First Links in the Markov Chain

Brian Hayes

American Scientist

Harvard SEAS

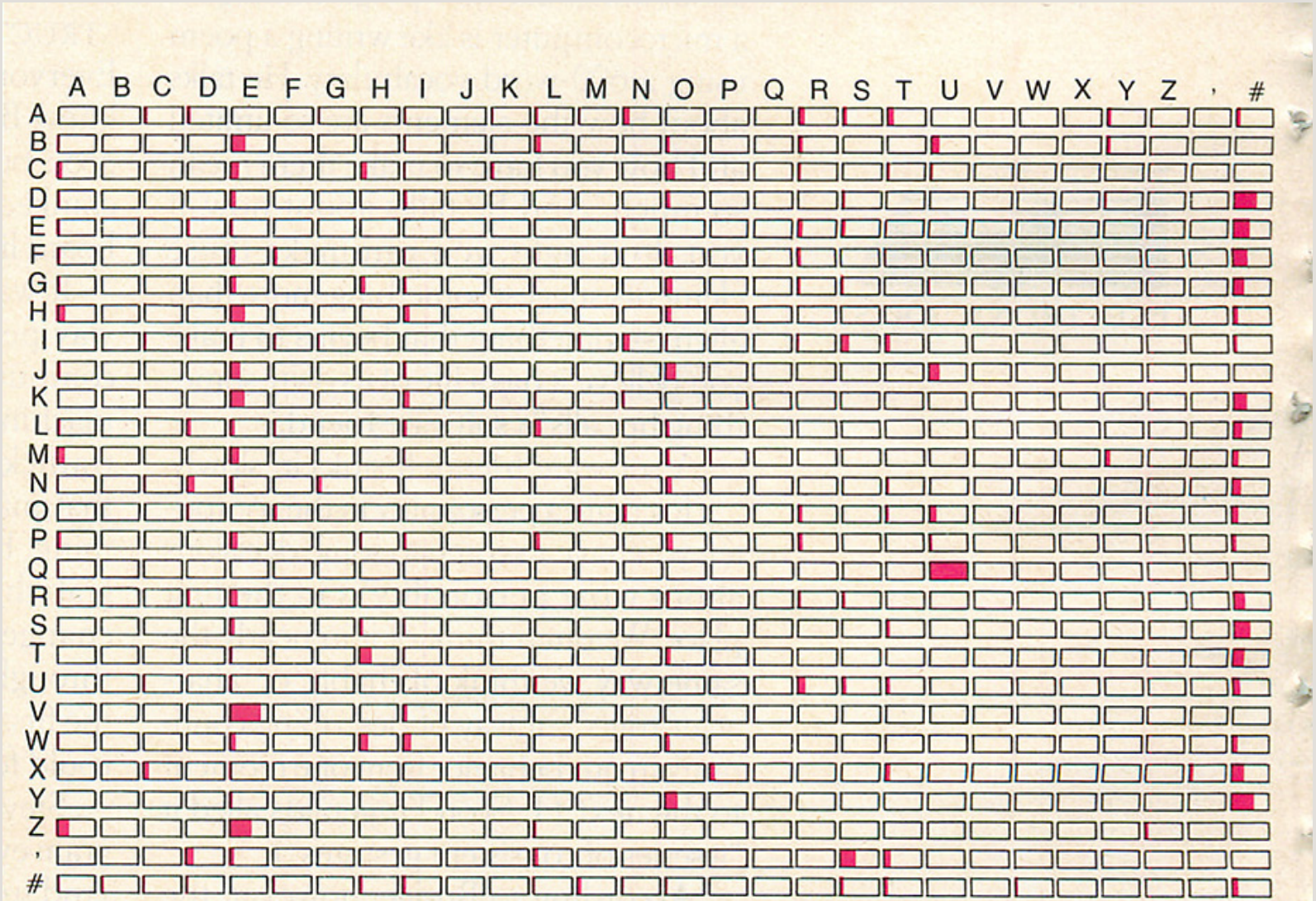
bit-player.org

Workshop on Deep Computation in Statistical Physics

Santa Fe Institute – 8 August 2013

These slides: **http://bit-player.org/extras/markov-sfi**

Drivel

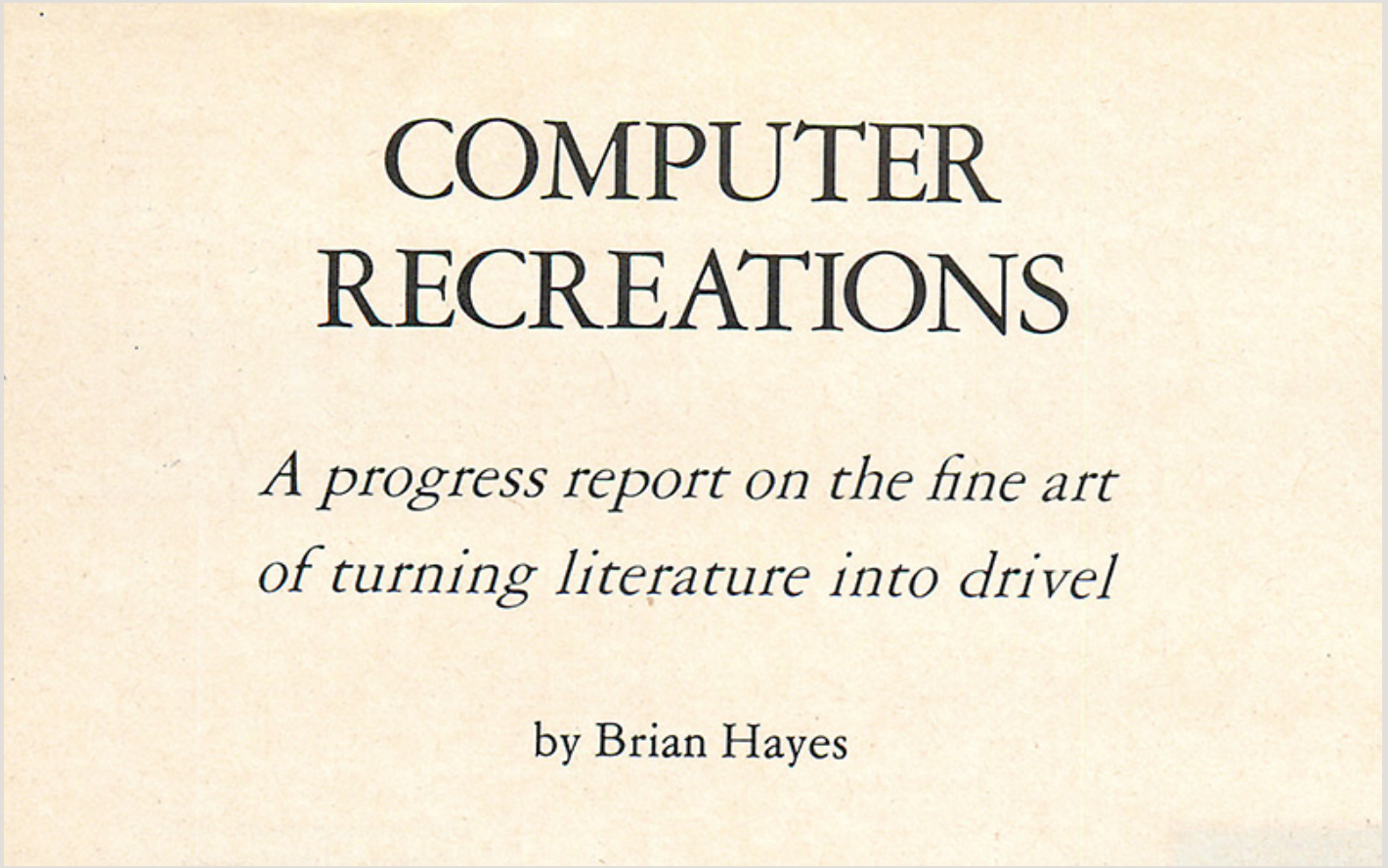


A second-order frequency table for Act III of **Hamlet**

eau	2	(Rousseau , château)
eaut	18	(beauty, beauties)
eau.	1	(Rousseau.)
eaub	1	(Chateaubriand)
eaux	4	(Bordeaux, tableaux)
eau,	1	(Rousseau,)

eau	2	(Rousseau, château)
eaut	18	(beauty, beauties)
eau.	1	(Rousseau.)
eaub	1	(Chateaubriand)
eaux	4	(Bordeaux, tableaux)
eau,	1	(Rousseau,)

auti	2	(beauties, cautious)
auty	18	(beauty)
auto	1	(autocrat, autograph)
auth	1	(author)
autu	4	(autumn)
aut.	1	(comme il faut.)



Scientific American, November 1983

Sergey Kapitsa (1928–2012)

The procedure for selecting a letter in the random-text program is known in probability theory as a Markov process, after the Russian mathematician Andrei A. Markov. Kapitza points out that Markov presented his first discussion of the process in terms of randomizing text. Markov's paper "On the Sequence of Letters in Eugene Onegin" asks to what extent Pushkin's poem remains Pushkin's when the letters are scrambled.

Примѣръ статистическаго изслѣдованія надъ текстомъ „Евгенія Онѣгина“ иллюстрирующій связь испытаній въ цѣпь.

А. А. Марковъ.

(Доложено въ засѣданіи Физико-Математическаго Отдѣленія 23 января 1913 г.).

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)

Questions at the Centennial

- Who was A. A. Markov?
- What led him to the mechanism that we call a Markov chain?
- Why hadn't anybody thought of it before?
- What did Pushkin's *Eugene Onegin* have to do with it?
- How did these ideas come down to us today?

Principal Sources

- David Link (a German scholar).
- Eugene Seneta (an Australian statistician).
- Oscar Sheynin (a Russian-German historian).

History

Peter the Great (1682–1725)

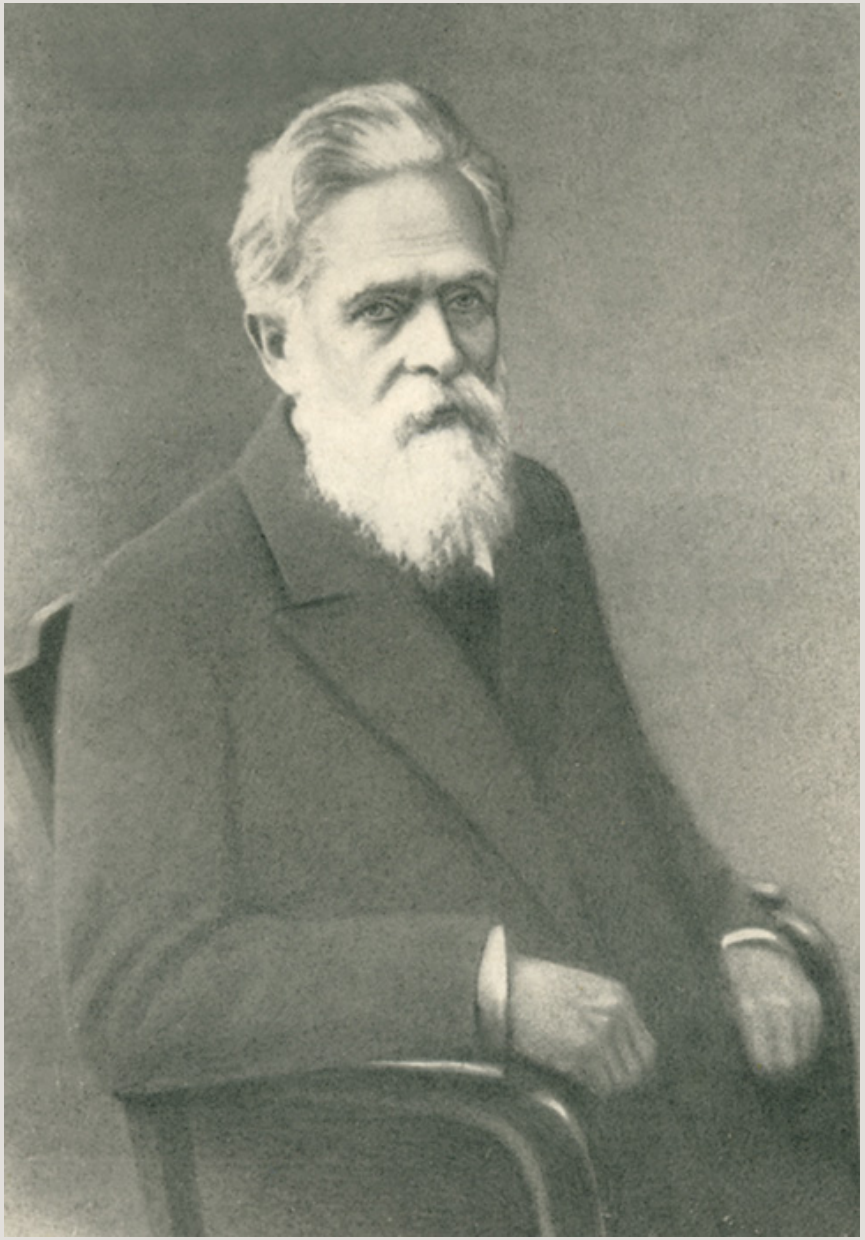
Imperial Saint Petersburg Academy of Sciences
(founded 1724)

Academy luminaries: Euler, Goldbach, Nicolas and Daniel Bernoulli

Pafnuty Lvovich Chebyshev (1821–1894)

А. А. Марков (1886).

Markov in 1886 (age 30)



Markov in 1918 (age 62)

Andrei Andreevich Markov

- 1856–1922.
- Child of a bureaucrat who later managed an aristocrat's estate.
- Married the lord's daughter.
- One son, A. A. Markov, born 1903.

Markov's Academic Career

- Student at St. Petersburg, with Chebyshev.
- Early work in number theory, continued fractions, PDEs.
- Doctorate 1884.
- University of Petersburg professor.
- Academy of Sciences academician.
- Retirement 1905 (age 49).
- Best-known work done after retirement.

Markov in Public Life

- Anti-tsarist gadfly ("Andrew the Furious").
- Excommunicated by request.
- The Maxim Gorky affair.
- 300 years of Romanovs; 200 years of *Ars Conjectandi*.

What led Markov to his chains?

- Logical outgrowth of work begun by Chebyshev on the law of large numbers.
- Connections with Bernoulli urn problems.
- The great quarrel with Nekrasov.

The Feud

A letter from Markov to Chuprov

" I note with astonishment that in the book of A. A. Chuprov, Essays on the Theory of Statistics, on page 195, P. A. Nekrasov, whose work in recent years represents an abuse of mathematics, is mentioned next to Chebyshev. — A. Markov "

Pavel Alekseevich Nekrasov (1853–1924)

- In the “Moscow School” of mathematics.
- From 1891 was Rector of Moscow University.
- Studied divinity before turning to mathematics.
- According to Loren Graham and Jean-Michel Kantor, a “name worshipper.”

Markov and Nekrasov

- St. Petersburg vs. Moscow.
- West vs. East.
- Anti-tsarist vs. monarchist.
- Secularist vs. ecclesiatic.

Also, an earlier spat over who borrowed from whom.

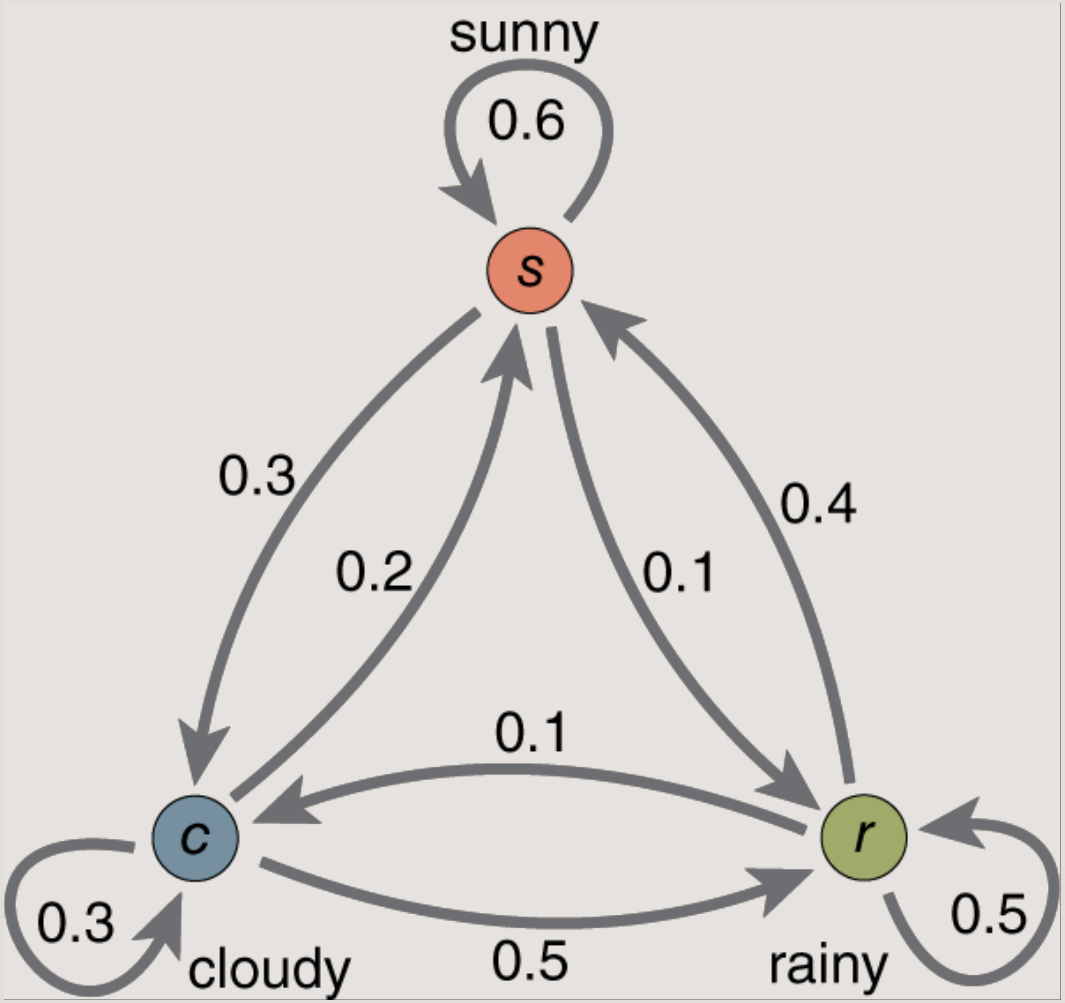
Nekrasov's argument for free will

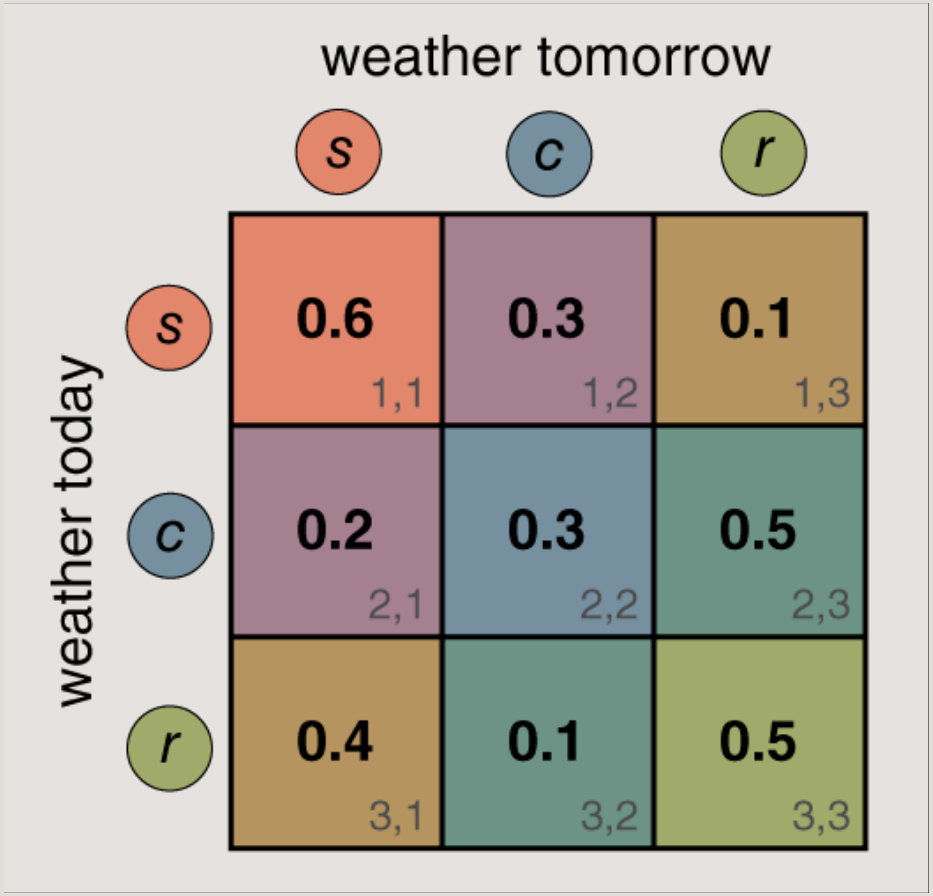
1. Voluntary acts are those without causal connections, like independent trials in probability theory.
2. The law of large numbers applies *only* to independent trials.
3. The law of large numbers is observed to hold in social statistics.
4. Therefore people act voluntarily, with free will.

Markov's refutation

- The goal was to show that independence is *not* necessary for LLN.
- 1906: Markov finds a special model in which LLN holds for dependent variables: the first Markov chains.
- 1907–11: Extends the theory to broader classes of systems.
- 1913: Applies the model to lexical analysis of *Eugene Onegin*.

A Markov model of the weather





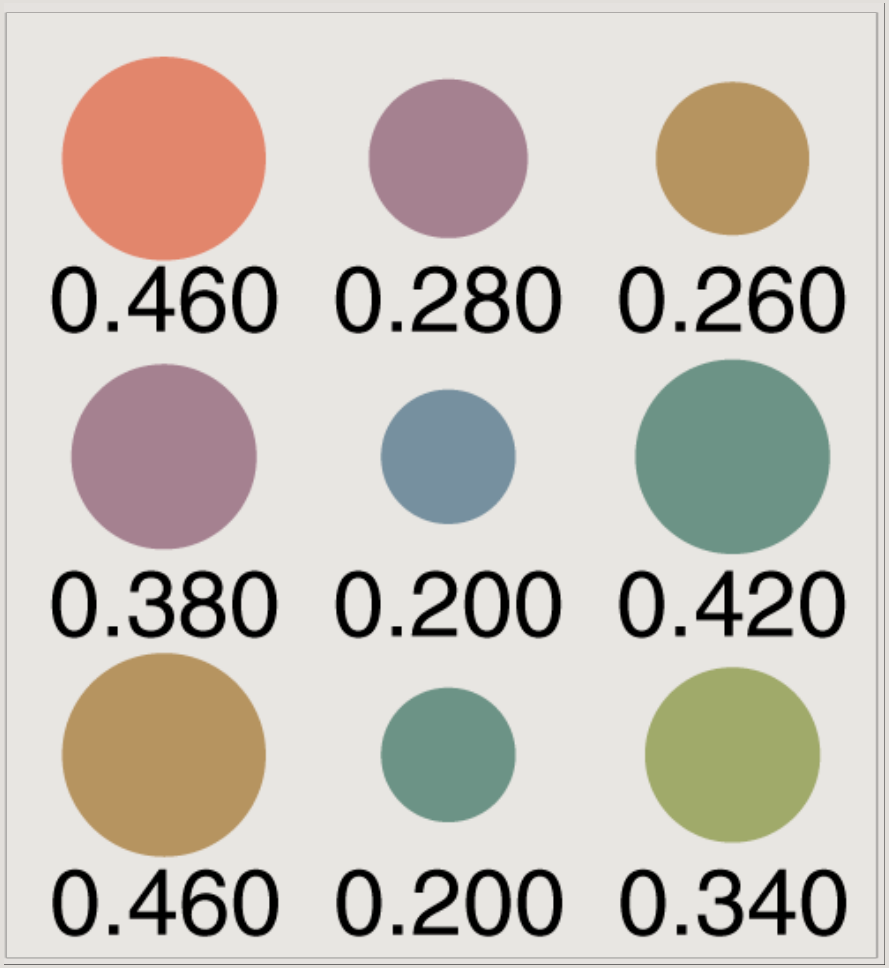
Stochastic matrices

$$0 < P_{i,j} < 1$$

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

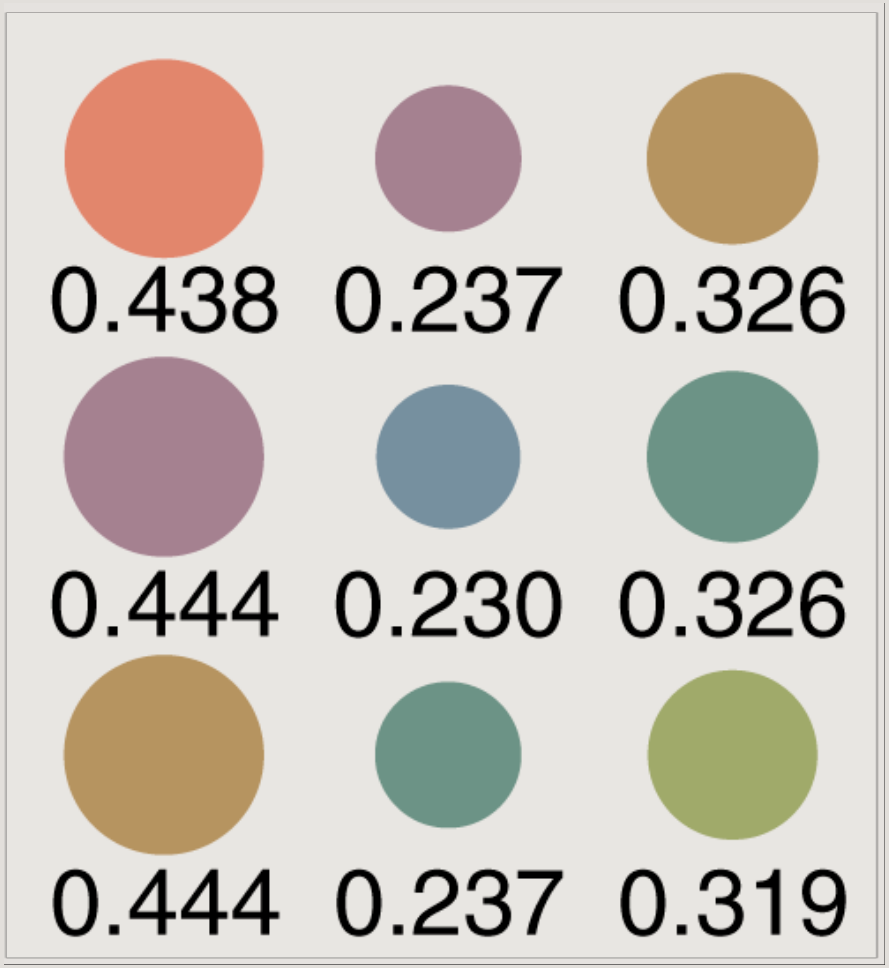
Powers of the transition matrix: 1

Powers of the transition matrix: 2

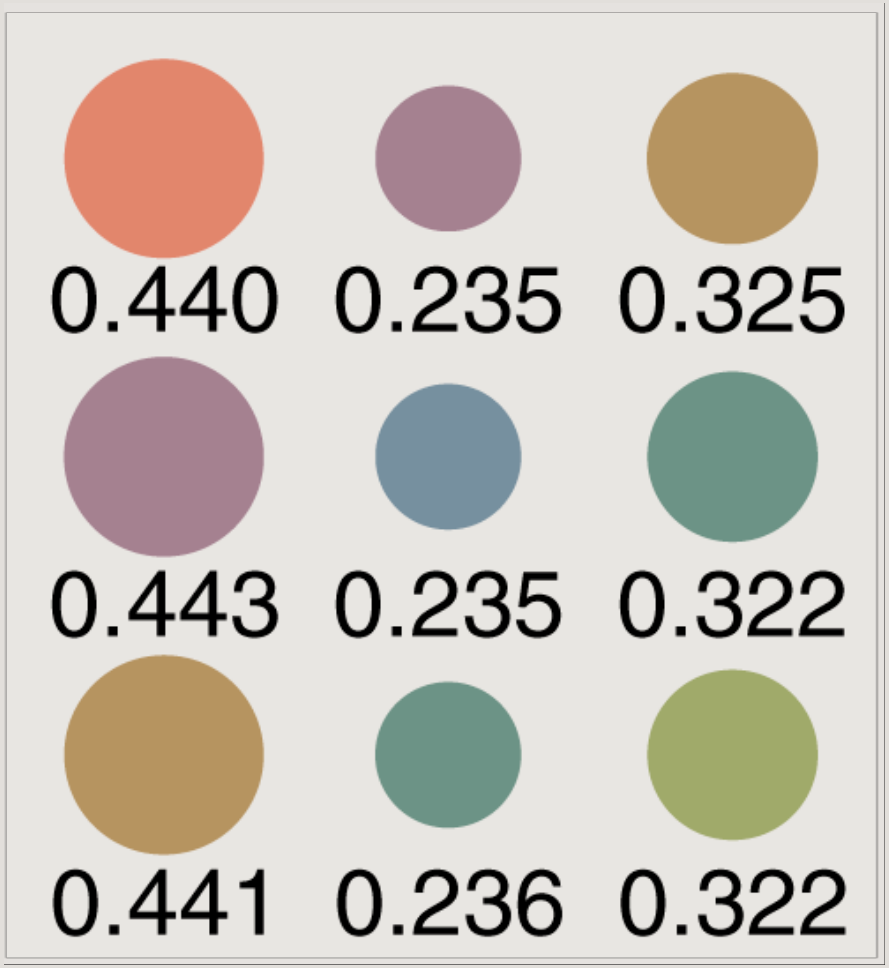


Powers of the transition matrix: 3

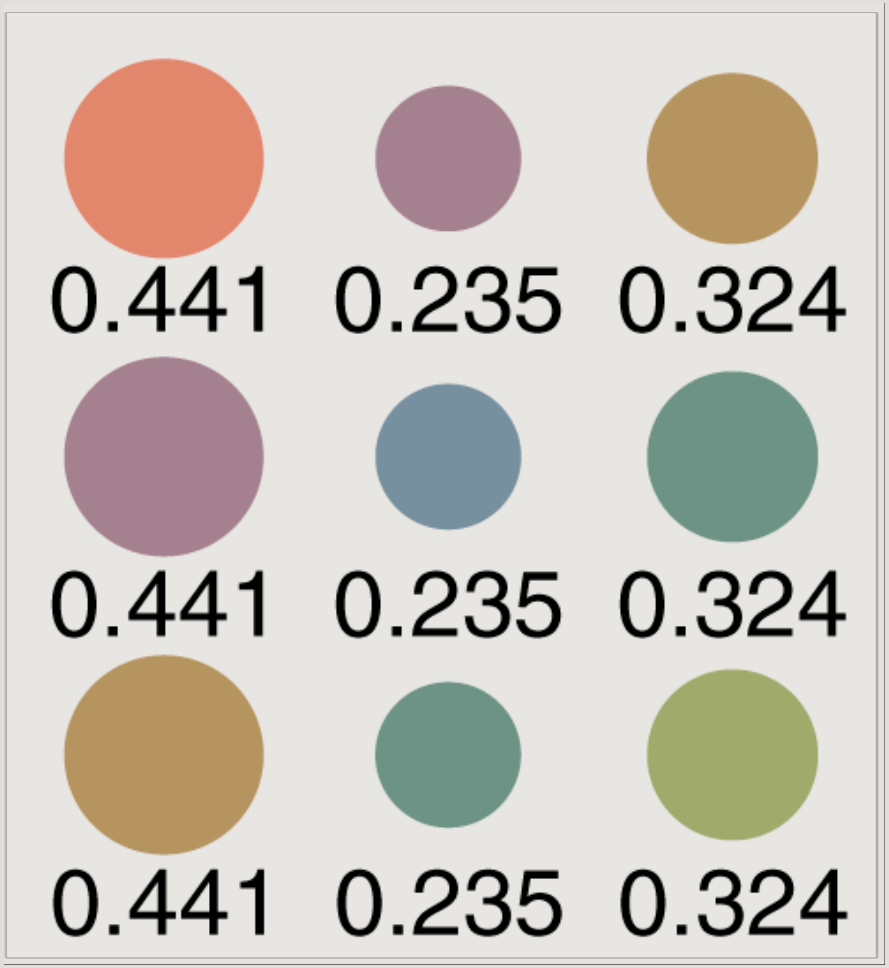
Powers of the transition matrix: 4



Powers of the transition matrix: 5



Powers of the transition matrix: 7



Proving LLN for Markov systems

- Framed as a problem on sums of random variables: $\bar{x}_n = \frac{1}{n}(x_1 + x_2 + x_3 + \cdots + x_n)$ for $x_i \in \mathbb{Z}$, where $E[x_i] = \mu$ and the variance σ^2 is finite.
- Want to show that $\bar{x}_n \to \mu$ as $n \to \infty$
- Chebyshev's inequality: $\Pr(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2}$.
- Chebyshev (and others) proved for *i.i.d.* variables, but lack of correlation was essential to the proof.
- Markov removed this restriction, allowing x_{i+1} to depend on x_i.
- Done by showing that upper and lower bounds on a solution converge to the same limit.

Markov's Conclusion

> *“ Thus, independence of quantities does not constitute a necessary condition for the existence of the law of large numbers. ”*
>
> — A. Markov (1906)

Was this really new?

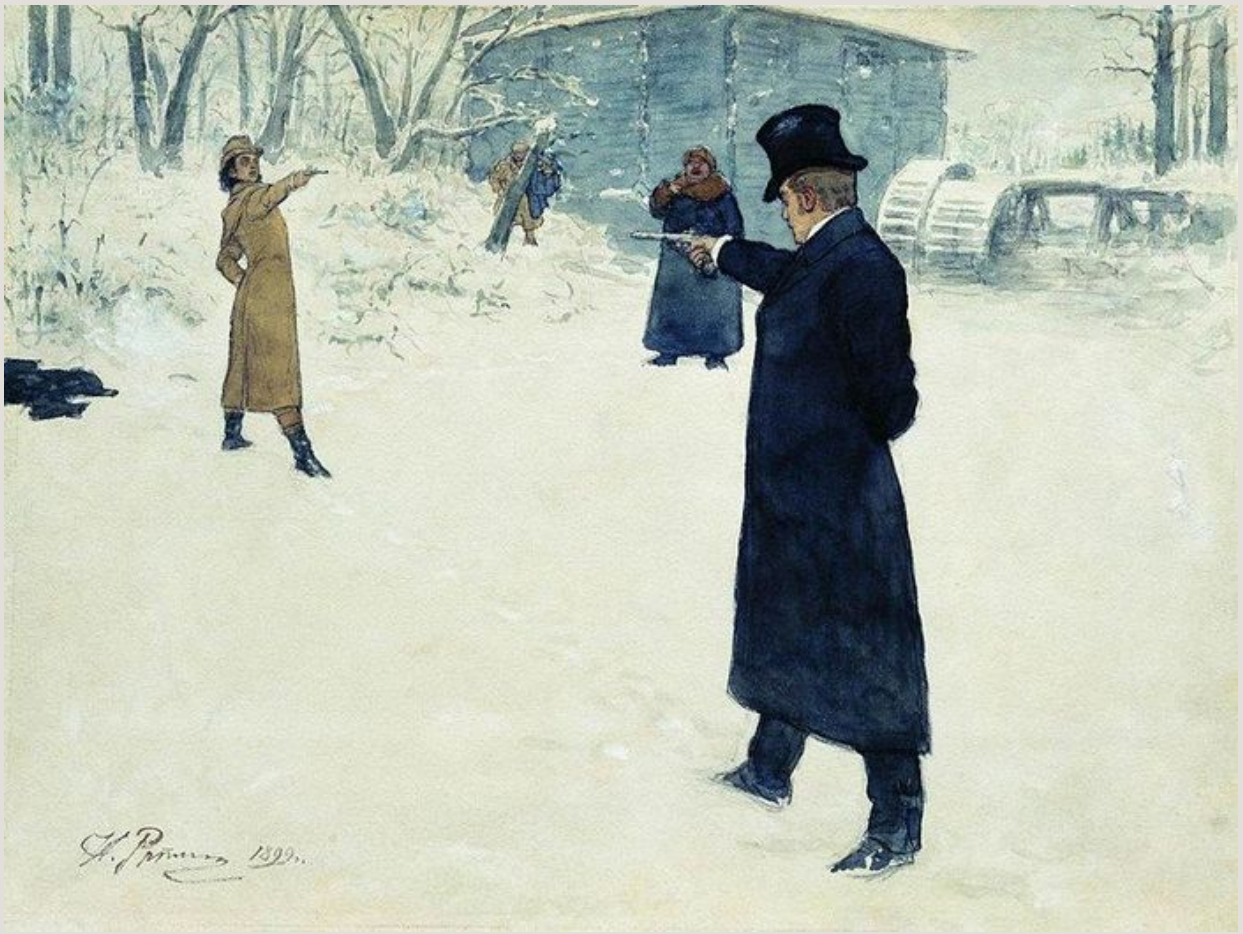
- Fermat and Pascal studied the gambler's ruin problem.
- Some urn models of Daniel Bernoulli can be seen as Markov chains.
- So can the Galton-Watson branching process.
- But not a issue that attracted much attention.
- Independence is very handy because of simple composition of probabilities.
- Without computing machinery, Markov's apparatus is not practical.

What happened next?

- Not seen as a new calculational tool or a new branch of probability theory; instead a footnote on the proof of LLN.
- No hint of applications.
- Nekrasov never backed down.

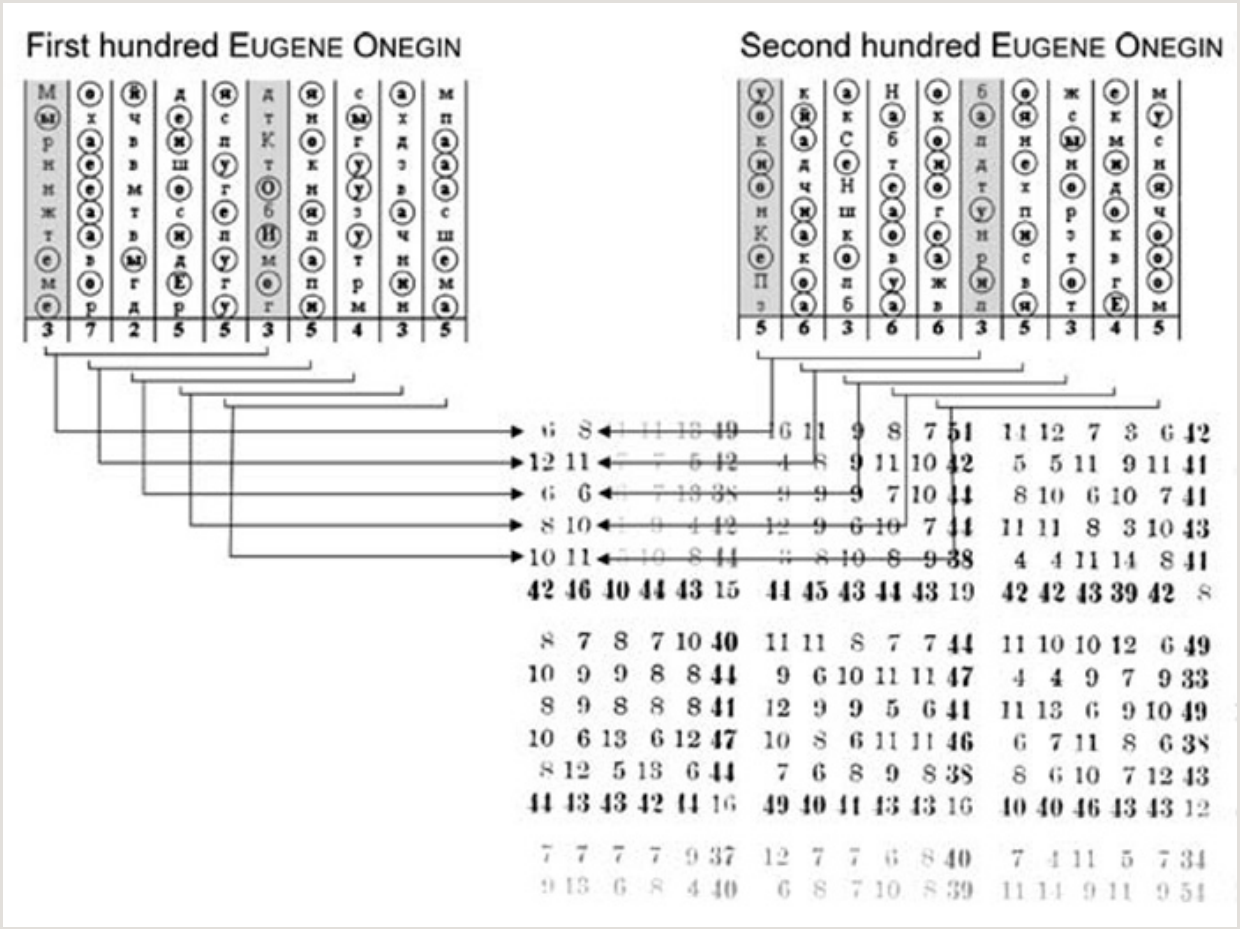
Eugene Onegin

From widow Clicquot and from Moët,
the draught whose blessings are agreed,
in frosted bottle, for the poet
is brought to table at full speed.
Bubbles like Hippocrene are spraying;
once, with its foaming and its playing,
(a simile of this and that)
it held me captive; tit for tat,
friends, recollect how I surrendered
my last poor lepton for a sup!
recall, by its bewitching cup,
how many follies were engendered;
how many lines of verse, and themes
for jokes, and rows, and merry dreams!

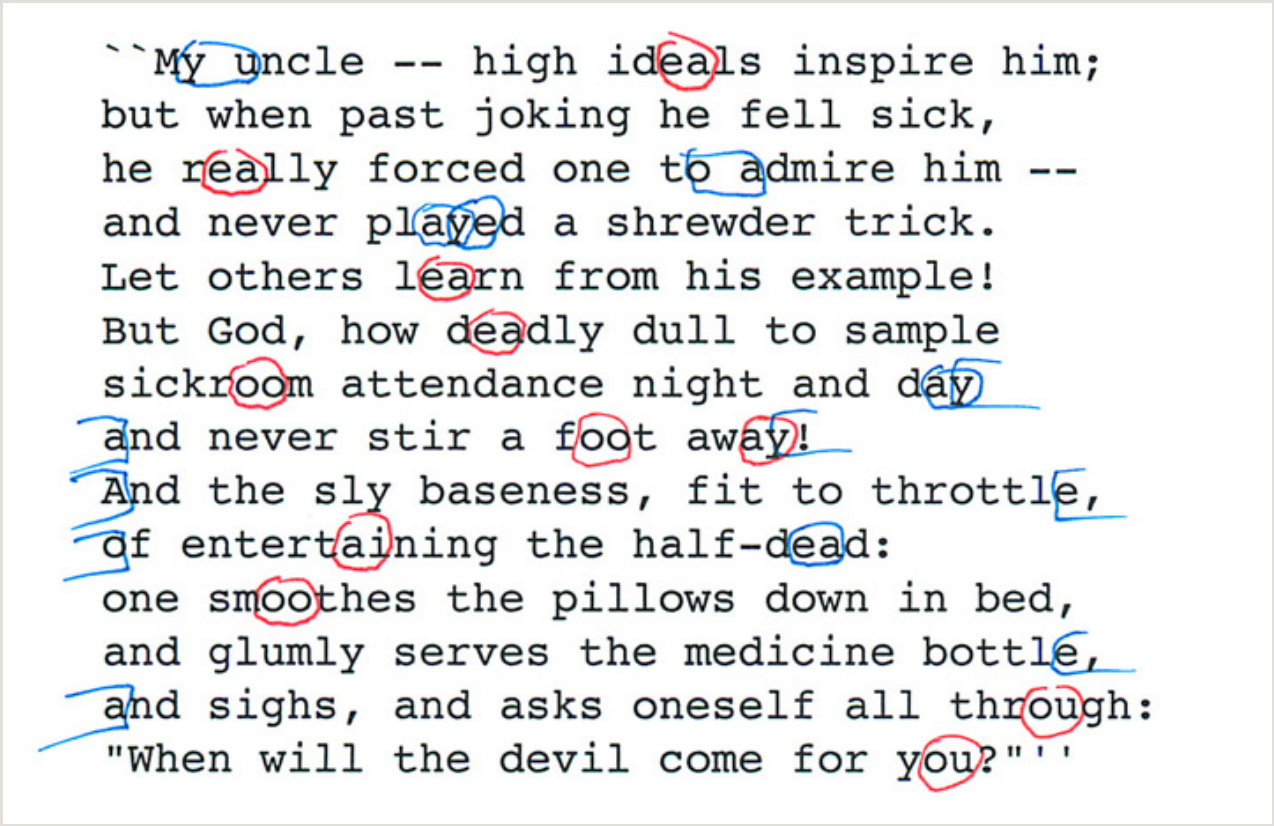


Markov's *Onegin* project

- First 20,000 characters (70 stanzas).
- Remove punctuation and word spaces.
- Count vowels and consonants.
- Count pairs *vv*, *vc*, *cv*, *cc*.
- Calculate moments (mean and variance).

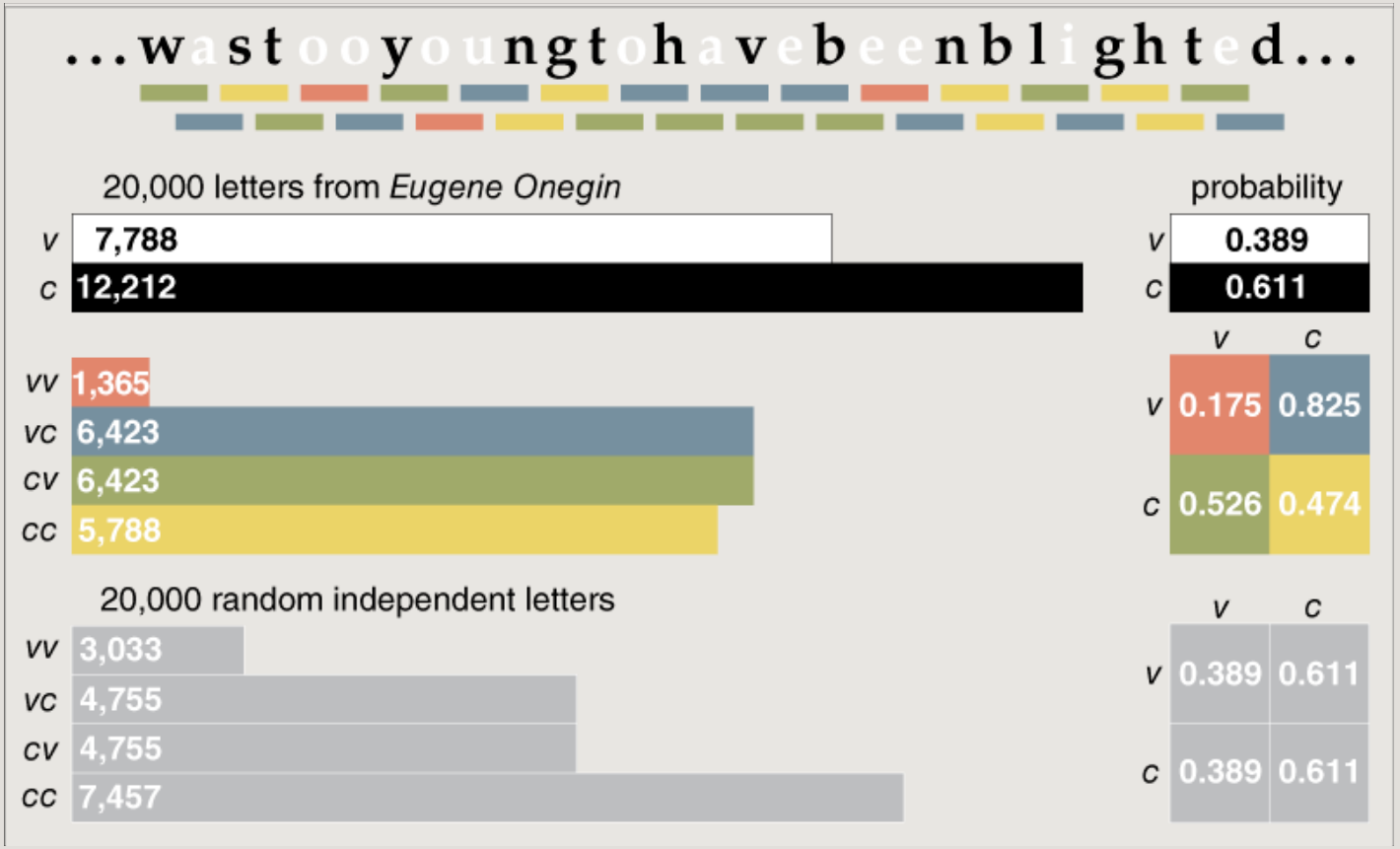

First hundred EUGENE ONEGIN
Second hundred EUGENE ONEGIN

Searching for vowel pairs



``My uncle -- high ideals inspire him;
but when past joking he fell sick,
he really forced one to admire him --
and never played a shrewder trick.
Let others learn from his example!
But God, how deadly dull to sample
sickroom attendance night and day
and never stir a foot away!
And the sly baseness, fit to throttle,
of entertaining the half-dead:
one smoothes the pillows down in bed,
and glumly serves the medicine bottle,
and sighs, and asks oneself all through:
"When will the devil come for you?"''

Results


...wastooyoungtohavebeenblighted...
20,000 letters from Eugene Onegin
v 7,788
c 12,212
vv 1,365
vc 6,423
cv 6,423
cc 5,788
20,000 random independent letters
vv 3,033
vc 4,755
cv 4,755
cc 7,457
probability
v 0.389
c 0.611
v c
v 0.175 0.825
c 0.526 0.474
v c
v 0.389 0.611
c 0.389 0.611

The Dénouement

What became of Markov?

- On the winning side, politically, after 1917.
- But still a malcontent: bootless.
- Nekrasov sunk into obscurity.

How did these ideas reach us?

- Except in Russia, not by people reading Markov's paper.
- Much of Markov's work is still available only in Russian.
- The Onegin paper was translated by Morris Halle in 1955, but never published.
- Translation by David Link and others, published in *Science in Context* (2006).

Example of a statistical investigation of the text of "Eugene Onegin" illustrating the dependence between samples in chain

A. A. Markov

(From Bulletin de l'académie Impériale des Sciences de St. Pétersbourg, 1913, pp. 153-162.)

Translated* by Morris Halle, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

The chain of transmission

- Early reviews by Polya and J. M. Keynes.
- 1928 ICM meeting in Bologna (Bernard Bru, 2003).
- Claude Shannon's use of Markov chains in defining the entropy of language, 1948.
- Change in emphasis: Proving LLN no longer of much interest.

MCMC

- The 1953 paper by Metropolis et al.: "Markov chain" does not appear.
- By 1964, Hammersley and Handscomb have a full chapter on Markov chains.
- Who made the connection, and when?

From Bill Press

Hi, Brian.
It's an interesting question, and I've never tried to find the answer. I'm guessing that Rosenbluth (who, I think, really did the work on Metropolis et al.) completely understood the relationship between detailed balance (the physics concept) and the ergodicity theorem, but that the Markov thing (per se) was simply subsumed in the fact that the physical laws being modeled were manifestly memoryless: state at time t implied state at time t+\Delta t.

Cheers,
Bill P.

Thanks

email: brian@bit-player.org

slides: **http://bit-player.org/extras/markov-sfi**

American Scientist article: **http://www.americanscientist.org/issues/pub/2013/2/first-links-in-the-markov-chain**

blog post: **http://bit-player.org/2013/driveling**

drivel generator: **http://bit-player.org/extras/drivel/drivel.html**