Sums of Random Variables (and Other "Concentrated" functions).

If $X_1, X_2 \dots X_n$ are iid then $Z_n = \frac{\sum_i X_i - n\mu}{n\sigma} \xrightarrow[n\to\infty]{} N(0,1) \sim \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

(Converges in distrib) means that $\Pr[Z_n \geq u] \to \Pr[N \geq u]$ $\forall u \in \mathbb{R}$.)

Can get a quantitative version of this [Berry Esseen].

Today get careful quantitative bounds that are useful for algorithm design

---

Basic Inequalities: [Markov] Any nonnegative r.v. X satisfies

$$\Pr[X \geq \lambda] \leq \frac{E[X]}{\lambda}.$$

[Chebyshev]: X has mean μ, variance $\sigma^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2$.

then $\Pr[|X-\mu| \geq \lambda] \leq \frac{\sigma^2}{\lambda^2}$.

If X was itself the sum of $X_1 + X_2 + \dots + X_n$, which were iid (say) ~~with mean~~ ~~and var~~ $Ber(1/2)$.

then $\mu = \underbrace{n \cdot \tfrac{1}{2}}_{\text{lin. of exp.}}$. $\sigma^2 = \underbrace{\sum_i \sigma_i^2}_{\text{indep}} = n/4$.

$$\Rightarrow \Pr[X \geq n/2 + t\sqrt{n}/2] \leq \frac{n}{n + t\sqrt{n}} \text{ by Markov. } \approx \frac{1}{1+(t/\sqrt{n})}.$$

$$\leq \frac{1}{t^2} \text{ by Chebyshev}$$

Note that Chebyshev just requires pairwise independence between r.v.s.

Similar to the analysis of Chebyshev, what about only 4-wise independence?

$$\Pr[|X-\mu| \geq \lambda] = \Pr[(X-\mu)^4 \geq \lambda^4] \leq \frac{E[(X-\mu)^4]}{\lambda^4}.$$

expectation of this

$$= \sum_i (X_i-\mu_i)^4 + c\sum_{i\neq j} (X_i-\mu_i)^3 (X_j-\mu_j) \dots + c' \sum_i (X_i-\mu_i)^2 (X_j-\mu_j)^2 + \dots$$

$$= \sum E[(X_i-\mu_i)^4] + 3n(n-1) \sum_{i\neq j} E[(X_i-\mu_i)^2] E[(X_j-\mu_j)^2]$$

$\Rightarrow$ for the Bin$(n, 1/2)$ case this will give us $\lesssim \frac{O(1)}{t^4}$.

To use full independence, we use the chernoff bound. Proved by Bernstein in 1920s
[and even the version proved by Chernoff was given to him by Herman Rubin]. (52)

Here's the version I remember:—

<u>Thm:</u> Let $X_1, X_2 \ldots X_n$ be indep random variables in $[0,1]$, Let $X = \sum X_i$.
Let $\mu := EX$. ~~and~~ then

$$\Pr[X \geq \mu + \lambda] \leq \exp\left(-\frac{\lambda^2}{2\mu+\lambda}\right)$$

$$\Pr[X \leq \mu - \lambda] \leq \exp\left(-\frac{\lambda^2}{3\mu}\right).$$

[There are many better versions, but this one suffices most of the time. We'll talk about improvements soon].

<u>Application</u>: if $X = \mathrm{Bin}(n, 1/2)$ then $\mu = n/2$. this says that

$$\Pr[X \geq n/2 + t\sqrt{n}/2] \leq \exp\left(-\frac{n/4 \cdot t^2}{n + t\sqrt{n}/2}\right) \quad \text{say } t \ll \sqrt{n}.$$

$$\leq \exp\left(-\frac{n/4\, t^2}{2n}\right) \leq \exp(-t^2/8).$$

$$\Rightarrow \leq 1/n^{10} \text{ if } t = \sqrt{80 \ln n}.$$

<u>Application</u>: if n balls into n bins, then known: $E[\text{max load}] = \Theta\left(\frac{\log n}{\log\log n}\right)$.

<u>This says</u>: Load on any bin i = Bin$(n, 1/n)$. $= E[X] = \mu := 1$.

$$\Rightarrow \Pr[X \geq 1+t] \leq \exp\left(-\frac{t^2}{2+t}\right) \leq \exp\left(-\frac{t}{3}\right) \quad t \geq 1$$

$$\leq \frac{1}{\mathrm{poly}(n)} \text{ if } t \approx \log n$$

$\Rightarrow$ union bound over all bins,

$$\Pr[\exists \text{ bin} \geq \Omega(\log n) \text{ balls}] \leq \frac{n}{\mathrm{poly}(n)} \ll 1$$

$\Rightarrow$ says that
max load $\leq O(\log n)$
with high probability.

[N.b. a better concentration bound can give you $\frac{\log n}{\log\log n}$.]

<u>Pf</u>: ~~Pr[X~~ Say want $\Pr[X \geq \mu + \lambda] = \Pr[X \geq m]$.
def this to be m. (short).

Also assume: $X_i \in \{0,1\}$. Will extend soon.

$\Pr[X_i = 1] = p_i$
$\Rightarrow \mu = \sum p_i$

$\Pr[X \geq m] = \Pr[e^{tX} \geq e^{tm}]$ for $t > 0$.

$$\leq \frac{E[e^{tX}]}{e^{tm}} \underset{\text{indep}}{=} e^{-tm} \cdot \prod_{i=1}^{n} E[e^{tX_i}]$$

$$= e^{-tm} \prod_{i=1}^{n} (1 - p_i + p_i e^t).$$

AM-GM.

$$= e^{-tm} \left(1 + \mu/n \, (e^t - 1)\right)^n$$

$1 + x \leq e^x$

$$\leq \exp\{-tm + \mu(e^t - 1)\}.$$

How to choose $t \in \mathbb{R}_+$ to minimize this?
Differentiating wrt t, we get that $e^{(\ )} \cdot [-m + \mu e^t] = 0 \Rightarrow e^t = m/\mu$.

$$\boxed{\leq \exp\{-m \ln(m/\mu) + (m - \mu)\}}$$

plug in $\ln(m/\mu) = \ln(1 + \lambda/\mu)$
$$\geq \frac{(\lambda/\mu)}{1 + (\lambda/2\mu)}$$

$\ln(1+x) \geq \frac{x}{1 + x/2}$

$$\leq \exp\left\{-(\lambda + \mu) \cdot \left(\frac{2\lambda}{2\mu + \lambda}\right) + \lambda\right\}$$

$$= \exp\left\{-\frac{\lambda^2}{2\mu + \lambda}\right\}.$$

∎

The proof for the "lower tail" is very similar

<u>Fact</u>: sometimes the bound in the box is "much" better.

$$\Pr[X \geq \mu + \lambda] \leq \left(\frac{e^\lambda}{(1 + \lambda/\mu)^{(\lambda + \mu)}}\right) = \left(\frac{e^\beta}{(1+\beta)^{(1+\beta)}}\right)^\mu \text{ if } \lambda = \mu\beta.$$

$\Rightarrow$ if $Bin(n, 1/n)$ then set $\beta = \lambda = c\frac{\log n}{\log\log n}$ we get $(1+\beta)^{1+\beta} \leq \frac{1}{poly(n)} \cdot e^{\beta}$.

$\Rightarrow Pr[X \geq 1+\beta] \leq \frac{1}{poly(n)}$. which is the right answer

---

Not discrete?
What about $X_i \in [0,1]$ instead of $X_i \in \{0,1\}$.

Use that since $EX_i = p_i$ then $E[f(X_i)]$ is maximized when you put p_i at 1 and $1-p_i$ at 0. → convex fn

---

Suppose variables are not independent?

- As long as X_i's are negatively correlated (some of the vars being "high" makes it more likely for others to be "low").

  ⇒ these bounds can hold.

- formally: $E[f(X_i : i \in A) g(X_j : j \in B)] \leq E[f(X_i : i \in A)] E[g(X_j : j \in B)]$.
  for all disjoint A, B, and f, g are monotone increasing functions

[Eg: if a bin i is occupied then bin $j \neq i$ is more likely to be empty.]
so if $X_i = \mathbb{1}(i \text{ is occupied})$ then X_i are negatively correlated [Ex.].

⇒ can argue that $\sum X_i$ = # of occupied bins is a sum of neg. correlated r.v.s.

Thm: can then show that $E[\prod_i e^{tX_i}] \leq \prod_i E[e^{tX_i}]$ ⇒ rest of proof goes thru.

⇒ 'Chernoff' bounds hold for these cases too!

[Bernstein bounds]: $X_i \in [0,1]$ indep, $\sigma^2 = \sum_i \sigma_i^2 = Var(X)$

$Pr[X \geq \mu + \lambda] \leq \exp\left(-\lambda^2 / (2\sigma^2 + \lambda)\right)$
$Pr[\quad \mu - \lambda] \leq$

BTW: what gives with this $\exp\left(-\frac{\lambda^2}{2\mu+\lambda}\right)$ why not $\exp\left(-\frac{\lambda^2}{2\mu}\right)$. etc.

Note: scaling can give no $[0,B]$ bounds.

Note: translation takes care of ~~pos~~ $[-B,B]$ bounds.

Not true! Look at n balls n bins case. Right side bound gives load $\sqrt{\lg n}$, false !!

Beyond Sums: what about $f(X_1, X_2 \ldots X_n)$? as long as f is "well behaved".

(1) want $X_1, X_2 \ldots X_n$ to be independent $\vec{X}$ ~~is~~ product space.

(2) want f not to depend too much on a single coordinate.

~~Def~~ [Lipschitz functions]: f is c_i-Lipschitz along coordinate i if

$$|f(x) - f(x(i \text{ ~~flipped~~ changed to something else}))| \le c_i$$

Thm: Let f be c_i-Lipschitz $\forall i$, X be product space. then

$$\Pr[f \ge Ef + \lambda] \le \exp\left(-\frac{\lambda^2}{2\sum c_i^2}\right).$$

Very weak in the case of n balls and n bins. f is the max load in the system. [Essentially depends on the # of variables, as opposed to ~~Bernstein~~-type, which are "dimension independent".]

Assume a little more: get

Def: f is 1-self bounding if ~~① f is c-Lipschitz along~~ each coordinate ← implied (by below)

~~②~~ $\exists$ functions f_i such that

(a) $0 \le f(X) - f_i(X_{-i}) \le 1 \qquad \forall X \in \Omega$

(b) $\sum_{i=1}^{n} (f(X) - f_i(X_{-i})) \le f(X) \qquad \forall X \in \Omega.$

Thm: if f is self-bounding then $\Pr[f - Ef \ge \lambda] \le \exp\left(-\frac{\lambda^2}{2Ef + \lambda}\right)$

~~$\Pr[Ef - Ef \le \lambda] \ge \exp(-\lambda^2$~~

[Good example?].

Matrix Chernoff: X_k are independent, symmetric matrices of dimension d.

Moreover: $X_k \succeq 0$ and $X_k \preceq I$. $\leftarrow$ all evs are in $[0,1]$.

Let $\mu_{min} = \lambda_{min}(\sum_i E(X_k))$
$\mu_{max} = \lambda_{max}(\sum E(X_k))$

then $Pr[\lambda_{max}(\sum_i X_k) \geq \mu_{max} + \gamma] \leq d \cdot \exp\left(-\frac{\gamma^2}{2\mu_{max}+\gamma}\right)$.

N.b: if matrices have only random values on diagonals, then this is saying that the sum of every diagonal entry ~~remains~~ ~~small~~ ~~around the~~ is never much more than the max expected diagonal. Obtainable by a union bound. Here we get this for all psd matrices.

Also exist Bernstein, Azuma, etc.

Other things to mention: –

- Sps X is a random point in $\{0,1\}^n$ $\quad$ $Bin(n, \frac{1}{2})$. $\quad$ $\mu(x) = \frac{1}{2^n}$. Say total vol = 1.

  then $\cdot$ $Pr[|X| \geq \frac{n}{2} + t\sqrt{n}]$ = volume of set outside $(\frac{n}{2} + t\sqrt{n})$

  The set $H_{\frac{n}{2}} = \{x \mid w(x) \leq \frac{1}{2}\}$. is half the cube, volume is $\frac{1}{2}$.

  but H_ε is almost 1.

  Most measure concentrated around the "equator".

- What about moment bounds? $\quad Pr[X \geq t] \leq \min_{k \geq 0} \frac{E[X^k]}{t^k}$.

  Phillips & Nelson show that: $\min_{k \geq 0} \frac{E[X^k]}{t^k} \leq \inf_{t > 0} \frac{E[e^{tX}]}{e^{t\lambda}}$

  for any non-negative r.v. $\quad$ moment $\quad$ Chernoff