Intro

This lecture will review everything you learned in 6.042.

- Basic tools in probability
- Expectations
- High probability events
- Deviations from expectation

Coupon collecting.

- n coupon types. Get a random one each round. How long to get all coupons?
- general example of waiting for combinations of events to happen.
- expected case analysis:
  - after get k coupons, each sample has $1 - k/n$ chance for new coupon
  - so wait for $(k+1)^{st}$ coupon has geometric distribution.
  - expected value of geo dist w/param p is $1/p$
  - so get harmonic sum
  - what standard tools did we use? using **conditional expectation** to study on phase; used **linearity of expectation** to add
  - expected time for all coupons: $n \ln n + O(n)$.

Stable Marriage

Problem:

- complete preference lists
- stable if no two unmarried (to each other) people prefer each other.
- med school
- always exists.

Proof by proposal algorithm:

- rank men arbitrarily
- lowest unmarried man proposes in order of preference
- woman accepts if unmarried or prefers new proposal to current mate.

Time Analysis:

- woman's state only improves with time
- only n improvements per woman
- while unattached man, proposals continue
- (some woman available, since every woman he proposed to is married now)
- must eventually all be attached

Stability Analysis

- suppose X-y are dissatisfied with pairing X-x, Y-y.
- X proposed to y first
- y prefers current Y to X.

Average case analysis

- nonstandard for our course
- random preference lists
- how many proposals?
- **principle of deferred decisions**
  - used intuitively already
  - random choices all made in advance
  - random choices made algorithm needs them.
- used while discussing autopartition, quicksort
- Proposal algorithm:
  - each proposal is random among unchosen women
  - still hard
  - Each proposal among all women
  - **stochastic domination**: X s.d. Y when $\Pr[X > z] \geq \Pr[Y > z]$ for all z.
  - done when all women get a proposal.
  - at each step $1/n$ chance women gts proposal
  - This is just coupon collection: $O(n \log n)$

Deviations from Expectation

Sometimes expectation isn't enough. Want to study *deviations*—**probability** and **magnitude** of deviation from expectation.
Example: balls in bins:

- n balls in n bins
- Expected balls per bin: 1 (not very interesting)
- What is max balls we expect to see in a bin?
- Start by bounding probability of many balls

$$\begin{aligned}
\Pr[k \text{ balls in bin } 1] &= \binom{n}{k}(1/n)^k(1-1/n)^{n-k} \\
&\leq \binom{n}{k}(1/n)^k \\
&\leq \left(\frac{ne}{k}\right)^k (1/n)^k \\
&= \left(\frac{e}{k}\right)^k
\end{aligned}$$

- So prob **at least** k balls is $\sum_{j\geq k}(e/j)^j = O((e/k)^k)$ (geometric series)
- $\leq 1/n^2$ if $k > (e\ln n)/\ln\ln n$
- What is probability **any** bin is over k? $1/n$ **union bound**.
- Now can bound expected max:
  - With probability $1-1/n$, max is $O(\ln n/\ln\ln n)$.
  - With probability $1/n$, max is bigger, but at most n
  - So, expected max $O(\ln n/\ln\ln n)$
- Typical approach: small expectation as small "common case" plus large "rare case"

Example: coupon collection/stable marriage.

- Probability didn't get k^{th} coupon after r rounds is $(1-1/n)^r \leq e^{-r/n}$
- which is $n^{-\beta}$ for $r = \beta n \ln n$
- so probability didn't get *some* coupon is at msot $n \cdot n^{-\beta} = n^{1-\beta}$ (using **union bound**)
- we say "time is $O(n\ln n)$ **with high probability**" because we can make probability $n^{-\beta}$ for **any** desired β by changing constant that doesn't affect assymptotic claim.

- sometime say "with high probability" when prove it for **some** $\beta > 1$ even if didn't prove it for all.
- Saying "almost never above $O(n \ln n)$" is a much stronger statement than saying "$O(n \ln n)$ on average."

Tail Bounds—Markov Inequality

At other times, don't want to get down and dirty with problem. So have developed set of bounding techniques that are basically problem independent.

- few assumptions, so applicable almost anywhere
- but for same reason, don't give as tight bounds
- the more you require of problem, the tighter bounds you can prove.

Markov inequality.

- $\Pr[Y \geq t] \leq E[Y]/t$
- $\Pr[Y \geq tE[Y]] \leq 1/t$.
- Only requires an expectation! So very widely applicable.

Application: $ZPP = RP \cap coRP$.

- If $RP \cap coRP$
  - just run both
  - if neither affirms, run again
  - Each iter has probability 1/2 to affirm
  - So expected iterations 2:
  - So ZPP.
- If ZPP
  - suppose expected time $T(n)$
  - Run for time $2T(n)$, then stop and give default answer
  - Probability of default answer at most 1/2 (Markov)
  - So, RP.
  - If flip default answer, $coRP$

On flip side, not very strong: balls in bins $\Pr[> \ln n] \leq 1/\ln n$.
Can make much stronger by generalizing: $\Pr[h(Y) > t] \leq E[h(Y)]/t$ for **any positive** h.