TÉCNICO LISBOA

Main corollary

$$\| Q^{*k} - U \|_{TV} = \frac{1}{2} \sum_{\pi \in S_n} \left| Q^{*k}(\pi) - \frac{1}{n!} \right|$$

We now have formulas for this. It equals $\frac{1}{2} \sum_{\pi \in S_n} \left| \frac{\binom{n+2^k-r(\pi)}{n}}{2^{nk}} - \frac{1}{n!} \right| =$

$$= \frac{1}{2} \sum_{r=1}^{n} b(n,r) \left| \frac{\binom{n+2^k-r}{n}}{2^{nk}} - \frac{1}{n!} \right| \quad \text{with } b(n,r) = \#\{\pi : r(\pi) = r\} = A(n,r-1)$$

↑ Eulerian number.

where

$$A(n,k) = \#\{\pi \mid d(\pi) = k\}$$

We know "everything" about these numbers. In particular we have the simple recurrence

$$A(n,k) = (n-k)\,A(n-1,k-1) + (k+1)\,A(n-1,k).$$

For $n = 52$ the first sum is hopeless: $52!$ is bigger than the number of atoms in the universe. The sum to 52 can easily be done on a computer.

For asymptotics

$$\frac{A(n,k)}{n!} = P(a < U_1 + \cdots + U_n < a+1) \qquad 0 \le a \le n-1$$

↑ chance a permutation at random has k descents

$U_1, \ldots, U_n$ are independent uniform random variables on $[0,1]$

→ this is very well understood

From Eriksson's lectures

To see the kind of analysis that is involved. Recall the separation distance

$$sep(k) = \max_{\pi \in S_n} 1 - \frac{Q^{*k}(\pi)}{U(\pi)} \quad \text{and} \quad \| Q^{*k} - \pi \|_{TV} \le sep(k)$$

For us $sep\ k = 1 - n! \frac{\binom{2^k}{n}}{2^{nk}}$ (because the maximum is achieved for $\pi = n \cdots 1$)

Now $n! \frac{\binom{2^k}{n}}{2^{nk}} = \frac{n!}{2^{nk}} \frac{2^k (2^k-1) \cdots (2^k-n+1)}{n!} = \frac{2^k}{2^k}\left(1 - \frac{1}{2^k}\right) \cdots \left(1 - \frac{n-1}{2^k}\right)$

$$= e^{\sum_{j=1}^{n-1} \log\left(1 - \frac{j}{2^k}\right)} \qquad \log(1-x) = -x + O(x^2)$$

$$= e^{-\sum_{j=1}^{n-1} \frac{j}{2^k} + O\left(\frac{j}{2^k}\right)^2}$$

$$= e^{-\binom{n}{2}\frac{1}{2^k} + O\left(\frac{k^3}{2^{2k}}\right)} \qquad k = 2\log_2 x + c$$

see beginning of Chap. 3 of notes.

$$\approx e^{-\frac{1}{2}c+1} + O\left(\frac{1}{n}\right)$$

Lecture 3 Today will talk about a completely different subject: how do the carries work in addition.

Consider a matrix

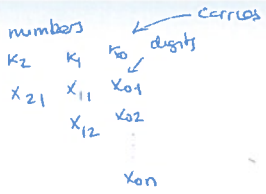
$$\begin{array}{c} \\ 1 \\ \\ q \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & & & & q \end{array} \\ \begin{pmatrix} 0 & \cdots & & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \\ & & & & \\ 0 & 1 & 1 & \cdots & 1 \end{pmatrix} \end{array}$$

two corners

in base b, the chance of a carry (picking digits at random) is $\binom{b}{2}/b^2 = \frac{1}{2} - \frac{1}{2b}$

For $b = 10$ this is 0.45. For $b = 2$ it is 0.25

Suppose we are adding n numbers



X_{ij} uniform on $0,1,\dots,b-1$ (base b)

There is dependence on what happened before.

Observation: The K process forms a Markov chain on $0,1,2,\dots,n-1$ (maximal carry is $n-1$ for n digits)

[Markov law is a probabilistic version of the laws of mechanics]

Amer Math Month - John Holte - "Carries, Combinatorics and an amazing matrix" is where I learned of this

$$P(i,j) = P(K^1=j \mid K=i) \qquad 0 \le i, j \le n-1$$

For $n=2$ the transition matrix is $\frac{1}{2b}\begin{pmatrix} b+1 & b-1 \\ b-1 & b+1 \end{pmatrix}$ (for $n=3$ it is already not symmetric)

$$\text{e.g.} \quad P(K^1=0 \mid K=0) = \frac{b+1}{2b} = \frac{1}{2} + \frac{1}{2b}$$
$$P(K^1=1 \mid K=0) = \frac{1}{2} - \frac{1}{2b}$$

For $n=3$ the matrix is $\frac{1}{6b^2}\begin{pmatrix} b^2+3b+2 & 4b^2-4 & b^2-3b+2 \\ b^2-1 & 4b^2+2 & b^2-1 \\ b^2-3b+2 & 4b^2-4 & b^2+3b+2 \end{pmatrix}$

one can write down an expression in general (see notes).

What's so amazing? A) $P(i,j)$ has real eigenvalues $1, \frac{1}{b}, \frac{1}{b^2}, \dots, \frac{1}{b^{n-1}}$

The eigenvalues of b-shuffling are $1, \frac{1}{b}, \frac{1}{b^2}, \dots, \frac{1}{b^{n-1}}$ (with certain multiplicities which are known) !

B) The stationary distribution of $P(i,j)$, $\pi(j)$ is $\frac{A(n,j)}{n!}$ where $A(n,j)$ is the Eulerian $\#\{\sigma \in S_n \mid \operatorname{desc}(\sigma)=j\}$

What is the symmetric group doing here?

Example: $n=2$ $\quad A(2,0) = A(2,1) = 1$

$$\pi(0) = \pi(1) = \frac{1}{2}$$

$n=3$ $\quad A(3,0) = 1 \quad A(3,1) = 4 \quad A(3,2) = 1$

$$\pi(0) = \frac{1}{6} \quad \pi(1) = \frac{2}{3} \quad \pi(2) = \frac{1}{6}$$

C) $P_a P_b = P_{ab}$ (good work trying to see this directly)

Clearly this must be related to card shuffling! What is the connection?

Let $k_0 = 0, k_1, k_2, \dots$ be the carries of n numbers mod b

Let $d_0 = 0, d_1, d_2, \dots$ be the number of descents when n cards are repeatedly b-shuffled (not obvious this is a Markov process at all)

Theorem (Diaconis, Fulman) $\quad P_a(k_1=a_1, \dots, k_\ell=a_\ell) = P_a(d_1=a_1, \dots, d_\ell=a_\ell)$ for any $n, b, \ell, a_1, \dots, a_\ell$.

First proof involved elaborate computations with symmetric functions but then we found a Math proof which I will now explain.

TÉCNICO LISBOA

$$\begin{array}{r} 3 \circ 3 \\ 9 \circ 2 \\ 8 \quad 0 \\ 6 \circ 6 \\ 6 \quad 2 \\ 2 \quad 4 \\ 5 \circ 9 \\ 8 \circ 7 \\ 4 \quad 1 \\ \hline 5\ 1 \end{array}$$

Tradenberg system of speed arithmetic

How do the dots go? base b, ind. dist. digits $X_i = \sum_0^i$ dot positions.

Fact: If left hand digit one i.i.d. uniform then so are the right hand digits. (when we add a uniformly distributed to a fixed one it is still uniformly distributed)

and $X_i = 1 \iff$ descent at RHS

So the dots are the up/down pattern in a random sequence.

This argument can be made into a proof of this theorem (see notes or paper in Am. Math Monthly) □

4 fact about the X_i process

- $P(X_i = 1) = \frac{\binom{b}{2}}{b^2} = \frac{1}{2} - \frac{1}{2b}$
- $P(X_i = X_{i+1} = \cdots = X_{i+d} = 1) = \frac{\binom{b}{d+1}}{b^{d+1}}$
- X_i is stationary, 1-dependent, determinantal just process (0 or 1 as in whether there is a tree or not in that position)

if $|i_j - i_{j+1}| \geq 2$ then X_{i_j}, X_{i_j+1} are independent

The central limit theorem holds for such processes, so if

$$S_k = X_1 + \cdots + X_k$$

$$E(S_k) = (k-1)\left(\frac{1}{2} - \frac{1}{2k}\right) \qquad \text{Var}(S_k) = \frac{k+1}{12}\left(1 - \frac{1}{b^2}\right)$$

$$\Rightarrow P\left(\frac{S_k - \text{Ev}}{\text{s.d}} \leq x\right) \longrightarrow \int_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt$$

determinantal means

$$P(X_1 = \varepsilon_1, \ldots, X_{n-1} = \varepsilon_{n-1}) = \frac{1}{b^n} \det\left(\frac{S_{i+1} - S_i + b - 1}{b-1}\right)$$ ← matrix is $(n+1) \times (n+1)$

where we have k ones among $\varepsilon_1, \ldots, \varepsilon_{n-1}$ in positions $s_1 < s_2 < \cdots < s_n$

$s_0 = 0, \quad s_{n+1} = n$

See the paper Barndorff, Diaconis, Kilman "On adding a list of numbers"

What one carries anyway? Answer: they are cocycles!

G finite grp, $H \leq G$ subgp. Let $X \subset G$ be coset representatives $G = \bigcup_{x \in X} xH$

The product of two coset representatives is not necessarily a coset representative.

e.g. $G = \mathbb{C}_{100}$ $H = C_{10} = \{0, 10, \ldots, 90\}$

Given $x, y \in X$ $\quad xy = \overline{xy}\, f(x,y)$ ← coset rep. of xy

for $f(x,y) \in H$ called a factor set

All we said about carries works in general for cocycles. See the paper mentioned above
All the papers above go this.

I.O.U. from last time: Said we'd prove $\frac{A(n,d)}{n!} = P(d \le U_1 + \cdots + U_n < d+1)$
(sum of i.d. uniforms on $[0,1]$)

Science in 1900 noticed the expansion on left studied by Euler and on the right by Laplace more than 100 years. Why?

$G = \mathbb{R}$, $H = \mathbb{Z}$ $\quad G/\mathbb{Z} = S^1$

U_1, U_2, U_3, ... i.i.d coset representatives

$V_1 = U_1$
$V_2 = U_1 + U_2$
$V_3 = U_1 + U_2 + U_3$
...

again there is a carry if there is a descent (left hand rule)

What did we learn about carries? (sums mod b)

$$E(K_j) = \frac{b-1}{2}\left(1 - \frac{1}{b^j}\right)$$

$$\mathrm{Var}(K_j) = \frac{b+1}{12}\left(1 - \frac{1}{b^{2j}}\right)$$

$T_k = K_1 + \cdots + K_k \sim$ normal (mean, s.d.)

What is the rate of convergence to the stationary distribution?

$$k = \lfloor 2\log_b x + \log_b 2 \rfloor$$

Then $\mathrm{sep}(k) = 1 - e^{-\frac{1}{2c}} + O(\frac{1}{x})$

There were open problems in Knuth's book that this process which we called carries with the above perspective.
the cut off phenomenon.

John Holte mentioned above also computed the eigenvectors of the matrix. Those left to the shuffling matrix

If $d(\pi) = \frac{n-1}{2}$ is right ev of b-shuffling.

Peak (π) is $\frac{n-1}{3}$ ev $\frac{1}{b}$, var $\frac{1}{b^2}$

Now subject

Sections of generating functions: $a_i \in \mathbb{R}$ sequence of numbers get $f(x) = \sum_{k=0}^{\infty} a_k x^k$.

These are often nice functions - rational/algebraic.

Suppose $f(x) = \frac{h(x)}{(1-x)^{n+1}}$ $\quad h(x) = \sum_{i=0}^{n} h_j x^j$

Examples: (1) $a_i = 1$ $\quad f(x) = \frac{1}{1-x}$

(2) $a_i = i$ $\quad f(x) = \frac{x}{(1-x)^2}$

(3) $a_k = k^n$ $\quad \sum k^n x^n = \frac{A(x)}{(1-x)^{n+1}}$ $\quad A(x) = \sum_{i=0}^{n+1} A(n,d)\, x^d$

this is the signed sum of Eulerian numbers.

(4) Hilbert series of an affine variety over $\mathbb{C}$

TÉCNICO LISBOA

Consider a fixed b.

$\sum a_{bk} x^k$ This is called a reaction of the gen. fct.

Then this is $\frac{h^{<b>}}{(1-x)^{n+1}}$ with $h^{<b>}$ is polynomial of degree n

$h_i^{<b>} = \sum c(i,j)\, b_j$ Linear combination

↑ This is the score matrix !!

Lecture 4 Move out the world of Markov chains.

Starting in the 50's Markov chains have revolutionized science. The kind of answers we have obtained for the simple examples we discussed before can open for many important examples.

Let f : symbol space $\longrightarrow \{a, b, c, \dots\}$

If the cypher is a substitution cypher the job is to get the right f.

Recall Transition matrix for written english $m(\alpha, \beta)$ = proportion of times β follows α. (40×40 matrix)

Plausibility Can assign a plausibility to f, $Pl(f) = \prod_{i=1}^{N} m(f(a_i), f(a_{i+1}))$

Problem: find f for which the plausibility is large

Idea: try Monte Carlo Markov chain:

- start w/ some guess f_0
- pick $s, s' \in$ symbol space uniformly and switch the values at s, s' — don't get stuck at local maxima.
- Compute $Pl(f_0)$, $Pl(f_1)$ If $Pl(f_1) > Pl(f_0)$ go to f_1
  If $Pl(f_1) < Pl(f_0)$ flip a coin of Prob $Pl(f_1)/Pl(f_0)$ if heads go to f_1 else stat at f_0

This is an important algorithm in scientific computing called the Metropolis algorithm.

Metropolis algorithm (Nick Metropolis - Sonke Ze) (w/ Teller, Teller, Rosenbluth)

$|X| < \infty$, $\pi(x) > 0$

Problem: Sample from π. Need some way of using coin on X.

Let $K(x,y)$ be a Markov chain (having nothing to do with π) $K(x,y) > 0 \iff K(y,x) > 0$

Let $A(x,y) = \frac{\pi(y) K(y,x)}{\pi(x) K(x,y)}$ be the acceptance probability (note this is independent of the normalizing constant)

$= \frac{\pi(y)}{\pi(x)}$ if $K(x,y)$ is symmetric

Algorithm: From X choose y from x

- choose y with probability $K(x,y)$
- calculate $A(x,y)$. If $A(x,y) \geq 1$ go to y
  If $A(x,y) < 1$ flip $A(x,y)$ coin. If heads, go to y. If tails stay at x.

what does this do? the matrix for the Metropolis Markov chain is

$$M(x,y) = \begin{cases} K(x,y) & \text{if } A(x,y) \geq 1 \\ K(x,y) A(x,y) & \text{if } A(x,y) < 1 \\ K(x,x) + \sum_{z : A(x,z) < 1} K(x,z)(1 - A(x,z)) & \end{cases}$$

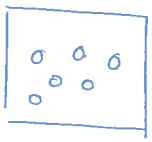
Theorem: M is a $\frac{\pi}{z}$-reversible Markov chain

If you want a life problem try to find anything about the Metropolis algorithm?

It is very hard and not much is known. This is one of the most used algorithms in Scientific Computing!

Original metropolis application: Hard disks in a box

Fix $n > 0$ disks and consider placing n disks of radius r in a box square so that they don't overlap.

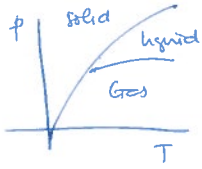


The set of all possible configs is a compact subset in $\mathbb{R}^{2n}$ and as such has a uniform distribution (normalized Lebesgue measure).

Problem: Pick a conf at random

open problem: how large/small must r be for betti #s to be trivial (e.g. connected, π_1).

Who cares?



phase space of many substances

why do they all look the same?

Kirk 1920s – there should be a phase transition if there are just repulsive forces.
Conjectured such in disks in a box. And these phase transitions.

area covered > 0.705 ⇒ disks all lined up

< 0.705 look random.

Metropolis invented in algorithm to sample from the uniform distribution for the disk problem.

In this case it goes: Start at x

pick disk at random and move its center in any dir ε. If OK go, if not stay.

To get estimates for one disk in one dimension is already extremely difficult.