration associated with the process \mathbf{Y} is $\{\mathcal{F}_{n+1} : n \in \mathbb{N}\}$. If $C \in \mathcal{S} \otimes \mathcal{S}$ then

$$\begin{aligned} \mathbb{P}(Y_{n+1} \in C \mid \mathcal{F}_{n+1}) &= \mathbb{P}[(X_{n+1}, X_{n+2}) \in C \mid \mathcal{F}_{n+1}] \\ &= \mathbb{P}[(X_{n+1}, X_{n+2}) \in C \mid X_n, X_{n+1}] = \mathbb{P}(Y_{n+1} \in C \mid Y_n) \end{aligned}$$

by the given assumption on \mathbf{X} . Hence \mathbf{Y} is a Markov process. Next,

$$\begin{aligned} \mathbb{P}[Y_{n+1} \in A \times B \mid Y_n = (x, y)] &= \mathbb{P}[(X_{n+1}, X_{n+2}) \in A \times B \mid (X_n, X_{n+1}) = (x, y)] \\ &= \mathbb{P}(X_{n+1} \in A, X_{n+2} \in B \mid X_n = x, X_{n+1} = y) = \mathbb{P}(y \in A, X_{n+2} \in B \mid X_n = x, X_{n+1} = y) \\ &= I(y, A)Q(x, y, B) \end{aligned}$$

The last result generalizes in a completely straightforward way to the case where the future of a random process in discrete time depends stochastically on the last k states, for some fixed $k \in \mathbb{N}$.

Examples and Applications

Recurrence Relations and Differential Equations

As noted in the introduction, Markov processes can be viewed as stochastic counterparts of deterministic recurrence relations (discrete time) and differential equations (continuous time). Our goal in this discussion is to explore these connections.

Suppose that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ is a stochastic process with state space (S, \mathcal{S}) and that \mathbf{X} satisfies the recurrence relation

$$X_{n+1} = g(X_n), \quad n \in \mathbb{N} \quad (16.1.34)$$

where $g : S \rightarrow S$ is measurable. Then \mathbf{X} is a homogeneous Markov process with one-step transition operator P given by $Pf = f \circ g$ for a measurable function $f : S \rightarrow \mathbb{R}$.

Proof

Clearly \mathbf{X} is uniquely determined by the initial state, and in fact $X_n = g^n(X_0)$ for $n \in \mathbb{N}$ where g^n is the n -fold composition power of g . So the only possible source of randomness is in the initial state. The Markov and time homogeneous properties simply follow from the trivial fact that $g^{m+n}(X_0) = g^n[g^m(X_0)]$, so that $X_{m+n} = g^n(X_m)$. That is, the state at time $m+n$ is completely determined by the state at time m (regardless of the previous states) and the time increment n . In particular, $Pf(x) = \mathbb{E}[g(X_1) \mid X_0 = x] = f[g(x)]$ for measurable $f : S \rightarrow \mathbb{R}$ and $x \in S$. Note that for $n \in \mathbb{N}$, the n -step transition operator is given by $P^n f = f \circ g^n$.

In the deterministic world, as in the stochastic world, the situation is more complicated in continuous time. Nonetheless, the same basic analogy applies.

Suppose that $\mathbf{X} = \{X_t : t \in [0, \infty)\}$ with state space $(\mathbb{R}, \mathcal{B})$ satisfies the first-order differential equation

$$\frac{d}{dt}X_t = g(X_t) \quad (16.1.35)$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous. Then \mathbf{X} is a Feller Markov process

Proof

Recall that *Lipschitz continuous* means that there exists a constant $k \in (0, \infty)$ such that $|g(y) - g(x)| \leq k|x - y|$ for $x, y \in \mathbb{R}$. This is a standard condition on g that guarantees the existence and uniqueness of a solution to the differential equation on $[0, \infty)$. So as before, the only source of randomness in the process comes from the initial value X_0 . Let $t \mapsto X_t(x)$ denote the unique solution with $X_0(x) = x$ for $x \in \mathbb{R}$. The Markov and homogenous properties follow from the fact that $X_{t+s}(x) = X_t(X_s(x))$ for $s, t \in [0, \infty)$ and $x \in S$. That is, the state at time $t+s$ depends only on the state at time s and the time increment t . The Feller properties follow from the continuity of $t \mapsto X_t(x)$ and the continuity of $x \mapsto X_t(x)$. The latter is the *continuous dependence on the initial value*, again guaranteed by the assumptions on g . Note that the transition operator is given by $P_t f(x) = f[X_t(x)]$ for a measurable function $f : S \rightarrow \mathbb{R}$ and $x \in S$.

In differential form, the process can be described by $dX_t = g(X_t) dt$. This essentially deterministic process can be extended to a very important class of Markov processes by the addition of a stochastic term related to Brownian motion. Such *stochastic differential equations* are the main tools for constructing Markov processes known as *diffusion processes*.

Processes with Stationary, Independent Increments

For our next discussion, we consider a general class of stochastic processes that are Markov processes. Suppose that $\mathbf{X} = \{X_t : t \in T\}$ is a random process with $S \subseteq \mathbb{R}$ as the set of states. The state space can be discrete (countable) or “continuous”. Typically, S is either \mathbb{N} or \mathbb{Z} in the discrete case, and is either $[0, \infty)$ or \mathbb{R} in the continuous case. In any case, S is given the usual σ -algebra \mathcal{S} of Borel subsets of S (which is the power set in the discrete case). Also, the state space (S, \mathcal{S}) has a natural reference measure λ , namely counting measure in the discrete case and Lebesgue measure in the continuous case. Let $\mathcal{F} = \{\mathcal{F}_t : t \in T\}$ denote the natural filtration, so that $\mathcal{F}_t = \sigma\{X_s : s \in T, s \leq t\}$ for $t \in T$.

The process \mathbf{X} has

1. *Independent increments* if $X_{s+t} - X_s$ is independent of \mathcal{F}_s for all $s, t \in T$.
2. *Stationary increments* if the distribution of $X_{s+t} - X_s$ is the same as the distribution of $X_t - X_0$ for all $s, t \in T$.

A difference of the form $X_{s+t} - X_s$ for $s, t \in T$ is an *increment* of the process, hence the names. Sometimes the definition of stationary increments is that $X_{s+t} - X_s$ have the same distribution as X_t . But this forces $X_0 = 0$ with probability 1, and as usual with Markov processes, it's best to keep the initial distribution unspecified. If \mathbf{X} has stationary increments in the sense of our definition, then the process $\mathbf{Y} = \{Y_t = X_t - X_0 : t \in T\}$ has stationary increments in the more restricted sense. For the remainder of this discussion, assume that $\mathbf{X} = \{X_t : t \in T\}$ has stationary, independent increments, and let Q_t denote the distribution of $X_t - X_0$ for $t \in T$.

$$Q_s * Q_t = Q_{s+t} \text{ for } s, t \in T.$$

Proof

For $s, t \in T$, Q_s is the distribution of $X_s - X_0$, and by the stationary property, Q_t is the distribution of $X_{s+t} - X_s$. By the independence property, $X_s - X_0$ and $X_{s+t} - X_s$ are independent. Hence $Q_s * Q_t$ is the distribution of $[X_s - X_0] + [X_{s+t} - X_s] = X_{s+t} - X_0$. But by definition, this variable has distribution Q_{s+t} .

So the collection of distributions $\mathbf{Q} = \{Q_t : t \in T\}$ forms a semigroup, with convolution as the operator. Note that Q_0 is simply point mass at 0.

The process \mathbf{X} is a homogeneous Markov process. For $t \in T$, the transition operator P_t is given by

$$P_t f(x) = \int_S f(x+y) Q_t(dy), \quad f \in \mathcal{B} \quad (16.1.36)$$

Proof

Suppose that $s, t \in T$ and $f \in \mathcal{B}$,

$$\mathbb{E}[f(X_{s+t}) | \mathcal{F}_s] = \mathbb{E}[f(X_{s+t} - X_s + X_s) | \mathcal{F}_s] = \mathbb{E}[f(X_{s+t}) | X_s] \quad (16.1.37)$$

since $X_{s+t} - X_s$ is independent of \mathcal{F}_s . Moreover, by the stationary property,

$$\mathbb{E}[f(X_{s+t}) | X_s = x] = \int_S f(x+y) Q_t(dy), \quad x \in S \quad (16.1.38)$$

Clearly the semigroup property of $\mathbf{P} = \{P_t : t \in T\}$ (with the usual operator product) is equivalent to the semigroup property of $\mathbf{Q} = \{Q_t : t \in T\}$ (with convolution as the product).

Suppose that for positive $t \in T$, the distribution Q_t has probability density function g_t with respect to the reference measure λ . Then the transition density is

$$p_t(x, y) = g_t(y - x), \quad x, y \in S \quad (16.1.39)$$

Of course, from the result above, it follows that $g_s * g_t = g_{s+t}$ for $s, t \in T$, where here $*$ refers to the convolution operation on probability density functions.

If $Q_t \rightarrow Q_0$ as $t \downarrow 0$ then \mathbf{X} is a Feller Markov process.

Thus, by the general theory sketched above, \mathbf{X} is a strong Markov process, and there exists a version of \mathbf{X} that is right continuous and has left limits. Such a process is known as a *Lévy process*, in honor of Paul Lévy.

For a real-valued stochastic process $\mathbf{X} = \{X_t : t \in T\}$, let m and v denote the mean and variance functions, so that

$$m(t) = \mathbb{E}(X_t), \quad v(t) = \text{var}(X_t); \quad t \in T \quad (16.1.40)$$

assuming of course that these exist. The mean and variance functions for a Lévy process are particularly simple.

Suppose again that \mathbf{X} has stationary, independent increments.

1. If $\mu_0 = \mathbb{E}(X_0) \in \mathbb{R}$ and $\mu_1 = \mathbb{E}(X_1) \in \mathbb{R}$ then $m(t) = \mu_0 + (\mu_1 - \mu_0)t$ for $t \in T$.
2. If in addition, $\sigma_0^2 = \text{var}(X_0) \in (0, \infty)$ and $\sigma_1^2 = \text{var}(X_1) \in (0, \infty)$ then $v(t) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)t$ for $t \in T$.

Proof

The proofs are simple using the independent and stationary increments properties. For $t \in T$, let $m_0(t) = \mathbb{E}(X_t - X_0) = m(t) - \mu_0$ and $v_0(t) = \text{var}(X_t - X_0) = v(t) - \sigma_0^2$. denote the mean and variance functions for the centered process $\{X_t - X_0 : t \in T\}$. Now let $s, t \in T$.

1. From the additive property of expected value and the stationary property,

$$m_0(t+s) = \mathbb{E}(X_{t+s} - X_0) = \mathbb{E}[(X_{t+s} - X_s) + (X_s - X_0)] = \mathbb{E}(X_{t+s} - X_s) + \mathbb{E}(X_s - X_0) = m_0(t) + m_0(s) \quad (16.1.41)$$

2. From the additive property of variance for independent variables and the stationary property,

$$v_0(t+s) = \text{var}(X_{t+s} - X_0) = \text{var}[(X_{t+s} - X_s) + (X_s - X_0)] = \text{var}(X_{t+s} - X_s) + \text{var}(X_s - X_0) = v_0(t) + v_0(s) \quad (16.1.42)$$

So m_0 and v_0 satisfy the *Cauchy equation*. In discrete time, it's simple to see that there exists $a \in \mathbb{R}$ and $b^2 \in (0, \infty)$ such that $m_0(t) = at$ and $v_0(t) = b^2 t$. The same is true in continuous time, given the continuity assumptions that we have on the process \mathbf{X} . Substituting $t = 1$ we have $a = \mu_1 - \mu_0$ and $b^2 = \sigma_1^2 - \sigma_0^2$, so the results follow,

It's easy to describe processes with stationary independent increments in discrete time.

A process $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ has independent increments if and only if there exists a sequence of independent, real-valued random variables (U_0, U_1, \dots) such that

$$X_n = \sum_{i=0}^n U_i \quad (16.1.43)$$

In addition, \mathbf{X} has stationary increments if and only if (U_1, U_2, \dots) are identically distributed.

Proof

Suppose first that $\mathbf{U} = (U_0, U_1, \dots)$ is a sequence of independent, real-valued random variables, and define $X_n = \sum_{i=0}^n U_i$ for $n \in \mathbb{N}$. Note that $\mathcal{F}_n = \sigma\{X_0, \dots, X_n\} = \sigma\{U_0, \dots, U_n\}$ for $n \in \mathbb{N}$. If $k, n \in \mathbb{N}$ with $k \leq n$, then $X_n - X_k = \sum_{i=k+1}^n U_i$ which is independent of \mathcal{F}_k by the independence assumption on \mathbf{U} . Hence \mathbf{X} has independent increments. Suppose in addition that (U_1, U_2, \dots) are identically distributed. Then the increment $X_n - X_k$ above has the same distribution as $\sum_{i=1}^{n-k} U_i = X_{n-k} - X_0$. Hence \mathbf{X} has stationary increments.

Conversely, suppose that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ has independent increments. Let $U_0 = X_0$ and $U_n = X_n - X_{n-1}$ for $n \in \mathbb{N}_+$. Then $X_n = \sum_{i=0}^n U_i$ for $n \in \mathbb{N}$. As before $\mathcal{F}_n = \sigma\{X_0, \dots, X_n\} = \sigma\{U_0, \dots, U_n\}$ for $n \in \mathbb{N}$. Since \mathbf{X} has independent increments, U_n is independent of \mathcal{F}_{n-1} for $n \in \mathbb{N}_+$, so (U_0, U_1, \dots) are mutually independent. If in addition, \mathbf{X} has stationary increments, $U_n = X_n - X_{n-1}$ has the same distribution as $X_1 - X_0 = U_1$ for $n \in \mathbb{N}_+$. Hence (U_1, U_2, \dots) are identically distributed.

Thus suppose that $\mathbf{U} = (U_0, U_1, \dots)$ is a sequence of independent, real-valued random variables, with (U_1, U_2, \dots) identically distributed with common distribution Q . Then from our main result above, the partial sum process $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ associated with \mathbf{U} is a homogeneous Markov process with one step transition kernel P given by

$$P(x, A) = Q(A - x), \quad x \in S, A \in \mathcal{S} \quad (16.1.44)$$

More generally, for $n \in \mathbb{N}$, the n -step transition kernel is $P^n(x, A) = Q^{*n}(A - x)$ for $x \in S$ and $A \in \mathcal{S}$. This Markov process is known as a *random walk* (although unfortunately, the term *random walk* is used in a number of other contexts as well). The idea is that at time n , the walker moves a (directed) distance U_n on the real line, and these steps are independent and identically distributed. If Q has probability density function g with respect to the reference measure λ , then the one-step transition density is

$$p(x, y) = g(y - x), \quad x, y \in S \quad (16.1.45)$$

Consider the random walk on \mathbb{R} with steps that have the standard normal distribution. Give each of the following explicitly:

1. The one-step transition density.
2. The n -step transition density for $n \in \mathbb{N}_+$.

Proof

1. For $x \in \mathbb{R}$, $p(x, \cdot)$ is the normal PDF with mean x and variance 1:

$$p(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y - x)^2\right]; \quad x, y \in \mathbb{R} \quad (16.1.46)$$

2. For $x \in \mathbb{R}$, $p^n(x, \cdot)$ is the normal PDF with mean x and variance n :

$$p^n(x, y) = \frac{1}{\sqrt{2\pi n}} \exp\left[-\frac{1}{2n}(y - x)^2\right], \quad x, y \in \mathbb{R} \quad (16.1.47)$$

In continuous time, there are two processes that are particularly important, one with the discrete state space \mathbb{N} and one with the continuous state space \mathbb{R} .

For $t \in [0, \infty)$, let g_t denote the probability density function of the Poisson distribution with parameter t , and let $p_t(x, y) = g_t(y - x)$ for $x, y \in \mathbb{N}$. Then $\{p_t : t \in [0, \infty)\}$ is the collection of transition densities for a Feller semigroup on \mathbb{N} .

Proof

Recall that

$$g_t(n) = e^{-t} \frac{t^n}{n!}, \quad n \in \mathbb{N} \quad (16.1.48)$$

We just need to show that $\{g_t : t \in [0, \infty)\}$ satisfies the semigroup property, and that the continuity result holds. But we already know that if U, V are independent variables having Poisson distributions with parameters $s, t \in [0, \infty)$, respectively, then $U + V$ has the Poisson distribution with parameter $s + t$. That is, $g_s * g_t = g_{s+t}$. Moreover, $g_t \rightarrow g_0$ as $t \downarrow 0$.

So a Lévy process $\mathbf{N} = \{N_t : t \in [0, \infty)\}$ with these transition densities would be a Markov process with stationary, independent increments and with sample paths that are right continuous and have left limits. We do know of such a process, namely the Poisson process with rate 1.

Open the Poisson experiment and set the rate parameter to 1 and the time parameter to 10. Run the experiment several times in single-step mode and note the behavior of the process.

For $t \in (0, \infty)$, let g_t denote the probability density function of the normal distribution with mean 0 and variance t , and let $p_t(x, y) = g_t(y - x)$ for $x, y \in \mathbb{R}$. Then $\{p_t : t \in [0, \infty)\}$ is the collection of transition densities of a Feller semigroup on \mathbb{R} .

Proof

Recall that for $t \in (0, \infty)$,

$$g_t(z) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{z^2}{2t}\right), \quad z \in \mathbb{R} \quad (16.1.49)$$

We just need to show that $\{g_t : t \in [0, \infty)\}$ satisfies the semigroup property, and that the continuity result holds. But we already know that if U, V are independent variables having normal distributions with mean 0 and variances $s, t \in (0, \infty)$, respectively, then $U + V$ has the normal distribution with mean 0 and variance $s + t$. That is, $g_s * g_t = g_{s+t}$. Moreover, we also know that the normal distribution with variance t converges to point mass at 0 as $t \downarrow 0$.

So a Lévy process $\mathbf{X} = \{X_t : t \in [0, \infty)\}$ on \mathbb{R} with these transition densities would be a Markov process with stationary, independent increments, and whose sample paths are continuous from the right and have left limits. In fact, there exists such a process with *continuous* sample paths. This process is Brownian motion, a process important enough to have its own chapter.

Run the simulation of standard Brownian motion and note the behavior of the process.

This page titled [16.1: Introduction to Markov Processes](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.