Markov, Zipf, Shannon

Statistical Approaches to Analyzing Inscriptions

Part I

Sitabhra Sinha

“The more complicated the system considered, the more simplified must its theoretical description be. . . .
A good theory of complicated systems should represent only a good ‘caricature’ of these systems, exaggerating the properties that are most difficult, and purposely ignoring all the remaining inessential properties.”

Frenkel (1946)

Image: prabook.com

Yakov Il’ich Frenkel

Modeling Complex systems a la Markov

Stochastic perspective

Treating the time-evolution of complex systems as a probabilistic process

Modeling the process by which a system switches from one possible state to another at discrete time steps

Andrey Andreyevich Markov (1856 –1922)

Image: Brian Hayes

Classical probability	Markov
Principle of independence	~~Principle of independence~~
Joint probability of A (*p*) and B (*q*) = *pq*	~~Joint probability of A (*p*) and B (*q*) = *pq*~~
Law of large numbers	Law of large numbers

Markov chain describes a set of states and transitions between them

- ❑ System belongs to any one of a set of states at a given time
- ❑ System evolution occurs as transitions between states.
- ❑ Future depends on the present.
- ❑ Future does *not* depend on the past (Memory-less).

Remember…

Without memory

Outcome of next coin toss ?

Image: justflipacoin.com

VS

Limited memory

Tomorrow's weather ?

Image: www.123rf.com

The probability of each flip is independent of the probability of the previous flips

The weather tomorrow may depend a bit on the weather today, but not what it was two weeks earlier

Let’s take a fun example!

Monopoly or "The Landlord's Game"

Image: Brian Hayes

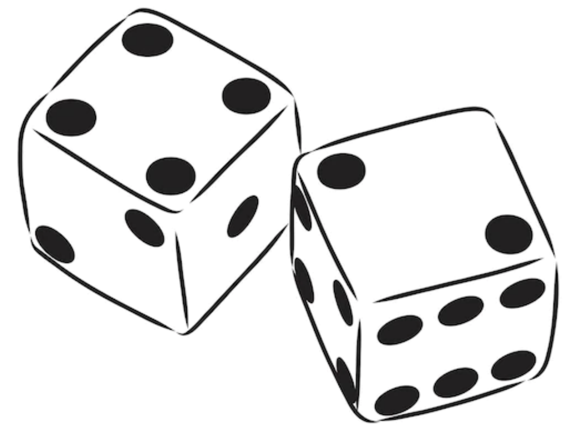
Image: wikimedia

Lizzie Magie

Originally designed to show the economic ill effects of monopoly

Monopoly

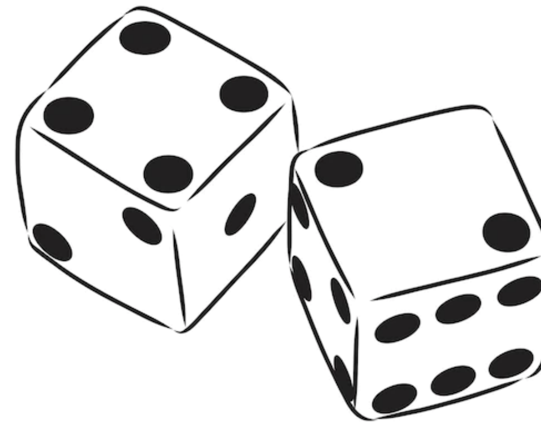
Begin here



Roll two die to decide how far to move

Image: Brian Hayes

Monopoly



Possibilities

1,1 = 2

1,2 or 2,1 = 3

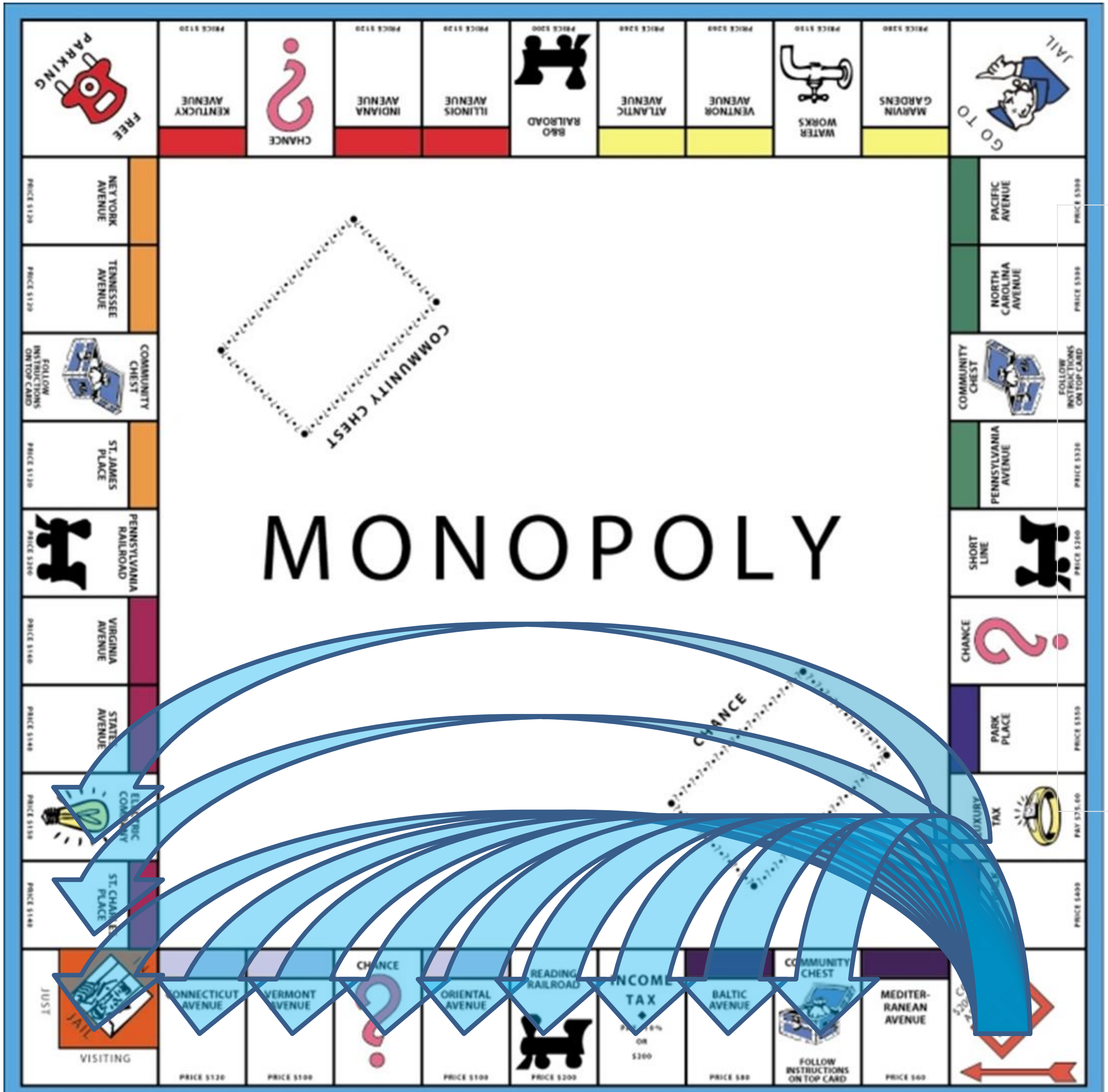
1,3 or 2,2 or 3,1 = 4

…

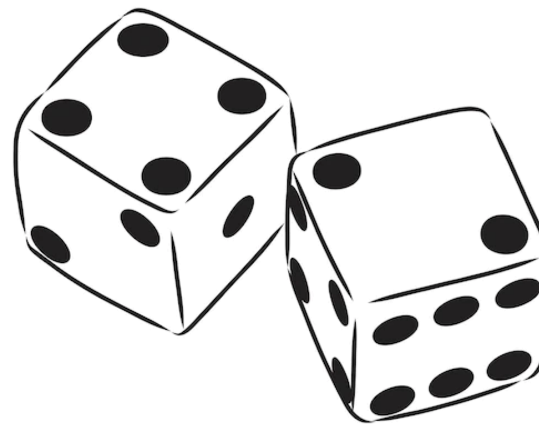
6,6 = 12

Image: Brian Hayes

Monopoly



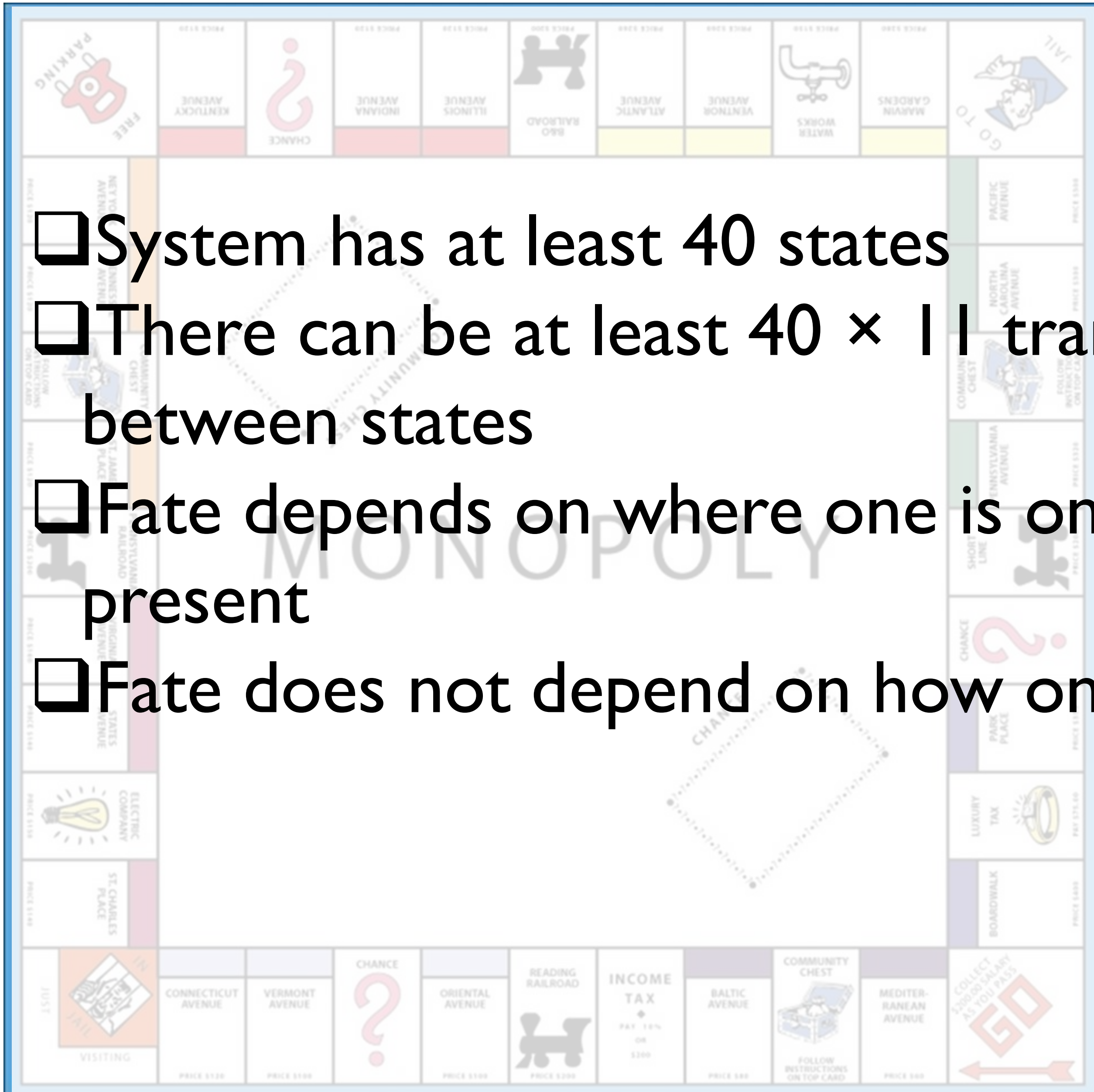
11 possible moves from each square



Possibilities
1,1 = 2
1,2 or 2,1 = 3
1,3 or 2,2 or 3,1 = 4
…
6,6 = 12

Image: Brian Hayes

Monopoly ≡ Markov chain



- System has at least 40 states
- There can be at least 40 × 11 transitions between states
- Fate depends on where one is on the board at present
- Fate does not depend on how one got there

A Simple Markovian model of the weather

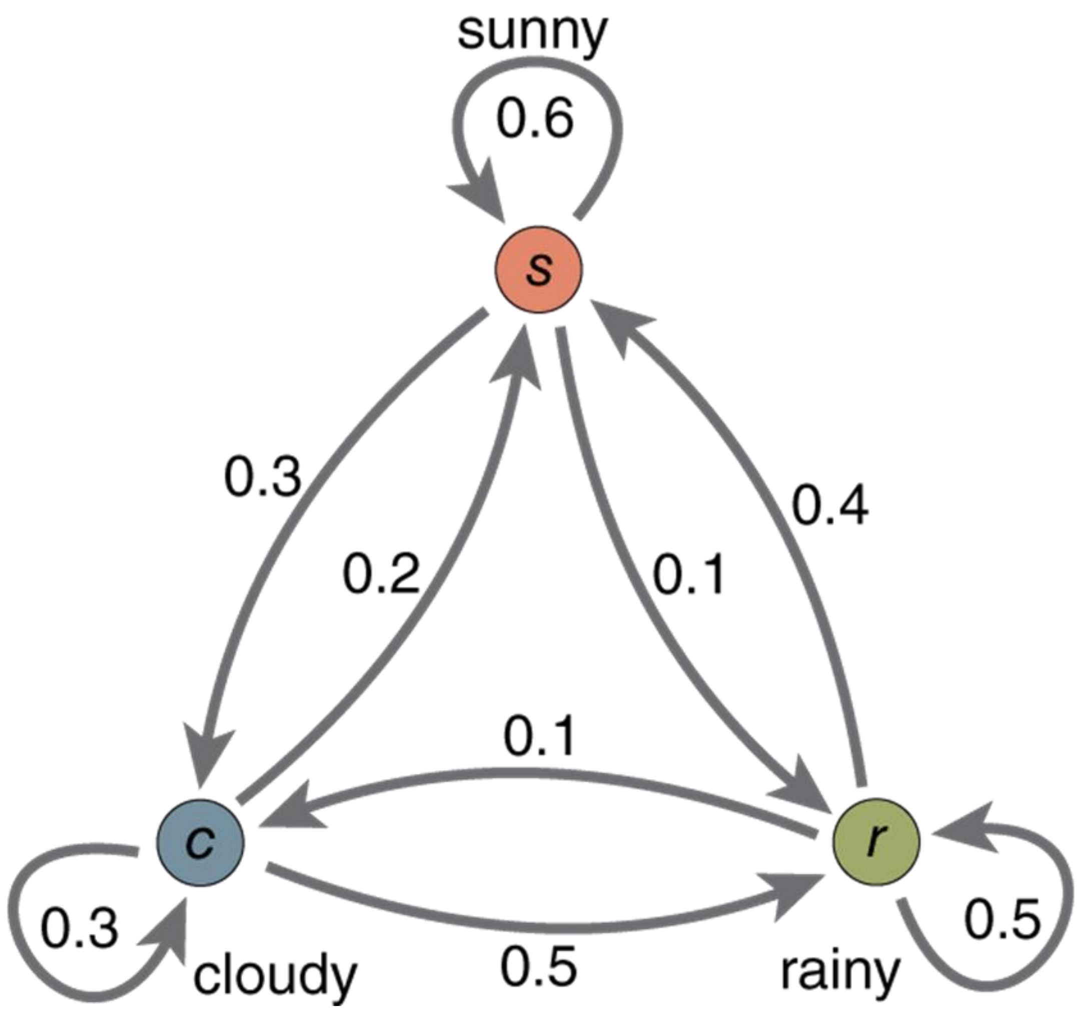
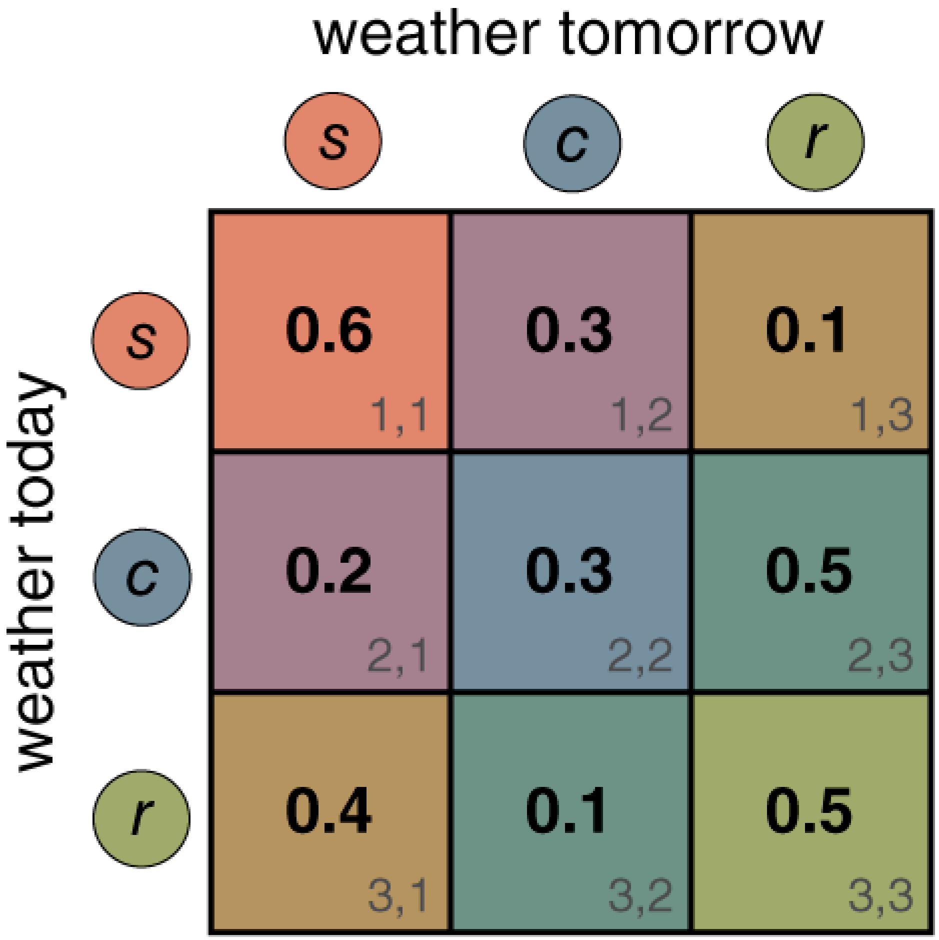


Image: Brian Hayes

Markov transition matrix



Matrix entries are probabilities

$$0 < P_{i,j} < 1$$
$$\sum_j P_{i,j} = 1$$

Each row sums to 1

Largest eigenvalue $\lambda_{max}=1$

with left eigenvector π such that

$$\pi P = \pi$$

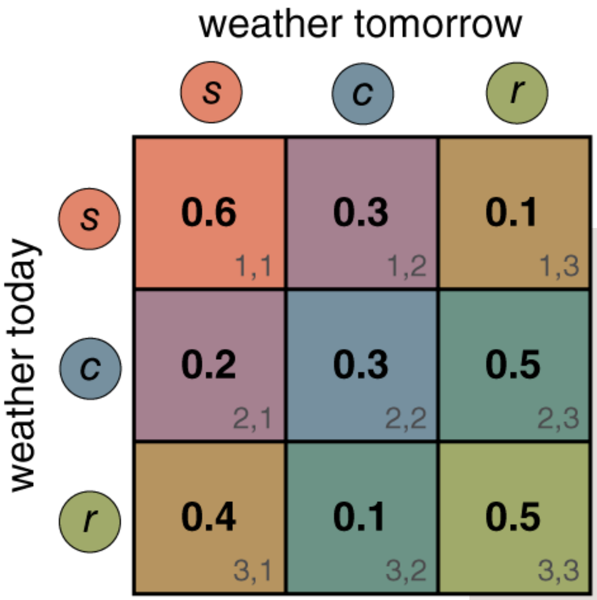
and

$$\lim_{n\to\infty} P^n = \pi$$

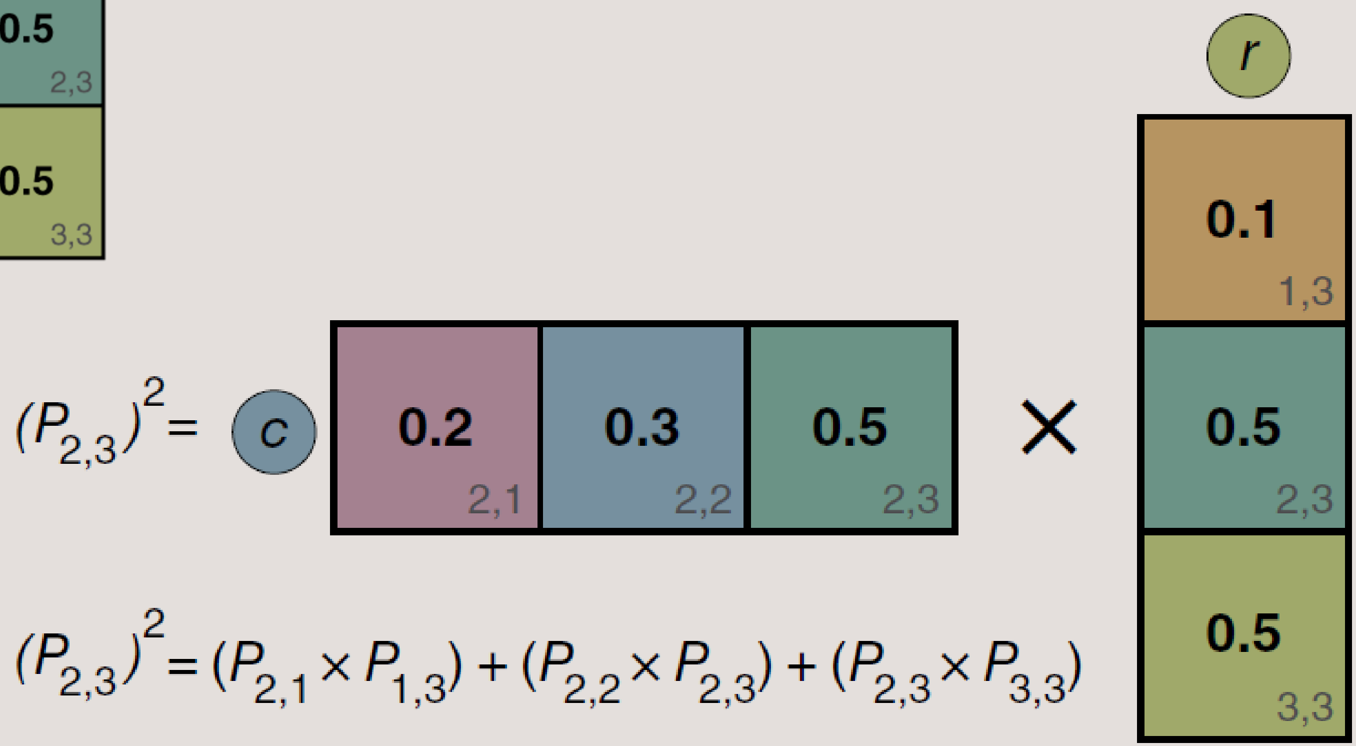
Image: Brian Hayes

Evolving the matrix

Following all two-step paths through the transition diagram produces a 2-day forecast – equivalent to multiplying the matrix by itself.



probability of rain in two days if it's cloudy today



$$(P_{2,3})^2 = (P_{2,1} \times P_{1,3}) + (P_{2,2} \times P_{2,3}) + (P_{2,3} \times P_{3,3})$$

$$(P_{2,3})^2 = (0.2 \times 0.1) + (0.3 \times 0.5) + (0.5 \times 0.5) = 0.42$$

Image: Brian Hayes

Evolving the matrix

Raising the matrix to higher powers produces longer term forecast – i.e., probabilities of longer sequences of transitions .

P 1-day forecast

0.600	0.300	0.100
0.200	0.300	0.500
0.400	0.100	0.500

P^2 2-day forecast

0.460	0.280	0.260
0.380	0.200	0.420
0.460	0.200	0.340

P^3 3-day forecast

0.436	0.248	0.316
0.436	0.216	0.348
0.452	0.232	0.316

P^4 4-day forecast

0.438	0.237	0.326
0.444	0.230	0.326
0.444	0.237	0.319

P^5 5-day forecast

0.440	0.235	0.325
0.443	0.235	0.322
0.441	0.236	0.322

P^7 7-day forecast

0.441	0.235	0.324
0.441	0.235	0.324
0.441	0.235	0.324

After just a few iterations, matrix converges to a stationary distribution where all rows are identical

Image: Brian Hayes

Markov and Poetry

An Example of Statistical Investigation of the Text "Eugene Onegin" Concerning the Connection of Samples in Chains

A. A. Markov

(Lecture at the physical-mathematical faculty, Royal Academy of Sciences, St. Petersburg, 23 January 1913)

Image: Brian Hayes

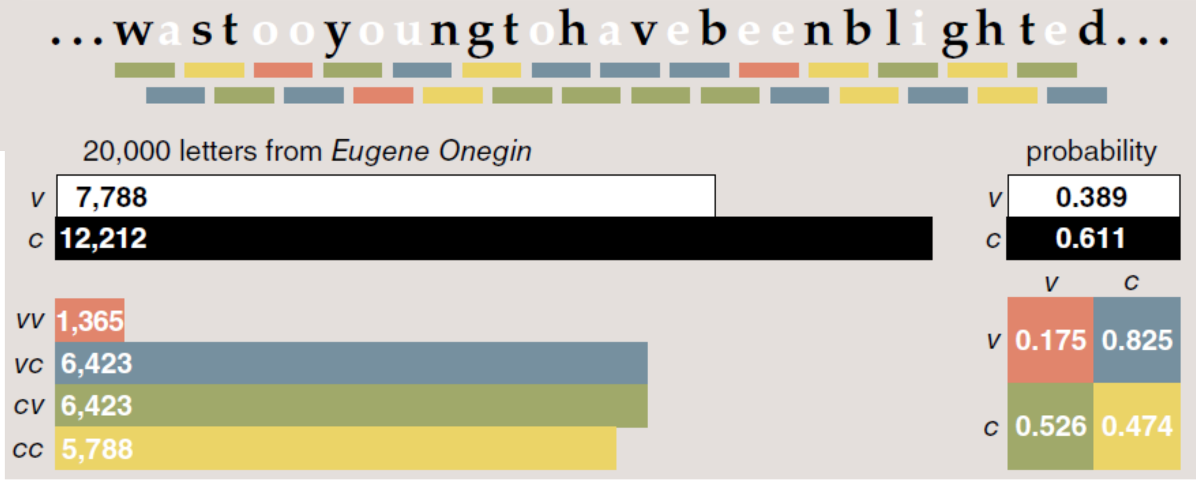
A machine called Pushkin

He was too young to have been blighted
by the cold world's corrupt finesse;
his soul still blossomed out, and lighted
at a friend's word, a girl's caress.
In heart's affairs, a sweet beginner,
he fed on hope's deceptive dinner;
the world's éclat, its thunder-roll,
still captivated his young soul.
He sweetened up with fancy's icing
the uncertainties within his heart;
for him, the objective on life's chart
was still mysterious and enticing—
something to rack his brains about,
suspecting wonders would come out.

Part of Markov's experiment on the statistics of language repeated with an English translation of Pushkin's *Eugene Onegin* (by Brian Hayes)

a single stanza (canto 2, verse 7)

Fragment of text without punctuation & spaces



Learning the bigram statistics

Image: Brian Hayes

Random text by Markov Chain

having statistical properties matching those in *Eugene Onegin*

First order

Theg sheso pa lyiklg ut. cout Scrpauscricre cobaives wingervet Ners, whe ilened te o wn taulie wom uld atimorerteansouroocono weveiknt hef ia ngry'sif farll t mmat and, tr iscond frnid riliofr th Gureckpeag

Third order

At oness, and no fall makestic to us, infessed Russion-bently our then a man thous al-ways, and toops in he roguestill shoed to dispric! Is Olga's up. Italked fore declaimsel the Juan's conven night toget nothem,

Fifth order

Meanwhile with jealousy bench, and so it was his time. But she trick. Let message we visits at dared here bored my sweet, who sets no inclination, and Homer, so prose, weight, my goods and envy and kin.

Seventh order

My sorrow her breast, over the dumb torment of her veil, with our poor head is stoop-ing. But now Aurora's crimson finger, your christening glow. Farewell. Evgeny loved one, honoured fate by calmly, not yet seeking?

In a *n*th-order random-text Markov model the states are n-grams, i.e., sequential blocks of n characters

Image: Brian Hayes

Shannon and English

- *Random characters.*
  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD
- *Sample from* $P^{(1)}$.
  OCRO HLI RGWR NMIELWIS EU LL NBBESEBYA TH EEI ALHENHTTPA OO BTTV
- *Sample from* $P^{(2)}$.
  ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE FUSO TIZIN ANDY TOBE SEACE CTISBE
- *Sample from* $P^{(3)}$.
  IN NO IST LAY WHEY CRATICT FROURE BERS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE
- *Sample from* $P^{(4)}$.
  THE GENERATED JOB PROVIDUAL BETTER TRAND THE DISPLAYED CODE ABOVERY UPONDULTS WELL THE CODERST IN THESTICAL IT TO HOCK BOTHE

Image:www.flowjuggle.com

Claude Shannon (1916-2001)

Shannon: Predicting an English text

Typically we possess an implicit knowledge of the statistics of a language we are familiar with, e.g., English – such that we can correctly predict subsequent letters in a sequence

```
(1) THE ROOM WAS NOT VERY LIGHT A SMALL OBLONG
(2) ----ROO------NOT-V-----I------SM----OBL----
```

Original text

Each letter correctly guessed replaced by -

```
(1) READING LAMP ON THE DESK SHED GLOW ON
(2) REA----------O------D----SHED-GLO--O--
```

Original text

Each letter correctly guessed replaced by -

```
(1) POLISHED WOOD BUT LESS ON THE SHABBY RED CARPET
(2) P-L-S-----O---BU--L-S--O------SH-----RE--C------
```

Original text

Each letter correctly guessed replaced by -

Do the sentences (1) contain more information than the sentences (2) because they have more letters filled in ?

NO

Autocorrect and predictive text

Autocorrection, also known as text replacement or replace-as-you-type is an automatic data validation function commonly found in word processors and text editing interfaces for smartphones and tablets

Autocomplete or word completion : when the writer writes the first letter or letters of a word, the program predicts one or more possible words as choices.

Image: www.iphonelife.com

Autocorrect and Autocomplete via Markov chains

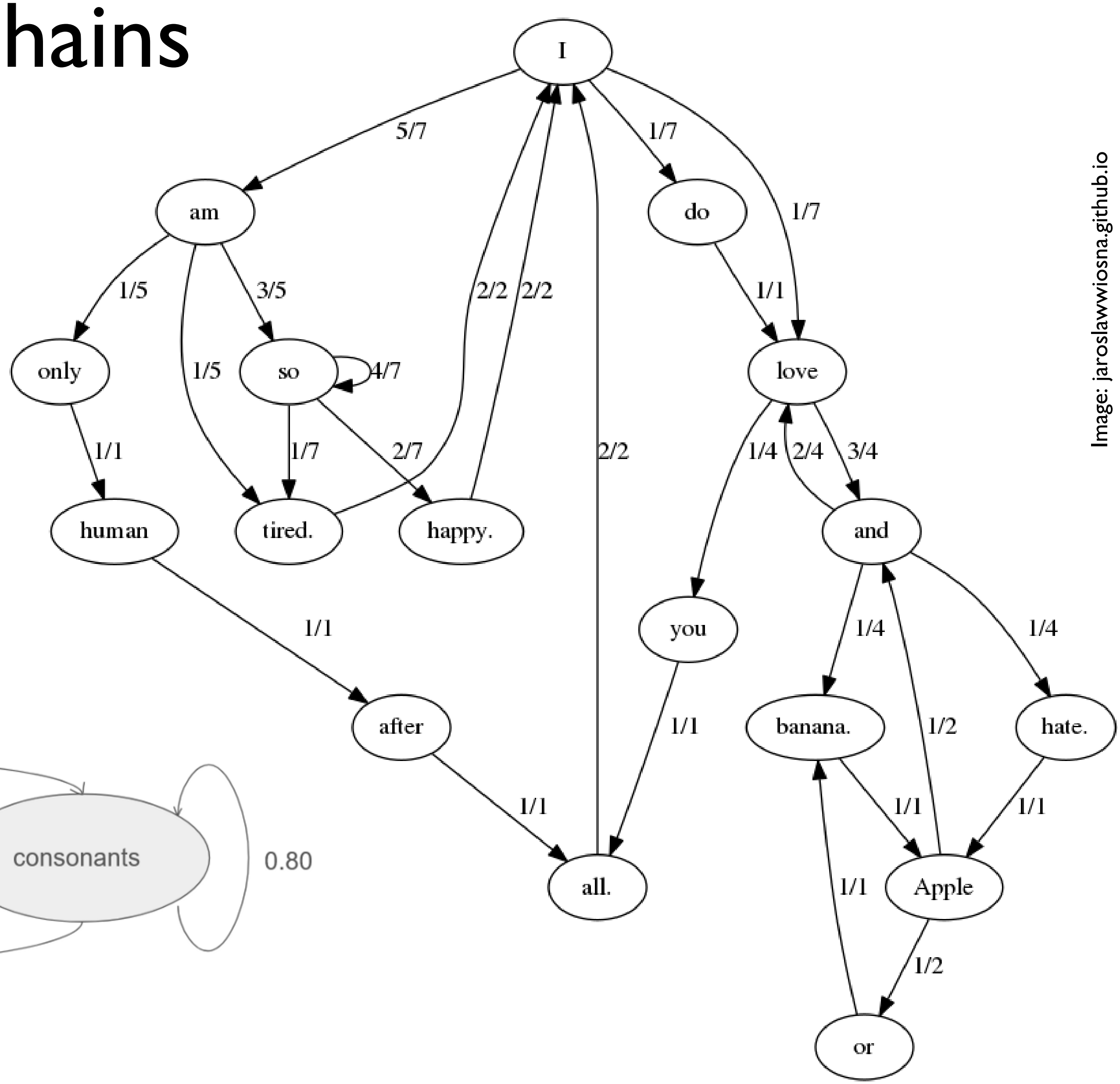


Image: jaroslawwiosna.github.io

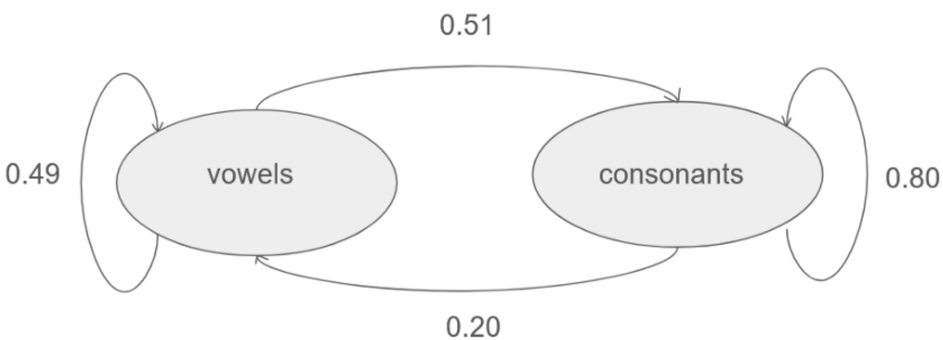


Image: towardsdatascience.com

Strong correlations between letters in a given word underlies our ability to solve

WORDLE

https://www.nytimes.com/games/wordle/index.html

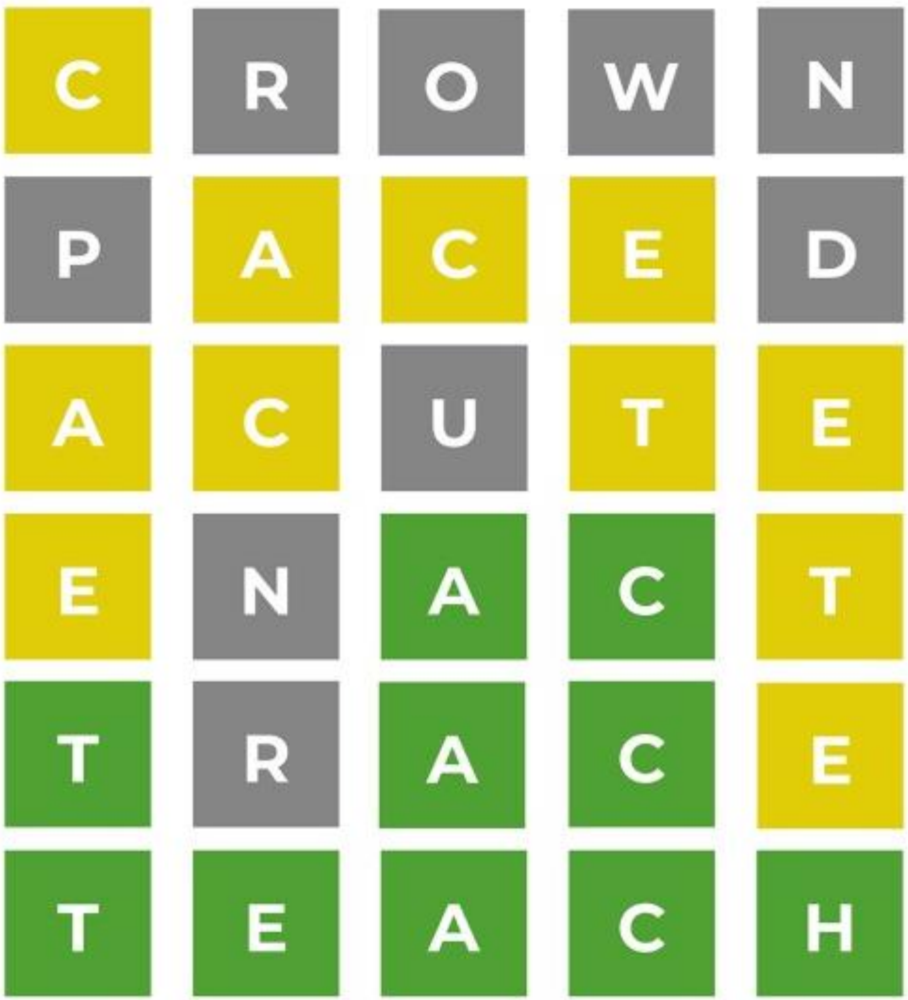


Image:www.edweek.org

An online 5-letter word guessing game which a player will have to correctly guess within 6 attempts

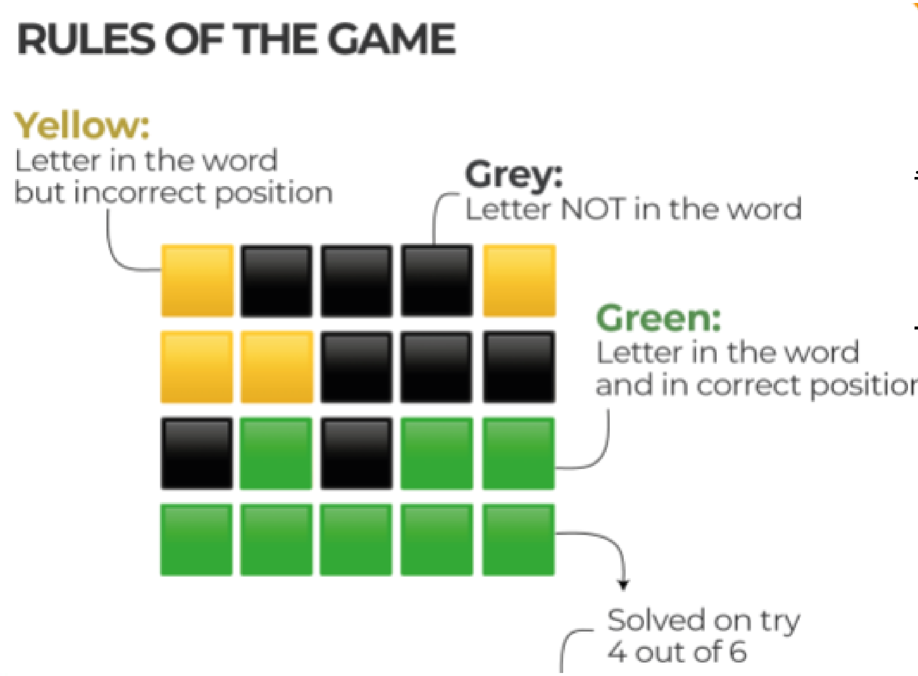


Image: www.aljazeera.com

Without correlations, would require 26^5 (~ 10^7) guesses in worst case