DON'S WORK ON BACKTRACKING

SEQUENTIAL IMPORTANCE SAMPLING

PITEÅ, JAN, 2018

EXAMPLE SEQUENTIAL IMPORTANCE SAMPLING

PROBLEM HOW MANY SELF AVOIDING PATHS ARE THERE FROM (0,0) TO (10,10) IN A 10X10 GRID?

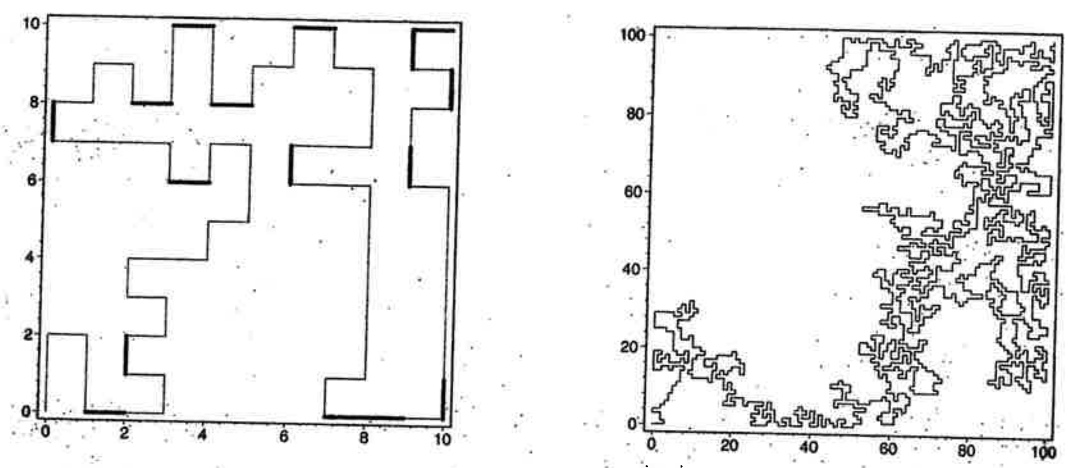
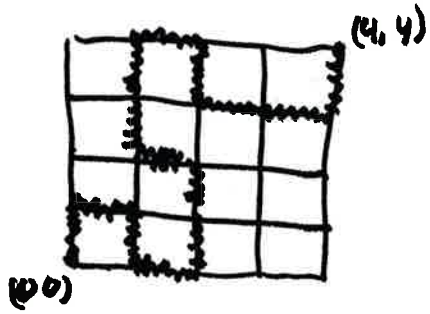


FIGURE 2. Left: A SAW crossing the 10 × 10 square. The thick steps have probability 1. That is, each of them is the only eligible step at the time when it is taken. Right: A SAW crossing the 100 × 100 square, obtained via Knuth's algorithm.

IDEA (KNUTH) BUILD ONE SEQUENTIALLY



$$P(\gamma) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{3}\cdot 1\cdot\frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{2}\cdot\frac{1}{3}\cdot 1\cdot 1\cdot\frac{1}{2}\cdot 1\cdot\frac{1}{2}\cdot 1$$

$$T(\gamma) = \begin{cases} 0 & \text{STUCK} \\ 1/P(\gamma) & \text{IF OK} \end{cases}$$

OBSERVE

$$E(T) = \sum_\gamma T(\gamma)\,P(\gamma) = \sum_\gamma \frac{1}{P(\gamma)}\,P(\gamma) = \#\text{ PATHS}$$

KNUTH FOUND

PATHS $= (1.6 \pm 0.3)\cdot 10^{24}$

AV LENGTH $= 92 \pm 5$

% THROWN (5,5) $= 81 \pm 0.5$ %

KNUTH (1976) MATHEMATICS AND COMPUTER SCIENCE: COPING WITH FINITENESS. SCIENCE

? DO THE NUMBERS MEAN ANYTHING ?

- HUGE VARIABILITY (AS USUAL FOR IMPORTANCE SAMPLES)
- SEEMS AS IF 10^6 SAMPLES DETERMINED BY 1 OR 2 #'S
- ESTIMATION OF ± S.D. OR EFFECTIVE SAMPLE SIZE PARTICULARLY QUESTIONABLE

IS THERE MORE INFORMATION

- THAT WE HAVE ?
- IN THE DATA ?

SOMETHING TO UNDERSTAND:

KNUTH: $\# = (1.6 \pm .03) \cdot 10^{24}$

EXACT: $\# = 156878030464750013214106$

$= 1.5687 \times 10^{24}$

H

SAME IDEA(S)

\# SOLUTIONS OF PUZZLES

\# GRAPHS WITH GIVEN DEGREE SEQUENCE

\# BINARY TABLES WITH GIVEN ROW COL SUMS

Table 1. Occurrence Matrix for Darwin's Finch Data

	Island																
Finch	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*
Large ground finch	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Medium ground finch	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0
Small ground finch	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0
Sharp-beaked ground finch	0	0	1	1	1	0	0	1	0	1	0	1	1	0	1	1	1
Cactus ground finch	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	0	0
Large cactus ground finch	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
Large tree finch	0	0	1	1	1	1	1	1	1	0	0	1	0	1	1	0	0
Medium tree finch	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Small tree finch	0	0	1	1	1	1	1	1	1	1	0	1	0	0	1	0	0
Vegetarian finch	0	0	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0
Woodpecker finch	0	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0
Mangrove finch	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Warbler finch	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

NOTE: Island name code: A = Seymour, B = Baltra, C = Isabella, D = Fernandina, E = Santiago, F = Rábida, G = Pinzón, H = Santa Cruz, I = Santa Fe, J = San Cristóbal, K = Española, L = Floreana, M = Genovesa, N = Marchena, O = Pinta, P = Darwin, Q = Wolf.

BUILD UP TABLES 1 COL AT TIME

IDEAS
- KEEP FROM GETTING STUCK USING GALE-RYSER
- PUT 1's IN ROWS WITH PROB ∝ ROW SUM

- CHEN, DIACONIS, HOLMES, LIU (2005) SEQUENTIAL MONTE CARLO METHODS FOR STATISTICAL ANALYSIS OF TABLES J.A.S.A.
- BLITZSTEIN, DIACONIS (2011) SEQUENTIAL IMPORTANCE SAMPLING ALGORITHM FOR GENERATING RANDOM GRAPHS WITH PRESCRIBED DEGREES J. INTERNET MATHS

QUESTIONS

- GIVEN SUCH A TASK, HOW CAN WE 'NOT GET STUCK' (CAN SOMEONE THINK OF A METHOD FOR 3-D S.A.W?)

- CAN WE PROVE ANYTHING (BOUNDING VARIANCE IS A BAD IDEA)

SUPPOSE WE ESTIMATE

$$I(f) = \int f \, d\mu \quad \text{BY} \quad I_n(f) = \frac{1}{n} \sum_{1}^{n} f(x_i) \, \rho(x_i)$$

$$x_i \text{ iid } \nu$$

$$\rho = \frac{d\mu}{d\nu}$$

"TH" LET $Q_n = \max_{1 \le i \le n} \rho(x_i) \Big/ \sum_{i=1}^{n} \rho(x_i)$

<u>IFF</u> $E(Q_n) < \epsilon \iff |I_n(f) - I(f)|$ SMALL

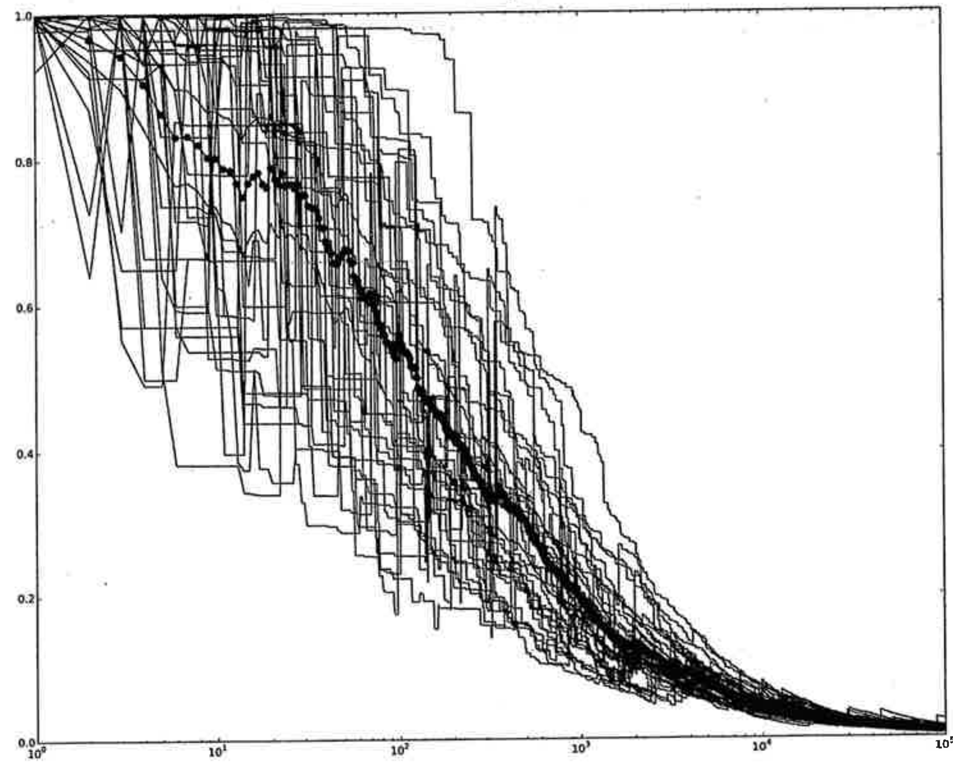


FIGURE 5. Performance of q_n for Knuth's self-avoiding walks on a 10×10 grid. The values of q_n, denoted by the thick dots, were estimated from 31 simulations of Q_n, which are depicted by the solid lines. Picture courtesy of Marc Coram.

- $Q_n, q_n > 0.2 \quad n \leq 10^3$
- SMALL AT 10^4
- TINY AT 10^5

SOME THEORY

$$Q_N = \max_{1 \le i \le N} \frac{\rho(X_i)}{\rho(X_1) + \cdots + \rho(X_N)}, \quad q_N = E(Q_N), \quad \epsilon_N = E|I_N(1) - 1|$$

WANT TO SAY "" q_N SMALL $\Leftrightarrow$ ϵ_N SMALL ""

THEOREM(A) $\Leftarrow$ IN COMPLETE GENERALITY: $\exists$ UNIVERSAL $C \ni \forall\ \nu, \mu$

$$q_n \le C \max\left\{ \frac{1}{n}, \frac{\log\log 1/\epsilon_n}{\log 1/\epsilon_n} \right\}$$

THEOREM(B) $\Rightarrow$ FOR EXPONENTIAL FAMILIES, ESTIMATING $Z(\beta)$

BY $\hat{Z}(\beta) = \frac{Z(\beta_0)}{n} \sum_{i=1}^{n} e^{-(\beta-\beta_0) H(X_i)}$ $\quad X_i$ i.i.d. P_{β_0}.

SET UP (NEED TO TAKE LIMITS)

$(\mathcal{X}_N, \lambda_N)$ MEASURE SPACES

$H_N : \mathcal{X}_N \to \mathbb{R}$, $\quad P_{\beta,N}(x) = Z_N^{-1}(\beta)\, e^{-\beta H_N(x)}$ W.R.T. λ_N

SUPPOSE HAVE $L_N \uparrow \infty$: $\lim \frac{\log Z_N(\beta)}{L_N} = p(\beta)$ "LIMITING FREE ENERGY"

THE MODEL HAS A k^{th} ORDER PHASE TRANSITION AT β IF FIRST $k-1$ DERIVATIVES OF $p(\beta)$ ARE IN C' BUT NOT THE k^{th}.

Th. IF $p(\beta)$ EXISTS AND IS DIFFERENTIABLE (NO FIRST ORDER PHASE TRANSITION)

$p(\beta_0)$ EXISTS

$|H_N(x)| \leq c L_N$ SOME FIXED $c>0$, All $x \in X_N$

THEN IF $\delta_{\lambda N} \ll \frac{1}{L_N^\theta}$, ANY $\theta > 0$,

$$\frac{\log \hat{Z}_{\lambda N}(\beta)}{L_N} \rightarrow p(\beta) \text{ in PROBABILITY}$$

- THIS IS OK FOR ISING + POTTS MODELS

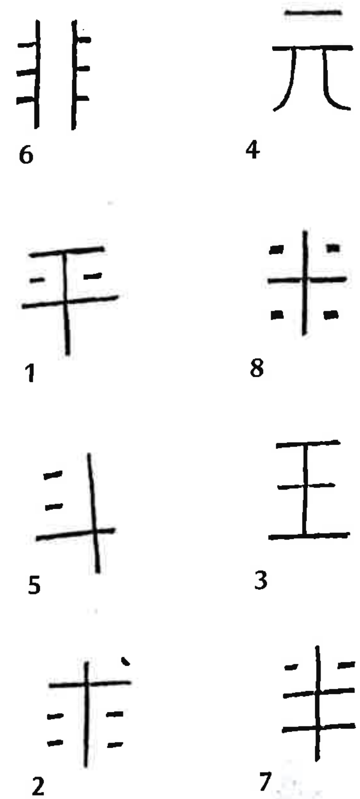
  VARIOUS EXPONENTIAL RANDOM GRAPH MODELS

  ⋮

- KNUTH'S MODEL IN THIS CLASS (DON'T KNOW IF NO FIRST ORDER TRANSITION)

BACKTRACKING (MANY MORE EXAMPLES)

EX: FROM 'MAGICAL MATHEMATICS' P.D + R. GRAHAM



CAN WE DO THIS IN ENGLISH?

<u>EX</u> HAT RAT HUT HAG
RUG RAG RUT HUG

HOW ABOUT WITH 4 (OR 5) LETTER WORDS?

DON'T WANT TO TRY ALL $(325)^5 \doteq 3.63 \times 10^{12}$ COMBINATION

BACKTRACKING: (THANKS TO S.P. HOLMES)

<u>START</u> $\binom{a}{b}$ $325 = \binom{26}{2}$ CHOICES (THAT EXTEND TO 5 LETTER WORDS IN $\mathcal{L}$)

<u>NEXT</u> $\binom{a}{b}\binom{k}{l}$ CHECK IF ak, al, bk, bl EXTEND TO 5 LETTER WORDS

⋮

ONLY NEED TO CHECK 111597 THINGS (VS 10^{12})

 ALAS, NO 5 LETTER SOLUTIONS (EVEN WITH BIG $\mathcal{L}$)

 <u>BUT</u> 261 4 LETTER SOLUTIONS

Good ones and bad ones (beginning of alphabet)

Inbox x

Susan Holmes

to me

A few Good ones (no underlining in red by my spelling program):

"bade" "bode" "code" "cade" "bate" "bote" "cote" "cate" "bads" "bods" "cods" "cads" "ba

"bane" "bone" "cone" "cane" "bate" "bote" "cote" "cate" "bans" "bons" "cons" "cans" "ba

"bana" "buna" "cuna" "cana" "bara" "bura" "cura" "cara" "bant" "bunt" "cunt" "cant" "bar

"bale" "bole" "hole" "hale" "bare" "bore" "hore" "hare" "balt" "bolt" "holt" "halt" "bart" "bo

"bane" "bine" "kine" "kane" "bate" "bite" "kite" "kate" "bans" "bins" "kins" "kans" "bats" "

* "bans" "bens" "mens" "mans" "bass" "bess" "mess" "mass" "bant" "bent" "ment" "mant"

"bane" "bene" "mene" "mane" "bate" "bete" "mete" "mate" "bans" "bens" "mens" "mans

** "bale" "bole" "mole" "male" "bare" "bore" "more" "mare" "balt" "bolt" "molt" "malt" "bart"

"bale" "bole" "mole" "male" "bate" "bote" "mote" "mate" "balt" "bolt" "molt" "malt" "batt" '

Really bad ones:

"bard" "bord" "lord" "lard" "baud" "boud" "loud" "laud" "barn" "born" "lorn" "larn" "baun" '

"bane" "bine" "line" "lane" "base" "bise" "lise" "lase" "bank" "bink" "link" "lank" "bask" "b

H.W. PBLM (PLEASE): FIND A 'NICE' 3-LETTER SOLUTION IN HEBREW, KOREAN, SWEDISH, FRENCH, PERSIAN ...

IDEA (DON): USE SEQUENTIAL IMPORTANCE SAMPLING TO ESTIMATE

- RUNNING TIME OF BACKTRACKING
- # SOLUTIONS OF FINAL PBLM
- # SOLUTIONS ON A LEVEL

CHOOSE RANDOM PATH IN BACKTRACKING TREE (SAY LENGTH $\ell+1$)

LET $D_1, D_2, \cdots D_\ell$ BE DEGREES ALONG THE WAY

$D_1 + D_1 D_2 + \cdots + D_1 D_2 \cdots D_\ell$ IS AN UNBIASED ESTIMATOR OF # VERTICES.

<u>EX</u> FOR 5-LETTER WORDS, 10,000 TRIALS OF $D_1 + D_1 D_2 + \cdots + D_1 D_2 \cdots D_\ell$ GIVE

MIN	L.Q.	MED.	MEAN	U.Q.	MAX
650	7150	14,950	113,140	62,075	4,147,325

RIGHT ANS 111,597

TO CONTINUE

- DON'S CHAPTER IN VOL 4B (+ FOLLOWING DANCING LINKS CHAPTER) HAS MANY DETAILS, BELLS + WHISTLES, EXAMPLES.
  THESE ARE WORTH ABSTRACTING
- LARGE STATISTICS LITERATURE ON IMPORTANCE SAMPLING AND PARTICLE FILTERS, BUILDS SETS OF PATHS, < KILL SOME ... / BREED SOME
  SEE THE SURVEY SECTION OF CHATERJEE-DIACONIS 'THE SAMPLE SIZE REQUIRED IN IMPORTANCE SAMPLING' ANN. APPL. PROBAB.
- CAN MAKE MODELS OF 'TYPICAL TREES' AND TRY TO PROVE SOMETHING (SVANTE, MICHAEL)
- IMPORTANCE SAMPLING SEEMS TO WORK MUCH BETTER IN PRACTICE THAN IN THEORY. ?WHY?
- WRITE MORE YOUNGSTERS!