Agenda

- Chaining together random variables
- Markov chains
- The long run of Markov chains

READING: Handouts on the class webpage
Multiple Random Variables

\texttt{rnorm}, \texttt{runif}, etc., give independent and identically distributed (IID) random variables

Most stochastic models don’t call for IID random variables

Varying distributions, dependence

How do we generate such things?
Try to arrange the variables in order of priority and/or time
Who someone votes for might change with their age or their race, but not vice versa
Try to arrange the variables in order of priority and/or time. Who someone votes for might change with their age or their race, but not vice versa. Generate the *exogenous* variables first.
Try to arrange the variables in order of priority and/or time.

Who someone votes for might change with their age or their race, but not vice versa.

Generate the **exogenous** variables first.

Then all the **endogenous** variables which only depend on exogenous ones.
Try to arrange the variables in order of priority and/or time. Who someone votes for might change with their age or their race, but not vice versa. Generate the exogenous variables first. Then all the endogenous variables which only depend on exogenous ones. Then all the variables which depend only on first-generation endogenous ones, etc.
Try to arrange the variables in order of priority and/or time.

Who someone votes for might change with their age or their race, but not vice versa.

Generate the **exogenous** variables first.

Then all the **endogenous** variables which only depend on exogenous ones.

Then all the variables which depend only on first-generation endogenous ones, etc.

You’ll see more of this with graphical models in 36-402.
Can have a sequence of variables going on in time, $X_1, X_2, \ldots X_n$
Earlier ones can cause later but not other way

$$p(X_1, X_2, \ldots X_n) = p(X_1)p(X_2|X_1)p(X_3|X_2, X_1)\ldots p(X_n|X_{n-1}, X_{n-2}, \ldots X_1)$$
The **Markov property**: Given the current state $X_t$, earlier states $X_{t-1}, X_{t-2}, \ldots$ are irrelevant to the future states $X_{t+1}, X_{t+2}, \ldots$
The **Markov property**: Given the current state $X_t$, earlier states $X_{t-1}, X_{t-2}, \ldots$ are irrelevant to the future states $X_{t+1}, X_{t+2}, \ldots$

\[ p(X_1, X_2, \ldots X_n) = p(X_1)p(X_2|X_1)p(X_3|X_2)\ldots p(X_n|X_{n-1}) \]
The **Markov property**: Given the current state $X_t$, earlier states $X_{t-1}, X_{t-2}, \ldots$ are irrelevant to the future states $X_{t+1}, X_{t+2}, \ldots$

\[ p(X_1, X_2, \ldots, X_n) = p(X_1)p(X_2|X_1)p(X_3|X_2)\ldots p(X_n|X_{n-1}) \]

This is an *assumption*, not a law of nature.
The **Markov property**: Given the current state $X_t$, earlier states $X_{t-1}, X_{t-2}, \ldots$ are irrelevant to the future states $X_{t+1}, X_{t+2}, \ldots$

\[ p(X_1, X_2, \ldots X_n) = p(X_1)p(X_2|X_1)p(X_3|X_2)\ldots p(X_n|X_{n-1}) \]

This is an *assumption*, not a law of nature

To simulate a Markov chain, we need to

- Draw the initial state $X_1$ from $p(X_1)$
- Draw $X_t$ from $p(X_t|X_{t-1})$ — inherently sequential
Inputs: number of steps, drawing function for initial distribution, drawing function for transition distribution

```r
rmrkov <- function(n,rinitial,rtransition) {
  x <- vector(length=n)
  x[1] <- rinitial()
  for (t in 2:n) {
    x[t] <- rtransition(x[t-1])
  }
  return(x)
}
```
Markov Chains

Each $X_t$ is discrete, not continuous
Represent $p(X_t|X_{t-1})$ in a transition matrix,
$q_{ij} = \Pr(X_t = j|X_{t-1} = i)$
Each row sums to 1 (stochastic matrix)
Each $X_t$ is discrete, not continuous
Represent $p(X_t|X_{t-1})$ in a transition matrix, $q_{ij} = \Pr(X_t = j|X_{t-1} = i)$
Each row sums to 1 (stochastic matrix)
Represent $p(X_1)$ as a vector $p_0$, summing to 1
Graph vs. matrix

\[ q = \begin{bmatrix} 0.5 & 0.5 \\ 0.75 & 0.25 \end{bmatrix} \]
rmarkovchain <- function(n,p0,q) {
  k <- length(p0)
  stopifnot(k==nrow(q),k==ncol(q),all.equal(rowSums(q),rep(1,time=k)))
  rinitial <- function() { sample(1:k,size=1,prob=p0) }
  rtransition <- function(x) { sample(1:k,size=1,prob=q[x,]) }
  return(rmarkov(n, rinitial, rtransition))
}

It runs:
> x <- rmarkovchain(1e4,c(0.5,0.5),q)
> head(x)
[1] 1 1 2 1 2 2

How do we know it works?
Your Basic Markov Chain

rmarkovchain <- function(n,p0,q) {
  k <- length(p0)
  stopifnot(k==nrow(q),k==ncol(q),all.equal(rowSums(q),rep(1,time=k)))
  rinitial <- function() { sample(1:k,size=1,prob=p0) }
  rtransition <- function(x) { sample(1:k,size=1,prob=q[x,]) }
  return(rmarkov(n,rinitial,rtransition))
}

It runs:

> x <- rmarkovchain(1e4,c(0.5,0.5),q)
> head(x)
[1] 1 1 2 1 2 2

How do we know it works?
Markov Chains
Invariance and the Long Run

Markov Property
Variations on the Theme

```r
> ones <- which(x[-1e4] == 1)
> twos <- which(x[-1e4] == 2)
> signif(table(x[ones + 1]) / length(ones), 3)
   1   2
0.489 0.511
> signif(table(x[twos + 1]) / length(twos), 3)
   1   2
0.752 0.248
```

vs. (0.5, 0.5) and (0.75, 0.25) ideally
Uses law of large numbers + conditional independence
Hidden Markov Model (HMM)

$X_t$ is Markov, but we see $Y_t = h(X_t) + \text{noise}$, not Markov

Example:

```r
> means <- c(10,-10)
> sds <- c(1,5)
> y <- rnorm(n=length(x),mean=means[x],sd=sds[x])
> signif(head(y),3)
[1] 11.00 10.00 -10.60 11.80 -16.30 -2.41
```

(noise and distortion might be much more complicated)
Interacting/coupled Markov chains: transition probability for chain 1 depends on its state and chain 2’s state
Variations

Interacting/coupled Markov chains: transition probability for chain 1 depends on its state and chain 2’s state
Continuous-time Markov chain: stay in the state for a random time, with exponential distribution, then take a chain step
Interacting/coupled Markov chains: transition probability for chain 1 depends on its state and chain 2’s state
Continuous-time Markov chain: stay in the state for a random time, with exponential distribution, then take a chain step
Semi-Markov chain: like CTMC, but non-exponential holding times
Variations

Interacting/coupled Markov chains: transition probability for chain 1 depends on its state and chain 2’s state
Continuous-time Markov chain: stay in the state for a random time, with exponential distribution, then take a chain step
Semi-Markov chain: like CTMC, but non-exponential holding times
Chain with complete connections: as in HMM, $Y_t = h(X_t) + \text{noise}$, but then $X_{t+1} = r(X_t, Y_t)$ (with no noise)
Invariant Distributions

$$p_1 = p_0q$$

$$p_2 = p_1q = p_0q^2$$

$$p_t = p_{t-1}q = p_0q^t$$

Fact: If the chain can go from any state to any other and back, and there are no fixed periods, then

$$p_t \to p_\infty = p_\infty q$$

$p_\infty = \text{left eigenvector of } q \text{ of eigenvalue 1}$

This is the **invariant distribution**
```r
> table(rmarkovchain(1e4,c(0.5,0.5),q))
  1  2
5999 4001
> table(rmarkovchain(1e4,c(0.5,0.5),q))
  1  2
5996 4004
> table(rmarkovchain(1e4,c(0,1),q))
  1  2
5989 4011
> table(rmarkovchain(1e4,c(1,0),q))
  1  2
5996 4004
```
Markov Chains
Invariance and the Long Run

> eigen(t(q))
$values
[1] 1.00 -0.25

$vectors
 [,1]        [,2]
[1,] 0.8320503 -0.7071068
[2,] 0.5547002  0.7071068

> eigen(t(q))$vectors[,1]/sum(eigen(t(q))$vectors[,1])
[1] 0.6 0.4
The Long Run of a Markov Chain

In the long run, all the $X_t$ come close to having the same distribution, the invariant distribution. They’re still dependent, though.

**Ergodic theorem:**

$$\frac{1}{n} \sum_{t=1}^{n} f(X_t) \to \sum_x p_\infty(x)f(x) = \mathbb{E}_{p_\infty}[f(X)]$$

Time averages converge on expected values.
Summary

1. Break complicated simulations into many draws from basic distributions
   - Make later draws depend on earlier ones
   - Use the Markov property when you can
2. Markov chains are the most basic non-trivial stochastic process
3. In the long run, Markov chains converge on their invariant distribution