Improved approximation obtained by paying attention to the edge of the histogram (RF) bars:

Normal approximation with continuity correction

\[ P(X \geq 1) = 1 - P(X < 0.5) = 1 - 0.219 = 0.781 \]

\[ = 0.781 \]

Correct answer: 0.761

much better approx.

Markov Chains

Recall the definition of a stochastic process:
Def. A sequence of r.v.s $X_1, X_2, \ldots$ is called a stochastic process with discrete time parameter $t = 1, 2, \ldots$. $X_0$ is the initial state of the process; $X_n$, $n \geq 1$, is the state of the process at time $t = n$. The simplest possible discrete-time stochastic process is an IID sequence of r.v.s $(X_1, X_2, \ldots)$.

Suppose that there is a parameter $\theta$ such that $(X_i | \theta)$ IID from some dist. depending on $\theta$. \[ \text{Q: Does this process have a memory?} \]
Example, machine with \( \Theta \) dial from \( \mathbb{R} \) to \( \{0, 1\} \), produces IID Bernoulli(\( \theta \))

_trials \( X_n \): The process \( (X_1, X_2, \ldots) \) does have a memory if \( \Theta \) is unknown to you: the information that 17 out of the first 20 trials were successes helps you to predict \( X_{21} \), because it's reasonable to conclude from \( X_1, \ldots, X_{20} \) that \( \Theta \) is around \( \frac{17}{20} = 0.85 \), so \( X_{21} \) will probably be a success. But the process \( \{(X_i; 10), i = 1, 2, \ldots\} \) has no memory once \( \Theta \) is known: information about
The first n trials is irrelevant to your prediction of $X_{n+1}$ if you know

$\theta$.

A white-noise (stochastic) process or a white noise time series is called a white-noise (stochastic) process or a white noise time series.

Q: What is the next level of complexity for discrete-time stochastic processes not from white noise? A: Allow $X_{n+1}$ to depend on $X_n$ but not on $X_{n-1}, X_{n-2}, \ldots$ (i.e., let the process have a short-term memory, i.e., time period back in the past).
From now on, I'll suppress the dependence of the process on t in the notation.

**Def.** A stochastic process is a **discrete-time** **(first-order) Markov chain** if for \( n = 1, 2, \ldots \); \( b \) any real number; and for all possible sequences of states \( x_1, x_2, \ldots \)

\[
P(\mathcal{X}_{n+1} \leq b \mid \mathcal{X}_1 = x_1, \ldots, \mathcal{X}_n = x_n) = P(\mathcal{X}_{n+1} \leq b \mid \mathcal{X}_n = x_n).\]

In other words, the only thing you need to know to simulate where the Markov chain is going next is where it is now.
(Can define higher-order Markov chains with memory of 2 or more time periods; we won't pursue that here.)

Def.

The set of values the Markov chain can take on is called its state space $\mathcal{S}$, which may be finite or infinite.

(Can also have Markov chains unfolding in continuous time, e.g. $X_t$ stock price at time $t$ in seconds, milliseconds, microseconds, ...; we also won't pursue that here.)

It's easy to write down the joint pdf of a Markov chain with finite

\[ \begin{array}{c}
\end{array} \]
Suppose you have a finite Markov chain with \( k \) states numbered \( 1, \ldots, k \). Let \( \mathbb{P}(X_{n+1} = j \mid X_n = i) \) be the transition probability from state \( i \) to state \( j \). A finite Markov chain is called a Markov chain with a finite state space if it has a finite number of states. A finite Markov chain is a Markov chain with a finite state space.
If \( P(X_{n+1} = j \mid X_n = i) \) is the same for all \( n \), the transition distribution is said to be stationary. If

\[
\sqrt{\text{time-homogeneous}}
\]

the Markov chain does have a stationary transition distribution, then the probability

\[
P_{ij} \triangleq P(X_{n+1} = j \mid X_n = i)
\]

completely characterize the Markov chain's behavior. Can arrange the \( P_{ij} \) to state \( \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{bmatrix} \).
All of the elements of $\mathbf{P}$ are non-negative (they're probabilities), and all of the row sums are 1 (because the chain has to go somewhere), i.e.

$$\sum_{j=1}^{k} p_{ij} = 1 \quad \text{for all } i = 1, \ldots, k.$$ 

A square matrix $\mathbf{P}$ with non-negative entries and row sums equal to 1 is called a stochastic matrix.

**Example:** Gene inheritance is Markovian. All we need to know to predict you is the genetic story of your parents.
Suppose that a gene of interest to you has two alleles, A and a. Then a state in the Markov chain is of the form \[ \left\{ \text{allele 1 \text{ allele 2 \text{ allele 3 \text{ allele 4}}} \right\}, \text{ for example } \{Aa, Aa\}. \] Ignoring order (because it's irrelevant in inheritance), there are 6 possible states: \{AA, AA\}, \{AA, Aa\}, \{AA, aa\}, \{Aa, Aa\}, \{Aa, aa\}, and \{aa, aa\}. 
Offspring gets $A$ or $a$ from parent 1 and $A$ or $a$ (independently) from parent 2.

Transition matrix:

<table>
<thead>
<tr>
<th></th>
<th>${AA, AA}$</th>
<th>${AA, Aa}$</th>
<th>${AA, aa}$</th>
<th>${Aa, Aa}$</th>
<th>${Aa, aa}$</th>
<th>${aa, aa}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${AA, AA}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>${AA, Aa}$</td>
<td>0.25</td>
<td>0.5</td>
<td>0</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>${AA, aa}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>${Aa, Aa}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>${Aa, aa}$</td>
<td>0.25</td>
<td>0.5</td>
<td>0</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>${aa, aa}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Example (random walk) You're watching a particle move around on the integers $\mathbb{Z} = \{ \ldots, -2, -1, 0, 1, 2, \ldots \}$ over time; here are the rules:

- wherever it is at time $t = n$, it moves left 1 unit with prob $p_1$,
- right 1 unit with prob $p_2$,
- and it stays where it is with prob $p_3$,

where $0 < p_i < 1$ and $\sum_{i=1}^{3} p_i = 1$. This is clearly a Markov chain (why?)

what is its transition matrix?
This is an example of a **tridiagonal matrix**, in which the only non-zero entries are on the diagonal and 1 diagonal either way from the main diagonal; since there are only 3 non-zero diagonals, \( P \) is said to be tridiagonal.
Moreover, all of the main diagonal entries are the same (p1); all of the entries below are also the same (p1); and all of the entries on diagonal above are also the same (p2).

Such matrices are called Toeplitz.

(Named after Otto Toeplitz, (1881-1940), a German mathematician who was fired by the Nazis from his university position in 1935 for being Jewish.)

Start this process, which is called a random walk, at 0 and let it go; where is the particle likely to be at time n, n large?
Suppose, for example, that $(p_1, p_2, p_3) = (0.1, 0.3, 0.6)$. Then you would expect the particle to drift off to $+\infty$. Similarly, $(p_1, p_2, p_3) = (0.5, 0.25, 0.25)$ should yield a drift to $-\infty$. $(p_1, p_2, p_3) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$?

Can show that as $n \to \infty$ every integer is visited infinitely many times, and the expected time you must wait for the chain to return to 0 (having started there) is also infinite.

The infinite random walk evidently has "too much freedom" to move around to get interesting results; let's bound it.
Bounded random walk

Restrict the Markov chain to \( S = \{ -k, -k+1, \ldots, -1, 0, 1, \ldots, k \} \), for some integer \( k \geq 1 \).

a: What to do at the boundaries?

One idea would be to wrap around: if you try to move to \((k+1)\), interpret that as a move to \(-k\); if you try to move to \(-(-k+1)\), move to \(+k\).

Transition matrix with \( k = 2 \):

\[
\begin{pmatrix}
-2 & p_2 & p_3 & 0 & 0 & p_1 \\
-1 & p_1 & p_2 & p_3 & 0 & 0 \\
0 & 0 & p_1 & p_2 & p_3 & 0 \\
1 & 0 & 0 & p_1 & p_2 & p_3 \\
2 & p_3 & 0 & 0 & p_1 & p_2 \\
\end{pmatrix}
\]

Another idea: keep trying until you make a legal move.
Back to a general finite Markov chain

Let \( p^{(m)}_{ij} \) \( \equiv p \) (chain move from \( i \) to \( j \) in \( m \) steps).

Theorem

\[
I = P(I_{n+m} = j | I_n = i)
\]

Finite Markov chain with stationary transition distributions & transition matrix \( P \rightarrow P^{(m)} \) is just the \((i,j)\) entry of the matrix \( P^m \), which is called the \( m \)-step transition matrix of the Markov chain.

Genetic example, continued.

\{AA, AA\} has the property that once the chain is in that state, it can't
so anywhere else; so does \( \{99, 99\} \). \( \square \)

This occurs for a state \( i \) when \( \pi_{ii} = 1 \).

Def. Any state with \( \pi_{ii} = 1 \) is called an absorbing state.

Notice that in this genetic Markov chain, states 2-5 all have positive probability of moving to state 1 in 2 steps, and the same is true of moving to state 6 in 2 steps. It follows that, if the chain is run long enough (simultaneously many generations), it will either end up
in state \{AA, AA\} or in state \{aa, aa\}.

Markov chains that settle down to a single long-run distribution are especially important in contemporary Bayesian computation; the long-run stable distribution is called the equilibrium distribution of the chain.

Important note on terminology: If call this distribution the stationary dist. of the chain, but this choice is unfortunate because they've already used stationary to mean something else.
DS: \[ \text{If } \pi_i \propto \left( x_{n+1} = j \mid x_n = i \right) \text{ for all } n, \text{ DS say that the transition distribution is stationary;} \]

other people call this time-homogeneous.

I'll use equilibrium distribution for the long-run behavior of Markov chains that settle down into a stable long-run story.

Where should the Markov chain start?

You can either initialize a Markov chain to a deterministic value or you can start random it off by making a draw from what's called the initial distribution of the Markov chain:
Any vector of non-negative numbers that add up to 1 is called a probability vector; any such vector whose components specify that a Markov chain will be in each possible state at time 1 is referred to as the initial distribution of the chain.

So: After 1 time step, the probability dist. over the Markov chain's possible states is \( \pi \); after 2 iterations the chain's dist. is \( \pi P \); after \((m+1)\) iterations its dist. is \( \pi P^m \); it would be nice if \( \pi P^m \) converged to a unique dist. as \( m \to \infty \).
be its equilibrium distribution.

Notice something interesting: if we choose \( \pi \) so that \( \pi \mathbb{P} = \pi \), then
\[
\pi \mathbb{P}^2 = (\pi \mathbb{P}) \pi = \pi \mathbb{P} = \pi, \quad \pi \mathbb{P}^3 = (\pi \mathbb{P}^2) \pi = \pi \mathbb{P} = \pi
\]
and so \( \lim_{m \to \infty} \pi \mathbb{P}^m = \pi \).

**Def.** Markov chain with transition matrix \( \mathbb{P} \) has a probability vector \( \pi \) such that \( \pi \mathbb{P} = \pi \) is an equilibrium dist. for the Markov chain under additional conditions on \( \mathbb{P} \), such an equilibrium dist. will be unique (we won't pursue that here).
How find \( \lambda \) so that \( \lambda \mathbf{P} = \mathbf{V} \)?

In linear algebra this is an example of an eigenvalue/eigenvector problem:

**Def.** Given a square matrix \( \mathbf{P} \), any vector \( \mathbf{v} \in \mathbb{R} \) satisfying \( \mathbf{P} \mathbf{v} = \lambda \mathbf{v} \) is called a right eigenvector of \( \mathbf{P} \) with eigenvalue \( \lambda \), and any vector \( \mathbf{v} \in \mathbb{R} \) satisfying \( \mathbf{v} \mathbf{P}^T = \frac{1}{\lambda} \mathbf{v} \) is called a left eigenvector of \( \mathbf{P} \) with eigenvalue \( \lambda \).
So, given a transition matrix $P$ for a Markov chain, an equilibrium dist. for the chain can be found by computing the left eigenvector $\lambda_k$, whose eigenvalue is 1, if such a vector exists. Most computer routines for eigenanalysis only give you right eigenvectors, but notice that if transpose

$$V_k^T P = V_k$$

then $$(V_k^T P)^T = V_k$$

$P^T V = V^T$, so we can just eigendecompose $P^T$ instead of $P$. 

$$P = V \Lambda V^T$$

where $\Lambda$ is the diagonal matrix of eigenvalues.
Genetic B's routine eigen gives (347) example, the following results: has continued two eigenvectors whose eigenvalues are 1: \((1\ 0\ 0)\) and \((0\ 0\ 1)\), corresponding to the two absorbing states.

This suggests that there is an entire family of equilibrium distributions of the form \((p, 0, 0, 0, 0, 0, 1-p)\) for \(0 \leq p \leq 1\); and Maple verifies this conjecture.

So the earlier guess is also correct: after many generations either one of \(\{AA, AA\}\) or \(\{aa, aa\}\) will be absorbing.
There is a special case in which a unique stationary distribution exists.

If you can find a positive integer $m \geq 1$ such that every element of $p^m$ is strictly positive, then $\lim_{n \to \infty} p^n$ is a matrix with all rows equal to the unique stationary distribution $\pi$, and no matter what the chain's initial distribution is, its distribution after $n$ steps converges to $\pi$ as $n \to \infty$. 