Background: Markov chains

Definition

If we have a sequence of random variables $X_1, X_2, \ldots X_n$ where the indices $1, 2, \ldots, n$ represent successive points in time, we can use the chain rule of probability to calculate the probability of the entire sequence:

$$p(X_1, X_2, \ldots X_n) = p(X_1) \cdot p(X_2|X_1) \cdot p(X_3|X_2, X_1) \cdot \ldots \cdot p(X_n|X_{n-1}, X_{n-2}, \ldots, X_2, X_1).$$

Markov chains simplify this expression by using the *Markov assumption*. The assumption is that given the entire past history, the probability distribution for the random variable at the next time step *only* depends on the current variable. Mathematically, the assumption is written like this:

$$p(X_{t+1}|X_t, X_{t-1}, \ldots, X_2, X_1) = p(X_{t+1}|X_t)$$

for all $t = 2, \ldots, n$. Under this assumption, we can write the first expression as

$$p(X_1, X_2, \ldots X_n) = p(X_1) \cdot p(X_2|X_1) \cdot p(X_3|X_2) \cdot p(X_4|X_3) \cdot \ldots \cdot p(X_n|X_{n-1}),$$

which is much simpler than the original. It consists of an initial distribution for the first variable, $p(X_1)$, and $n-1$ transition probabilities. We usually make one more assumption: that the transition probabilities do not change with time. Hence, the transition from time t to time $t+1$ depends only on the value of X_t.

Examples of Markov chains

Discrete Markov chain

Suppose you have a secret number (make it an integer) between 1 and 5. We will call it your initial number at step 1. Now for each time step, your secret number will change according to the following rules:

1. Flip a coin.
2.   a. If the coin turns up heads, then increase your secret number by one (5 increases to 1).
       b. If the coin turns up tails, then decrease your secret number by one (1 decreases to 5).
3. Repeat n times, and record the evolving history of your secret number.

Before the experiment, we can think of the sequence of secret numbers as a sequence of random variables, each taking on a value in $\{1, 2, 3, 4, 5\}$. Assume that the coin is fair, so that with each flip, the probability of heads and tails are both 0.5.

Does this game qualify as a true Markov chain? Suppose your secret number is currently 4 and that the history of your secret numbers is $(2, 1, 2, 3)$. What is the probability that on the next step, your secret number will be 5? What about the other four possibilities? Because of the rules of this game, the probability of the next transition will depend only on the fact that your current number is 4. The numbers further back in your history are irrelevant, so this is a Markov chain.

This is an example of a discrete Markov chain, where the possible values of the random variables come from a discrete set. Those possible values (secret numbers in this example) are called states of the chain. The states are usually numbers, as in this example, but they can represent anything. In one common example, the states describe the weather on a particular day, which could be labeled as 1-fair, 2-poor.

Random walk (continuous)

Now let's look at a continuous example of a Markov chain. Say $X_t = 0$ and we have the following transition model: $p(X_{t+1}|X_t = x_t) = \mathrm{N}(x_t, 1)$. That is, the probability distribution for the next state is Normal with variance 1 and mean equal to the current state. This is often referred to as a "random walk." Clearly, it is a Markov chain because the transition to the next state X_{t+1} only depends on the current state X_t.
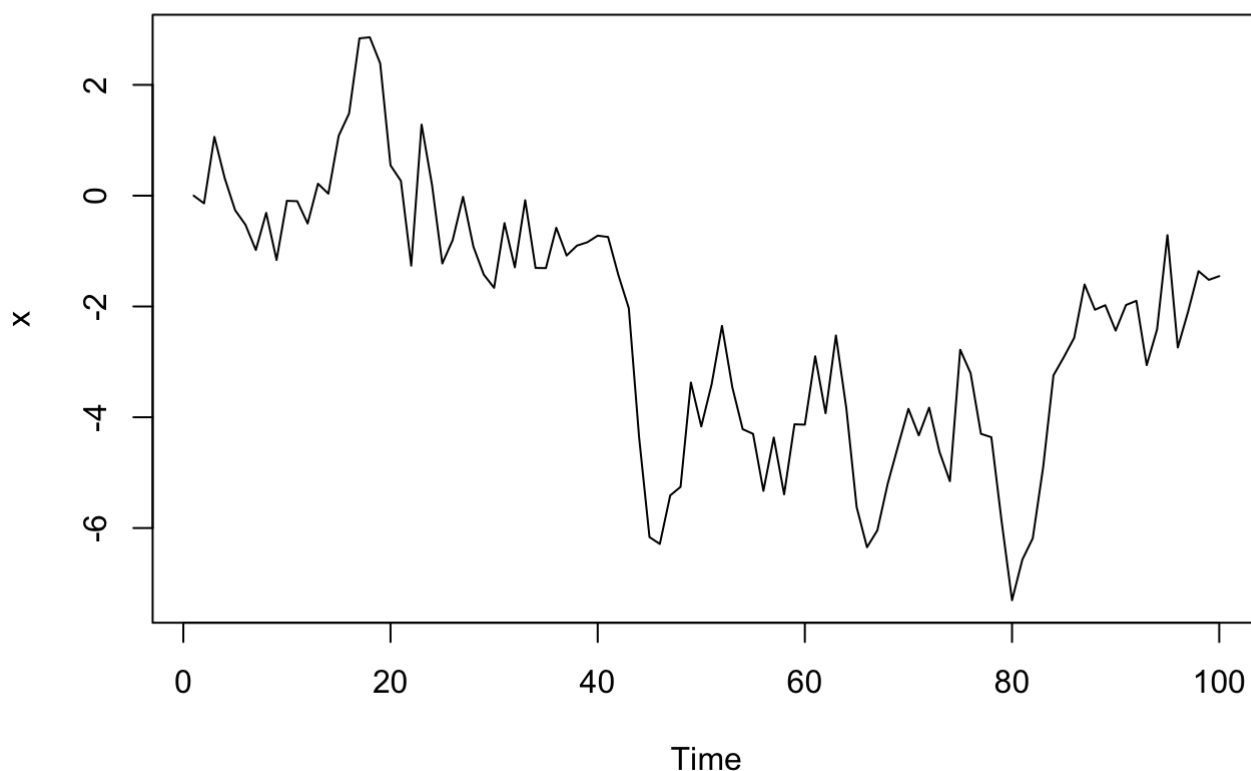
This example is straightforward to code in `R` :

```r
set.seed(34)

n = 100
x = numeric(n)

for (i in 2:n) {
  x[i] = rnorm(1, mean=x[i-1], sd=1.0)
}

plot.ts(x)
```


Transition matrix

Let's return to our example of the discrete Markov chain. If we assume that transition probabilities do not change with time, then there are a total of 25 $\left(5^2\right)$ potential transition probabilities. Potential transition probabilities would be from State 1 to State 2, State 1 to State 3, and so forth. These transition probabilities can be arranged into a matrix Q:

$$Q = \begin{pmatrix} 0 & .5 & 0 & 0 & .5 \\ .5 & 0 & .5 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & .5 & 0 & .5 \\ .5 & 0 & 0 & .5 & 0 \end{pmatrix}$$

where the transitions *from* State 1 are in the first row, the transitions from State 2 are in the second row, etc. For example, the probability $p(X_{t+1} = 5 \mid X_t = 4)$ can be found in the fourth row, fifth column.

The transition matrix is especially useful if we want to find the probabilities associated with multiple steps of the chain. For example, we might want to know $p(X_{t+2} = 3 \mid X_t = 1)$, the probability of your secret number being 3 two steps from now, given that your number is currently 1. We can calculate this as $\sum_{k=1}^{5} p(X_{t+2} = 3 \mid X_{t+1} = k) \cdot p(X_{t+1} = k \mid X_t = 1)$, which conveniently is found in the first row and third column of Q^2.

We can perform this matrix multiplication easily in `R` :

```
Q = matrix(c(0.0, 0.5, 0.0, 0.0, 0.5,
             0.5, 0.0, 0.5, 0.0, 0.0,
             0.0, 0.5, 0.0, 0.5, 0.0,
             0.0, 0.0, 0.5, 0.0, 0.5,
             0.5, 0.0, 0.0, 0.5, 0.0),
           nrow=5, byrow=TRUE)

Q %*% Q # Matrix multiplication in R.  This is Q^2.
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 0.50 0.00 0.25 0.25 0.00
## [2,] 0.00 0.50 0.00 0.25 0.25
## [3,] 0.25 0.00 0.50 0.00 0.25
## [4,] 0.25 0.25 0.00 0.50 0.00
## [5,] 0.00 0.25 0.25 0.00 0.50
```

```
(Q %*% Q)[1,3]
```

```
## [1] 0.25
```

Therefore, if your secret number is currently 1, the probability that the number will be 3 two steps from now is .25.

Stationary distribution

Suppose we want to know the probability distribution of the your secret number in the distant future, say $p(X_{t+h} \mid X_t)$ where h is a large number. Let's calculate this for a few different values of h.

```
Q5 = Q %*% Q %*% Q %*% Q %*% Q # h=5 steps in the future
round(Q5, 3)
```

```
##        [,1]  [,2]  [,3]  [,4]  [,5]
## [1,] 0.062 0.312 0.156 0.156 0.312
## [2,] 0.312 0.062 0.312 0.156 0.156
## [3,] 0.156 0.312 0.062 0.312 0.156
## [4,] 0.156 0.156 0.312 0.062 0.312
## [5,] 0.312 0.156 0.156 0.312 0.062
```

```
Q10 = Q %*% Q %*% Q %*% Q %*% Q %*% Q %*% Q %*% Q %*% Q # h=10 steps in the future
round(Q10, 3)
```

```
##        [,1]  [,2]  [,3]  [,4]  [,5]
## [1,] 0.248 0.161 0.215 0.215 0.161
## [2,] 0.161 0.248 0.161 0.215 0.215
## [3,] 0.215 0.161 0.248 0.161 0.215
## [4,] 0.215 0.215 0.161 0.248 0.161
## [5,] 0.161 0.215 0.215 0.161 0.248
```

```
Q30 = Q
for (i in 2:30) {
  Q30 = Q30 %*% Q
}
round(Q30, 3) # h=30 steps in the future
```

```
##        [,1]  [,2]  [,3]  [,4]  [,5]
## [1,] 0.201 0.199 0.200 0.200 0.199
## [2,] 0.199 0.201 0.199 0.200 0.200
## [3,] 0.200 0.199 0.201 0.199 0.200
## [4,] 0.200 0.200 0.199 0.201 0.199
## [5,] 0.199 0.200 0.200 0.199 0.201
```

Notice that as the future horizon gets more distant, the transition distributions appear to converge. The state you are currently in becomes less important in determining the more distant future. If we let h get really large, and take it to the limit, all the rows of the long-range transition matrix will become equal to $(.2, .2, .2, .2, .2)$. That is, if you run the Markov chain for a very long time, the probability that you will end up in any particular state is $1/5 = .2$ for each of the five states. These long-range probabilities are equal to what is called the **stationary** distribution of the Markov chain.

The stationary distribution of a chain is the initial state distribution for which performing a transition will not change the probability of ending up in any given state. That is,

```
c(0.2, 0.2, 0.2, 0.2, 0.2) %*% Q
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  0.2  0.2  0.2  0.2  0.2
```

One consequence of this property is that once a chain reaches its stationary distribution, the stationary distribution will remain the distribution of the states thereafter.

We can also demonstrate the stationary distribution by simulating a long chain from this example.

```
n = 5000
x = numeric(n)
x[1] = 1 # fix the state as 1 for time 1
for (i in 2:n) {
  x[i] = sample.int(5, size=1, prob=Q[x[i-1],]) # draw the next state from the intergers 1 to
  5 with probabilities from the transition matrix Q, based on the previous value of X.
}
```

Now that we have simulated the chain, let's look at the distribution of visits to the five states.

```
table(x) / n
```

```
## x
##      1      2      3      4      5
## 0.1996 0.2020 0.1980 0.1994 0.2010
```

The overall distribution of the visits to the states is approximately equal to the stationary distribution.

As we have just seen, if you simulate a Markov chain for many iterations, the samples can be used as a Monte Carlo sample from the stationary distribution. This is exactly how we are going to use Markov chains for Bayesian inference. In order to simulate from a complicated posterior distribution, we will set up and run a Markov chain whose stationary distribution is the posterior distribution.

It is important to note that the stationary distribution doesn't always exist for any given Markov chain. The Markov chain must have certain properties, which we won't discuss here. However, the Markov chain algorithms we'll use in future lessons for Monte Carlo estimation are guaranteed to produce stationary distributions.

Continuous example

The continuous random walk example we gave earlier does not have a stationary distribution. However, we can modify it so that it does have a stationary distribution.

Let the transition distribution be $p(X_{t+1}|X_t = x_t) = N(\phi x_t, 1)$ where $-1 < \phi < 1$. That is, the probability distribution for the next state is Normal with variance 1 and mean equal to ϕ times the current state. As long as ϕ is between -1 and 1, then the stationary distribution will exist for this model.

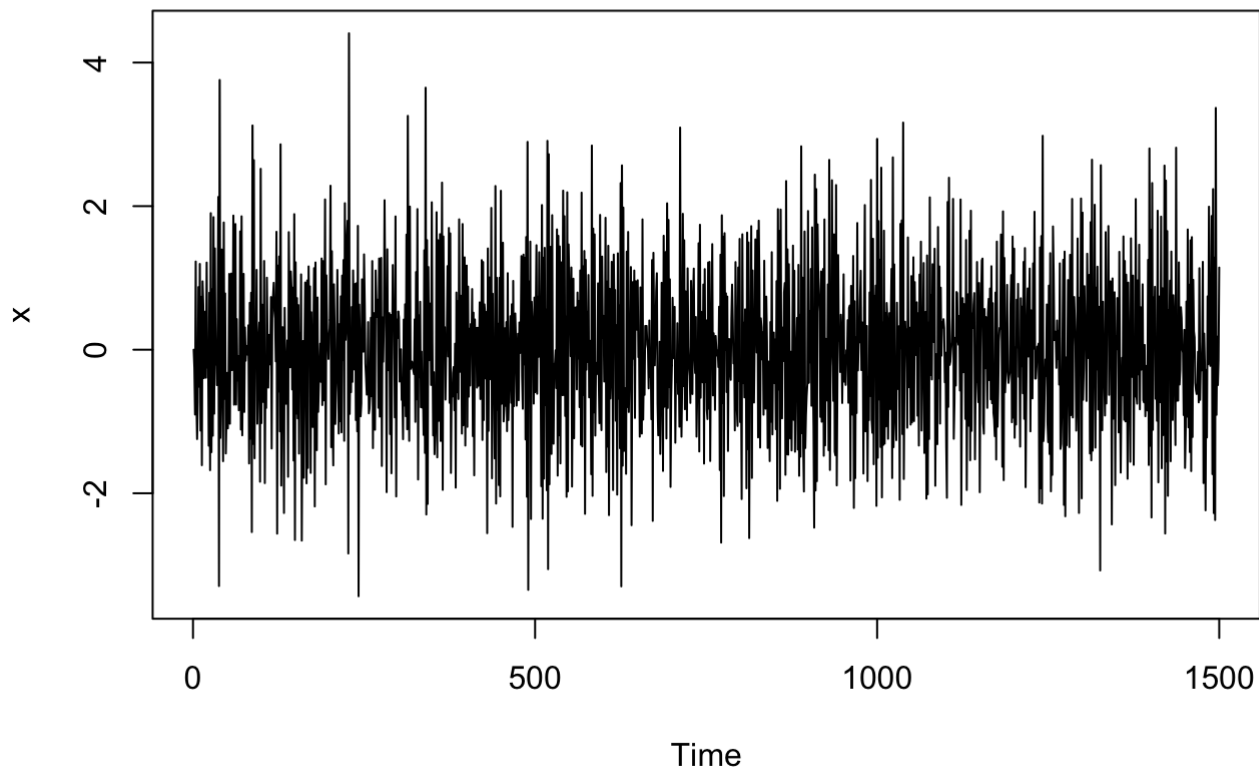Let's simulate this chain for $\phi = -0.6$.

```
set.seed(38)

n = 1500
x = numeric(n)
phi = -0.6

for (i in 2:n) {
  x[i] = rnorm(1, mean=phi*x[i-1], sd=1.0)
}

plot.ts(x)
```

The theoretical stationary distribution for this chain is normal with mean 0 and variance $1/(1 - \phi^2)$, which in our example approximately equals 1.562. Let's look at a histogram of our chain and compare that with the theoretical stationary distribution.

```
hist(x, freq=FALSE)
curve(dnorm(x, mean=0.0, sd=sqrt(1.0/(1.0-phi^2))), col="red", add=TRUE)
legend("topright", legend="theoretical stationary\ndistribution", col="red", lty=1, bty="n")
```

Histogram of x
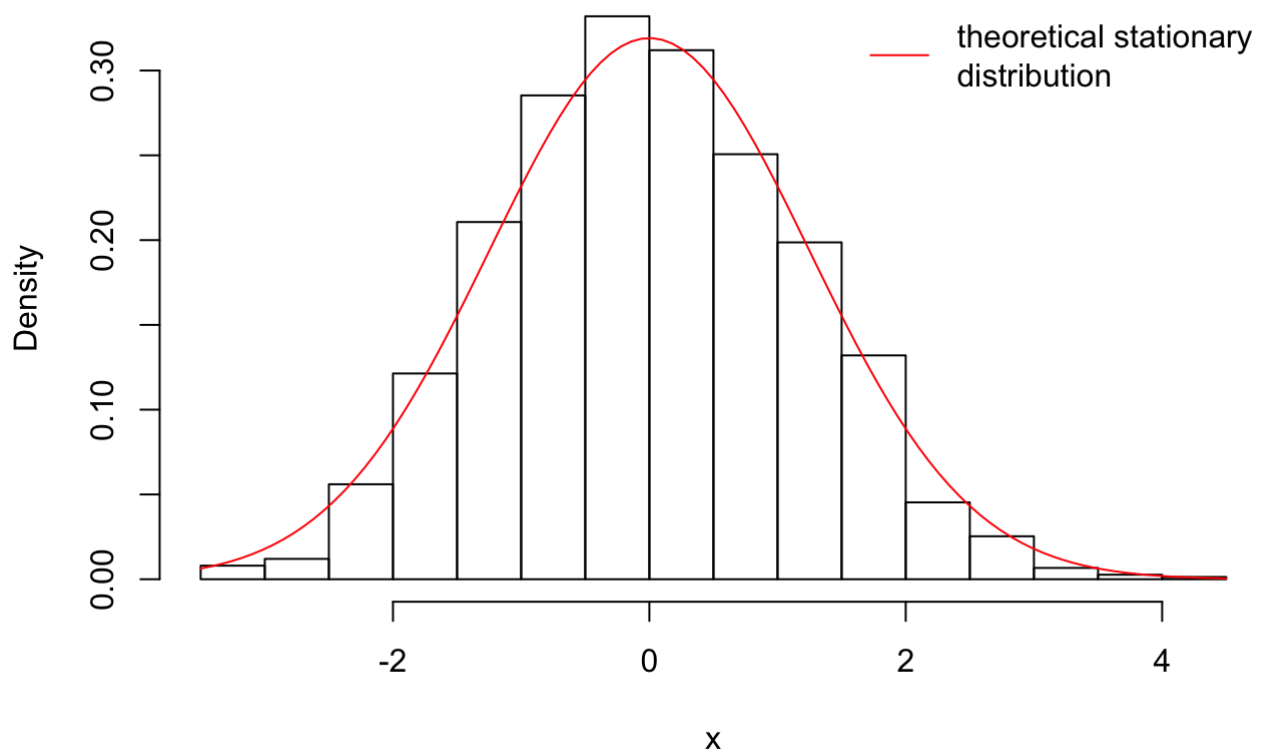
It appears that the chain has reached the stationary distribution. Therefore, we could treat this simulation from the chain like a Monte Carlo sample from the stationary distribution, a normal with mean 0 and variance 1.562.

Because most posterior distributions we will look at are continuous, our Monte Carlo simulations with Markov chains will be similar to this example.