Exponential tail bounds

Goal: Obtain bounds on $\Pr[X \geq \mu(1+\delta)]$

Chebyshev inequality gives a bound that decays as $\frac{\sigma^2}{\mu^2\delta^2}$

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n} X_i \geq \mu(1+\delta)\right] \leq \frac{\sigma^2}{n\mu^2\delta^2}$$

Can we obtain a better bound?

Chernoff bound

Let X be a rv with $\mathbb{E}X = \mu < \infty$. Then,

$$\Pr[X \geq t] \leq \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda t}} \quad \text{for all } \lambda > 0.$$

$$\Pr[X \leq t] \leq \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda t}} \quad \text{for all } \lambda < 0$$

Proof: for all $\lambda > 0$, $\Pr[X \geq t] = \Pr[\lambda X \geq \lambda t]$

$$= \Pr[e^{\lambda X} \geq e^{\lambda t}]$$

$$\overset{\text{Markov}}{\leq} \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda t}}$$

Since this is true for all $\lambda > 0$,

$$\Pr[X > t] \leq \inf_{\lambda > 0} \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda t}}$$

Set $\psi_X(\lambda) = \log \mathbb{E} e^{\lambda X}$

$$\therefore \Pr[X > t] \leq \inf_{\lambda \geq 0} e^{\psi_X(\lambda) - \lambda t}$$

$$= e^{\inf_{\lambda \geq 0} (\psi_X(\lambda) - \lambda t)}$$

$$= e^{-\psi_X^*(t)}$$

where

$$\psi_X^*(t) = -\inf_{\lambda \geq 0} \left[\psi_X(\lambda) - \lambda t\right]$$

$$= \sup_{\lambda \geq 0} \left[\lambda t - \psi_X(\lambda)\right] \rightarrow \text{called the Cramer transform of } \psi_X$$

Now, let $Y_n = \sum_{i=1}^{n} X_i$, where $X_1 \dots X_n$ are iid

$$\mathbb{E} e^{\lambda Y_n} = \mathbb{E}\, e^{\lambda \sum_{i=1}^{n} X_i} = \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right]$$

$$= \prod_{i=1}^{n} \mathbb{E} e^{\lambda X_i} = \left(\mathbb{E}\, e^{\lambda X}\right)^n$$

$$= e^{n \psi_X(\lambda)}$$

$$\psi_{Y_n}(\lambda) = n\, \psi_X(\lambda)$$

$$\psi^*_{Y_n}(t) = \sup_{\lambda \geq 0} \left[\lambda t - n\, \psi_X(\lambda)\right]$$

$$= n \sup_{\lambda \geq 0} \left[\lambda \frac{t}{n} - \psi_X(\lambda)\right]$$

$$= n\, \psi^*_X(t/n)$$

$\therefore$ For iid $X_1 X_2 \cdots X_n$,

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n} X_i \geq t\right] = \Pr\left[\sum_{i=1}^{n} X_i \geq nt\right]$$

$$\leq e^{-n\,\psi_X^*(nt/n)}$$

$$= e^{-n\,\psi_X^*(t)}$$

$\Rightarrow$ If $\psi_X^*(t)$ exists for all $t>0$, then

$\Pr\left[\frac{1}{n}\sum_{i=1}^{n} X_i > t\right]$ decays exponentially fast (in n)

As a consequence,

$$\sum_{n=1}^{\infty} Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu) > \varepsilon\right] \leq \sum_{n=1}^{\infty} e^{-n\psi_{X-\mu}(\varepsilon)}$$

$$< \infty$$

$$\Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s} \mu \qquad \text{SLLN}$$

[Here, we use the following result without proof:

If $\sum_{n=1}^{\infty} Pr\left[|X_n - X| > \varepsilon\right] < \infty \quad \forall \varepsilon > 0$, then

$$X_n \xrightarrow{a.s} X$$

See the book by Bruce Hajek]

Example: $\mathcal{N}(0, \sigma^2)$

$$\mathbb{E} e^{\lambda X} = e^{\lambda^2 \sigma^2 / 2}$$

$$\psi_X^*(t) = \sup_{\lambda \geq 0} \underbrace{\left[\lambda t - \frac{\lambda^2 \sigma^2}{2} \right]}_{g(\lambda, t)}$$

g is a concave fn of λ

$$\frac{\partial g}{\partial \lambda} = 0 \Rightarrow \lambda = \frac{t}{\sigma^2}$$

$$\psi_X^*(t) = t^2 / 2\sigma^2$$

By Chernoff bound, $\Pr[X > t] \leq e^{-t^2/2\sigma^2}$

For $X_1 \dots X_n$ iid $\mathcal{N}(0,\sigma^2)$,

$$\Pr\Big[\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i > t}_{A}\Big] \le e^{-nt^2/2\sigma^2}$$

$$\Pr\Big[\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i < -t}_{B}\Big] \le e^{-nt^2/2\sigma^2}$$

$$\Pr\Big[\Big|\frac{1}{n}\sum_{i=1}^{n} X_i\Big| > t\Big] = \Pr[A \cup B]$$

$$\le \Pr[A] + \Pr[B]$$

(Union bound)

$$\le 2e^{-nt^2/2\sigma^2}$$

Chernoff bd

Example 2 : $Y \sim Ber(p)$

$$Y = \begin{cases} 1 & wp \quad p \\ 0 & wp \quad 1-p \end{cases}$$

$$X = Y - \mathbb{E}Y$$

$$\mathbb{E}Y = p$$

$$X = \begin{cases} 1-p & wp \quad p \\ -p & wp \quad 1-p \end{cases}$$

$$\mathbb{E}e^{\lambda X} = e^{\lambda(1-p)} p + e^{\lambda(-p)}(1-p)$$

$$= e^{-\lambda p}\left[pe^{\lambda} + (1-p)\right]$$

$$\psi_X(\lambda) = \ln \mathbb{E} e^{\lambda X} = -\lambda p + \ln\left[pe^{\lambda} + (1-p)\right]$$

$$\psi_X^*(t) = \sup_{\lambda \geq 0} \underbrace{\left[\lambda t + \lambda p - \ln\left[pe^{\lambda} + (1-p)\right]\right]}_{g(\lambda, t)}$$

$$\frac{\partial g}{\partial \lambda} = 0 \Rightarrow t + p - \frac{pe^{\lambda}}{pe^{\lambda} + (1-p)} = 0$$

$$t + p = \frac{pe^{\lambda}}{(1-p) + pe^{\lambda}}$$

$$e^{\lambda^*} = \frac{(t+p)(1-p)}{(1-(t+p))p}$$

$$\lambda^* = \ln\left[\frac{(t+p)(1-p)}{(1-(t+p))p}\right]$$

$$\psi_X^*(t) = \lambda^*(t+p) - \ln\left[pe^{\lambda^*} + 1 - p\right]$$

$$= (t+p)\ln\left[\frac{(t+p)(1-p)}{[1-(t+p)]p}\right] - \ln\left[1 - p + \frac{(t+p)(1-p)}{1-(t+p)}\right]$$

$$= (t+p)\ln\left(\frac{t+p}{p}\right) + (t+p)\ln\left[\frac{1-p}{1-(t+p)}\right]$$

$$- \ln\left[(1-p)\left(1 + \frac{t+p}{1-(t+p)}\right)\right]$$

$$= (t+p)\ln\left(\frac{t+p}{p}\right) + (t+p)\ln\left[\frac{1-p}{1-(t+p)}\right]$$

$$- \ln\left[\frac{1-p}{1-(t+p)}\right]$$

$$= (t+p)\ln\left(\frac{t+p}{p}\right) + (1-(t+p))\ln\left(\frac{1-(t+p)}{1-p}\right)$$

$$= D_{KL}(p_x \parallel q_x) \qquad p_x = (t+p,\ 1-(t+p))$$

$$= D_{KL}(t+p \parallel p) \qquad q_x = (p,\ (1-p))$$

If $X_1 X_2 - X_n$ are iid $\sim X = \begin{cases} 1-p & \text{wp } p \\ -p & \text{wp } 1-p \end{cases}$

$$\Pr\left[\frac{1}{n} \sum_{i=1}^{n} X_i \geq t \right] \leq e^{-n D_{KL}(t+p \| p)}$$

<u>Fact</u> : $D_{KL}(p(1+\delta) \| p) \geq \frac{\delta^2 p}{4}$ $\quad \delta \in \left[-\frac{1}{2}, \frac{1}{2}\right]$

$$\Pr\left[\frac{1}{n} \sum_{i=1}^{n} X_i \leq -t \right] \leq e^{-n D_{KL}(p-t \| p)}$$

$$\Pr\left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i \right| \geq p\delta \right] \leq 2 e^{-n\delta^2 p/4}$$

Application 1 : Random graphs

Erdős - Rényi graph : $V = \{1, 2, \dots n\}$

$(i,j) \in E$ wp p

$(i,j) \notin E$ wp $1-p$

Let d_i be the degree of vertex i

$$d_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{1}_{\{(i,j) \in E\}}$$

1 wp p

0 wp $1-p$

$$\mathbb{E} d_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \underbrace{\mathbb{E}\, \mathbb{1}_{\{(i,j) \in E\}}}_{p}$$

$$= (n-1)p$$

$$\Pr\left[\ d_i \in (n-1)p(1\pm\delta) \quad \text{for all } i\right]$$

$$\mathcal{E}_i$$

Instead, $\Pr\left[\ \{d_i \geq (n-1)p(1+\delta) \text{ or } d_i \leq (n-1)p(1-\delta)\} \text{ for at most one } i\right]$

$$= \Pr\left[\bigcup_{i=1}^{n} \mathcal{E}_i\right] \leq \sum_{i=1}^{n} \Pr[\mathcal{E}_i]$$

$$= n \Pr[\mathcal{E}_1] \leq n\, 2\, e^{-n\delta^2 p/4}$$

$$\mathcal{E}_1 = \{\ d_i \geq (n-1)p(1+\delta) \text{ or } d_i \leq (n-1)p(1-\delta)\}$$

$$\downarrow$$

$$\sum_{j=2}^{n} 1_{\{\ \}}$$

SUBGAUSSIAN RANDOM VARIABLES

Recall that if $X \sim \mathcal{N}(0, \sigma^2)$, then

$$\log \mathbb{E} e^{\lambda X} = \psi_X(\lambda) = \frac{\lambda^2 \sigma^2}{2}$$

$$\Pr[|X| \geq t] \leq 2e^{-t^2/2\sigma^2}$$

We say that a random variable is subgaussian with variance factor ν^2 if

$$\psi_X(\lambda) \leq \frac{\lambda^2 \nu^2}{2} \qquad \forall \lambda \in \mathbb{R}$$

$$\mathbb{E} e^{\lambda X} \leq e^{\lambda^2 \nu^2/2}$$

$$X \in \mathcal{G}(\nu^2)$$

Claim: If $X \in \mathcal{G}(\gamma^2)$, then

$\operatorname{Var}(X) \leq \gamma^2$

Proof:

$$\mathbb{E}\, e^{\lambda X} \leq \underbrace{e^{\lambda^2 \gamma^2/2}}_{}$$

MGF $\mathcal{N}(0, \gamma^2) = \mathbb{E}\, e^{\lambda Z}$

$$\mathbb{E}\left[1 + \lambda X + \frac{\lambda^2 X^2}{2!} + \frac{\lambda^3 X^3}{3!} + \cdots\right] \leq \mathbb{E}_Z\left[1 + \lambda Z + \frac{\lambda^2 Z^2}{2!} + \cdots\right]$$

$$\cancel{1} + \underbrace{\lambda\, \mathbb{E}X}_{0} + \frac{\lambda^2\, \mathbb{E}X^2}{2!} + \lambda^3 \frac{\mathbb{E}X^3}{3!} + \cdots \leq \cancel{1} + \underbrace{\lambda\, \mathbb{E}Z}_{0} + \frac{\lambda^2\, \mathbb{E}Z^2}{2!} + \frac{\lambda^3\, \mathbb{E}Z^3}{3!} + \cdots$$

Divide both sides by λ^2

& set $\lambda = 0$ (let $\lambda \to 0$)

$$\frac{\mathbb{E}X^2}{2!} \leq \frac{\mathbb{E}Z^2}{2!}$$

$$\mathbb{E}X^2 \leq \mathbb{E}Z^2$$

$$\mathrm{Var}(X) \leq \gamma^2$$

Theorem (Equivalent conditions for subgaussianity)

Let X be a rv with $EX = 0$. Then, the foll conditions are equivalent:

1. $\Pr[|X| > t] \leq 2e^{-t^2/K_1^2} \quad \forall t > 0$

2. $E|X|^p \leq (K_2\sqrt{p})^p \quad \forall p \geq 1$

3. $\mathbb{E}e^{\lambda^2 X^2} \leq e^{K_3^2\lambda^2} \quad \forall -\frac{1}{K_3} \leq \lambda \leq \frac{1}{K_3}$

4. $\mathbb{E}\, e^{X^2/K_4^2} \leq 2$

5. $\mathbb{E}\, e^{\lambda X} \leq e^{K_5^2\lambda^2} \quad \forall \lambda \in \mathbb{R}$

Reference: Vershynin, "High-dimensional probability"

Proof : ① ⇒ ② $\Pr[\,|X| > t\,] \le e^{-t^2/k_1^2} \quad \forall\, t \ge 0$

Exercise : If Y is a non-negative r.v. $\mathbb{E}Y = \mu < \infty$.

$$\mu = \int_0^\infty \Pr[Y > t]\, dt$$

$$\mathbb{E}|X|^p = \int_0^\infty \Pr[\,|X|^p > t\,]\, dt$$

$$= \int_0^\infty \Pr[\,|X| > t^{1/p}\,]\, dt$$

$$\le \int_0^\infty 2\, e^{-t^{2/p}/k_1^2}\, dt$$

Let $\frac{t^{2/p}}{K_1^2} = y$ $\qquad dy = \frac{t^{2/p - 1}}{K_1^2} \left(\frac{2}{p}\right) dt$

$$dt = K_1^2 \frac{p}{2} (K_1^2)^{p/2-1} y^{p/2-1} \, dy$$

$$= \frac{p}{2} (K_1^2)^{p/2} y^{p/2-1} \, dy$$

$$\mathbb{E}|X|^p \leq 2 \int_0^\infty \frac{p}{2} K_1^p y^{p/2-1} e^{-y} \, dy$$

$$= p K_1^p \int_0^\infty y^{p/2-1} e^{-y} \, dy$$

$$= p K_1^p \Gamma(p/2)$$

$$\leq p K_1^p (c p/2)^{p/2-1} = (c_2 K_1 \sqrt{p})^p$$

② $\Rightarrow$ ⑤ $\quad \mathbb{E}|X|^p \le (K_2\sqrt{p})^p$

$$\mathbb{E}e^{\lambda^2 X^2} = \sum_{i=0}^{\infty} \frac{\mathbb{E}(\lambda^2 X^2)^i}{i!}$$

$$= \sum_{i=0}^{\infty} \lambda^{2i} \frac{\mathbb{E}X^{2i}}{i!}$$

$$\le \sum_{i=0}^{\infty} \lambda^{2i} \frac{(K_2\sqrt{2i})^{2i}}{i!}$$

$$\le \sum_{i=0}^{\infty} (\lambda K_2\sqrt{2})^{2i} \frac{i^i}{i!}$$

$$\le \sum_{i=0}^{\infty} (\lambda c K_2)^{2i}$$

(Stirling's approx)

$$\leq \frac{1}{1-(\lambda c K_2)^2} \quad \text{for } |\lambda| < \frac{1}{cK_2}$$

For $x \in [0, 1/2]$, $\frac{1}{1-x} \leq e^{+2x}$

$$\mathbb{E} e^{\lambda^2 X^2} \leq e^{2(\lambda c K_2)^2} \quad |\lambda c K_2| < \frac{1}{2}$$

$$\leq e^{\lambda^2 K_3^2}$$

(3) $\Rightarrow$ (4) $\quad \mathbb{E} e^{\lambda^2 X^2} \leq e^{\lambda^2 K_3^2}$

$\Rightarrow \exists K_4$ st $\mathbb{E} e^{X^2/K_4^2} \leq 2$

4 $\Rightarrow$ 1

$$\Pr[|X| \geq t] = \Pr[X^2 \geq t^2]$$

$$= \Pr\left[e^{X^2/K_4^2} \geq e^{t^2/K_4^2}\right]$$

$$\leq \frac{\mathbb{E}\, e^{X^2/K_4^2}}{e^{t^2/K_4^2}} \qquad \text{Markov inequality}$$

$$\leq 2e^{-t^2/K_4^2} \qquad \text{from } 4$$

③ $\Rightarrow$ ⑤ $\qquad \mathbb{E}\, e^{\lambda^2 X^2} \le e^{K_3^2 \lambda^2} \qquad |\lambda| \le \frac{1}{K_3}$

<u>Claim</u>: $\quad e^{x} \le x + e^{x^2} \quad$ for all $x \in \mathbb{R}$

$$\mathbb{E}\, e^{\lambda X} \le \mathbb{E}\left(\lambda X + e^{\lambda^2 X^2}\right)$$

$$\le e^{K_3^3 \lambda^2} \qquad |\lambda| \le \frac{1}{K_3}$$

for $\quad |K_3 \lambda| > 1,$

$$2\lambda X \le K_3^2 \lambda^2 + \frac{X^2}{K_3^2}$$

$$\Updownarrow$$

$$0 \le \left(K_3 \lambda - \frac{X}{K_3}\right)^2$$

$$\mathbb{E} e^{\lambda X} \le \mathbb{E}\left[e^{K_3 \lambda^2/2} e^{+X^2/2K_3^2} \right]$$

$$\le e^{K_3^2 \lambda^2/2} \, \mathbb{E} e^{+X^2/2K_3^2}$$

$$\le e^{K_3^2\lambda^2/2} \, e^{K_3^2 \times 1/2K_3^2}$$

$$= e^{K_3^2\lambda^2/2} \, e^{1/2} \le e^{K_3^2\lambda^2/2}$$

$$\le e^{K_3^2\lambda^2}$$

since $K_3^2\lambda^2 \ge 1$

$\forall \; |\lambda| \ge \frac{1}{K_3}$

from ③

$\mathbb{E} e^{\bar{\lambda}^2 X^2} \le e^{\bar{\lambda}^2 K_3^2}$

$|\bar{\lambda}| \le \frac{1}{K_3}$

$\bar{\lambda}^2 = \frac{1}{2K_3^2}$

<u>Exercise</u>: Prove ③ ⇒ ①

(Use Chernoff bounds & then carefully choose λ)

Useful inequalities :

(1) $e^x \geq 1+x \quad \forall x \in \mathbb{R}$

$$f(x) = e^x - 1 - x$$

$$f(0) = 0$$

$$f'(x) = e^x - 1 \begin{cases} \geq 0 & \text{for } x \geq 0 \\ < 0 & \text{for } x \leq 0 \end{cases}$$

$\Rightarrow$ f is increasing for $x \in [0, \infty)$

f is decreasing for $x \in (-\infty, 0]$

$\Rightarrow$ $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

(2) $e^{-x} \leq 1 - x + x^2/2 \quad$ for $x \geq 0$

$$f(x) = e^{-x} - 1 + x - x^2/2$$

$$f(0) = 0$$

$$f'(x) = -e^{-x} + 1 - x$$

$$\leq -e^{-x} + e^{-x} \qquad \forall x \geq 0 \qquad \text{from } ①$$

$$\leq 0$$

$\Rightarrow$ f is decreasing in $[0, \infty)$

$\Rightarrow$ $f(x) \leq 0 \qquad \forall x \geq 0$

③ $$e^x \leq x + e^{x^2} \qquad \forall x \in \mathbb{R}$$

$$\equiv \quad xe^{-x} + e^{x^2 - x} \geq 1 \qquad \forall x \in \mathbb{R}$$

For $x \geq 0$,

$$xe^{-x} + e^{x^2 - x} \geq x(1-x) + 1 + x^2 - x \quad (\text{from } ①)$$

$$= 1$$

for $x < 0$, set $y = -x \quad (y > 0)$

$$-y\, e^{y} + e^{y^2 + y} = e^{y}\left(-y + e^{y^2}\right)$$

$$\geq e^{y}\left(-y + 1 + y^2\right) \qquad \text{from } ①$$

$$\geq e^{y}\left(1-y+y^2/2\right) \quad \text{as } \frac{y^2}{2} \geq 0$$

$$\geq e^{y}\, e^{-y} \quad \forall y \geq 0 \quad (\text{from } ②)$$

$$= 1$$

Examples : Bernoulli & Gaussian

(1) Bounded random variables

$$\Pr\left[X \in [\alpha, \beta]\right] = 1$$

Assume $\mathbb{E}X = 0$. $\Rightarrow$ $\alpha \leq 0$ & $\beta \geq 0$

$$\mathbb{E}\, e^{X^2/K_4^2} \leq 2 \qquad \text{for some } K_4 \in (0, \infty)$$

$$\mathbb{E}\, e^{X^2/K_4^2} \leq \exp\left[\max\{\alpha^2, \beta^2\}/K_4^2\right] \leq 2$$

$$K_4^2 = \frac{\max\{\alpha^2, \beta^2\}}{\log 2}$$

$\Rightarrow$ X is subgaussian

② Exponential random variable $\alpha > 0$

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{\alpha} e^{-x/\alpha}, & x \geq 0. \end{cases}$$

$$\underline{f_X(x)} = \frac{1}{2\alpha} e^{-|x|/\alpha} \qquad x \in \mathbb{R}$$

④ $$\mathbb{E} e^{X^2/K_4^2} = \int_{-\infty}^{\infty} \frac{1}{2\alpha} e^{-|x|/\alpha} e^{x^2/K_4^2} \, dx$$

$$= 2 \frac{1}{2\alpha} \int_0^{\infty} e^{x^2/K_4^2} e^{-x/\alpha} \, dx$$

$$= \frac{1}{\alpha} \int_0^{\infty} \exp\left[\frac{x^2}{K_4^2} - \frac{x}{\alpha} + \frac{K_4^2}{4\alpha^2} - \frac{K_4^2}{4\alpha^2} \right] dx$$

$$= \frac{1}{\alpha} e^{-K_4^2/4\alpha^2} \underbrace{\int_0^\infty \exp\left[\left(\frac{x}{K_4} - \frac{K_4}{2\alpha}\right)^2\right] dx}_{\infty}$$

$$= \infty$$

Not subgaussian

In fact, gamma distribution is NOT subgaussian.

APPLICATION : Dimensionality reduction using random projections.

$$\underline{x}_1, \overbrace{\underline{x}_2, \dots, \underline{x}_M} \qquad \underline{x}_i \in \mathbb{R}^N$$

$$f : \mathbb{R}^N \rightarrow \mathbb{R}^n$$

$$\| f(x_i) - f(x_j) \| \approx \| \underline{x}_i - \underline{x}_j \| \quad \forall\, i, j$$

$$(1-\varepsilon) \| x_i - x_j \| \le \| f(\underline{x}_i) - f(\underline{x}_j) \| \le (1+\varepsilon) \| x_i - x_j \|$$

$$O(NM)$$

$$O(nM)$$

Johnson–Lindenstrauss lemma

$m = 2^N$

$\varepsilon > 0$, $\quad \underline{x}_1, \underline{x}_2, \ldots \underline{x}_m \in \mathbb{R}^N \qquad n = \Theta\left(\frac{\log m}{\varepsilon^2}\right)$

$\exists\ f : \mathbb{R}^N \to \mathbb{R}^n$ st

$$(1-\varepsilon)\|\underline{x}_i - \underline{x}_j\|_2 \le \|f(\underline{x}_i) - f(\underline{x}_j)\|_2 \le (1+\varepsilon)\|\underline{x}_i - \underline{x}_j\|_2$$

Proof: Let $\underset{n\times N}{\bar{G}}$ with iid $\mathcal{N}(0,1)$

$$G = \frac{1}{\sqrt{n}}\bar{G}$$

Take any $\underline{x} \in \mathbb{R}^N$

$$\underline{y} = G\underline{x}$$

$1 \le i \le n \qquad y(i) = \sum_{j=1}^{N} g_{ij}\, x(j) = \sum_{j=1}^{N} \frac{\bar{g}_{ij}}{\sqrt{n}}\, x(j)$

$$\mathbb{E}\, y(i) = 0$$

$$\mathbb{E}\, y^2(i) = \frac{1}{n} \|\underline{x}\|^2$$

$$\mathbb{E}\, \|\underline{y}\|^2 = \sum_{i=1}^{n} \mathbb{E}\, y^2(i) = \|\underline{x}\|^2$$

<u>Claim</u>: $\Pr\Big[\underbrace{(1-\varepsilon)\|x\|^2 \le \|\underline{y}\|^2 \le (1+\varepsilon)\|x\|^2}_{A(x)}\Big] \ge 1 - \delta_n$

$$\ge 1 - 2^{-\theta(n\varepsilon^2)}$$

$$\Pr[A(\underline{x})] \ge 1 - 2^{-\theta(n\varepsilon^2)}$$

Fix $\underline{x}$. $\Pr\left[\|G\underline{x}\|^2 > (1+\varepsilon)\|x\|^2 \text{ OR } \|Gx\|^2 < (1-\varepsilon)\|x\|^2 \right] \leq 2^{-\theta(n\varepsilon^2)}$

for any $\underline{x}_i$, $\underline{x}_j$,

$$\Pr\left[\|G\underline{x}_i - G\underline{x}_j\|^2 > (1+\varepsilon)\|\underline{x}_i - \underline{x}_j\|^2 \text{ OR } \|Gx_i - Gx_j\|^2 < (1-\varepsilon)\|\underline{x}_i - x_j\|^2 \right]$$

$$\leq 2^{-\theta(n\varepsilon^2)}$$

$S = \{\underline{x}_1, \underline{x}_2 \dots \underline{x}_m\}$

$$\Pr\left[\|G\underline{x}_i - G\underline{x}_j\|^2 > (1+\varepsilon)\|\underline{x}_i - \underline{x}_j\|^2 \text{ OR } \|Gx_i - Gx_j\|^2 < (1-\varepsilon)\|\underline{x}_i - x_j\|^2 \text{ for any } \underline{x}_i, x_j \in S \right]$$

$$\leq \binom{m}{2} 2^{-\alpha n\varepsilon^2} \leq m^2 2^{-\alpha n\varepsilon^2} \leq \frac{1}{2}$$

as long as $n > \frac{\beta \log M^2}{\varepsilon^2}$

$n > c \frac{\log M}{\varepsilon^2}$

Need to s.t

$$\Pr\left[\ \|G\underline{x}\|^2 > (1+\varepsilon)\|x\|^2 \ \text{ OR } \ \|Gx\|^2 < (1-\varepsilon)\|x\|^2\right] \le 2^{-\theta(n\varepsilon^2)}$$

$$\mathbb{E}\|Gx\|^2 = \|x\|^2$$

$$Gx = y$$

$$\|Gx\|^2 = \|y\|^2 = \sum_{i=1}^{n} y^2(i)$$

$y^2(i)$ is Chi squared

$y(i) \sim \mathcal{N}(0, \|x\|^2/n)$

Use Chernoff bound & find best λ

Ex:

$$\Pr\left[\ \|y\|^2 > (1+\varepsilon)\mathbb{E}\|y\|^2 \ \text{ OR } \ \|y\|^2 < (1-\varepsilon)\mathbb{E}\|y\|^2\right]$$

$$= \Pr\left[\ \left|\|y\|^2 - \mathbb{E}\|y\|^2\right| > \varepsilon\|y\|^2\right] \le 2^{-\alpha n\varepsilon^2}$$

Subexponential / Subgamma distributions

If X is a r.v. Then, the foll are equivalent

($\forall i \neq j$, $\exists$ universal const c_1, c_2

$c_1 k_j \leq k_i \leq c_2 k_j$)

① $\Pr[\,|X| > t\,] \leq 2e^{-t/k_1}$ $\quad \forall t \geq 0$

② $\mathbb{E}|X|^p \leq (k_2 p)^p$ $\quad \forall p \geq 1$

③ $\mathbb{E}e^{\lambda|X|} \leq e^{k_3\lambda}$ $\quad |\lambda| \leq \frac{1}{k_3}$

④ $\mathbb{E}\,e^{|X|/k_4} \leq 2$

⑤ $\mathbb{E}e^{\lambda X} \leq e^{\lambda^2 k_5^2}$ $\quad |\lambda| \leq \frac{1}{k_6}$

Suppose that X is subgaussian.

$$Y = X^2 \qquad |Y| = Y = X^2$$

$$\mathbb{E} e^{X^2/K_4^2} \leq 2$$

$$\mathbb{E} e^{|Y|/K_4^2} \leq 2 \Rightarrow Y \text{ is subexponential.}$$

Suppose X is a r.v & $Y = X^2$ is subexponential

$$|Y| = Y = X^2$$

$$\mathbb{E} e^{|Y|/K_4} \leq 2$$

$$\mathbb{E} e^{X^2/K_4} \leq 2 \Rightarrow X \text{ is subgaussian}$$

If X & Y are independent & subgaussian,

$$X \in \mathcal{G}(\gamma_1) \qquad Y \in \mathcal{G}(\gamma_2)$$

$$\mathbb{E} e^{\lambda XY} \leq \mathbb{E} e^{\lambda(X^2+Y^2)/2}$$

$$= \mathbb{E} e^{\lambda X^2/2} \; \mathbb{E} e^{\lambda Y^2/2}$$

$$\leq e^{\lambda^2 K_1^2/4} \; e^{\lambda^2 K_2^2/4} \qquad |\lambda| \leq \min\left\{\frac{1}{K_1}, \frac{1}{K_2}\right\}$$

$$= e^{\lambda^2 (K_1^2+K_2^2)/4} \qquad |\lambda| \leq \frac{1}{K_3}$$

$\Rightarrow XY$ is subexponential.

WKT $\quad \mathbb{E} e^{\lambda X} \le e^{\lambda^2 K_5^2} \quad$ for $\quad |\lambda| \le \frac{1}{K_6}$

$$\psi_X(\lambda) \le \lambda^2 K_5^2 \qquad \text{— " —}$$

$$\psi_X^*(\lambda) = \sup_{\lambda \geq 0} \left[\lambda t - \psi_X(\lambda) \right]$$

$$\psi_X^*(t) \geq \sup_{0 \le \lambda \le \frac{1}{K_6}} \Big[\underbrace{\lambda t - \lambda^2 K_5^2}_{g(\lambda, t)} \Big]$$

$$\frac{\partial g}{\partial \lambda} = 0 \quad \Rightarrow \quad t - 2\lambda^* K_5^2 = 0$$

$$\lambda^* = \frac{t}{2K_5^2} \qquad \text{if} \quad \frac{t}{2K_5^2} \le \frac{1}{K_6}$$

$$t \le \frac{2K_5^2}{K_6}$$

$$\lambda^* = \frac{1}{K_6} \qquad \text{if} \quad t > \frac{2K_5^2}{K_6}$$

$$\psi_X^*(t) \geq \begin{cases} t^2/4K_5^2 & \text{if} \quad t \leq \dfrac{2K_5^2}{K_6} \\[2ex] \dfrac{t}{K_6} - \dfrac{K_5^2}{K_6^2} & \quad t > \dfrac{2K_5^2}{K_6} \end{cases}$$

$$\Pr\left[|X| > t\right] \leq \begin{cases} 2\exp\left[-\dfrac{t^2}{4K_5^2}\right] & \text{if } t \leq \dfrac{2K_5^2}{K_6} \rightarrow \text{small deviation regime (subgauss. tail)} \\[2ex] \underbrace{2\exp\left[\dfrac{K_5^2}{K_6^2}\right]\exp\left[-\dfrac{t}{K_6}\right]}_{\forall t \geq 0} & \text{else} \end{cases}$$

$\downarrow$

large deviation (exp. tail)

Suppose $X_1 X_2 - X_n$ independent

$$X_i \sim \text{Subexp}\left(k_5^{(i)} \; k_6^{(i)} \right)$$

$$\mathbb{E} e^{\lambda X_i} \le e^{\lambda^2 (k_5^{(i)})^2} \qquad \text{for } |\lambda| \le \frac{1}{k_6^{(i)}}$$

What can we say about $\sum_{i=1}^n X_i$?

$$X = \sum_{i=1}^n X_i$$

$$\mathbb{E} e^{\lambda X} = \mathbb{E} e^{\lambda \sum_{i=1}^n X_i} = \mathbb{E}\left[\prod_{i=1}^n e^{\lambda X_i} \right]$$

$$\le \prod_{i=1}^n e^{\lambda^2 (k_5^{(i)})^2} \qquad |\lambda| \le \frac{1}{k_6^{(i)}} \; \forall i$$

$$|\lambda| \le \min_i \frac{1}{k_6^{(i)}}$$

$$= e^{\lambda^2 \sum_{i=1}^n (k_5^{(i)})^2}$$

$$X \sim \text{Subexp}\left\{ \sqrt{\sum_{i=1}^n (k_5^{(i)})^2}, \; \max_i k_6^{(i)} \right\}$$

$$\Pr\left[\,\left|\sum_{i=1}^{n} X_i\right| > t\right] \le \begin{cases} 2\exp\left[\dfrac{-t^2}{4\sum_{i=1}^{n}\left(k_5^{(i)}\right)^2}\right] & t \le \dfrac{2\sum_i \left(k_6^{(i)}\right)^2}{\max_i k_6} \\ 2\exp\left[\dfrac{\sum_i \left(k_5^{(i)}\right)^2}{\max_i \left(k_6^{(i)}\right)^2} - \dfrac{t}{\max_i k_6^{(i)}}\right] & \end{cases}$$

Eg: Source coding

$$X^n \longrightarrow \boxed{ENC} \longrightarrow C^{nR} \longrightarrow \boxed{DEC} \longrightarrow \hat{X}_n$$

$(X_1 \dots X_n)$ $\qquad$ $(C_1 \dots C_{nR})$

$X_i \sim iid(p_X)$

Naive encoding: $R = \lceil \log_2 |\mathcal{X}| \rceil$

Want: R to be small

$$Pr[\hat{X}^n \neq X^n] \leq 2^{-0.001 n}$$

$$T_\varepsilon = \{ x^n : n p_X(a)(1-\varepsilon) \leq \underbrace{n_a(x^n)}_{\#\text{ times } a \text{ occurs in } x^n} \leq n p_X(a)(1+\varepsilon), \ \forall a \in \mathcal{X} \}$$

$x^n = (\alpha \beta \alpha \alpha \alpha \gamma \delta \gamma)$

$n_\alpha(x^n) = 4$

$n_\beta(x^n) = 1$

$n_\gamma(x^n) = 2$

$n_\delta(x^n) = 1$

Compression scheme: List all sequences in T_ε & assign a binary string/index to each sequence

$\lceil \log_2 |T_\varepsilon| \rceil$

ENC: if $X^n \in T_\varepsilon$, output index

if $X^n \notin T_\varepsilon$, output $00 \cdots 0$

DEC: Output sequence corresp to index (in T_ε)

$$nR = \lceil \log_2 |T_\varepsilon| \rceil$$

$$\Pr[\hat{X}^n \neq X^n] = \Pr[X^n \notin T_\varepsilon]$$

$$\Pr[X^n \notin T_\varepsilon] = \Pr[\; n_a(X^n) > n p_X(a)(1+\varepsilon) \text{ OR } n_a(X^n) < n p_X(a)(1-\varepsilon) \text{ for some } a \in \mathcal{X} \;]$$

$$\left[\text{for any } a, \quad \Pr[\; n_a(X^n) > n p_X(a)(1+\varepsilon) \quad \text{OR} \quad] \le 2^{-c\varepsilon^2 n} \right]$$

$$\Pr[X^n \notin T_\varepsilon] \le \sum_{a \in \mathcal{X}} \Pr[\; n_a(X^n) > n p_X(a)(1+\varepsilon) \text{ OR } n_a(X^n) < n p_X(a)(1-\varepsilon) \;]$$

$$\le |\mathcal{X}| \, 2^{-c\varepsilon^2 n} \le 2^{-c' n \varepsilon^2}$$