Random processes

1 Introduction

Random processes, also known as stochastic processes, allow us to model quantities that evolve in time (or space) in an uncertain way: the trajectory of a particle, the price of oil, the temperature in New York, the national debt of the United States, etc. In these notes we introduce a mathematical framework that allows to reason probabilistically about such quantities.

2 Definition

We denote random processes using a tilde over an upper case letter $\tilde{X}$. This is not standard notation, but we want to emphasize the difference with random variables and random vectors.

Formally, a random process $\tilde{X}$ is a function that maps elements in a sample space $\Omega$ to real-valued functions.

**Definition 2.1 (Random process).** Given a probability space $(\Omega, \mathcal{F}, P)$, a random process $\tilde{X}$ is a function that maps each element $\omega$ in the sample space $\Omega$ to a function $\tilde{X}(\omega, \cdot) : T \to \mathbb{R}$, where $T$ is a discrete or continuous set.

There are two possible interpretations for $\tilde{X}(\omega, t)$:

- If we fix $\omega$, then $\tilde{X}(\omega, t)$ is a deterministic function of $t$ known as a realization of the random process.
- If we fix $t$ then $\tilde{X}(\omega, t)$ is a random variable, which we usually just denote by $\tilde{X}(t)$.

We can consequently interpret $\tilde{X}$ as an infinite collection of random variables indexed by $t$. The set of possible values that the random variable $\tilde{X}(t)$ can take for fixed $t$ is called the state space of the random process. Random processes can be classified according to the indexing variable or to their state space.

- If the indexing variable $t$ is defined on $\mathbb{R}$, or on a semi-infinite interval $(t_0, \infty)$ for some $t_0 \in \mathbb{R}$, then $\tilde{X}$ is a **continuous-time** random process.
• If the indexing variable \( t \) is defined on a discrete set, usually the integers or the natural numbers, then \( X \) is a **discrete-time** random process. In such cases we often use a different letter from \( t \), such as \( i \), as an indexing variable.

• If \( X(t) \) is a discrete random variable for all \( t \), then \( X \) is a **discrete-state** random process. If the discrete random variable takes a finite number of values that is the same for all \( t \), then \( X \) is a **finite-state** random process.

• If \( X(t) \) is a continuous random variable for all \( t \), then \( X \) is a **continuous-state** random process.

Note that there are continuous-state discrete-time random processes and discrete-state continuous-time random processes. Any combination is possible.

The underlying probability space \((\Omega, \mathcal{F}, P)\) mentioned in the definition completely determines the stochastic behavior of the random process. In principle we can specify random processes by defining the probability space \((\Omega, \mathcal{F}, P)\) and the mapping from elements in \(\Omega\) to continuous or discrete functions, as illustrated in the following example. As we will discuss later on, this way of specifying random processes is only tractable for very simple cases.

---

**Example 2.2** (Puddle). Bob asks Mary to model a puddle probabilistically. When the puddle is formed, it contains an amount of water that is distributed uniformly between 0 and 1 gallon. As time passes, the water evaporates. After a time interval \( t \) the water that is left is \( t \) times less than the initial quantity.

Mary models the water in the puddle as a continuous-state continuous-time random process \( \tilde{C} \). The underlying sample space is \((0, 1)\), the \( \sigma \) algebra is the corresponding Borel \( \sigma \) algebra (all possible countable unions of intervals in \((0, 1)\)) and the probability measure is the uniform probability measure on \((0, 1)\). For a particular element in the sample space \( \omega \in (0, 1) \)

\[
\tilde{C}(\omega, t) := \frac{\omega}{t}, \quad t \in [1, \infty),
\]

where the unit of \( t \) is days in this example. Figure 1 shows different realizations of the random process. Each realization is a deterministic function on \([1, \infty)\).

Bob points out that he only cares what the state of the puddle is each day, as opposed to at any time \( t \). Mary decides to simplify the model by using a continuous-state discrete-time random process \( \tilde{D} \). The underlying probability space is exactly the same as before, but the time index is now discrete. For a particular element in the sample space \( \omega \in (0, 1) \)

\[
\tilde{D}(\omega, i) := \frac{\omega}{i}, \quad i = 1, 2, \ldots
\]

Figure 1 shows different realizations of the continuous random process. Note that each realization is just a deterministic discrete sequence.
Figure 1: Realizations of the continuous-time (left) and discrete-time (right) random process defined in Example 2.2.

When working with random processes in a probabilistic model, we are often interested in the joint distribution of the process sampled at several fixed times. This is given by the \( n \)th-order distribution of the random process.

**Definition 2.3** (\( n \)th-order distribution). The \( n \)th-order distribution of a random process \( \tilde{X} \) is the joint distribution of the random variables \( \tilde{X}(t_1), \tilde{X}(t_2), \ldots, \tilde{X}(t_n) \) for any \( n \) samples \( \{t_1, t_2, \ldots, t_n\} \) of the time index \( t \).

**Example 2.4** (Puddle (continued)). The first-order cdf of \( \tilde{C}(t) \) in Example 2.2 is

\[
F_{\tilde{C}(t)}(x) := P(\tilde{C}(t) \leq x) = P(\omega \leq tx) = \begin{cases} 
\int_{u=0}^{tx} du = tx & \text{if } 0 \leq x \leq \frac{1}{\tau}, \\
1 & \text{if } x > \frac{1}{\tau}, \\
0 & \text{if } x < 0.
\end{cases}
\]

We obtain the first-order pdf by differentiating,

\[
f_{\tilde{C}(t)}(x) = \begin{cases} 
t & \text{if } 0 \leq x \leq \frac{1}{\tau}, \\
0 & \text{otherwise.}
\end{cases}
\]
If the $n$th order distribution of a random process is shift-invariant, then the process is said to be strictly or strongly stationary.

**Definition 2.5** (Strictly/strongly stationary process). A process is stationary in a strict or strong sense if for any $n \geq 0$ if we select $n$ samples $t_1, t_2, \ldots, t_n$ and any displacement $\tau$ the random variables $\tilde{X}(t_1), \tilde{X}(t_2), \ldots, \tilde{X}(t_n)$ have the same joint distribution as $\tilde{X}(t_1 + \tau), \tilde{X}(t_2 + \tau), \ldots, \tilde{X}(t_n + \tau)$.

The random processes in Example 2.2 are clearly not strictly stationary because their first-order pdf and pmf are not the same at every point. An important example of strictly stationary processes are independent identically-distributed sequences, presented in Section 4.1.

As in the case of random variables and random vectors, defining the underlying probability space in order to specify a random process is usually not very practical, except for very simple cases like the one in Example 2.2. The reason is that it is challenging to come up with a probability space that gives rise to a given $n$-th order distribution of interest. Fortunately, we can also specify a random process by directly specifying its $n$-th order distribution for all values of $n = 1, 2, \ldots$. This completely characterizes the random process. Most of the random processes described in Section 4, e.g. independent identically-distributed sequences, Markov chains, Poisson processes and Gaussian processes, are specified in this way.

Finally, random processes can also be specified by expressing them as functions of other random processes. A function $\tilde{Y} := g(\tilde{X})$ of a random process $\tilde{X}$ is also a random process, as it maps any element $\omega$ in the sample space $\Omega$ to a function $\tilde{Y}(\omega, \cdot) := g(\tilde{X}(\omega, \cdot))$. In Section 4.4 we define random walks in this way.

### 3 Mean and autocovariance functions

The expectation operator allows to derive quantities that summarize the behavior of the random process through weighted averaging. The mean of the random vector is the mean of $\tilde{X}(t)$ at any fixed time $t$.

**Definition 3.1** (Mean). The mean of a random process is the function

$$
\mu_{\tilde{X}}(t) := \mathbb{E}\left(\tilde{X}(t)\right).
$$

Note that the mean is a deterministic function of $t$. The autocovariance of a random process is another deterministic function that is equal to the covariance of $\tilde{X}(t_1)$ and $\tilde{X}(t_2)$ for any two points $t_1$ and $t_2$. If we set $t_1 := t_2$, then the autocovariance equals the variance at $t_1$. 

4
Definition 3.2 (Autocovariance). The autocovariance of a random process is the function
\[ R_{\tilde{X}}(t_1, t_2) := \text{Cov}\left(\tilde{X}(t_1), \tilde{X}(t_2)\right). \]  
(8)

In particular,
\[ R_{\tilde{X}}(t, t) := \text{Var}\left(\tilde{X}(t)\right). \]  
(9)

Intuitively, the autocovariance quantifies the correlation between the process at two different time points. If this correlation only depends on the separation between the two points, then the process is said to be wide-sense stationary.

Definition 3.3 (Wide-sense/weakly stationary process). A process is stationary in a wide or weak sense if its mean is constant
\[ \mu_{\tilde{X}}(t) := \mu \]  
(10)

and its autocovariance function is shift invariant, i.e.
\[ R_{\tilde{X}}(t_1, t_2) := R_{\tilde{X}}(t_1 + \tau, t_2 + \tau) \]  
(11)

for any \( t_1 \) and \( t_2 \) and any shift \( \tau \). For weakly stationary processes, the autocovariance is usually expressed as a function of the difference between the two time points,
\[ R_{\tilde{X}}(s) := R_{\tilde{X}}(t, t + s) \]  
for any \( t \).  
(12)

Note that any strictly stationary process is necessarily weakly stationary because its first and second-order distributions are shift invariant.

Figure 2 shows several stationary random processes with different autocovariance functions. If the autocovariance function is only nonzero at the origin, then the values of the random processes at different points are uncorrelated. This results in erratic fluctuations. When the autocovariance at neighboring times is high, the trajectory random process becomes smoother. The autocorrelation can also induce more structured behavior, as in the right column of the figure. In that example \( \tilde{X}(i) \) is negatively correlated with its two neighbors \( \tilde{X}(i - 1) \) and \( \tilde{X}(i + 1) \), but positively correlated with \( \tilde{X}(i - 2) \) and \( \tilde{X}(i + 2) \). This results in rapid periodic fluctuations.

4 Important random processes

In this section we describe some important examples of random processes.
Figure 2: Realizations (bottom three rows) of Gaussian processes with zero mean and the autocovariance functions shown on the top row.
4.1 Independent identically-distributed sequences

An independent identically-distributed (iid) sequence $\tilde{X}$ is a discrete-time random process where $\tilde{X} (i)$ has the same distribution for any fixed $i$ and $\tilde{X} (i_1), \tilde{X} (i_2), \ldots, \tilde{X} (i_n)$ are mutually independent for any $n$ fixed indices and any $n \geq 2$. If $\tilde{X} (i)$ is a discrete random variable (or equivalently the state space of the random process is discrete), then we denote the pmf associated to the distribution of each entry by $p_{\tilde{X}}$. This pdf completely characterizes the random process, since for any $n$ indices $i_1, i_2, \ldots, i_n$ and any $n$:

$$p_{\tilde{X}(i_1), \tilde{X}(i_2), \ldots, \tilde{X}(i_n)} (x_{i_1}, x_{i_2}, \ldots, x_{i_n}) = \prod_{i=1}^{n} p_{\tilde{X}} (x_i). \quad (13)$$

Note that the distribution that does not vary if we shift every index by the same amount, so the process is strictly stationary.

Similarly, if $\tilde{X} (i)$ is a continuous random variable, then we denote the pdf associated to the
distribution by \( f_{\tilde{X}} \). For any \( n \) indices \( i_1, i_2, \ldots, i_n \) and any \( n \) we have

\[
f_{\tilde{X}(i_1), \tilde{X}(i_2), \ldots, \tilde{X}(i_n)} (x_{i_1}, x_{i_2}, \ldots, x_{i_n}) = \prod_{i=1}^{n} f_{\tilde{X}} (x_i). \tag{14}
\]

Figure 3 shows several realizations from iid sequences which follow a uniform and a geometric distribution.

The mean of an iid random sequence is constant and equal to the mean of its associated distribution, which we denote by \( \mu \),

\[
\mu_{\tilde{X}} (i) := \mathbb{E} \left( \tilde{X} (i) \right) \tag{15}
\]

\[
= \mu. \tag{16}
\]

Let us denote the variance of the distribution associated to the iid sequence by \( \sigma^2 \). The autocovariance function is given by

\[
R_{\tilde{X}} (i, j) := \mathbb{E} \left( \tilde{X} (i) \tilde{X} (j) \right) - \mathbb{E} \left( \tilde{X} (i) \right) \mathbb{E} \left( \tilde{X} (j) \right) \tag{17}
\]

\[
= \begin{cases} \sigma^2, \\ 0. \end{cases} \tag{18}
\]

This is not surprising, \( \tilde{X} (i) \) and \( \tilde{X} (j) \) are independent for all \( i \neq j \), so they are also uncorrelated.

### 4.2 Gaussian process

A random process \( \tilde{X} \) is Gaussian if the joint distribution of the random variables \( \tilde{X} (t_1), \tilde{X} (t_2), \ldots, \tilde{X} (t_n) \) is Gaussian for all \( t_1, t_2, \ldots, t_n \) and any \( n \geq 1 \). An interesting feature of Gaussian processes is that they are fully characterized by their mean and autocovariance function. Figure 2 shows realizations of several discrete Gaussian processes with different autocovariances.

### 4.3 Poisson process

In Lecture Notes 2 we motivated the definition of Poisson random variable by deriving the distribution of the number of events that occur in a fixed time interval under the following conditions:

1. Each event occurs independently from every other event.
2. Events occur uniformly.

3. Events occur at a rate of $\lambda$ events per time interval.

We now assume that these conditions hold in the semi-infinite interval $[0, \infty)$ and define a random process $\tilde{N}$ that counts the events. To be clear $\tilde{N}(t)$ is the number of events that happen between 0 and $t$.

By the same reasoning as in Example 3.7 of Lecture Notes 2, the distribution of the random variable $\tilde{N}(t_2) - \tilde{N}(t_1)$, which represents the number of events that occur between $t_1$ and $t_2$, is a Poisson random variable with parameter $\lambda(t_2 - t_1)$. This holds for any $t_1$ and $t_2$. In addition the random variables $\tilde{N}(t_2) - \tilde{N}(t_1)$ and $\tilde{N}(t_4) - \tilde{N}(t_3)$ are independent as along as the intervals $[t_1, t_2]$ and $(t_3, t_4)$ do not overlap by Condition 1. A Poisson process is a discrete-state continuous random process that satisfies these two properties. Poisson processes are often used to model events such as earthquakes, telephone calls, decay of radioactive particles, neural spikes, etc. In Lecture Notes 2, Figure 5 shows an example of a real scenario where the number of calls received at a call center is well approximated as a Poisson process (as long as we only consider a few hours). Note that here we are using the word event to mean something that happens, such as the arrival of an email, instead of a set within a sample space, which is the meaning that it usually has elsewhere in these notes.

**Definition 4.1** (Poisson process). A Poisson process with parameter $\lambda$ is a discrete-state continuous random process $\tilde{N}$ such that

1. $\tilde{N}(0) = 0$.

2. For any $t_1 < t_2 < t_3 < t_4$ $\tilde{N}(t_2) - \tilde{N}(t_1)$ is a Poisson random variable with parameter $\lambda(t_2 - t_1)$.

3. For any $t_1 < t_2 < t_3 < t_4$ the random variables $\tilde{N}(t_2) - \tilde{N}(t_1)$ and $\tilde{N}(t_4) - \tilde{N}(t_3)$ are independent.

We now check that the random process is well defined, by proving that we can derive the joint pmf of $\tilde{N}$ at any $n$ points $t_1 < t_2 < \ldots < t_n$ for any $n \geq 0$. To alleviate notation let $p(\tilde{\lambda}, x)$ be the value of the pmf of a Poisson random variable with parameter $\tilde{\lambda}$ at $x$, i.e.

$$p(\tilde{\lambda}, x) := \frac{\tilde{\lambda}^x e^{-\tilde{\lambda}}}{x!}. \quad (19)$$
Figure 4: Events corresponding to the realizations of a Poisson process $\tilde{N}$ for different values of the parameter $\lambda$. $\tilde{N}(t)$ equals the number of events up to time $t$. 

$\lambda = 0.2$ 

$\lambda = 1$ 

$\lambda = 2$
We have

\[
p_{\tilde{N}(t_1),\ldots,\tilde{N}(t_n)}(x_1,\ldots,x_n) = P \left( \tilde{N}(t_1) = x_1, \ldots, \tilde{N}(t_n) = x_n \right) = P \left( \tilde{N}(t_1) = x_1, \tilde{N}(t_2) - \tilde{N}(t_1) = x_2 - x_1, \ldots, \tilde{N}(t_n) - \tilde{N}(t_{n-1}) = x_n - x_{n-1} \right) = P \left( \tilde{N}(t_1) = x_1 \right) P \left( \tilde{N}(t_2) - \tilde{N}(t_1) = x_2 - x_1 \right) \ldots P \left( \tilde{N}(t_n) - \tilde{N}(t_{n-1}) = x_n - x_{n-1} \right) = p(\lambda t_1, x_1) p(\lambda(t_2-t_1), x_2-x_1) \ldots p(\lambda(t_n-t_{n-1}), x_n-x_{n-1}) .
\]

In words, we have expressed the event that \(\tilde{N}(t_i) = x_i\) for \(1 \leq i \leq n\) in terms of the random variables \(\tilde{N}(t_1)\) and \(\tilde{N}(t_i) - \tilde{N}(t_{i-1})\), \(2 \leq i \leq n\), which are independent Poisson random variables with parameters \(\lambda t_1\) and \(\lambda(t_i-t_{i-1})\) respectively.

Figure 4 shows several sequences of events corresponding to the realizations of a Poisson process \(\tilde{N}\) for different values of the parameter \(\lambda\) (\(\tilde{N}(t)\) equals the number of events up to time \(t\)). Interestingly, the interarrival time of the events, i.e. the time between contiguous events, always has the same distribution: it is an exponential random variable. This allows to simulate Poisson processes by sampling from an exponential distribution. Figure 4 was generated in this way.

**Lemma 4.2** (Interarrival times of a Poisson process are exponential). Let \(T\) denote the time between two contiguous events in a Poisson process with parameter \(\lambda\). \(T\) is an exponential random variable with parameter \(\lambda\).

The proof is in Section A of the appendix. Figure 8 in Lecture Notes 2 shows that the interarrival times of telephone calls at a call center are indeed well modeled as exponential.

The following lemma, which derives the mean and autocovariance functions of a Poisson process is proved in Section B.

**Lemma 4.3** (Mean and autocovariance of a Poisson process). The mean and autocovariance of a Poisson process equal

\[
E(\tilde{X}(t)) = \lambda t , \tag{24}
\]

\[
R_{\tilde{X}}(t_1, t_2) = \lambda \min \{t_1, t_2\} . \tag{25}
\]

The mean of the Poisson process is not constant and its autocovariance is not shift-invariant, so the process is neither strictly nor wide-sense stationary.
Example 4.4 (Earthquakes). The number of earthquakes with intensity at least 3 on the Richter scale occurring in the San Francisco peninsula is modeled using a Poisson process with parameter 0.3 earthquakes/year. What is the probability that there are no earthquakes in the next ten years and then at least one earthquake over the following twenty years?

We define a Poisson process $\tilde{X}$ with parameter 0.3 to model the problem. The number of earthquakes in the next 10 years, i.e. $\tilde{X}(10)$, is a Poisson random variable with parameter $0.3 \cdot 10 = 3$. The earthquakes in the following 20 years, $\tilde{X}(30) - \tilde{X}(10)$, are Poisson with parameter $0.3 \cdot 20 = 6$. The two random variables are independent because the intervals do not overlap.

\[
P\left(\tilde{X}(10) = 0, \tilde{X}(30) \geq 1\right) = P\left(\tilde{X}(10) = 0, \tilde{X}(30) - \tilde{X}(10) \geq 1\right) = P\left(\tilde{X}(10) = 0\right) P\left(\tilde{X}(30) - \tilde{X}(10) \geq 1\right) = e^{-3} \left(1 - e^{-6}\right) = 4.97 \times 10^{-2}.
\]

The probability is 4.97%.

4.4 Random walk

A random walk is a discrete-time random process that is used to model a sequence that evolves by taking steps in random directions. To define a random walk formally, we first define an iid sequence of steps $\tilde{S}$ such that

\[
\tilde{S}(i) = \begin{cases} 
+1 & \text{with probability } \frac{1}{2}, \\
-1 & \text{with probability } \frac{1}{2}.
\end{cases}
\]

We define a random walk $\tilde{X}$ as the discrete-state discrete-time random process

\[
\tilde{X}(i) := \begin{cases} 
0 & \text{for } i = 0, \\
\sum_{j=1}^{i} \tilde{S}(j) & \text{for } i = 1, 2, \ldots
\end{cases}
\]

We have specified $\tilde{X}$ as a function of an iid sequence, so it is well defined. Figure 5 shows several realizations of the random walk.

$\tilde{X}$ is symmetric (there is the same probability of taking a positive step and a negative step) and begins at the origin. It is easy to define variations where the walk is non-symmetric.
and begins at another point. Generalizations to higher dimensional spaces— for instance to model random processes on a 2D surface— are also possible.

We derive the first-order pmf of the random walk in the following lemma, proved in Section C of the appendix.

**Lemma 4.5 (First-order pmf of a random walk).** The first-order pmf of the random walk $\tilde{X}$ is

$$p_{\tilde{X}(i)}(x) = \begin{cases} (\frac{i+x}{2})^{\frac{1}{2}} & \text{if } i + x \text{ is even and } -i \leq x \leq i \\ 0 & \text{otherwise.} \end{cases} \quad (32)$$

The first-order distribution of the random walk is clearly time-dependent, so the random process is not strictly stationary. By the following lemma, the mean of the random walk is constant (it equals zero). The autocovariance, however, is not shift invariant, so the process is not weakly stationary either.

**Lemma 4.6 (Mean and autocovariance of a random walk).** The mean and autocovariance of the random walk $\tilde{X}$ are

$$\mu_{\tilde{X}}(i) = 0, \quad (33)$$

$$R_{\tilde{X}}(i,j) = \min \{i,j\} . \quad (34)$$
Proof.

\[
\mu_{\tilde{X}} (i) := E \left( \tilde{X} (i) \right) = E \left( \sum_{j=1}^{i} \tilde{S} (j) \right) = \sum_{j=1}^{i} E \left( \tilde{S} (j) \right) \quad \text{by linearity of expectation}
\]

\[
= 0.
\]

\[
R_{\tilde{X}} (i, j) := E \left( \tilde{X} (i) \tilde{X} (j) \right) - E \left( \tilde{X} (i) \right) E \left( \tilde{X} (j) \right)
\]

\[
= E \left( \sum_{k=1}^{i} \sum_{l=1}^{j} \tilde{S} (k) \tilde{S} (l) \right)
\]

\[
= E \left( \sum_{k=1}^{\min\{i,j\}} \tilde{S} (k)^2 + \sum_{k=1}^{i} \sum_{l=1}^{j} \tilde{S} (k) \tilde{S} (l) \right)
\]

\[
= \sum_{k=1}^{\min\{i,j\}} 1 + \sum_{k=1}^{i} \sum_{l=1}^{j} E \left( \tilde{S} (k) \right) E \left( \tilde{S} (l) \right)
\]

\[
= \min \{i, j\}, \quad \text{(43)}
\]

where (42) follows from linearity of expectation and independence. \( \square \)

The variance of \( \tilde{X} \) at \( i \) equals \( R_{\tilde{X}} (i, i) = i \) which means that the standard deviation of the random walk scales as \( \sqrt{i} \).

---

**Example 4.7 (Gambler).** A gambler is playing the following game. A fair coin is flipped sequentially. Every time the result is heads the gambler wins a dollar, every time it lands on tails she loses a dollar. We can model the amount of money earned (or lost) by the gambler as a random walk, as long as the flips are independent. This allows us to estimate that the expected gain equals zero or that the probability that the gambler is up 6 dollars or more after the first 10 flips is

\[
P (\text{gambler is up } $6 \text{ or more}) = p_{\tilde{X}(10)} (6) + p_{\tilde{X}(10)} (8) + p_{\tilde{X}(10)} (10)
\]

\[
= \binom{10}{8} \frac{1}{2^{10}} + \binom{10}{9} \frac{1}{2^{10}} + \frac{1}{2^{10}}
\]

\[
= 5.47 \times 10^{-2}. \quad \text{(46)}
\]
4.5 Markov chains

We begin by defining the Markov property, which is satisfied by any random process for which the future is conditionally independent from the past given the present.

**Definition 4.8 (Markov property).** A random process satisfies the Markov property if
\[
\tilde{X}(t_{i+1}) \text{ is conditionally independent of } \tilde{X}(t_1), \ldots, \tilde{X}(t_i) \text{ given } \tilde{X}(t_i) \text{ for any } t_1 < t_2 < \ldots < t_i < t_{i+1}.
\]
If the state space of the random process is discrete, then for any \(x_1, x_2, \ldots, x_{i+1}\)
\[
p_{\tilde{X}(t_{i+1})|\tilde{X}(t_1), \tilde{X}(t_2), \ldots, \tilde{X}(t_i)}(x_{n+1}|x_1, x_2, \ldots, x_n) = p_{\tilde{X}(t_{i+1})|\tilde{X}(t_i)}(x_{i+1}|x_i).
\]
If the state space of the random process is continuous (and the distribution has a joint pdf),
\[
f_{\tilde{X}(t_{i+1})|\tilde{X}(t_1), \tilde{X}(t_2), \ldots, \tilde{X}(t_i)}(x_{i+1}|x_1, x_2, \ldots, x_i) = f_{\tilde{X}(t_{i+1})|\tilde{X}(t_i)}(x_{i+1}|x_i).
\]
Any iid sequence satisfies the Markov property, since all conditional pmfs or pdfs are just equal to the marginals. The random walk also satisfies the property, since once we fix where the walk is at a certain time \(i\) the path that it took before \(i\) has no influence in its next steps.

**Lemma 4.9.** The random walk satisfies the Markov property.

*Proof.* Let \(\tilde{X}\) denote the random walk defined in Section 4.4. Conditioned on \(\tilde{X}(j) = x_i\) for \(j \leq i\), \(\tilde{X}(i+1)\) equals \(x_i + \hat{S}(i+1)\). This does not depend on \(x_1, \ldots, x_{i-1}\), which implies (47).

A Markov chain is a random process that satisfies the Markov property. In these notes we will consider discrete-time Markov chains with a finite state space, which means that the process can only take a finite number of values at any given time point. To specify such a Markov chain, we only need to define the pmf of the random process at its starting point (which we will assume is at \(i = 0\)) and its transition probabilities. This follows from the Markov property, since for any \(n \geq 0\)
\[
p_{\tilde{X}(0), \tilde{X}(1), \ldots, \tilde{X}(n)}(x_0, x_1, \ldots, x_n) := \prod_{i=0}^{n} p_{\tilde{X}(i)|\tilde{X}(0), \ldots, \tilde{X}(i-1)}(x_i|x_0, \ldots, x_{i-1})
\]
\[
= \prod_{i=0}^{n} p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_i|x_{i-1}).
\]
If these transition probabilities are the same at every time step (i.e. they are constant and do not depend on \( i \)), then the Markov chain is said to be \textbf{time homogeneous}. In this case, we can store the probability of each possible transition in an \( s \times s \) matrix \( T_X \), where \( s \) is the number of states.

\[
(T_X)_{jk} := p_X(i+1) | X(i) (x_j | x_k).
\]

In the rest of this section we will focus on time-homogeneous finite-state Markov chains. The transition probabilities of these chains can be visualized using a state diagram, which shows each state and the probability of every possible transition. See Figure 6 below for an example.

To simplify notation we define an \( s \)-dimensional vector \( \tilde{p}_X(i) \) called the \textbf{state vector}, which contains the marginal pmf of the Markov chain at each time \( i \),

\[
\tilde{p}_X(i) :=
\begin{bmatrix}
p_X(i)(x_1) \\
p_X(i)(x_2) \\
\vdots \\
p_X(i)(x_s)
\end{bmatrix}.
\]

Each entry in the state vector contains the probability that the Markov chain is in that particular state at time \( i \). It is \textit{not} the value of the Markov chain, which is a random variable.

The initial state space \( \tilde{p}_X(0) \) and the transition matrix \( T_X \) suffice to completely specify a time-homogeneous finite-state Markov chain. Indeed, we can compute the joint distribution of the chain at any \( n \) time points \( i_1, i_2, \ldots, i_n \) for any \( n \geq 1 \) from \( \tilde{p}_X(0) \) and \( T_X \) by applying (50) and marginalizing over any times that we are not interested in. We illustrate this in the following example.

---

\textbf{Example 4.10 (Car rental).} A car-rental company hires you to model the location of their cars. The company operates in Los Angeles, San Francisco and San Jose. Customers regularly take a car in a city and drop it off in another. It would be very useful for the company to be able to compute how likely it is for a car to end up in a given city. You decide to model the location of the car as a Markov chain, where each time step corresponds to a new customer taking the car. The company allocates new cars evenly between the three cities. The transition probabilities, obtained from past data, are given by
Figure 6: State diagram of the Markov chain described in Example (4.10) (top). Each arrow shows the probability of a transition between the two states. Below we show three realizations of the Markov chain.
San Francisco Los Angeles San Jose
(0.6 0.1 0.3) San Francisco
(0.2 0.8 0.3) Los Angeles
(0.2 0.1 0.4) San Jose

To be clear, the probability that a customer moves the car from San Francisco to LA is 0.2, the probability that the car stays in San Francisco is 0.6, and so on.

The initial state vector and the transition matrix of the Markov chain are
\[
\vec{p}_{\tilde{X}(0)} := \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \quad T_{\tilde{X}} := \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.2 & 0.1 & 0.4 \end{bmatrix}.
\]

State 1 is assigned to San Francisco, state 2 to Los Angeles and state 3 to San Jose. Figure 6 shows a state diagram of the Markov chain. Figure 6 shows some realizations of the Markov chain.

The researcher now wishes to estimate the probability that the car starts in San Francisco and is in San Jose after the second customer. This is given by
\[
p_{\tilde{X}(0),\tilde{X}(2)} (1, 3) = \sum_{i=1}^{3} p_{\tilde{X}(0),\tilde{X}(1),\tilde{X}(2)} (1, i, 3)
= \sum_{i=1}^{3} p_{\tilde{X}(0)} (1) p_{\tilde{X}(1)|\tilde{X}(0)} (i|1) p_{\tilde{X}(2)|\tilde{X}(1)} (3|i)
= \left(\vec{p}_{\tilde{X}(0)}\right) \sum_{i=1}^{3} (T_{\tilde{X}})_{i1} (T_{\tilde{X}})_{3i}
= 0.6 \cdot 0.2 + 0.2 \cdot 0.1 + 0.2 \cdot 0.4 \approx 7.33 \cdot 10^{-2}.
\]

The probability is 7.33%.

The following lemma provides a simple expression for the state vector at time \(i\) \(\vec{p}_{\tilde{X}(i)}\) in terms of \(T_{\tilde{X}}\) and the previous state vector.

**Lemma 4.11** (State vector and transition matrix). For a Markov chain \(\tilde{X}\) with transition matrix \(T_{\tilde{X}}\)
\[
\vec{p}_{\tilde{X}(i)} = T_{\tilde{X}} \vec{p}_{\tilde{X}(i-1)}.
\]
If the Markov chain starts at time 0 then

\[ \vec{p}_{\tilde{X}(i)} = T^i_{\tilde{X}} \vec{p}_{\tilde{X}(0)}, \quad (59) \]

where \( T^i_{\tilde{X}} \) denotes multiplying \( i \) times by matrix \( T_{\tilde{X}} \).

**Proof.** The proof follows directly from the definitions,

\[
\vec{p}_{\tilde{X}(i)} := \begin{bmatrix}
p_{\tilde{X}(i)}(x_1) \\
p_{\tilde{X}(i)}(x_2) \\
\vdots \\
p_{\tilde{X}(i)}(x_s)
\end{bmatrix} = \begin{bmatrix}
\sum_{j=1}^s p_{\tilde{X}(i-1)}(x_j) p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_1|x_j) \\
\sum_{j=1}^s p_{\tilde{X}(i-1)}(x_j) p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_2|x_j) \\
\vdots \\
\sum_{j=1}^s p_{\tilde{X}(i-1)}(x_j) p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_s|x_j)
\end{bmatrix}
\]

\[ = \begin{bmatrix}
p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_1|x_1) & p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_1|x_2) & \cdots & p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_1|x_s) \\
p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_2|x_1) & p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_2|x_2) & \cdots & p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_2|x_s) \\
\vdots & \vdots & \ddots & \vdots \\
p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_s|x_1) & p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_s|x_2) & \cdots & p_{\tilde{X}(i)|\tilde{X}(i-1)}(x_s|x_s)
\end{bmatrix}
\begin{bmatrix}
p_{\tilde{X}(i-1)}(x_1) \\
p_{\tilde{X}(i-1)}(x_2) \\
\vdots \\
p_{\tilde{X}(i-1)}(x_s)
\end{bmatrix}
\]

\[ = T_{\tilde{X}} \vec{p}_{\tilde{X}(i-1)} \quad (61) \]

Equation (59) is obtained by applying (58) \( i \) times and taking into account the Markov property.

---

**Example 4.12** (Car rental (continued)). The company wants to estimate the distribution of locations right after the 5th customer has used a car. Applying Lemma 4.11 we obtain

\[ \vec{p}_{\tilde{X}(5)} = T^5_{\tilde{X}} \vec{p}_{\tilde{X}(0)} \]

\[
= \begin{bmatrix}
0.281 \\
0.534 \\
0.185
\end{bmatrix}.
\]

The model estimates that after 5 customers more than half of the cars are in Los Angeles.

---

The states of a Markov chain can be classified depending on whether the Markov chain may eventually stop visiting them or not.
Definition 4.13 (Recurrent and transient states). Let $\tilde{X}$ be a time-homogeneous finite-state Markov chain. We consider a particular state $x$. If

$$P \left( \tilde{X} (j) = s \text{ for some } j > i \mid \tilde{X} (i) = s \right) = 1$$  \hspace{1cm} (64)

then the state is recurrent. In words, given that the Markov chain is at $x$, the probability that it returns to $x$ is one. In contrast, if

$$P \left( \tilde{X} (j) \neq s \text{ for all } j > i \mid \tilde{X} (i) = s \right) > 0$$  \hspace{1cm} (65)

the state is transient. Given that the Markov chain is at $x$, there is nonzero probability that it will never return.

The following example illustrates the difference between recurrent and transient states.

Example 4.14 (Employment dynamics). A researcher is interested in modeling the employment dynamics of young people using a Markov chain.

She determines that at age 18 a person is either a student with probability 0.9 or an intern with probability 0.1. After that she estimates the following transition probabilities:

$$
\begin{bmatrix}
0.8 & 0.5 & 0 & 0 \\
0.1 & 0.5 & 0 & 0 \\
0.1 & 0 & 0.9 & 0.4 \\
0 & 0 & 0.1 & 0.6
\end{bmatrix}
$$

The Markov assumption is obviously not completely precise, someone who has been a student for longer is probably less likely to remain a student, but such Markov models are easier to fit (we only need to estimate the transition probabilities) and often yield useful insights.

The initial state vector and the transition matrix of the Markov chain are

$$
\vec{p}_{\tilde{X}(0)} := \begin{bmatrix} 0.9 \\ 0.1 \\ 0 \\ 0 \end{bmatrix}, \quad T_{\tilde{X}} := \begin{bmatrix} 0.8 & 0.5 & 0 & 0 \\
0.1 & 0.5 & 0 & 0 \\
0.1 & 0 & 0.9 & 0.4 \\
0 & 0 & 0.1 & 0.6 \end{bmatrix}.
$$  \hspace{1cm} (66)

Figure 7 shows the state diagram and some realizations of the Markov chain.
Figure 7: State diagram of the Markov chain described in Example (4.14) (top). Below we show three realizations of the Markov chain.
States 1 (student) and 2 (intern) are transient states. Note that the probability that the Markov chain returns to those states after visiting state 3 (employed) is zero, so

\[
P\left(\tilde{X}(j) \neq 1 \text{ for all } j > i \mid \tilde{X}(i) = 1\right) \geq P\left(\tilde{X}(i + 1) = 3 \mid \tilde{X}(i) = 1\right) = 0.1 > 0,
\]

\[
P\left(\tilde{X}(j) \neq 2 \text{ for all } j > i \mid \tilde{X}(i) = 2\right) \geq P\left(\tilde{X}(i + 2) = 3 \mid \tilde{X}(i) = 2\right) = 0.4 \cdot 0.1 > 0.
\]

In contrast, states 3 and 4 (unemployed) are recurrent. We prove this for state 3 (the argument for state 4 is exactly the same):

\[
P\left(\tilde{X}(j) \neq 3 \text{ for all } j > i \mid \tilde{X}(i) = 3\right) = P\left(\tilde{X}(j) = 4 \text{ for all } j > i \mid \tilde{X}(i) = 3\right) = \lim_{k \to \infty} P\left(\tilde{X}(i + 1) = 4 \mid \tilde{X}(i) = 3\right) \prod_{j=1}^{k} P\left(\tilde{X}(i + j + 1) = 4 \mid \tilde{X}(i + j) = 4\right) = \lim_{k \to \infty} 0.1 \cdot 0.6^k = 0.
\]

In this example, it is not possible to reach the states student and intern from the states employed or unemployed. Markov chains for which there is a possible transition between any two states (even if it is not direct) are called irreducible.

**Definition 4.15 (Irreducible Markov chain).** A time-homogeneous finite-state Markov chain is irreducible if for any state \( x \), the probability of reaching every other state \( y \neq x \) in a finite number of steps is nonzero, i.e. there exists \( m \geq 0 \) such that

\[
P\left(\tilde{X}(i + m) = y \mid \tilde{X}(i) = x\right) > 0.
\]

One can easily check that the Markov chain in Example 4.10 is irreducible, whereas the one in Example 4.14. An important result is that all states in an irreducible Markov chain are recurrent.

**Theorem 4.16 (Irreducible Markov chains).** All states in an irreducible Markov chain are recurrent.

The result is proved in Section D of the appendix.

We end this section by defining the period of a state.
Figure 8: State diagram of a Markov chain where states the states have period two.

**Definition 4.17** (Period of a state). Let $\tilde{X}$ be a time-homogeneous finite-state Markov chain and $x$ a state of the Markov chain. The period $m$ of $x$ is the smallest integer such that it is only possible to return to $x$ in a number of steps that is a multiple of $m$, i.e. $km$ for some positive integer $k$ with nonzero probability.

Figure 8 shows a Markov chain where the states have a period equal to two. Aperiodic Markov chains do not contain states with periods greater than one.

**Definition 4.18** (Aperiodic Markov chain). A time-homogeneous finite-state Markov chain $\tilde{X}$ is aperiodic if all states have period equal to one.

The Markov chains in Examples 4.10 and 4.14 are both aperiodic.

**A  Proof of Lemma 4.2**

We begin by deriving the cdf of $T$,

$$F_T(t) := P(T \leq t) = 1 - P(T > t) = 1 - P(\text{no events in an interval of length } t) = 1 - e^{-\lambda t}$$

because the number of points in an interval of length $t$ follows a Poisson distribution with parameter $\lambda t$. Differentiating we conclude that

$$f_T(t) = \lambda e^{-\lambda t}.$$  

**B  Proof of Lemma 4.3**

By definition the number of events between 0 and $t$ is distributed as a Poisson random variables with parameter $\lambda t$ and hence its mean is equal to $\lambda t$. 

23
The autocovariance equals

\[ R_{\tilde{X}}(t_1, t_2) := E\left( \tilde{X}(t_1) \tilde{X}(t_2) \right) - E\left( \tilde{X}(t_1) \right) E\left( \tilde{X}(t_2) \right) \]

(82)

\[ = E\left( \tilde{X}(t_1) \tilde{X}(t_2) \right) - \lambda^2 t_1 t_2. \]

(83)

By assumption \( \tilde{X}(t_1) \) and \( \tilde{X}(t_2) - \tilde{X}(t_1) \) are independent so that

\[ E\left( \tilde{X}(t_1) \tilde{X}(t_2) \right) = E\left( \tilde{X}(t_1) \right) E\left( \tilde{X}(t_2) - \tilde{X}(t_1) \right) + E\left( \tilde{X}(t_1)^2 \right) \]

(84)

\[ = \lambda^2 t_1 (t_2 - t_1) + \lambda t_1 + \lambda^2 t_1^2 \]

(85)

\[ = \lambda^2 t_1 t_2 + \lambda t_1. \]

(87)

C Proof of Lemma 4.5

Let us define the number of positive steps \( S_+ \) that the random walk takes. Given the assumptions on \( \tilde{S} \), this is a binomial random variable with parameters \( i \) and \( 1/2 \). The number of negative steps is \( S_- := i - S_+ \). In order for \( \tilde{X}(i) \) to equal \( x \) we need for the net number of steps to equal \( x \), which implies

\[ x = S_+ - S_- \]

(88)

\[ = 2S_+ - i. \]

(89)

This means that \( S_+ \) must equal \( \frac{i + x}{2} \). We conclude that

\[ p_{\tilde{X}(i)}(i) = P\left( \sum_{j=0}^{i} \tilde{S}(i) = x \right) \]

(90)

\[ = \left( \begin{array}{c} i \\ \frac{i + x}{2} \end{array} \right) \frac{1}{2^i} \quad \text{if } \frac{i + x}{2} \text{ is an integer between 0 and } i. \]

(91)

D Proof of Theorem 4.16

In any finite-state Markov chain there must be at least one state that is recurrent. If all the states are transient there is a nonzero probability that it leaves all of the states forever, which is not possible. Without loss of generality let us assume that state \( x \) is recurrent. We
will now prove that another arbitrary state $y$ must also be recurrent. To alleviate notation let

$$p_{x,x} := P \left( \tilde{X} (j) = x \text{ for some } j > i \mid \tilde{X} (i) = x \right),$$

(92)

$$p_{x,y} := P \left( \tilde{X} (j) = y \text{ for some } j > i \mid \tilde{X} (i) = x \right).$$

(93)

The chain is irreducible so there is a nonzero probability $p_m > 0$ of reaching $y$ from $x$ in at most $m$ steps for some $m > 0$. The probability that the chain goes from $x$ to $y$ and never goes back to $x$ is consequently at least $p_m \left(1 - p_{y,x}\right)$. However, $x$ is recurrent, so this probability must be zero! Since $p_m > 0$ this implies $p_{y,x} = 1$.

Consider the following event:

1. $\tilde{X}$ goes from $y$ to $x$.
2. $\tilde{X}$ does not return to $y$ in $m$ steps after reaching $x$.
3. $\tilde{X}$ eventually reaches $x$ again at a time $m' > m$.

The probability of this event is equal to $p_{y,x} \left(1 - p_m\right) p_{x,x} = 1 - p_m$ (recall that $x$ is recurrent so $p_{x,x} = 1$). Now imagine that steps 2 and 3 repeat $k$ times, i.e. that $\tilde{X}$ fails to go from $x$ to $y$ in $m$ steps $k$ times. The probability of this event is $p_{y,x} \left(1 - p_m\right)^k p_{x,x}^k = \left(1 - p_m\right)^k$. Taking $k \to \infty$ we have that the probability that $\tilde{X}$ does not eventually return to $x$ must be zero.