Chernoff inequality (or bound).

Thm (Chernoff 1950s): Let X be a binary rnd. variable with $P(X=1) =: p$. Let $X_1, \dots X_n$ be identical to X & mutually independent.
(called i.i.d.)

Consider the sum $S := \sum_{i=1}^{n} X_i$ & $\delta \in (0,1)$.

Then,

$$P(S < (1-\delta)\cdot E[S]) < \left(e^{-E[X]\cdot \delta^2/2}\right)^n.$$

↳ This is the strongest ineq. till now, as the decay is exponential in $n :=$ #repetitions of X.

Pf: • Idea: Reduce it to Markov's somehow?

• Let $u := E[S] = E[\Sigma X_i] = n \cdot E[X]$.

• $P(S < (1-\delta)\cdot u) = P(e^{-tS} > e^{-t(1-\delta)u})$

where $t > 0$ is a parameter & $e :=$ base of natural log.

• Markov's $\Rightarrow P(e^{-tS} > e^{-t(1-\delta)u}) < E[e^{-tS}] / e^{-t(1-\delta)u}$

▷ $E[e^{-tS}] = E[\prod e^{-tX_i}] = \prod E[e^{-tX_i}] = E[e^{-tX}]^n$

$= (p \cdot e^{-t} + (1-p)\cdot 1)^n = (1 - p\cdot(1-e^{-t}))^n$

$\leq e^{-np\cdot(1-e^{-t})}$ [$\because 1-\varepsilon \leq e^{-\varepsilon}$.]

$$\Rightarrow P(S < (1-\delta)\cdot u) < e^{-u(1-e^{-t})+ut(1-\delta)}$$
$$= \left(e^{t(1-\delta)+e^{-t}-1}\right)^u .$$

▷ $t(1-\delta)+e^{-t}-1$ is minimized when the derivative $1-\delta-e^{-t}=0$. [$\Rightarrow t = \log 1/(1-\delta) > 0$.]

- Substituting:

▷ $P(S<(1-\delta)u) < \left(e^{-\delta}/(1-\delta)^{1-\delta}\right)^u \leq \left(e^{-\delta^2/2}\right)^u$.

[$\because (1-\delta)\log(1-\delta) = (1-\delta)\cdot\left(-\delta-\frac{\delta^2}{2}-\cdots\right) = -\delta+\frac{\delta^2}{2}+\cdots$.] □

Corollary: $P(S > (1+\delta)\cdot u) < \left(e^{\delta}/(1+\delta)^{1+\delta}\right)^u$.

Pf: Similar; work with $P(e^{tS} > e^{t(1+\delta)u})$. □

↳ Bottomline: $P(S < (1-\delta).E[S] \text{ OR } S > (1+\delta).E[S])$
$< (\text{exponentially small in } n)$.

↳ This makes Chernoff bound the most popular bound in CS applications!

- This wasn't the case in the weak law of large numbers. But, there we needed only 2-wise independence (& not mutual independence!).

k-wise Independence

- **Defn:** • Let $\{X_i \mid i \in [n]\}$ be a family of rnd. variables. Call it k-wise independent if

$\forall x_i \in \mathbb{R}$, $\forall J \subseteq [n]$ with $|J| \leq k$:

$$P\left(\bigcap_{j \in J} [X_j = x_j]\right) = \prod_{j \in J} P(X_j = x_j).$$

- If $k=2$, call them pairwise independent.
- If $k=n$, " " mutually " .

▷ k-wise indep $\Rightarrow$ ℓ-wise indep., $\forall \ell \leq k$.

Qn: k-wise indep. $\Rightarrow$ (k+1)-wise indep.?

Eg. Let X, Y be two indep. rnd. variables.
Then, $\{X, Y, X+Y, X-Y\}$ is a family that is:

- pairwise (or 2-wise) indep.
- not 3-wise indep.
- not mutually indep

- To understand the asymptotics of Chernoff bound, consider, eg. Toss a coin n times.
Let $S := \#H$'s. $E[S] = n/2$.

$$P\left(S \le \frac{n}{2} - \underbrace{\sqrt{n \log n}}_{\delta := 2\sqrt{\frac{\log n}{n}}}\right) < e^{-\frac{n}{2} \cdot \frac{\delta^2}{2}} = e^{-\log n} = 1/n.$$

linear decay

- But, for $\delta = 1/2$:

$$P(S < \tfrac{n}{2} \cdot \tfrac{1}{2}) < e^{-\frac{n}{2} \cdot (1/2)^2/2} = e^{-n/16}$$

↖ exponential decay

Application to Probability Boosting

- Suppose A is an algorithm that, on any input $x \in \{0,1\}^n$, gives the wrong answer (Yes/No) with probability $\leq 1/3 =: p$.

Qn: Could you make the error arbitrarily small, say $< 1/2^n$?

Design a new algorithm A_m:

- Repeat A m times independently. (say m is odd)
- Output the majority vote.

Analyse:

- Let $S := \#(\text{correct answers})$. ▷ $E[S] = 2m/3$.

▷ $P(A_m \text{ errs on } x) = P(S < m/2) = P(S < \frac{3}{4} \cdot E[S])$ ← $\delta = 1/4$

$\leq e^{-(2m/3)\cdot \delta^2/2} = e^{-m/48}$.

⇒ Repeating A $m := 48n$ times, gives a decay $= e^{-n}$.

↳ Is very useful in randomized algorithms in practice!

- We saw a lot of theory on concentration ineqs. Let's take a break with an interesting CS example:

- In CS, multi-server computation has millions of clients trying to access the servers.

Qn: What's the (expected) load on a server? Could it also be millions?

Probability Tool (Union & Factorial estimates)

- Let $\mathcal{E}$ be an event in prob. space (Ω, P). To estimate $P(\mathcal{E})$, express $\mathcal{E}$ as a union of easier events $\mathcal{E}_i$, $i \in [n]$; they may overlap.

▷ $P(\mathcal{Q}) \le \sum_{i\in[n]} P(\mathcal{Q}_i)$ [equality if $\mathcal{Q}_i$'s disjoint]

$\le n \cdot \max_{i\in[n]} P(\mathcal{Q}_i)$.

Theorem: Let n clients randomly access n servers.
Whp, max. load on the servers $\le 6\lg n/\lg\lg n =: \ell$.

Pf:

- Define, $\mathcal{Q}$: some server has $>\ell$ clients.
  $\mathcal{Q}_j$: j-th " " " ".

  ▷ $\mathcal{Q} = \bigcup_{j\in[n]} \mathcal{Q}_j$. $\Rightarrow P(\mathcal{Q}) \le n \cdot P(\mathcal{Q}_1)$.

- $P(\mathcal{Q}_1) = \sum_{i>\ell} P(\text{server-1 has } i \text{ clients})$

exp. smaller than #clients!

$$= \sum_{\ell < i \le n} \binom{n}{i} \cdot \left(\frac{1}{n}\right)^i \cdot \left(1-\frac{1}{n}\right)^{n-i} \le \sum_{\ell < i \le n} \frac{1 \cdot \left(1-\frac{1}{n}\right) \cdots \left(1-\frac{i-1}{n}\right)}{i!} \cdot 1$$

$$< n/\ell!$$

$$\Rightarrow P(\ell) < n^2/\ell!.$$

▷ [Stirling's estimate] $\ell! \approx (\ell/e)^\ell \cdot \sqrt{2\pi\ell}$. (Why?)

$$\Rightarrow \ell! \ge \left(\ell \lg n / \lg\lg n\right)^\ell \cdot \sqrt{2\pi\ell} = 2^{\ell(\lg\lg n - \lg\lg\lg n)} \cdot \sqrt{2\pi\ell}$$

$$> 2^{5\lg n} \cdot \sqrt{2\pi\ell} > n^5.$$

$\Rightarrow P(\ell) < n^2/\ell! < 1/n^3$. $\Rightarrow$ Load on each server is almost always $\le 6\lg n/\lg\lg n$ ← (exp. smaller than n) □