Markov chains: examples

Markov chains: examples

Yahtzee. In the game of Yahtzee, you have five six-sided dice.

Markov chains: examples

Yahtzee. In the game of Yahtzee, you have five six-sided dice.

You get three turns to try to achieve certain sets of dice. After the first and second turns, you can keep any of the thrown dice and re-roll the others.

Markov chains: examples

Yahtzee. In the game of Yahtzee, you have five six-sided dice.

You get three turns to try to achieve certain sets of dice. After the first and second turns, you can keep any of the thrown dice and re-roll the others.

Suppose our goal is to get all the faces to be the same (this is called Yahtzee) and we procede in the most effective manner.

Here's an example.

Here's an example.

On the first roll, we might roll 11234.

Here's an example.

On the first roll, we might roll 11234.

Since we have two 1s, we keep those and roll the other three dice.

Here's an example.

On the first roll, we might roll 11234.

Since we have two 1s, we keep those and roll the other three dice.

Suppose those come up 155.

Here's an example.

On the first roll, we might roll 11234.

Since we have two 1s, we keep those and roll the other three dice.

Suppose those come up 155.

Since we already have two 1s, we keep this new 1 and roll the remaining two once more.

Here’s an example.

On the first roll, we might roll 11234.

Since we have two 1s, we keep those and roll the other three dice.

Suppose those come up 155.

Since we already have two 1s, we keep this new 1 and roll the remaining two once more.

If the two came up 11, we would have achieved Yahtzee.

What is the probability that we can achieve Yahtzee in (at most) three rolls?

What is the probability that we can achieve Yahtzee in (at most) three rolls?

We can model this with a Markov chain.

What is the probability that we can achieve Yahtzee in (at most) three rolls?

We can model this with a Markov chain.

At any point, we have zero, 1, 2, 3, 4, or 5 identical faces. So these can be our states.

What is the probability that we can achieve Yahtzee in (at most) three rolls?

We can model this with a Markov chain.

At any point, we have zero, 1, 2, 3, 4, or 5 identical faces. So these can be our states.

When we achieve state 5, we stay there, so 5 is the sole absorbing state.

What is the probability that we can achieve Yahtzee in (at most) three rolls?

We can model this with a Markov chain.

At any point, we have zero, 1, 2, 3, 4, or 5 identical faces. So these can be our states.

When we achieve state 5, we stay there, so 5 is the sole absorbing state.

Then, we can work out the probability of transitioning from each state to each other state.

What is the probability that we can achieve Yahtzee in (at most) three rolls?

We can model this with a Markov chain.

At any point, we have zero, 1, 2, 3, 4, or 5 identical faces. So these can be our states.

When we achieve state 5, we stay there, so 5 is the sole absorbing state.

Then, we can work out the probability of transitioning from each state to each other state.

These calculations are summarized in the following transition matrix:

$$P = \begin{pmatrix} 0 & \frac{120}{1296} & \frac{900}{1296} & \frac{250}{1296} & \frac{25}{1296} & \frac{1}{1296} \\ 0 & \frac{120}{1296} & \frac{900}{1296} & \frac{250}{1296} & \frac{25}{1296} & \frac{1}{1296} \\ 0 & 0 & \frac{120}{216} & \frac{80}{216} & \frac{15}{216} & \frac{1}{216} \\ 0 & 0 & 0 & \frac{25}{36} & \frac{10}{36} & \frac{1}{36} \\ 0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We start in the zero state and wish to achieve state 5 in at most three rolls.

Looking at the third power of P, we find the probability that we are looking for as the rightmost entry in the first row: this is the probability of transitioning from the 0 state to the 5 state in three or fewer rolls:

$$P^3 = \begin{pmatrix} 0 & \frac{125}{157464} & \frac{26875}{104976} & \frac{3419375}{7558272} & \frac{115625}{472392} & \frac{347897}{7558272} \\ 0 & \frac{125}{157464} & \frac{26875}{104976} & \frac{3419375}{7558272} & \frac{115625}{472392} & \frac{347897}{7558272} \\ 0 & 0 & \frac{125}{729} & \frac{7625}{17496} & \frac{7375}{23328} & \frac{5359}{69984} \\ 0 & 0 & 0 & \frac{15625}{46656} & \frac{11375}{23328} & \frac{8281}{46656} \\ 0 & 0 & 0 & 0 & \frac{125}{216} & \frac{91}{216} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

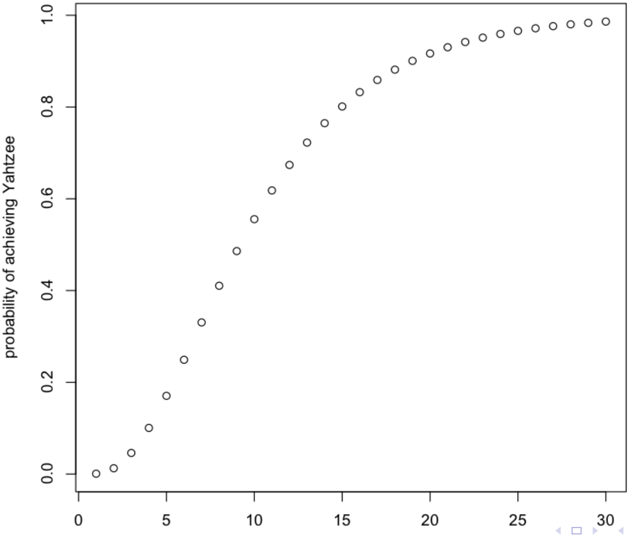
Hence, the probability of achieving Yahtzee is

$$\frac{347897}{7558272} = \frac{1}{22} + \frac{1}{1742} + \frac{1}{22594635} + \cdots \approx 0.0460286425$$

so we expect success about once every 22 attempts.

In general, we can find the probability of achieving Yahtzee in m rolls by looking at the right-most entry in the first row of P^m.

Here is a plot of the probability of achieving Yahtzee with from 1 to 30 rolls.



Probabilistic model of natural language text

A number of authors have applied Markov chain ideas to the analysis of text (e.g., novels).

Probabilistic model of natural language text

A number of authors have applied Markov chain ideas to the analysis of text (e.g., novels).

We can take a text, and remove all punctuation and formatting, leaving only letters and spaces.

Probabilistic model of natural language text

A number of authors have applied Markov chain ideas to the analysis of text (e.g., novels).

We can take a text, and remove all punctuation and formatting, leaving only letters and spaces.

Then we can consider every two-letter sequence (bigram) and count how many times each appears (e.g., how many times an 'a' is followed by an 'a', by a 'b', etc.).

Probabilistic model of natural language text

A number of authors have applied Markov chain ideas to the analysis of text (e.g., novels).

We can take a text, and remove all punctuation and formatting, leaving only letters and spaces.

Then we can consider every two-letter sequence (bigram) and count how many times each appears (e.g., how many times an 'a' is followed by an 'a', by a 'b', etc.).

From this we can empirically create a matrix P where, for i and j letters of the alphabet, P_{ij} is the probability that letter i is followed by letter j.

That is, we can treat the text as a Markov chain, with each letter representing a different state, and we move from state to state as we move through the text.

That is, we can treat the text as a Markov chain, with each letter representing a different state, and we move from state to state as we move through the text.

(This kind of analysis goes back to Markov himself, who analysed $20,000$ letters of Pushkin's *Eugene Onegin* in 1913.)

That is, we can treat the text as a Markov chain, with each letter representing a different state, and we move from state to state as we move through the text.

(This kind of analysis goes back to Markov himself, who analysed $20,000$ letters of Pushkin's *Eugene Onegin* in 1913.)

What's the point? One application is that, given a few of these transition matrices for a particular author, you can average them and create a matrix that captures something of an author's way of writing.

That is, we can treat the text as a Markov chain, with each letter representing a different state, and we move from state to state as we move through the text.

(This kind of analysis goes back to Markov himself, who analysed $20,000$ letters of Pushkin's *Eugene Onegin* in 1913.)

What's the point? One application is that, given a few of these transition matrices for a particular author, you can average them and create a matrix that captures something of an author's way of writing.

Then, given a piece of text, for whom the authorship is in question, you can measure the piece's transition matrix and compare to known authors' matrices.

Then, given a piece of text, for whom the authorship is in question, you can measure the piece's transition matrix and compare to known authors' matrices.

There has, apparently, been some success.

Then, given a piece of text, for whom the authorship is in question, you can measure the piece's transition matrix and compare to known authors' matrices.

There has, apparently, been some success.

One can also apply the same idea to words, rather than letters. (One difficulty is the much reduced sample size.)