Chebyshev.

- Remind variance, standard deviation. \( \sigma_X^2 = E[(X - \mu_X)^2] \)
- \( E[XY] = E[X]E[Y] \) if independent
- variance of independent variables: sum of variances
- \( \Pr[|X - \mu| \geq t\sigma] = \Pr[(X - \mu)^2 \geq t^2\sigma^2] \leq 1/t^2 \)
- binomial distribution. variance \( np(1-p) \). stdev \( \sqrt{n} \).
- requires (only) a mean and variance. less applicable but more powerful than markov
- Balls in bins: err \( /1 \ln^2 n \).

Two point sampling.

- pseudorandom generators. Motivation. Idea of randomness as (complexity theoretic) resource like space or time.
- pairwise independent vars.
- generating over \( Z_p \).
- pairwise sufficient for chebyshev.
- Suppose \( RP \) algorithm using \( n \) bits.
- What do with \( 2n \) bits?
- two direct draws: error prob. 1/4.
- pseudorandom generators gives error prob. 1/t for \( t \) trials.
- \( \mu = t/2. \sigma = \sqrt{t}/2. \)
- error if no cert, i.e. \( Y - E[Y] \geq t/2, \) prob. 1/t.

Chernoff Bound

Intro

- Markov: \( \Pr[f(X) > z] < E[f(X)]/z. \)
- Chebyshev used \( X^2 \) in \( f \)
- other functions yield other bounds
- Chernoff most popular

Theorem:
• Let $X_i$ poisson (ie independent 0/1) trials, $E[\sum X_i] = \mu$

$$Pr[X > (1 + \delta)\mu] < \left[\frac{e^{\delta}}{(1 + \delta)(1+\delta)}\right]^\mu.$$  

• note independent of $n$, exponential in $\mu$.

Proof.

• For any $t > 0$,

$$Pr[X > (1 + \delta)\mu] = Pr[\exp(tX) > \exp(t(1 + \delta)\mu)]$$

$$< \frac{E[\exp(tX)]}{\exp(t(1 + \delta)\mu)}$$

• Use independence.

$$E[\exp(tX)] = \prod E[\exp(tX_i)]$$

$$E[\exp(tX_i)] = p_i e^t + (1 - p_i)$$

$$= 1 + p_i(e^t - 1)$$

$$\leq \exp(p_i(e^t - 1))$$

$$\prod \exp(p_i(e^t - 1)) = \exp(\mu(e^t - 1))$$

• So overall bound is

$$\frac{\exp((e^t - 1)\mu)}{\exp(t(1 + \delta)\mu)}$$

True for any $t$. Plug in $t = \ln(1 + \delta)$.

• This in turn less than $e^{-\mu\delta^2/4}$ for $\delta < 2e - 1$. (Less than $2^{-(1+\delta)\mu}$ for larger $\delta$).

• By same argument on $\exp(-tX)$,

$$Pr[X < (1 - \delta)\mu] < \left[\frac{e^{-\delta}}{(1 - \delta)(1-\delta)}\right]^\mu$$

bound $e^{-\delta^2/2}$.

Summary, Probability of deviation by relative error $\delta < 1$ is at most $e^{-\delta^2\mu/3}$ in each direction. Large $\delta$ gives $2^{-(1+\delta)\mu}$

• Trails off when $\delta \approx \sqrt{\mu}$, meaning absolute error is “expected” to be $\sqrt{\mu}$

• (note variance is less than $\mu$. Compare Chebyshev).

• If $\mu = \Omega(\log n)$, probability of constant deviation is $O(1/n)$, Useful if polynomial number of events.
**Remark:** bound applies to any vars distributed in range $[0, 1]$.

Basic applications:

- $cn \log n$ balls in $c$ bins. max matches average (unlike $n$ balls in $n$ bins).
- Set balancing (book p. 73). minimize max bias. get $4\sqrt{n \ln n}$.

Zillions of Chernoff applications; will see next time.

- Summary, Probability of deviation by relative error $\delta < 1$ is at most $e^{-\delta^2 \mu/3}$ in each direction. Large $\delta$ gives $2^{-(1+\delta)\mu}$
  - Trails off when $\delta \approx \sqrt{\mu}$, meaning absolute error is “expected” to be $\sqrt{\mu}$
  - (note variance is less than $\mu$. Compare Chebyshev).
  - If $\mu = \Omega(\log n)$, probability of constant deviation is $O(1/n)$, Useful if polynomial number of events.

- Basic applications:
  - Set balancing. minimize max bias. $4\sqrt{n \ln n}$.
  - $cn \log n$ balls in $c$ bins. max matches average.