Probability and Computing

June 15, 2025

Maths

- $1 + x \leq \exp(x)$ for $x \in \mathbb{R}$
- $\frac{\exp(x) + \exp(-x)}{2} \leq \exp(x^2/2)$ [proof by Taylor expansion]
- $\binom{n}{k} \in \left[(n/k)^k, (en/k)^k\right]$
- Stirling: $\sqrt{2\pi m}(\frac{m}{e})^m \leq m! \leq 2\sqrt{2\pi m}(\frac{m}{e})^m$

Problem sheet Qs

S1Q2: test matrix equivalence $A = B$ by $Ax = Bx$ for a random vector x

S1Q5: probability $A[i]$ is smallest in randomly ordered A is $1/|A|$

S1Q6: care defining events: $\mathbb{P}$(all even) should be $\mathbb{P}(6$ before $1, 3, 5)$ which can be calculated, and equals $1/4$ by geometric sum...

S2Q5: remember idea of "Markov on $n - X$"

S2Q6: Friendship: Jensen + $\mathbb{E}[\#\text{friends of random } U] = \mathbb{E}[d_U^2]$ by swapping sums in u, w st $(w, u) \in E$. strict inequality: 3 friends in a row.

S3Q1: for $\mathbb{P}(\max_i X_i < k) \leq \varepsilon$, for $X_i =$ length of longest sequence starting at i, only consider set of indices $i = jk+1$, so $\max X_i < k$ implies that $\max_{jk+1} X_i < k$ also (nothing longer than k, so nothing can stretch across two groups), now these are independent....

S3Q3: triples of vertices - take care to check whether independent for Chernoff or not

S2Q4: remember Chernoff bound has μ in both sides of inequality

S2Q5: median of X_i: Chernoff on indicators $X_i \geq$target val, aim for $\sum Y_i \geq n/2$

S4Q2: for stat dists, don't forget $\sum_i \pi_1 = 1$ might simplify!

S4Q3: cat + mouse: even-length walk len $\leq 2n$: prove by bipartiteness => odd cycle; expected time to eat: random walk on product $G \times G$, middle $w$of even

length walk is "closest" $(w, w) \in G \times G$ to (u, v), so $\mathbb{E}$[meet] $\mathbb{E}[T$ u to v$] \geq \mathbb{E}[T$ u to v along $P]$, but $\mathbb{E}[T] = \sum_{\text{paths}P} \mathbb{E}[T_p$

hitting time to a set = min hitting time to each element <= hitting time to a specific element

1 Intro

RP class: L if $\exists$ a poly time deterministic alg $A(x, r)$, for $r \in \{0,1\}^{p(|x|)}$ (poly p) st $\forall x$ if $x \in L$, $A(x, r)$ accepts for at least half of the r's, and if $x \notin L$ $A(x, r)$ rejects for all r's.

co-RP: above, but with reject/accept swapped

ZPP: $\mathrm{RP} \cap \mathrm{coRP}$ - zero error, will give correct answer, but in expected poly time

BPP: L in bounded-error probabilistic poly time: if $\exists$ a poly-time det $A(x, r)$ for $r \in \{0,1\}^{p(|x|)}$ st $\forall x$:

- if $x \in L$, $A(x, r)$ accepts for $\geq 3/4$ of the r's
- if $x \notin L$, $A(x, r)$ rejects for $\geq 3/4$ of the r's

note constant $3/4$ is arbitrary in $(1/2, 1)$.

and $\mathrm{RP} \subseteq \mathrm{BPP}$ by running the alg twice.

Examples

- string inequality
- min-size cut-set

Techniques

- equivalence of polynomials
- working mod a finite field $\mathbb{F}_p$

2 Linearity of expectation

Expectation is linear

PTAS: $\forall \varepsilon$ there is a poly time (in n) that finds a solution within $1 - \varepsilon$ of opt (for maximisation problems)

FPTAS: $\forall \varepsilon, n$ PTAS but poly in n and $1/\varepsilon$.

APX-hard means there is no PTAS

independence, pairwise indep

geometric dist

Harmonic numbers

Jensen's inequality: $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$ if f convex, X has finite expectation

[proof only for differentiable]

Examples

- max-cut: $C \subseteq V$ cut size is $\#$ of edges $C \to V \backslash C$
  - alg: put $v \in C$ with prob $1/2$
- max-3-sat
  - same naive as above
- coupon collector
  - total time = sum of X_i =time to ith new coupon (doesn't matter which one), which is geom
- Random quicksort

Techniques

- splitting $\mu := \mathbb{E}[X]$ into $\mathbb{E}[X|X > \mu](1-p) + \mathbb{E}[X|X \leq \mu]p$ where $p := \mathbb{P}[X \leq \mu]$ and bounding each term, to get a bound on p
- **probabilistic method**: if the $\mathbb{P}$ of simething is > 0 or $\mathbb{E} > 0$ then there must exist a...
- derandomisation with conditional expectation
  - want succinct way of calculating the conditional expectations, possibly way to simplify out terms that are same on both sides
- turning n RVs into 2^n pairwise indep RVs by $Y_S = \bigoplus_{i \in S} X_i$

3 Tail bounds

Variance, moments

Covariance

pairwise indep: $\text{Var}[\sum] = \sum \text{Var}$

Examples

- success of Max-Cut - know $\mathbb{E}[\#\text{cut}]$ - a direct bound on $\mathbb{P}[\#\text{cut} \geq m/4]$ useless with Markov, do $\#$not cut instead.
- Coupon collecting:
  - $\mathbb{E}$ time to all = sum of n (indep) Geom's
  - so Variance with Chebyshev is easy
- Coin flips: $\#$ of heads in n - v. standard Chernoff
- Set balancing: minimise $\max_i \sum_j A_{ij} b_j$ where $b_j \in \{-1, 1\}$- simple argument - pick b_ju.a.r, use simpler Chernoff bound
- Balls and bins: maximum load - one bin at a time, Chernoff on $\#$ in bin, then split $\mathbb{E}$[max load], using fact bounded by n.
- powering BPP: use Chernoff, on $X_i = i$th trial is correct (which has $p \geq 3/4$)
- BPP derandomisation existence: for any n, there is a deterministic version of A (i.e. an r) st A is correct on all x of length n. Use the probabilistic method (summing over all x) + powering with $k = 24(|x| + 1)$ to show there is an r st (powered)$A(x, r)$ is correct on all x of that length.

Tricks:

- colour a graph's edges [or vertices] u.a.r in k colours

Techniques

- Union bound: $\mathbb{P}[\bigcup_i A_i] \leq \sum_i \mathbb{P}[A_i]$ [equality if mutually exclusive]
- Markov's inequality $X \geq 0$: $\mathbb{P}[X \geq a] \leq \mathbb{E}[X]/a$
- Markov with exponential: any X $\mathbb{P}[X \geq x] = \mathbb{P}[\exp(tX) \geq \exp(tx)] \leq \mathbb{E}[\exp(tX)]/\exp(tx)$, optimise over t
- Chebyshev: any X: $3\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \text{Var}(X)/a^2$
- Chernoff: **0-1, indep** (may have different Bernoulli probs) X_i, $X = \sum_i X_i, \mu = \mathbb{E}[X] = \sum_i \mathbb{E}[X_i]$ (so μ is not the mean of an individual term):

$$\text{for } \delta > 0\text{: } \mathbb{P}[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

$$\text{for } \delta \in (0,1)\text{: } \mathbb{P}[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$

  - [proof by Markov + exp, $\mathbb{E}[\exp(\sum X_i)] = \prod \mathbb{E}[\exp(tX_i)]$ by indep, can write down $\mathbb{E}[\exp(tX_i)]$, rest is maths bounds

- Simpler Chernoff bounds: for **0-1 indep**, $\delta \in (0,1)$:

$$\begin{aligned}\mathbb{P}[X \geq (1+\delta)\mu] &\leq \exp(-\mu\delta^2/3)\\ \mathbb{P}[X \leq (1-\delta)\mu] &\leq \exp(-\mu\delta^2/2) \leq \exp(-\mu\delta^2/3)\\ \mathbb{P}[|X-\mu| \geq \delta\boldsymbol{\mu}] &\leq 2\exp(-\mu\delta^2/3)\end{aligned}$$

- Chernoff for $\{-1,1\}$ fair, indep RVs: $\mathbb{P}[\sum_i X_i \geq a] \leq \exp(-a^2/2n)$
- Useful maths bounds:

4 Markov chains (same as before)

Markov chain: set of states Ω, transition matrix P - rows add up to 1

Markov property: X_t depends only on X_{t-1}

Transition matrix (as time-homogeneous)

irreducible: $\exists t \geq 0$ st $P^t_{i,j} > 0$

hitting times: $H_{i,j} = \min\{t > 0 : X_t = j \mid X_0 = i\}$, $h_{i,j} := \mathbb{E}H_{i,j}$ is expected hitting time, $r^t_{ij} = \mathbb{P}(H_{i,j} = t)$ is prob we hit j at time t

recurrent: $\sum_{t\geq 1} r^t_{i,i} = 1$, otherwise transient

positive recurrent if $h_{i,i} < \infty$, otherwise null recurrent (equiv to $\mathbb{P}(H_{i,i} = \infty) = 0$)

finite irreducible Markov chains are positive recurrent, and further $h_{i,j} < \infty$ for all $i, j \in \Omega$

period of a state i: $\gcd\{m > 0 : P^m_{i,i} > 0\}$, aperiodic if period $= 1$

ergodic: aperiodic and positive recurrent

finite + irreducible + aperiodic => ergodic [proof: ε=smallest entry > 0, d=max distance (finite as irred.), then prob of visit at least ε^d, write out sum...]

stationary distribution π st $\pi = \pi P \iff \forall i \;\; \pi_j = \sum_i \pi_i P_{i,j}$ (π_j based on states i that can transition to j)

On a finite, irreducible aperiodic Markov chain, there is a unique stationary distribution π st $\pi_i = \lim_{t\to\infty} P^t_{j,i} = 1/h_{i,i}$ (where the limit exists and does not depend on j)

time-reversible wrt π if $\pi_i P_{i,j} = \pi_j P_{j,i}$, if this holds then π is a stationary distribution.

cover time of a graph: $\max_{v\in V} \mathbb{E}$[time to visit all (other) vertices $|X_0 = v$]

Examples

- Randomised 2-SAT
  - using that $\mathbb{E}$[hitting time] of random process which is **not** a Markov chain is $\geq$ to that for an actual Markov chain (reflecting,symmetric random walk on $0...n$)

- reflecting, symmetric random walk on $0...n$: hitting time to n is $h_j = n^2 - j^2$ - solve from $0 \to n$
- Randomised 3-SAT
  - similar random walk idea, but bounds not very good - random assignment is around $n/2$, chain has preference to 0
  - so restart every $3n$ times [i.e. geometric]
- Random walks on *finite, undirected, connected* graphs
  - $d(v) =$ degree of v
  - random walk is aperiodic $\iff G$ is not bipartite
  - if G is aperiodic=non-bipartite, then the unique stationary dist is [det bal]
$$\pi_v = \frac{d(v)}{2|E|} = \frac{1}{h_{u,u}} \text{ (by def)}$$
  - if G is connected, non-bipartite, u, v adjacent in G then $h_{u,v} < 2|E|$ [$h_{u,u} = \frac{1}{d(u)} \sum_{v \in N(u)} (1 + h_{v,u})$, so $h_{v,u} < \sum \cdots = 2|E|$
  - cover time of a non-bipartite connected graph with $|V| = n, |E| = m$ is $\leq 4nm$
    - spanning tree, do DFS on it - so get a cycle which takes all edges twice, so cover time $\leq$ time to cover spanning tree (1 of many routes) $\leq (2|V| - 2)2|E|$
- $s - t$ connectivity: Q: are s and t connected
  - Random ans: do a random walk from s for $4n^3$steps - if reach t, YES, else NO
  - correctness: no false positives, failure is $\mathbb{P}(H_{st} > 4n^3 | s, t \text{ connected})$. WLOG assume all components of G non-bipartite (if not, add a triangle to a lone edge), then use cover time $4nm \leq 2n^3$ + Markov bound for prob $\geq 1/2$

Techniques

- comparing random process to a Markov chain
- random restarts
  - outer loop:
    - pick a random assignment
    - inner loop: repeat, up to fix number of times, some random-walk style work - e.g. pick unsat clause, flip literal in it
  - idea:
    - if the random-walk aspect has a tendency to get worse over time, then we can reset it.
  - analysis

* bound probability q that inner loop succeeds
* use geometric dist with q to find $\mathbb{E}$ number of outer loop runs needed to succeed
* then use Markov with $p = 1/2$ to get bound for $\mathbb{P}[X \geq 2/q] \leq \frac{1/q}{2/q} = 1/2$

5 Monte Carlo (notes 4.3)

Monte carlo method: random draws, sum over them, apply Chernoff

Randomised approximation scheme (**RAS**) for f is rand alg A that takes instance I, error tolerance $\varepsilon \in (0,1)$ st $\forall I, \varepsilon$ we have

$$\mathbb{P}\left(|A(I,\varepsilon) - f(I)| \leq \varepsilon f(I)\right) \geq 3/4$$

anything in $(1/2, 1)$ is equivalent - i.e. an $\varepsilon - \delta$ approx [Chernoff on $X_i = |z_i - f(I)| \leq \varepsilon f(I)$, where we return the median $z_i =$ output of ith FPRAS run

A **FPRAS** is a RAS st that the running time of A is bounded by a polynomial in $|I|$and $1/\varepsilon$.

ε, δ-approximation:

- A takes input $I, \varepsilon \in (0,1), \delta \in (0,1)$ st $\forall I, \varepsilon, \delta$

$$\mathbb{P}\left(|A(I,\varepsilon) - f(I)| \leq \varepsilon f(I)\right) \geq 1 - \delta$$

- and running time of A is bounded by a poly in $|I|, 1/\varepsilon, \log(1/\delta)$

chernoff for ε, δ-approximation: $X_1, ..., X_m$ iid indicator variables, $\mu = \mathbb{E}[X_i]$ if $m \geq \frac{3\log(2/\delta)}{\varepsilon^2 \mu}$ then

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| \geq \varepsilon\mu\right) \leq \delta$$

$\#P$: given Σ is a finite alphabet,

- a counting problem $f : \Sigma^* \to \mathbb{Z}_{\geq 0}$ is in FP (functional poly) time if it can be computed in poly time
- f is in $\#P$ if $\exists p$ poly, ϕ, a poly-time checkable predicate st $\forall x \in \Sigma^*$

$$f(x) = |\{w \in \Sigma^* : |w| \leq p(x) \wedge \phi(x,w) = 1\}|$$

Examples

- DNF counting is FPRAS (using union-of-sets with $U=$ all truth assignemnts, $H_i=$ assignments that satisfy C_i)

Techniques

- Union of sets: t subsets H_i of $U, n := |U| < \infty$, and want to estimate $|H|, H := \bigcup_i H_i$.
  - **Requirements:**
    - we know $|H_i|$,
    - can sample unif from H_i,
    - can check if $x \in U$ is in H_i,
    - all in time poly in $t + \log |U|$
  - concoct $W = \{(i, x) : 1 \leq i \leq t, x \in H_i\}$, $H' = \{(i, x) \in W : (j, x) \notin W \text{ for } j < i\}$, $|H'| = |H|$, $|W| = \sum_i |H_i|$
  - can sample from W by picking i then $x \in H_i$
  - sample from H' by sampling from W and checking if in $H_j$for $j < i$
  - need $m = 3 \log(2/\delta) t/\varepsilon^2$ samples for an ε, δ-approximation
  - so is a FPRAS ($\delta = 1/4$) with large enough m, st m poly in n

6 More sampling

total variation distance between π_1, π_2 on finite S:

$$\|\pi_1 - \pi_2\|_{TV} = \frac{1}{2} \sum_{s \in S} |\pi_1(s) - \pi_2(s)|$$

$$= \max_{A \subseteq S} |\pi_1(A) - \pi_2(A)|$$

sampling problem: there is a distribution $\mathcal{D}(x)$ on $\Omega(x)$ for each $x \in \Sigma^*$

ε-approximate sampler: A that takes x, such that output dist sat $\|A(x) - \mathcal{D}(x)\|_{TV} \leq \varepsilon$

PAUS: Poly almost uniform sampler: an ε-approximate sampler if runtime is poly in $|x|, 1/\varepsilon$

FPAUS: ε-approx sampler, runtime poly in $|x|, \log 1/\varepsilon$

self-reproducible: working on set of graphs, can express $\Omega(G)$ in terms of $\Omega(G')$ for G' smaller than G

Examples

- FPRAS for counting indep sets, if have a PAUS for sampling them: by splitting into ratios $f(G_i)/f(G_{i-1})$ - then PAUS sample elements of $f(G_{i-1})$, and check [poly] if they are independent in G_i (one extra vertex)

7 MCMC and mixing

Metropolis algorithm: (Ω, E) connected, undirected graph, max degree Δ. π a distribtution on Ω, $\pi_x > 0$ for all $x \in \Omega$. fix $C > \Delta$, π is the unique stat dist of the chain with transition matrix P:

$$P_{x,y} = \begin{cases} \frac{1}{C} \min\left(1, \frac{\pi_y}{\pi_x}\right) & (x,y) \in E \\ 0 & x \neq y, (x,y) \notin E \\ 1 - \sum_{z:z\neq x} P_{x,z} & x = y \end{cases}$$

[proof by det bal + ergodic finite]

$d_{x,TV}(t) := \|P^t_{x,\cdot} - \pi\|_{TV}$ is non-increasing function of t [triangle equality],

$d_{TV}(t) := \max_{x\in\Omega} d_{x,TV}(t)$

Mixing time from state x: $\tau_x(\varepsilon) := \min\{t | d_{x,TV}(t) \leq \varepsilon\}$, for the whole chain: $\tau(\varepsilon) := \max_{x\in\Omega} \tau_x(\varepsilon)$

rapidly mixing: $\tau(\varepsilon)$ is poly in $\log 1/\varepsilon$ and the size of the problem

coupling: a Markov chain (X_t, Y_t) on $\Omega\times\Omega$ st X_t, Y_t are both (individually/marginally) copies of the original chain, and $X_t = Y_t \implies X_s = Y_s$ for all $s \geq t$

[2x completely indep chains would not satisfy this, as could/would diverge, but 2x indep until meeting would satisfy, but isn't useful]

Coupling lemma: coupling (X_t, Y_t) of a chain M_t with a uniq stat dist: if $t : [0,1] \to \mathbb{Z}_{\geq 0}$ is st $\forall x, y \in \Omega$

$$\mathbb{P}(X_{t(\varepsilon)} \neq Y_{t(\varepsilon)} | X_0 = x, Y_0 = y) \leq \varepsilon$$

then $\|P^{t(\varepsilon)}_{x,\cdot} - P^{t(\varepsilon)}_{y,\cdot}\|_{TV} \leq \varepsilon$, and the mixing time $\tau(\varepsilon) \leq t(\varepsilon)$

[$\mathbb{P}(M_{t(\varepsilon)} \in A | M_0 = x) = \mathbb{P}(X_{t(\varepsilon)} \in A | X_0 = x \wedge Y_0 = y)$, since Y_0no effect on X

$\geq \mathbb{P}(X_{t(\varepsilon)} = Y_{t(\varepsilon)} \wedge Y_{t(\varepsilon)} \in A | ...)$, union bound on $1-$, use symmetry of Y also a copy of M. choosing y from stat dist gives mixing time]

integral distance metric: a metric, but distances are integers

nb Markov's inequality says $\mathbb{P}(X_t \neq Y_t) = \mathbb{P}(d(X_t, Y_t) \geq 1) \leq \mathbb{E}[d(X_t, Y_t)]$

Coupling contraction lemma: $d : \Omega \times \Omega \to \mathbb{N}$ be an integral distance on Ω, M_t has a uniq stat dist, coupling X_t, Y_t. if $\exists \beta \in (0,1)$ $\forall x, y \in \Omega$

$$\mathbb{E}[d(X_1, Y_1) | X_0 = x, Y_0 = y] \leq \beta d(x, y)$$

then $\tau(\varepsilon) \leq \left\lceil \log \frac{D}{\varepsilon} \frac{1}{\log \frac{1}{\beta}} \right\rceil$, where $D := \max_{x,y\in\Omega} d(x,y)$. [v simple induction]

edge-weighted graph (H, d): connected graph H, with a distance $d(x,y)$ on every edge x, y in H

(H, d) is **minimal**: $\forall (x, y) \in E$ $d(x, y)$ is the length of the shortest path from $x \to y$ in H

can convert edge-weighted to minimal by removing edges that are not shortest path (still connected!)

define integral metric on a minimal edge-weighted graph by shortest dist.

path coupling: a coupling (X_t, Y_t) st $\forall (x, y) \in E\, \mathbb{E}[d(X_1, Y_1)|(X_0, Y_0) = (x, y)] \leq \beta d(x, y)$

path coupling extension lemma: M_t with finite state space Ω. (H, d)minimal edge-weighted, fix $\beta \in (0, 1)$. If (X_t, Y_t) is a path coupling, then there is a coupling $(\hat{X}_t, \hat{Y}_t)$ st $\forall (x, y) \in \Omega^2$ $\mathbb{E}[d(\hat{X}_1, \hat{Y}_1)|(\hat{X}_0, \hat{Y}_0) = (x, y)] \leq \beta d(x, y)$

[v. boring proof, basically define new coupling as probability of all paths between points]

Path coupling lemma: a path coupling (X_t, Y_t) of a chain M_t with a uniq stat dist, also (H, d): then the mixing time satisfies $\tau(\varepsilon) \leq \left\lceil \log \frac{D}{\varepsilon} \frac{1}{\log \frac{1}{\beta}} \right\rceil$, where $D := \max_{x,y \in \Omega} d(x, y)$

[proof: combine above]

Examples

- unif dist on indep sets on G ($\Omega = \mathcal{I}(G)$):
  - randomly choose vertex, try to move to "xor" sum of $X_i \oplus \{v\}$ if indep, else stay
  - irreducible -> move to empty and back, aperiodic because self-loops
- Hard-core Gibbs measure: use Metropolis, note can skmplify ratios π_y/π_x since you know they are adjacent
- card shuffling: M_t: pick a card u.a.r, move to top, stat dist is uniform mixing is "moving the same card to the top". mixing is coupon collector
- token ring: indep until meet, so difference is a random walk on 0....n
- binary trees: lazy random walk. Mixing: pick 1 of X, Y u.a.r, move that one u.a.r, when on same level, make the same direction choices - i.e. levels then states
  - so total time $\leq h_{r,l} + h_{l,r}$ where r=root, l = a leaf: 1 starting closer to root -> leaf for level, and back again for state
- colouring of graph by chain that picks v,c, w.p. 0.5 no change, else colour v with c: coupling: choose same v,c, use contraction lemma
- proper colourings (no touching same-coloured vertices) - chain is random v,c, change if allowed; use path coupling (extend Ω to all (incl improper) couplings, as chain will stay proper once proper); H edges is "differ on exactly 1 vertex", so distance = Hamming; coupling:

- choose v, c;
- if there is a neighbour w coloured differently by X_t, Y_t:
  - if $X_t(w) = c$ and $Y_t(w) = d$, set $X_{t+1}(v) = d, Y_{t+1}(v) = c$
  - if $X_t(w) = d$ and $Y_t(w) = c$, set $X_{t+1}(v) = c, Y_{t+1}(v) = d$
  - now v, w are different colours in both X_{t+1}, Y_{t+1}, so more proper, but mixing same or worse
- else update v, c in X_{t+1}, Y_{t+1}
- little bit of prob to work out how distance changes ($\beta = 1 - \frac{1}{nq}$)

8 Martingales

Martingale: sequence $(X_n)_{n\geq 1}$ st $\forall n \geq 0$ $\mathbb{E}[X_{n+1} \mid X_0, ..., X_n] = X_n$, and $\mathbb{E}[|X_n|] < \infty$

$(Z_n)_{n\geq 1}$ is a **martingale wrt a sequence** $(X_n)_{n\geq 1}$ if $\forall n \geq 0$ Z_n is a function of $X_0, ..., X_n$, $\mathbb{E}[|Z_n|] < \infty$, $\mathbb{E}[Z_{n+1} \mid X_0, ..., X_n] = Z_n$

$\mathbb{E}[X_n] = \mathbb{E}[X_0]$

supermartingale: $\leq$

submartingale: $\geq$

Doob martingale: sequence $X_0, ..., X_m$w/ bounded exp., Y dep on $X_0, ..., X_m$, then $Z_i := \mathbb{E}[Y|X_0, ..., X_i]$ is a martingale sequence

simple tower property: $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$ for any rv Y

full tower property: $\mathbb{E}[X|\mathcal{F}_1] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_2] \mid \mathcal{F}_1]$, where $\mathcal{F}_1 \subseteq \mathcal{F}_2$ - i.e. $\mathcal{F}_1$ has less information - $\sigma(X_1, ..., X_n) \subseteq \sigma(X_1, ..., X_{n+1})$

stopping time τ for $Z_0, ...$ if $\forall n$ $\tau = n$ depends only on $Z_0, ..., Z_n$

stopped martingale: $Z^\tau = (Z_{\tau\wedge n})_{n\geq 1} := Z_\tau$ if $\tau \leq n$ else Z_n

optional stopping theorem: (Z_n) a martingale wrt (X_n), τ a stopping time wrt (X_n), and 1 of:

1. τ is bounded
2. $\exists c$ $\forall n$ $|Z_{\tau\wedge n}| \leq c$, or
3. $\mathbb{E}[T] < \infty$ and $\exists c$ $\forall n$ $\mathbb{E}\left[|Z_{\tau\wedge n+1} - Z_{\tau\wedge n} \mid X_1, ..., X_n|\right] \leq c$

then $\mathbb{E}[Z_\tau] = \mathbb{E}[Z_0]$. [no proof]

Wald's equation: $(X_n)_{n\geq 1}$ non-negative, IID RVs, τ a stopping time for $(X_n)_{n\geq 1}$. If $\mathbb{E}[|\tau|], \mathbb{E}[|X_n|]$ are finite, then $\mathbb{E}[\sum_{i=1}^{\tau} X_i] = \mathbb{E}[\tau]\mathbb{E}[X]$. [OST cond #3 on $Z_t = \sum_{i=1}^{t}(X_i - \mathbb{E}[X])$, note $X_i \geq 0$!]

Azuma-Hoeffding: $(X_n)_{n\geq 0}$a martingale, $|X_i - X_{i-1}| \leq c_i$ then $\forall \lambda > 0$:

$$\mathbb{P}(X_n - X_0 \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}\right)$$
$$\mathbb{P}(X_n - X_0 \leq -\lambda) \leq \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}\right)$$

Proof:

- $\mathbb{P}\left(\exp(t(X_n - X_0) \geq \exp \lambda t\right) \leq \exp(-\lambda t)\mathbb{E}[\exp(t(X_n - X_0)]$ by Chernoff/Markov + exp
- $Y_i := X_i - X_{i-1}$, $\mathbb{E}[\exp(t(X_n - X_0)] = \exp[\prod \exp(tY_i)]$ by telescoping
- $\mathbb{E}\left[\prod e^{tY_i} | X_0, .., X_{n-1}\right] = \left(\prod_1^{n-1} e^{tY_i}\right) \mathbb{E}[e^{tY_n} | X_0, ..., X_{n-1}]$ by T.O.K.
- nasty mathsy bound on $\exp tY_i \leq \exp\left((tc_i)^2/2\right) + \text{const}Y_i$
- induct on n to bound full product with cond exp's, $t = \lambda / \sum c_i^2$ to minimise above

McDiarmid's inequality: if $\exists c$ st $|f(x_1, ..., x_i, ..., x_n) - f(x_1, ..., x_i', ..., x_n)| \leq c$ for all i then for indep RVs $X_1, ..., X_n$ and $\lambda > 0$

$$\mathbb{P}\left(|f(X_1, ..., X_n) - \mathbb{E}[f(X_1, ..., X_n)| \geq \lambda\right) \leq 2\exp(-2\lambda^2/(nc^2))$$

[no proof]

Examples

- Gambler's fortune Z_t with each game/step being 'fair', stopping when it reaches one of $[-b, +a]$:
  - $\mathbb{P}$[hits a before b]: use OST cond #2 on Z_T
  - $\mathbb{E}[T]$: use OST cond #3 on $Y_t := Z_t^2 - t$, plus $\mathbb{P}$[hits a before b]
  - concentration of winnings: if maximum win per bet is bounded (e.g. by $c_i = 10$), then Az-H gives high-probability bounds in $[-k, k]$, $k = O(\sqrt{n \log n})$
- balls and bins: X_i: bin i'th ball falls into, $Y =$ # empty at end, $Z_i := \mathbb{E}[Y \mid X_1, ..., X_i]$ is a *Doob* martingale
- v. simple Wald example - $\mathbb{E}$ sum of all rolls until the 1st 5 [use $\mathbb{E}[T] = 6$ by Geom, $\mathbb{E}[X_i] = 3.5$, none of which depends on 5!]
- Chromatic number: min # of colors $\chi(G)$ for proper colouring - use a martingale of $\mathbb{E}[\chi(G)|G_1, ..., G_i]$, for subgraphs $G_i = G \cap \{1, ..., i\}$, then Azuma-Hoeffding - change at most 1
- pattern matching of a pattern in a string $X_1....X_n$, each char selected u.a.r. - given pattern length k, each new character can be in at most k patterns, then Az-H
- # of empty bins: n balls into n bins, #empty also Az-H, as each ball can change # by at most 1

9 Lovász local lemma

dependency graph of a set of events $\{A_1, ..., A_n\}$ is a graph $G = (V, E)$ where $V = \{A_1, ..., A_n\}$ and A_i is mutually independent of $\{A_j : (A_i, A_j) \notin E\}$.

Lovász local lemma: if $\{A_1, ..., A_n\}$ is a set of "bad" events, $p < 1$ st:

1. $\forall i\ \mathbb{P}(A_i) \leq p < 1$,
2. the maximum degree of the dependency graph is $\leq d$,
3. $4dp \leq 1$ [$\implies p \leq 1/(4d) < 1/2$]

then $\mathbb{P}(\bigcap A_i^C) > 0$ - i.e. the probability of "all good" is > 0, so such an instance exists

So if ϕ is a k-CNF formula in which every clause shares variables with at most $2^{k-3} - 1$ other clauses, it is satisfiable.

Proof:

define $F_S = \bigwedge_{i \in S} \bar{A}_i$ for $S \subseteq \{1, ...n\}$.

We prove that $\forall S\ \mathbb{P}(F_S) > 0$ and $\forall i \in [n] \backslash S$, $\mathbb{P}[A_i \mid F_S] \leq 2p$.

Induct on $|S|$, base case is straightforward.

Induction: first property straightforward by conditioning; second property by partitioning S into neighbours of A_i, and the rest.