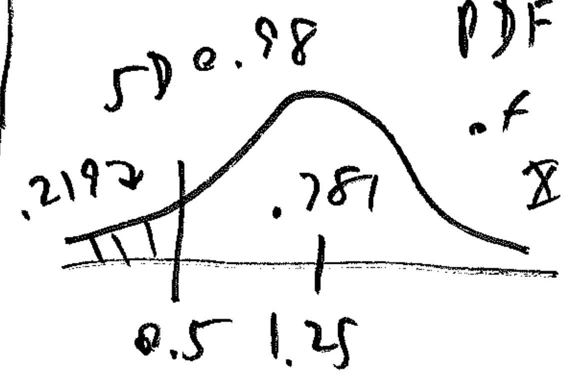
Improved approximation obtained by paying attention to the edges of the histogram (PMF) bars:

Normal approximation with continuity correction



$$\frac{0.5-1.25}{0.98} \doteq -0.77$$

$P(X \geq 1) \doteq 1 - P(X' < 0.5)$

$\doteq 1 - .219$

$\doteq 0.781$ (correct answer 0.76; much better approx.)

(4 Jan 19)

Markov Chains

Recall the definition of a stochastic process:

Def. A sequence of rvs $X_1, X_2, \ldots$ is called a stochastic process with discrete time parameter $t = 1, 2, \ldots$.

X_1 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 rvs $(X_1, X_2, \ldots)$.

Suppose that there's a parameter θ such that $(X_i \mid \theta) \overset{IID}{\sim}$ from some dist. depending on θ.

$\boxed{Q:}$ Does this process have a memory?

Example, revisited Machine with θ dial from 0 to 1, produces IID Bernoulli(θ) trials X_i. Recall that the process $(X_1, X_2, \ldots)$ does have a memory for you if θ 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 θ is around $\frac{17}{20} \doteq 0.85$, so X_{21} will probably be a success. But the process

$\{(X_i \mid \theta), i = 1, 2, \ldots\}$ has no memory once θ is known: information about the first n trials is irrelevant to your prediction of X_{n+1} if you know θ. An IID process $(X_i \mid \theta) \overset{IID}{\sim}$ is called a white-noise (stochastic) process or a white noise time series.

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

From now on, I'll suppress the dependence of the process on θ in the notation.

$\boxed{\text{Def.}}$ A discrete-time stochastic process is a (first-order) <u>Markov chain</u> if for $n = 1, 2, \ldots$; b any real number; and for all possible sequences of states $x_1, x_2, \ldots$

$$P(X_{n+1} \le b \mid X_1 = x_1, \ldots, X_n = x_n)$$

$$= P(X_{n+1} \le b \mid X_n = x_n).$$

In other words, the only thing you need to know to simulate where the Markov chain is going <u>next</u> is <u>where it is now</u>.

(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 a Markov chain can take on is called its *state space* 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 = seconds, milliseconds, microseconds, ...; we also won't pursue that here.)

It's easy to write down the joint PF of a Markov chain with finite S:

Consequences

Def. A Markov chain with a finite state space is called a <u>finite</u> Markov chain.

① $(X_1, X_2, \ldots)$ finite Markov chain →

$P(X_1 = x_1, \ldots, X_n = x_n) =$

$P(X_1 = x_1) \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot$

$P(X_3 = x_3 \mid X_2 = x_2) \cdot \ldots$

$P(X_n = x_n \mid X_{n-1} = x_{n-1})$.

Def. Suppose you have a finite Markov chain with k possible states numbered $1, \ldots, k$ (k integer ≥ 2) $\to \{P(X_{n+1} = j \mid X_n = i),$ $i, j = 1, \ldots, k,$ $n = 1, 2, \ldots\}$ ~~are~~ is called the <u>transition</u> <u>distribution</u> of the Markov chain.

If $P(X_{n+1}=j \mid X_n = i)$ is the same for all n, the transition distribution is said to be stationary. ← (DS) (bad name)

(time-homogeneous)

If the Markov chain does have a ~~stationary~~ transition distribution, then the probabilities

$p_{ij} \triangleq P(X_{n+1}=j \mid X_n = i)$ completely characterize the Markov chain's behavior.

Can arrange the p_{ij} in a matrix called the transition matrix.

$$
P_{k \times k} = \begin{array}{c|c}
 & \text{to state} \\
\text{from state} & \begin{array}{cccc} 1 & 2 & \cdots & k \end{array} \\
\begin{array}{c} 1 \\ 2 \\ \vdots \\ \vdots \\ k \end{array} & \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{bmatrix}
\end{array}
$$

All of the elements of $\underline{P}$ are non-negative (they're probabilities), and all of the row sums are 1 (because the chain has to go somewhere), i.e.

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

matrix versus quaternion

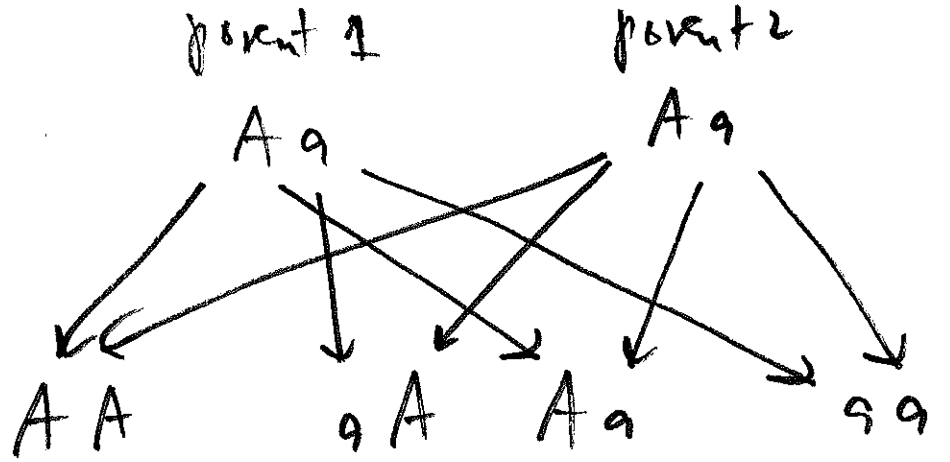
Def. A square matrix $\underset{k \times k}{\underline{P}}$ with non-negative entries and all row sums equal to 1 is called a stochastic matrix.

Example ~~(dominant/recessive)~~ Gene inheritance is Markovian: all we need to know to predict your genetic makeup at birth is the genetic story of your parents (your grand parents, ..., are irrelevant).

~~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

$\left\{ \begin{array}{cccc} \text{allele 1} & \text{allele 2} & \text{allele 1} & \text{allele 2} \\ \text{from} & \text{from} & \text{from} & \text{from} \\ \text{parent} & \text{parent} & \text{parent} & \text{parent 2} \\ 1 & 1, & 2 & \end{array} \right\}$, 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 $\{aa, aa\}$.



One possible inheritance sequence

offspring gets A or a from parent 1 and A or a (independently) from parent 2, each (A or a) with probability $\frac{1}{2}$.

Transition matrix

$$
\begin{array}{c|cccccc}
\text{from} \downarrow \ \text{to} \to & \{AA,AA\} & \{AA,Aa\} & \{AA,aa\} & \{Aa,Aa\} & \{Aa,aa\} & \{aa,aa\} \\
\hline
\{AA,AA\} & 1 & 0 & 0 & 0 & 0 & 0 \\
\{AA,Aa\} & \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\
\{AA,aa\} & 0 & 0 & 0 & 1 & 0 & 0 \\
\{Aa,Aa\} & \frac{1}{16} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{16} \\
\{Aa,aa\} & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\{aa,aa\} & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
$$

Example (random walk) You're watching a particle move around on the integers $S = \{ \ldots, -2, -1, 0, 1, 2, \ldots \}$ 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 —— p_3,

and it stays where it is with prob p_2,

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?

$$
\begin{array}{c|c}
 & \text{to} \rightarrow \quad \cdots \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad \cdots \\
\begin{array}{c} \text{from} \downarrow \\ \vdots \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} &
\left[\begin{array}{c|c|c|c|c|c|c}
\ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\hline
\cdots & p_2 & p_3 & 0 & 0 & 0 & \cdots \\
\hline
\cdots & p_1 & p_2 & p_3 & 0 & 0 & \cdots \\
\hline
\cdots & 0 & p_1 & p_2 & p_3 & 0 & \cdots \\
\hline
\cdots & 0 & 0 & p_1 & p_2 & p_3 & \cdots \\
\hline
\cdots & 0 & 0 & 0 & p_1 & p_2 & \cdots \\
\hline
\iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}\right]
\end{array}
= \underline{\underline{P}}
$$

This is an example of a <u>band matrix</u>, in which the only non-zero entries are on the main diagonal and 1 diagonal either way from the main diagonal; since there are only 3 non-zero diagonals, $\underline{\underline{P}}$ is said to be <u>tridiagonal</u>.

Moreover, all of the main diagonal entries are the same (p_2); all of the entries 1 diagonal below are also the same (p_1); and all of the entries 1 diagonal above are also the same (p_3).

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). (died of tuberculosis at 58)

Q:

---

start this process, which is called a random walk, at 0 & let it go; where is the particle likely to be at time n, n large?

A: 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 $-\infty$. $(p_1, p_2, p_3) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$?

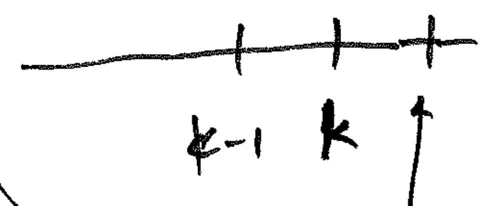
Can show that as $n \to \infty$ every integer is visited infinitely many times, and the expected time you must wait for the chain to return to 0 (having started there) is also infinite.

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 $S = \{ -k, -k+1, \ldots, -1, 0, 1, \ldots, k-1, k \}$ for some integer $k \geq 1$.

Q: What to do at the boundaries?



(k+1 not possible)

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 $+k$. (move.2)

Transition matrix with $k=2$

$$
\begin{array}{c|c}
 & \text{to} \\
\text{from} &
\begin{array}{c|ccccc}
 & -2 & -1 & 0 & 1 & 2 \\
\hline
-2 & p_2 & p_3 & 0 & 0 & p_1 \\
-1 & p_1 & p_2 & p_3 & 0 & 0 \\
0 & 0 & p_1 & p_2 & p_3 & 0 \\
1 & 0 & 0 & p_1 & p_2 & p_3 \\
2 & p_3 & 0 & 0 & p_1 & p_2
\end{array}
\end{array}
$$

(move.2)

$\boxed{R}$ demo

Another idea: keep trying until you make a legal move (move.1)

Back to a general finite Markov chain

Let $p_{ij}^{(m)} = P\left(\begin{array}{l}\text{chain moves} \\ \text{from } i \text{ to } j \\ \text{in } \boxed{m} \text{ steps}\end{array}\right)$

Theorem $= P(X_{n+m} = j \mid X_n = i)$

Finite Markov chain with stationary transition distributions & transition matrix $\underline{\underline{P}} \rightarrow p_{ij}^{(m)}$ is just the (i, j) entry of the matrix $\underline{\underline{P}}^m$, which is called the <u>m-step transition matrix</u> of the Markov chain.

Genetic example, continued

$\{AA, AA\}$ has the property that once the chain is in that state, it can't go anywhere else; so does $\{99, 99\}$ (34)

This occurs for a state i when $p_{ii} = 1$.

Def. Any state with $p_{ii} = 1$ is called an <u>absorbing state</u>. Notice that in this genetic Markov chain, states 1, 2, 4 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 (simulating many generations), it will either end up in state $\{AA, AA\}$ or in state $\{aa, aa\}$.

Markov chains that settle down to a single stable 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

JS 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: (If $P(X_{n+1} = j \mid X_n = i)$ is the same for all n, DS say) that the transition distribution is stationary; other people call this time-homogeneous.

I'll use equilibrium distribution for the long-run behaviour 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 it off by making a random draw from what's called the initial distribution of the Markov chain:

Def | Any vector $\underset{\sim}{v}$ of non-negative numbers that add up to 1 is called a <u>probability vector</u>; any such vector whose components specify that a Markov chain will be in each possible state at time 1 is referred to as the <u>initial distribution</u> of the chain.

So: | After 1 timestep (iteration) the probability dist. over the Markov chain's possible states is $\underset{\sim}{v}$; after 2 iterations the chain's dist. is $\underset{1\sim}{v} \underset{k=k}{P}$; after $(m+1)$ iterations its dist. is $\underset{\sim}{v}\,\underset{=}{P}^m$; it would be nice if $\underset{\sim}{v}\,\underset{=}{P}^m$ converged to a unique dist. as $m \to \infty$: this would be its <u>equilibrium</u> <u>distribution</u>.

Notice something interesting: if we choose $\underset{\sim}{v}$ so that $\underset{\sim}{v}\underset{=}{P} = \underset{\sim}{v}$, then

$$\underset{\sim}{v}\underset{=}{P}^2 = (\underset{\sim}{v}\underset{=}{P})\underset{=}{P} = \underset{\sim}{v}\underset{=}{P} = \underset{\sim}{v}, \quad \underset{\sim}{v}\underset{=}{P}^3 = (\underset{\sim}{v}\underset{=}{P}^2)\underset{=}{P}$$

$= \underset{\sim}{v}\underset{=}{P} = \underset{\sim}{v}$; and so $\lim_{m \to \infty} \underset{\sim}{v}\underset{=}{P}^m = \underset{\sim}{v}$

Def. Markov chain with transition matrix $\underset{=}{P}$ + any probability vector $\underset{\sim}{v}$ such that $\underset{\sim}{v}\underset{=}{P} = \underset{\sim}{v}$ is an equilibrium dist. for the Markov chain

under additional conditions on $\underset{=}{P}$, such an equilibrium dist. will be unique (we won't fully pursue that here).

How find $\underset{\sim}{v}$ so that $\underset{\sim}{v}\,\underset{=}{P} = \underset{\sim}{v}$?

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

<u>Def.</u>] Given a square matrix $\underset{k\times k}{\underset{=}{P}}$,

any vector $\underset{k\times 1}{\underset{\sim}{v}_R}$ satisfying $\underset{k\times k}{\underset{=}{P}}\;\underset{k\times 1}{\underset{\sim}{v}_R} = \lambda_R\,\underset{k\times 1}{\underset{\sim}{v}_R}$

is called a <u>right</u> <u>eigenvector</u> of $\underset{=}{P}$ with

<u>eigenvalue</u> λ_R, and any vector $\underset{1\times k}{\underset{\sim}{v}_L}$

satisfying $\underset{1\times k}{\underset{\sim}{v}_L}\;\underset{k\times k}{\underset{=}{P}} = \underset{1\times 1}{\lambda_L}\;\underset{1\times k}{\underset{\sim}{v}_L}$ is called

a <u>left</u> <u>eigenvector</u> of $\underset{=}{P}$ with

eigenvalue λ_L.

So, given a transition matrix ${}_k\underset{=}{P}{}_k$ for a Markov chain, an equilibrium dist. for the chain can be found by computing the left eigenvector ${}_1\underset{\sim}{v}{}_{L\,k}$ where eigenvalue is 1, if such a vector exists.

$$\boxed{\underset{\sim}{v}{}_L \underset{=}{P} = \underset{\sim}{v}{}_L}$$

Most computer routines for eigenanalysis only give you right eigenvectors, but notice that if

$${}_1\underset{\sim}{v}{}_{L\,k}\underset{=}{P}{}_k = {}_1\underset{\sim}{v}{}_{L\,k} \quad \text{then} \quad \left({}_1\underset{\sim}{v}{}_{L\,k}\underset{=}{P}{}_k\right)^T = {}_k\underset{\sim}{v}{}_{L\,1}^T$$

transpose

$$= {}_k\underset{=}{P}{}_k^T\, \underset{\sim}{v}{}_{L\,1}^T = {}_k\underset{\sim}{v}{}_{L\,1}^T,$$

so we can just eigendecompose $\underset{=}{P}^T$ instead of $\underset{=}{P}$.

Genetic example, continued

$\boxed{R}$'s routine eigen gives the following results: P^T has two eigenvectors whose eigenvalues are 1: $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, corresponding to the two absorbing states.

This suggests that there's an entire family of equilibrium distributions of the form $(p, 0, 0, 0, 0, 1-p)^T$ for $0 \leq p \leq 1$; and $\boxed{Wa}$ can be used to verify 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.

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