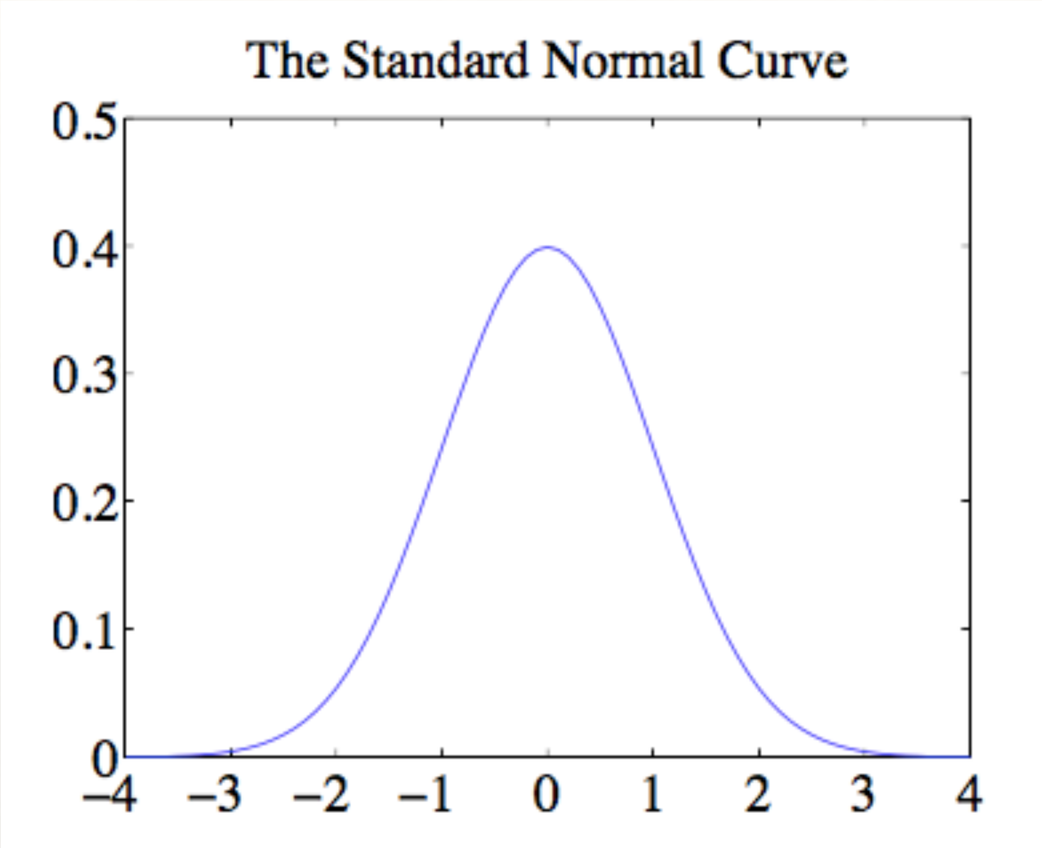
October 5, 2017

CS 361: Probability & Statistics

Random variables

Standard normal distribution



Recall our work with histograms and datasets

This was what our standard normal data looked like when we plotted it

Useful Facts: 6.9 *standard normal distribution*

1. The mean of the standard normal distribution is 0.
2. The variance of the standard normal distribution is 1.

These results are easily established by looking up (or doing!) the relevant integrals; they are relegated to the exercises.

Normal distribution

If we have a random variable X with a mean of μ and a standard deviation of σ and

$$\frac{X - \mu}{\sigma}$$

is a standard normal random variable, then X is called a **normal random variable** and its density is given by

$$p(x) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right) \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

Normal distribution

Useful Facts: 6.10 *normal distribution*

The probability density function

$$p(x) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right) \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right).$$

has

1. mean μ
2. and variance σ.

These results are easily established by looking up (or doing!) the relevant integrals; they are relegated to the exercises.

These random variables and distributions are also commonly called **Gaussian** distributions or random variables

Central limit theorem

Normal distributions occur extremely often in real datasets. Anything that behaves like a binomial distribution with a lot of trials will produce a normal distribution. More on this in a minute

Another result which we will not prove is the **central limit theorem** which states that if you add together many independent random variables with the same distribution, no matter the distribution, the resulting sum will be a random variable which has a distribution close to the normal distribution

Things like height, your IQ, etc can be viewed as the sum of bunch of independent random variables which explains why these phenomena usually follow a normal distribution

Normal data

Data that is drawn from a normal distribution tends to be close to the mean in terms of the number of standard deviations. Integrating the standard normal density, we see, that if we observe a bunch of standard normal data:

$$\frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \exp\left(-\frac{x^2}{2}\right) dx \approx .68$$

Around 68% of the data will be within 1 standard deviation of the mean

$$\frac{1}{\sqrt{2\pi}} \int_{-2}^{2} \exp\left(-\frac{x^2}{2}\right) dx \approx .95$$

Around 95% of the data will be within 2 standard deviations of the mean

$$\frac{1}{\sqrt{2\pi}} \int_{-3}^{3} \exp\left(-\frac{x^2}{2}\right) dx \approx .99$$

Around 99% of the data will be within 3 standard deviations of the mean

Binomial approximation

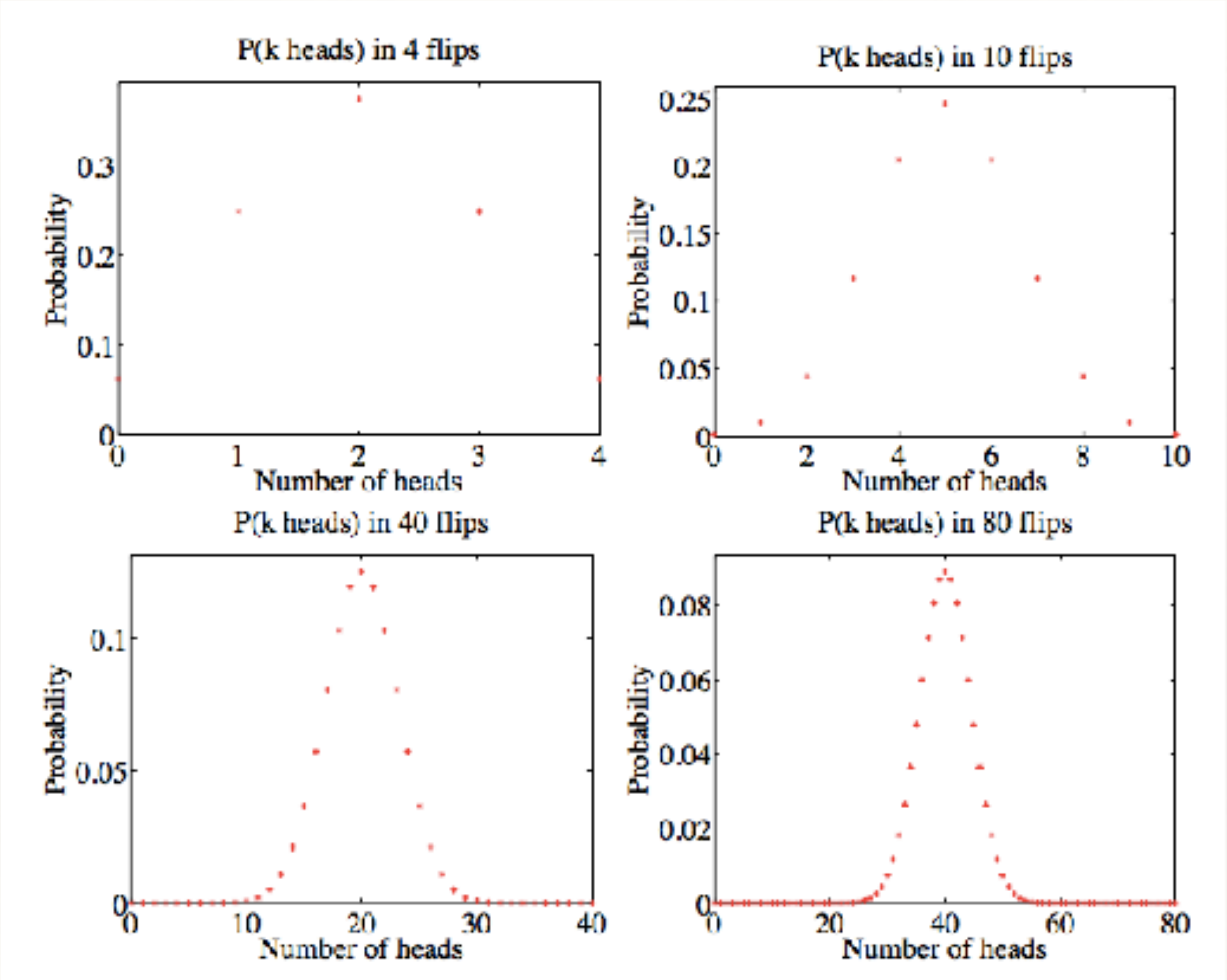
The binomial distribution is pretty straightforward with one caveat. Suppose we have a very large number of trials. If N is the number of trials, p is the probability of a success and $q = 1-p$ is the probability of a failure, then the probability distribution of X the random variable which counts the number of successes h is

$$P(h) = \frac{N!}{h!(N-h)!} p^h q^{(N-h)}$$

The caveat is that factorials grow rather quickly, so computing this probability can be difficult for large N as it can give us numerical overflows

We will construct an approximation that allows us to evaluate the probability that the number of successes lies within some range

Binomial approximation



Simulating some coin flips or looking at enough real-world binomially distributed data gives the impression that perhaps the binomial distribution looks enough like the normal distribution for that to potentially be useful

Binomial approximation

Finally, recall the definition of a Bernoulli trial and random variable: it takes value 0 with probability 1-p and value 1 with probability p

A Binomial random variable gives the probability of *h* successes in *N* Bernoulli trials or the sum of *N* independent Bernoulli random variables

The central limit theorem told us that the distribution of the sum of independent random variables is approximately normally distributed

All roads point to the normal distribution as being a good approximation for the Binomial distribution

Approximating with a normal

It turns out this is not an accident. So long as p isn't too close to 1 or 0 and N is large, the binomial distribution is well approximated by a normal distribution

If *h* is the number of heads in *N* coin flips where heads occurs with probability *p* and tails occurs with probability *q=1-p* and we write

$$x = \frac{h - Np}{\sqrt{Npq}}$$

Then for large *N*, the probability distribution of *x*, *P(x)*, is well approximated by the standard normal density *p(x)*

$$p(x) = \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left(\frac{-x^2}{2}\right)$$

in the sense that

$$P(\{x \in [a, b]\}) \approx \int_a^b \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left(\frac{-u^2}{2}\right) du$$

Markov Chains

Motivation

Consider this problem. You choose to flip a fair coin until you see two heads in a row at which point you will stop flipping.

What is the probability that you will flip the coin exactly twice and stop?

$$P(\{2\text{ flips}\}) = P(\{HH\}) = P(H)P(H) = 1/4$$

What is the probability that you flip the coin exactly three times?

$$P(\{THH\}) = 1/8$$

How about 4 times?

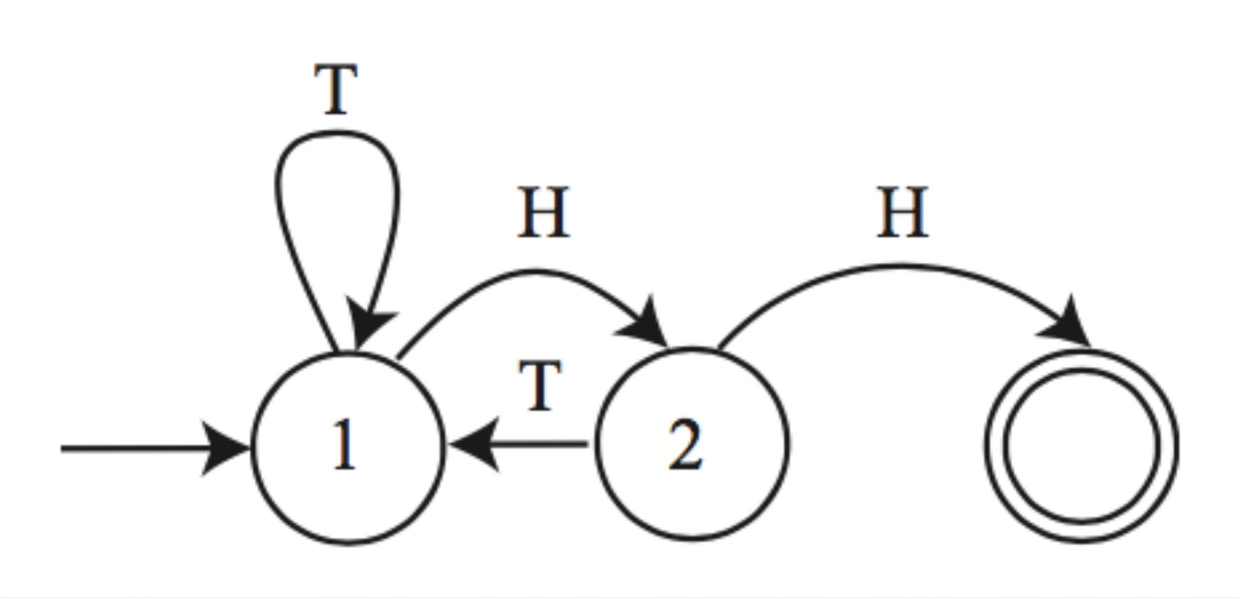
$$P(\{HTHH, TTHH\}) = 2/16$$

Calculating the probability that we stop after N flips is going to be pretty laborious

Finite state diagrams

Useful to model a problem like this in terms of thinking about states and transitions

Start state indicated by an arrow

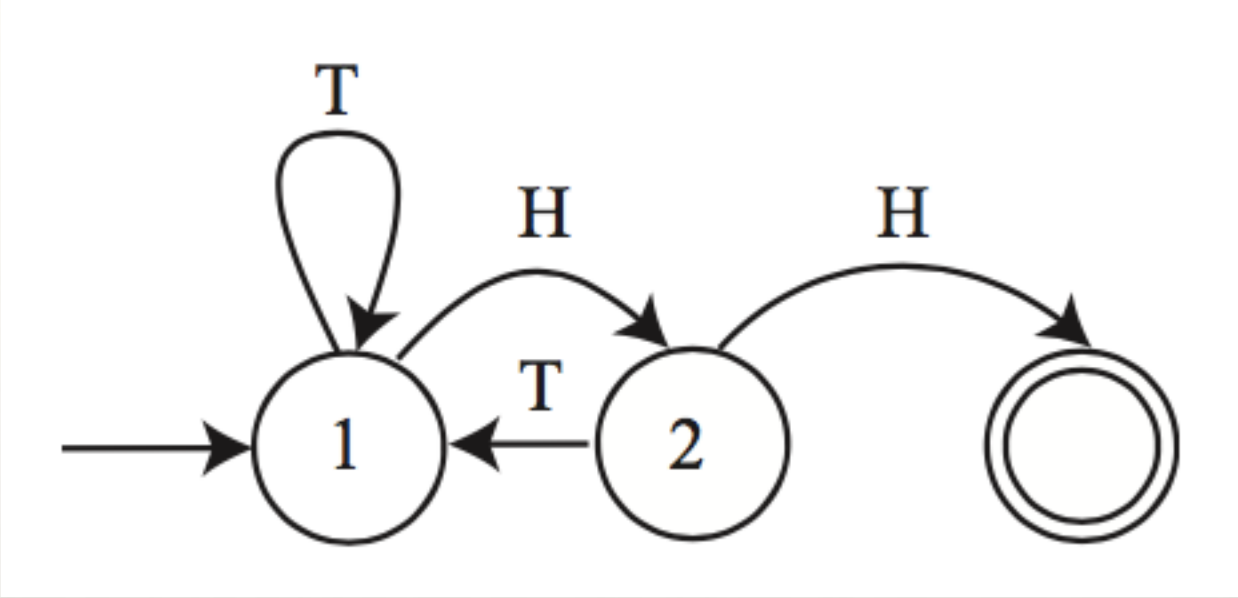


An end state indicated by an extra circle

Each node represents a state

Something happens, we observe, then transition to a new state by following the appropriate edge on the diagram

Finite state diagrams



Many problems admit a formulation in terms of being describable as a random sequence of states where we can ask questions like "what is the probability of being in a given state after some number of time steps?"

Recurrence relation

Consider this problem. You choose to flip a fair coin until you see two heads in a row at which point you will stop flipping. What is P(N) the probability that you stop after exactly N flips?

Write Σ_N for a string of length *N* that ends in HH

Then for N > 2, either the string starts with T and $\Sigma_N = T\Sigma_{N-1}$

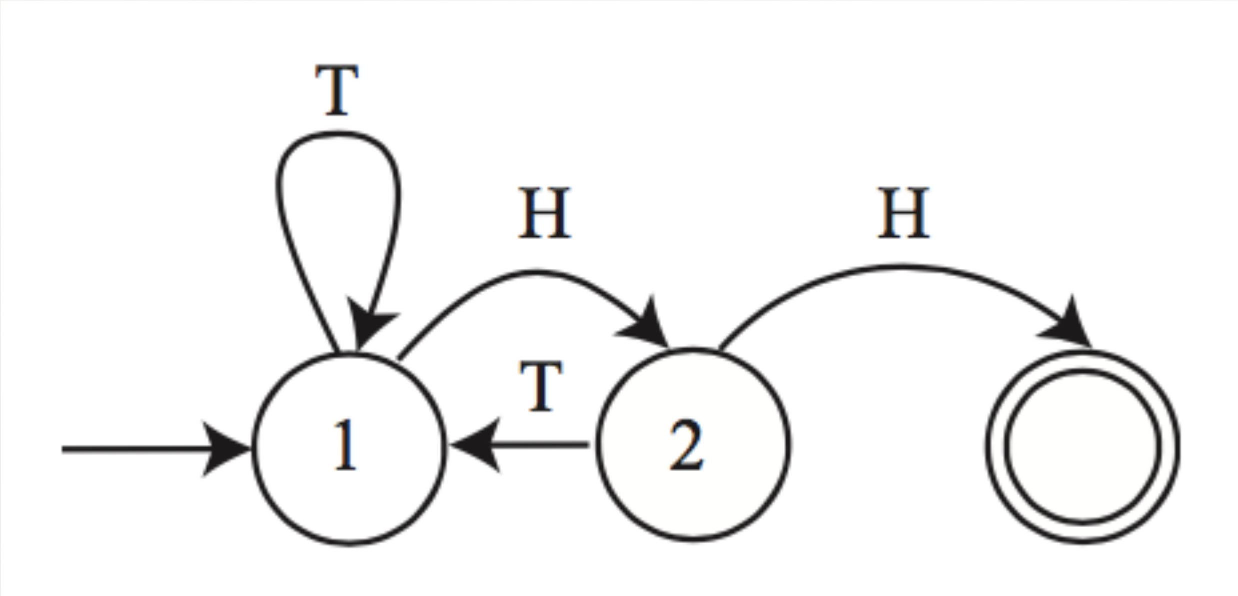
Or it starts with HT and $\Sigma_N = HT\Sigma_{N-2}$

Base cases:

P(1) = 0 P(2) = 1/4

$$
\begin{aligned}
P(N) &= P(T)P(N-1) + P(HT)P(N-2) \\
&= (1/2)P(N-1) + (1/4)P(N-2)
\end{aligned}
$$

Finite state machine



Get from state 1 to end in exactly N moves

If N < 2, it can't be done

If N=2 take edge H and then edge H

Else

Take edge T and then get from 1 to end in N-1 moves

Or take edge H, then T, and then get from 1 to end in N-2 moves

$$\begin{aligned} P(N) &= P(T)P(N-1) + P(HT)P(N-2) \\ &= (1/2)P(N-1) + (1/4)P(N-2) \end{aligned}$$

Motivation 2: gambler's ruin

Suppose you're playing a game of chance and you win with probability p and lose with probability 1-p. You bet $1 in every round of play, if you win, you get your bet back plus $1, if you lose, you lose your $1

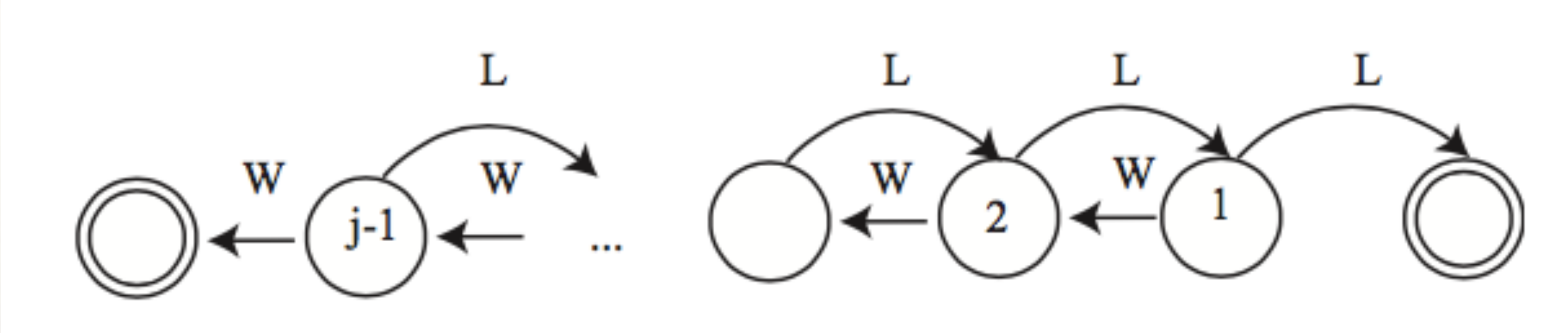
Assume you have $s to start. You'll keep playing until you've lost all your money or you've accumulated $j

Write p_s to represent the probability that you leave with nothing ("ruined") given that you start with $s

What is p_0? p_0 =1

What is p_j? p_j = 0

Gambler's ruin



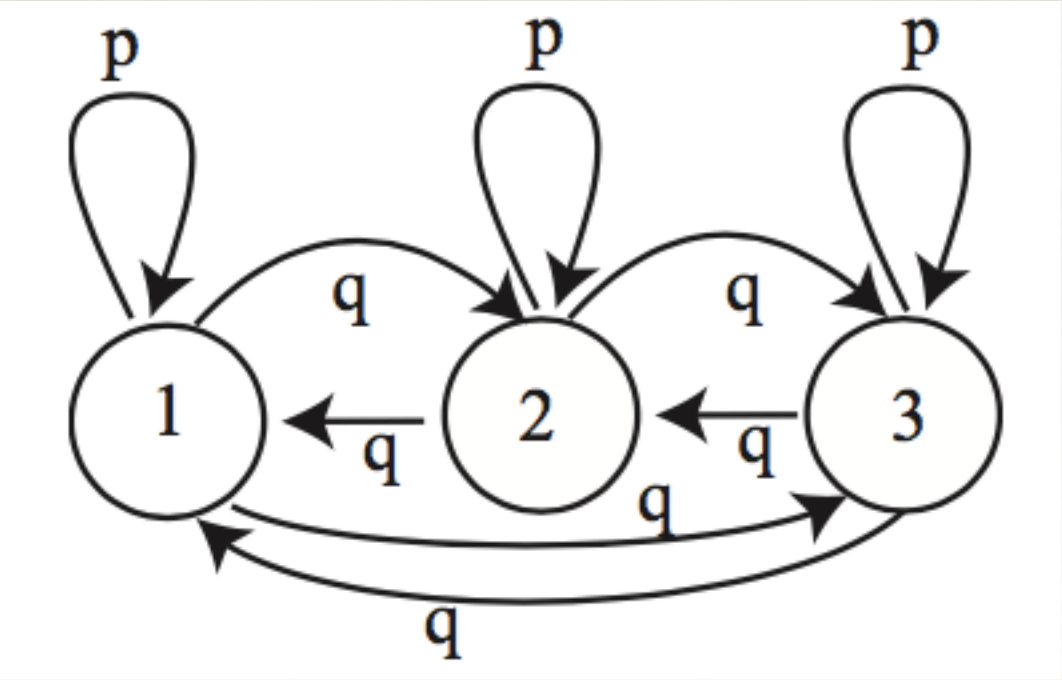
Write a recurrence relation for p_s

$$p_s = p p_{s+1} + (1-p) p_{s-1}$$

Motivation 3: viruses

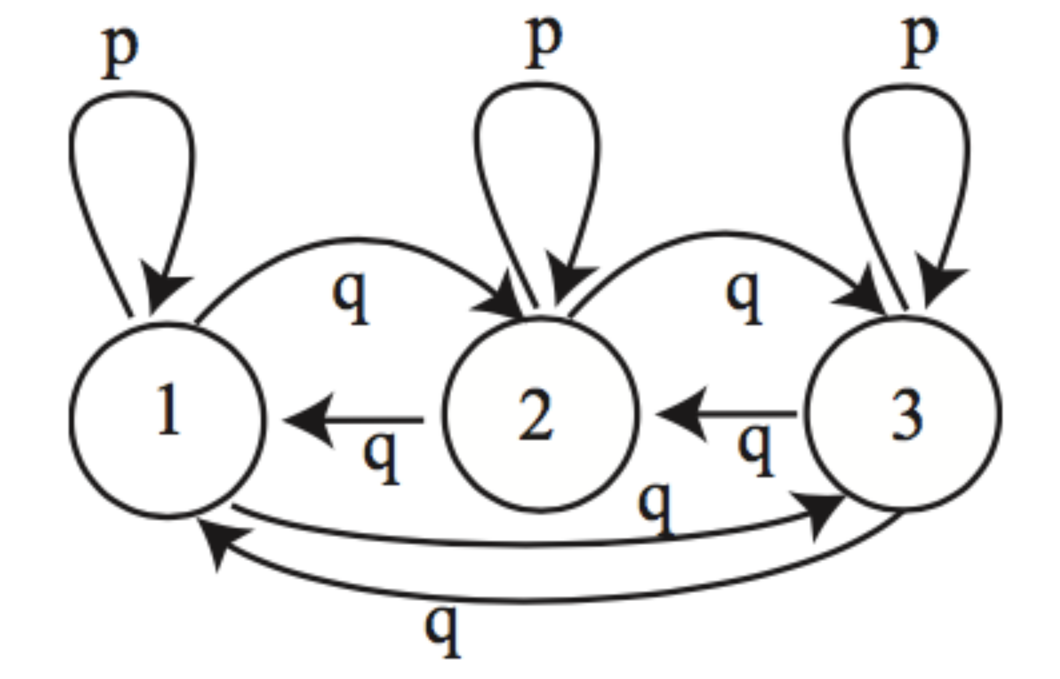
These kinds of finite state problems don't need to have end states

Consider a virus that can exist in one of 3 strains. At the end of the year it mutates with probability *a* and if it does it changes to one of the other strains uniformly at random. Then with p = (1-a) and q=a/2 this state diagram is appropriate



Note that the edges are now transition probabilities

Markov Chains



Picture a bug being placed on one of the nodes of this weighted directed graph. Its initial placement is determined by some probability distribution. Then, it walks from node to node according to the transition probabilities indicated in the arrows.

As it walks, the states that it visits (state 1 through state k) are a sequence of random variables. The random variable X_n=j if the bug was on state *j* at time step *n*

Markov chains

If we have such a setup—a distribution for the initial state, a set of transition probabilities for each state, and a sequence of random variables that take values indicating the state of the process at time *n*—and one additional constraint is met we have a Markov chain

The additional constraint is that the probability that the process is in state *j* at time *n* depends only on the state that the process is in at time *n-1*

$$P(X_n = j|\text{all previous states visited}) = P(X_n = j|X_{n-1})$$

Markov chains

Any model built by putting probabilities on the transitions of a finite state diagram is a Markov Chain but this isn't the only way to build or represent them

Another representation is to use a matrix to encode the transition probabilities from state *i* to state *j*

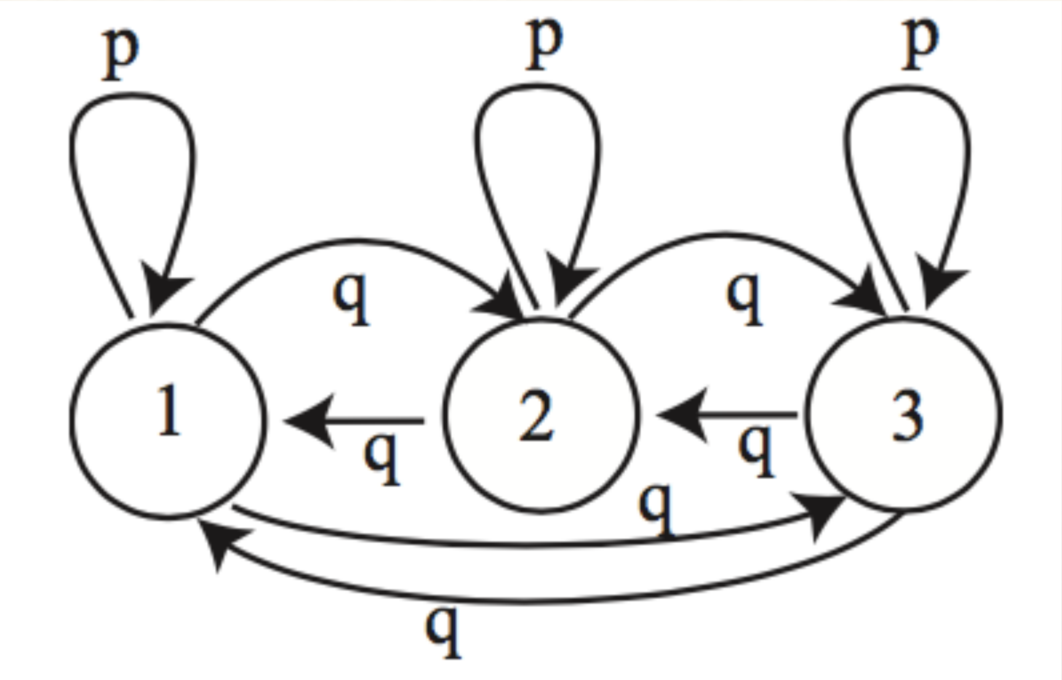
Define a matrix *P* such that

$$p_{ij} = P(X_n = j | X_{n-1} = i)$$

Note that this matrix will satisfy $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$

Example

Write the transition matrix for the virus example with a=0.2



$p = 1\text{-}a$

$q = a/2$

Recall:

$p_{ij} = P(X_n = j | X_{n-1} = i)$

$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$