MEASURE CONCENTRATION

1. CLASSICAL BOUNDS:

- Concentration of measures relates to controlling the tail probability: $P[|x| \geq t]$, x is a R.V.
- A standard way is to use higher order moments of the random variable x
- Examples: markov, chebyshev, chernoff

  markov $\downarrow$ First; chebyshev $\downarrow$ second; chernoff $\downarrow$ M.G.F

  Hoeffding $\downarrow$ sub Gaussian, Bernstein $\downarrow$ sub exponential

2. Markov inequality:

Let $x \geq 0$ be a r.v. with finite mean. Then

$$E(x) = \int_0^{\infty} x\, p(x)\, dx = \int_0^{t} x\, p(x)\, dx + \underbrace{\int_t^{\infty} x\, p(x)\, dx}_{\geq 0}$$

$$\geq \int_t^{\infty} x\, p(x)\, dx \geq t \int_0^{t} p(x)\, dx$$

$$= t\, P[x \geq t]$$

$$\therefore\ P[x \geq t] \leq \frac{E(x)}{t} \longrightarrow (1)$$

3. Chebyshev inequality:

Let x be a r.v. with finite variance. Let μ be the mean of x (ie) $\mu = E(x)$.

Then, $|x-\mu| \geq t \Rightarrow (x-\mu)^2 \geq t^2$

- By Markov: $P[(x-\mu)^2 \geq t^2] \leq \frac{E(x-\mu)^2}{t^2}$

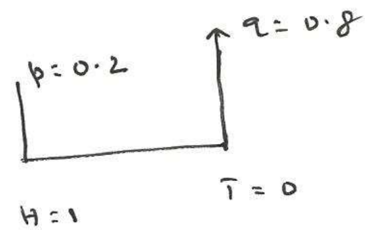
$$= VAR(x) / t^2 \longrightarrow (2)$$

$$\therefore \; P[\,|x-\mu| \geq t\,] \leq \frac{Var(x)}{t^2} \text{ for all } t > 0.$$

4. Examples : Markov

1) A coin falls head with probability $p = 0.2$. Toss it 20 times. What is the bound on the probability that it falls head at least 16 times?

Ans: BINAMIAL distribution:



- $E(x) = 0 * .8 + 1 * .2 = 0.2 = \frac{1}{5}$
- $E[x = 20, H] = np = 20 * \frac{1}{5} = 4$
- Markov: $P[x \geq 16] \leq \frac{E(x)}{16} = \frac{4}{16} = \frac{1}{4}$
- Actual probability: $= \sum_{k=16}^{20} \binom{20}{k} p^k q^{20-k}$

$$= \sum_{k=16}^{20} \binom{20}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^k \approx 1.38 * 10^{-18}$$

$\Rightarrow$ markov bound is a loose bound.

---

2) Consider $x = 0$ with $p = \frac{24}{25}$

$= 5$ with $q = \frac{1}{25}$

$$E(x) = 0 * \frac{24}{25} + 5 * \frac{1}{25} = \frac{5}{25} = \frac{1}{5}$$

Markov: $P[x \geq 5] \leq \frac{E(x)}{5} = \frac{1}{25}$

This is exact and hence there exists distribution for which Markov is tight.

---

3) Consider $x = 0$ with Prob $= 1 - 1/k^2$

$= k$ with Prob $= \frac{1}{k^2}$

$$E(x) = 0 * \left(1 - \frac{1}{k^2}\right) + k \cdot \frac{1}{k^2} = \frac{1}{k}$$

Markov: $P[x \geq k] \leq \frac{E(x)}{k} = \frac{1}{k^2} \Rightarrow$ Exact

5. Example chebyshev

1) . $P[x = 0] = 0.8$ $= 4/5$, $P[x = 1] = 0.2$ $= 1/5$

. $E(x) = 0 * 0.8 + 1 * 0.2 = 0.2 = 1/5 = p$

. $Var(x) = \left(0 - \frac{1}{5}\right)^2 * \frac{4}{5} + \left(1 - 1/5\right)^2 * \frac{1}{5}$

$= \frac{1}{25} * \frac{4}{5} + \frac{16}{25} \cdot \frac{1}{5} = \frac{\cancel{20}^{4}}{25 * \cancel{5}} = \frac{4}{25} = pq$

. $Var(x = 20) = npq = \cancel{20}^{4} * \frac{4}{\cancel{25}_{5}} = \frac{16}{5}$

. chebyshev: $P[x \geq 16] = P[x - 4 \geq 12]$

$= P(x - E(x) \geq 12)$

$\leq \frac{Var(x=20)}{(12)^2}$

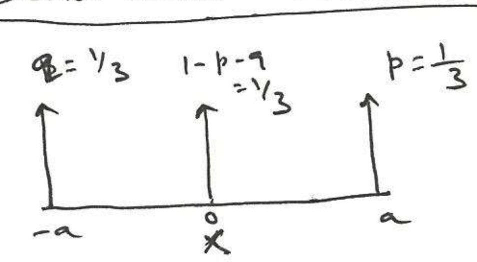
$= \frac{\cancel{16}^{4}}{5} \cdot \frac{1}{\cancel{12}_{3} \times \cancel{12}_{3}} = \frac{1}{45}$

$= 0.02222$

$\Rightarrow$ Better than markov.

---

2)



$a > 0$

$E(x) = 0$

$Var(x) = a^2 \cdot \frac{1}{3} + a^2 \cdot \frac{1}{3} = \frac{2a^2}{3}$

$P[|x| \geq t] \leq \frac{Var(x)}{t^2} = \frac{2a^2}{3t^2}$

$P[|x| \geq a] \leq \frac{2 \cdot a^2}{3 \cdot a^2} = \frac{2}{3}$

Note:- markov and chebyshev are tight and cannot be improved.

P.T.O

6. Extensions: Use higher order moments.

- Apply Markov to $|x-\mu|^k$ to get

$$P[\,|x-\mu| \geq t\,] = P[\,|x-\mu|^k \geq t^k\,]$$
$$\leq \frac{E(|x-\mu|^k)}{t^k} \quad \text{for } t > 0 \quad \rightarrow (3)$$

- This is based on k^{th} central moment
- We can use a polynomial function of $|x-\mu|^k$
- We can use moment generating function (MGF)
- Let x be a r.v. with MGF in the neighborhood of the origin. That is: $\exists\, b > 0$:

$$\phi(\lambda) = E[\,e^{\lambda(x-\mu)}\,]$$

exists for all $\lambda \in [0, b]$, ~~[illegible]~~

- Apply Markov to $Y = e^{\lambda(x-\mu)}$ to get

$$P[\,|x-\mu| \geq t\,] = P[\,e^{\lambda(x-\mu)} \geq e^{\lambda t}\,]$$
$$\leq \frac{E[\,e^{\lambda(x-\mu)}\,]}{e^{\lambda t}} \quad \rightarrow (4)$$

- By optimizing w.r. to $\lambda \in [0, b]$, we obtain <u>chernoff bound</u> - a tight bound:

$$\log P[\,|x-\mu| \geq t\,] \leq -\sup_{0 \leq \lambda \leq b} [\,\lambda t - \log E[e^{\lambda(x-\mu)}]\,] \quad \rightarrow (5)$$

<u>Note:-</u>
- The moment bound in (3) is never worse than ~~[illegible]~~ (5)
- But chernoff is easy analytically.

7. Chernoff-bound for Gaussian r.v.

- Let $X \sim N(\mu, \sigma^2)$
- $$E[e^{\lambda X}] = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{\lambda x} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$
- $$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2} + \lambda x\right]}\, dx$$
- Expressing the exponent as a perfect square and simplifying, verify that (Home work)

$$\boxed{\text{M.G.F}} \longrightarrow \boxed{E[e^{\lambda X}] = e^{\left(\mu\lambda + \frac{\sigma^2\lambda^2}{2}\right)}} \quad \forall \lambda \in \mathbb{R} \rightarrow (6)$$

- Apply (5):

$$\sup_{\lambda \in \mathbb{R}} \left[\lambda t - \log E\left[e^{\lambda(X-\mu)}\right]\right]$$

$$= \sup_{\lambda \in \mathbb{R}} \left[\lambda t - \frac{\sigma^2\lambda^2}{2}\right] \quad \left[\text{since } E\left[e^{\lambda(X-\mu)}\right] = e^{\frac{\sigma^2\lambda^2}{2}} \text{ Verify: Home work}\right]$$

$$= \left[\frac{t^2}{\sigma^2} - \frac{\sigma^2 t^2}{2\sigma^4}\right]$$

$$= \left[\frac{t^2}{\sigma^2} - \frac{t^2}{2\sigma^2}\right] = \frac{t^2}{2\sigma^2}$$

$$\boxed{\begin{aligned} &\frac{d}{d\lambda}\left[\lambda t - \frac{\sigma^2\lambda^2}{2}\right] \\ &= t - \sigma^2\lambda = 0 \\ &\Rightarrow \lambda = \frac{t}{\sigma^2} \end{aligned}}$$

$\therefore$ By (5)

$$\boxed{\text{upper Deviation inequality}} \quad P[X \geq \mu + t] \leq e^{-\frac{t^2}{2\sigma^2}} \longrightarrow (7)$$

- This bound is sharp up to a polynomial factor.
- Tail bounds obtained using Chernoff method depends on the growth rate of MGF.

8. Sub Gaussian Variables: Definition

- A r.v. X with mean $\mu = E(X)$ is called sub Gaussian (SG), if $\exists$ a $\sigma > 0$: $\forall \lambda \in R$

$$E\left[e^{\lambda(X-\mu)}\right] \le e^{\frac{\sigma^2\lambda^2}{2}} \longrightarrow (8)$$

- $\sigma > 0$ is called sub Gaussian parameter
- If $X \sim N(\mu, \sigma^2)$, then clearly X is SG
- By symmetry, $-X$ is sub Gaussian iff X is. Thus, we get a lower deviation inequality.
- Combining: for any SG r.v., we get the concentration inequality:

$$\boxed{\text{Concentration}} \rightarrow \quad P\left[|X-\mu| \ge t\right] \le 2e^{-\frac{t^2}{2\sigma^2}}, \; t \in R \longrightarrow (9)$$

- A large family of r.v. are sub Gaussian:

9). Examples: Let X be r.v. such that

1) $$P[X=1] = P[X=-1] = 1/2$$

called Rademacher variable.

- $$E[e^{\lambda X}] = \frac{1}{2}\left[e^{-\lambda} + e^{\lambda}\right]$$

$$= \frac{1}{2}\left[\sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right]$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!}$$

$$\le 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{2^k \cdot k!} \qquad (\because (2k)! \le 2^k\, k!)$$

$$= e^{\frac{\lambda^2}{2}}. \longrightarrow (10)$$

. That is x is SG with $\sigma = 1$.

2) Any bounded r.v. is SG:

Let $x \in [a, b]$ with zero mean. Let ~~x~~ y be a copy of x - independent copy

$$E_x[e^{\lambda x}] = E_x\left[e^{\lambda[x - E_y(y)]}\right]$$

$$\leq E_{x,y}\left[e^{\lambda(x-y)}\right] \rightarrow (11)$$

. Follows from Convexity of exp and Jensen's inequality: If f is convex:

$$\begin{cases} f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \\ \Rightarrow f(E(x)) \leq E[f(x)] \end{cases}$$

$$\Rightarrow e^{\lambda[x - E_y(y)]} \leq E_y\left[e^{\lambda(x-y)}\right] \Rightarrow (12)$$

Symmetri-zation

. Let z be an independent Rademacher variable.
. Then, the distribution of $(x-y)$ is the same as $z(x-y)$.

$$\cdot\ E_{x,y}\left[e^{\lambda(x-y)}\right] = E_{x,y}\left[E_z\, e^{\lambda z(x-y)}\right]$$

$$\leq E_{x,y}\left[e^{\frac{\lambda^2 (x-y)^2}{2}}\right] \rightarrow (13)$$

where we keep x, y fixed and use (10) with λ replaced by $\lambda(x-y)$.

. Recall $|x-y| \leq (b-a)$ and we get

$$E_{x,y}\left[e^{\frac{\lambda^2 (x-y)^2}{2}}\right] \leq e^{\frac{\lambda^2 (b-a)^2}{2}} \rightarrow (14)$$

(ii) x is SG with $\sigma = (b-a)$

3) <u>Property</u>: If x_1 and x_2 are independent Sub-Gaussian with paramaters σ_1^2 and σ_2^2, then $x_1 + x_2$ is Sub Gaussian with parameter $\sigma_1^2 + \sigma_2^2$. (Home work)

9. <u>HOEFFDING INEQUALITY</u> : <u>Sum of independent SG r.v.s</u>: Let x_i, $1 \le i \le n$ be independent r.v.s with mean, $E(x_i) = \mu_i$ and are sub Gaussian with parameters σ_i, $1 \le i \le n$. Then for all $t \geq 0$

$$P\left[\sum_{i=1}^{n}(x_i - \mu_i) \geq t\right] \le \exp\left(-\frac{t^2}{2\sum_{i=1}^{n}\sigma_i^2}\right) \to (15)$$

. This is often stated for the special case of bounded r.v. If $x_i \in [a, b]$, $1 \le i \le n$, then

$$P\left[\sum_{i=1}^{n}(x_i - \mu_i) \geq t\right] \le e^{\left[-\frac{t^2}{2n(b-a)^2}\right]} \to (16)$$

10. <u>Equivalent characterization of SG r.v.</u>

<u>Theorem</u>: Let x be a zero mean r.v.. Then the following are equivalent:

Proof: Refer to chapter 2 M. Wainwright

1) $\exists \sigma > 0$: $E[e^{\lambda x}] \le e^{-\frac{\lambda^2 \sigma^2}{2}}$, $\lambda \in \mathbb{R}$

2) $\exists c \geq 1$ and $Z \sim N(0, \sigma^2)$:

$$P[|x| \geq s] \le c\, P[|Z| \geq s], \quad \forall s \geq 0$$

3) $\exists \theta \geq 0$: $E[x^{2k}] = \frac{(2k)!}{2^k\, k!}\theta^{2k}$, $k = 1, 2, \ldots$

<u>Note</u>:- (1) is the definition of SG. (2) follows from SG is dominated by G. (3) gives control over moments.

4) $E\left[e^{\frac{\lambda x^2}{2\sigma^2}}\right] \le \frac{1}{\sqrt{1-\lambda}}$ for $\lambda \in [0, 1)$

11) Sub exponential r. v.

- The concept of SG is restrictive: $\sigma > 0$ and

$$E[e^{\lambda(x-\mu)}] \le e^{-\frac{\sigma^2\lambda^2}{2}} \quad \forall \lambda \in \mathbb{R}$$

- Relaxing the condition on λ gives raise to a new class called Sub-Exponential (SE) Variables.

Definition: A r.v. x is SE with mean μ if $\exists$ non-negative parameters (ν, b) : $b > 0$

$$E[e^{\lambda(x-\mu)}] \le e^{\frac{\nu^2\lambda^2}{2}} \quad \text{for } |\lambda| \le \frac{1}{b} \rightarrow (17)$$

special case

- Setting $\nu = \sigma$ and $b = 0 \Rightarrow |\lambda| \le \infty$ (ie) $\lambda \in \mathbb{R}$ and hence SG is SE but not viceversa

12) Example SE that is not an SG.

Let $Z \sim N(0,1)$ and $x = Z^2$. For $\lambda < 1/2$

$$E[e^{\lambda(x-1)}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda(z^2-1)} e^{-\frac{z^2}{2}} dz$$

$$= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}}$$

$\therefore$ For $\lambda > 1/2$, x is not SG.

Note: Existence of MGF near the mean is actually equivalent to the definition of SE family

To Verify: $\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \le e^{2\lambda^2} = e^{\frac{4\lambda^2}{2}}$ for $|\lambda| \le \frac{1}{4}$

which shows that x is SE with parameters $(\nu, b) = (2, 4)$.

- As with SG, the control on MGF which when combined with chernoff, yield deviation and concentration inequalities for SE family.

13) <u>Sub Exponential tail</u>. Let X be SE r.v. with (ν, b) as parameters. Then

$$P[X \geq \mu + t] \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}} & \text{for } 0 \leq t \leq \frac{\nu^2}{b} \\ e^{-\frac{t}{2b}} & \text{for } t > \frac{\nu^2}{b} \end{cases} \quad \rightarrow (18)$$

- That is, for t small, the bound is SG in nature with quadratic t, but for larger t the exponent in the bound scales linearly in t.

- By symmetry:

$$P[|X-\mu| \geq t] \leq \begin{cases} 2e^{-\frac{t^2}{2\nu^2}} & \text{for } 0 \leq t \leq \frac{\nu^2}{b} \\ 2e^{-t/2b} & \text{for } t > \frac{\nu^2}{b} \end{cases} \quad \rightarrow (19)$$

<u>Proof</u>: Assume centering and $\mu = 0$ and follow chernoff:

Combine the definition of SE in (17) and Markov:

$$\begin{aligned} P[X \geq t] &= P[e^{\lambda X} \geq e^{\lambda t}] \\ &\leq \frac{E[e^{\lambda X}]}{e^{\lambda t}} = e^{-\lambda t} E[e^{+\lambda X}] \\ &= e^{-\lambda t} \cdot e^{\frac{\nu^2 \lambda^2}{2}} \quad (\because (17) \text{ used}) \\ &= e^{g(\lambda, t)} \quad \text{for } \lambda \in [0, \tfrac{1}{b}] \end{aligned}$$

where $g(\lambda, t) = \frac{\nu^2 \lambda^2}{2} - \lambda t$

- unconstrained min. of $g(\lambda, t)$ occurs at $\lambda^* = \frac{t}{\nu^2}$ for each fixed $t \geq 0$.
- For $0 \leq t < \frac{\nu^2}{b} \Rightarrow$ it is also constrained min (since $0 \leq \lambda^* \leq \frac{\nu^2}{b} \cdot \frac{1}{\nu^2} = \frac{1}{b}$) and

$$g^*(t) = \frac{-t}{2\nu^2}$$

- For $t \geq \frac{\nu^2}{b}$, $g(\cdot, t)$ is monotonically decreasing in $[0, \lambda^*]$ and the constrained min. occurs at the boundary point $\lambda^+ = \frac{1}{b}$.
- Then $g^*(t) = g(\lambda^+, t) = -\frac{t}{b} + \frac{\nu^2}{2b^2}$

$$\leq -\frac{t}{2\alpha} \quad \text{for } t \geq \frac{\nu^2}{b}.$$

---

14) Note: . As in Example in section 12, SE property can be verified by explicit computing or bounding MGF.
- If this becomes difficult, the alternative is to control on the polynomial of the moments of x.
- It is here Bernstein's condition becomes handy.

14 Bernstein's Condition: with parameter b holds if $\mu = E(x)$, $\sigma^2 = E(x-\mu)^2$ and

$$|E(x-\mu)^k| \leq \frac{1}{2} k! \, \sigma^2 \, b^{k-2} \longrightarrow (20)$$

- A sufficient condition is that x is bounded.

- Using this we can get sharper tail bounds and is applicable for unbounded r.v. Hence, its wide applicability.

- Let (x, μ, σ^2) satisfy Bernstein's Condition with parameter b. Then, [expanding $e^{\lambda(x-\mu)}$]

$$E\left[e^{\lambda(x-\mu)}\right] = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!} E(x-\mu)^k$$

Bernstein bound $\rightarrow$ But $E\left[(x-\mu)^k\right] \leq \frac{1}{2} k!\, \sigma^2 b^{k-2}, \quad k = 3, 4, \dots$

$$\Rightarrow E\left[e^{\lambda(x-\mu)}\right] \leq 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2} \sum_{k=3}^{\infty} \left[|\lambda| b\right]^{k-2}$$

- For $|\lambda| < \frac{1}{b}$, sum of the geometric series gives

$$E\left[e^{\lambda(x-\mu)}\right] \leq 1 + \frac{\lambda^2\sigma^2/2}{1-b|\lambda|}$$

$$\leq e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}} \quad (\because 1+t \leq e^t)$$

- Hence

$$E\left[e^{\lambda(x-\mu)}\right] \leq e^{\frac{\lambda^2(\sqrt{2}\sigma)^2}{2}} \quad \forall \lambda \leq \frac{1}{2b} \rightarrow (21)$$

Note: $\frac{\lambda^2\sigma^2}{2(1-b|\lambda|)} \leq \frac{\lambda^2\sigma^2}{2} \cdot \frac{1}{1-b\cdot\frac{1}{2b}} = \frac{\lambda^2\sigma^2}{2} \cdot 2$

$$= \frac{\lambda^2(\sqrt{2}\sigma)^2}{2}$$

$\Rightarrow$ x is SE with $(\nu, b) = (\sqrt{2}\sigma, 2b)$

15) <u>Bernstein type bound</u>: for any r.v. satisfying Bernstein condition (20), we have

$$E\left[e^{\lambda(x-\mu)}\right] \le e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}} \quad \text{for } |\lambda| < \frac{1}{b}$$

→ (22)

and tail bound

$$P\left[|x-\mu| \geq t\right] \le 2e^{-\frac{t^2}{2(\sigma^2+bt)}} \quad \text{for } t \geq 0$$

→ (23)

<u>Note</u>: (22) was proved above. Using this bound on mgf, tail bound in (23) follows by setting $\lambda = \frac{t}{bt+\sigma^2} \in [0, \frac{1}{b})$ in the chernoff bound and simplifying - Home work.

16) <u>Note on the applicability</u>: Consequences for bounded r.v: $|x-\mu| \le b$

<u>First</u>: use boundedness to show that it is SG with parameter b and apply Hoeffding in section 9

<u>Second</u>: show that bounded r.v. satisfy Bernstein condition and above bound in section 15. This tail bound shows that for small t, x is SG with parameter σ as opposed to b that would arise in Hoeffding.

. Since $\sigma^2 = E(x-\mu)^2 \le b^2$, this bound is never worse. For $\sigma^2 << b^2$, it is better. This happens when the r.v. takes large values occasionally with small variance.

17) <u>Closure property for SE</u>: SE property is preserved under addition of independent

r.v. with parameters transform in a simple way.

- Let $\{x_k\}$ be a sequence of n ind. SE r.v with (ν_k, b_k) and $\mu_k = E(x_k)$ as parameters

Then, m.g.f. is

$$E\left[e^{\lambda \sum_{k=1}^{n}(x_k - \mu_k)} \right] = \prod_{k=1}^{n} E\left[e^{\lambda(x_k-\mu_k)} \right]$$

$$\leq \prod_{k=1}^{n} e^{\lambda^2 \nu_k^2/2}$$

for $|\lambda| \leq \max_k \left\{ \frac{1}{b_k} \right\}$ where independence and defn. of SE r.v. are used.

- That is, $\sum_{k=1}^{n}(x_k - \mu_k)$ is SE with (ν_*, b_*)

$$b_* = \max_k \{b_k\} \text{ and } \nu_* = \left(\sum_{k=1}^{n} \nu_k^2 \right)^{1/2}$$

- using the result in Section (13): we get the tail bound:

$$P\left[\frac{1}{n} \sum_{k=1}^{n}(x_k - \mu_k) \geq t \right] \leq \begin{cases} e^{-\frac{nt^2}{2\nu_*}}, & 0 \leq t \leq \frac{\nu_*^2}{b_*} \\ e^{-\frac{nt}{2b_*}}, & t > \frac{\nu_*^2}{b_*} \end{cases}$$

→ (24)

18) Example: χ^2 Variables

- $z_k \sim N(0,1)$, $z_1, z_2, \ldots z_n$ are i.i.d.

$\boxed{b_k = 4 \text{ for all } k}$

- $Y = \sum_{k=1}^{2} z_k^2 \sim \chi^2_n$

$\nu_k = 2$
$\left(\sum \frac{\nu_k^2}{n}\right)^{1/2} = \left(\frac{4n^2}{n}\right)^{1/2} = 2\sqrt{n}$

As in Example in Section (12): Y is SE with $(\sigma, b) = (2\sqrt{n}, 4)$.

Hence

$$P\left[\left|\frac{1}{n}\sum_{k=1}^{n}(z_k^2-1)\right|>t\right] \le 2e^{-\frac{nt^2}{8}} \quad \forall\, t\in(0,1) \longrightarrow (25)$$

19) ~~Proo~~ Johnson-Lindenstrauss embedding:

d-guest dimension

- Let $\{u_1, u_2, \dots u_N\} \in \mathbb{R}^d$ where d is large
  (ie) N objects in d-dimensional space.
- For large d, it requires too much space to store the data

m-Host dimension

- Seek a map $f: \mathbb{R}^d \to \mathbb{R}^m$ with $n << d$ that preserves the essential features of the data by storing only $\{f(u_1), f(u_2) \dots f(u_N)\} \in \mathbb{R}^n$
- One goal is to preserve the pairwise distances and interested in f:

$$(u_i \ne u_j) \quad (1-\delta)\|u_i-u_j\|_2^2 \le \left(\|f(u_i)-f(u_j)\|^2\right) \le (1+\delta)\|u_i-u_j\|^2 \longrightarrow (26)$$

for some $\delta\in(0,1)$ with $n << d$.

- (26) is called near isometric embedding.
- Solution: we can realize (26) with large probability so long as $m = \Omega\left(\frac{\log N}{\delta^2}\right)$ and scales logarithmically as the number of points and does not depend on ambient dimension d.

Proof: Form $X \in \mathbb{R}^{m \times d}$ with $x_{ij} \overset{iid}{\sim} N(0,1)$ and define $f : \mathbb{R}^d \to \mathbb{R}^n$ as $f(u) = \frac{1}{\sqrt{m}} X u$.

- Let x_i be the i^{th} row of X and $u \neq 0$.
- Since $x_i \sim N(0, I_d)$, $\langle x_i, \frac{u}{\|u\|_2} \rangle$ which is a linear combination of x_{ij} with components of the unit vector $\frac{u}{\|u\|_2}$ is clearly $N(0,1)$ (Verify: Homework)
- Define

$$Y = \frac{\|Xu\|_2^2}{\|u\|_2^2}$$

$$= \sum_{i=1}^{m} \langle x_i, \frac{u}{\|u\|_2} \rangle^2$$

$$\sim \chi^2_m \quad \text{(Verify: Homework)}$$

- Apply tail bound in (25) in Section 18.

$$P\left[\left| \frac{\|Xu\|^2}{m\|u\|_2} - 1 \right| \geq \delta \right] \leq 2e^{-\frac{m\delta^2}{8}} \quad \text{for } \delta \in (0,1)$$

- From the definition of f and rearranging:

$$P\left[\left| \frac{\|f(u)\|_2^2}{\|u\|_2^2} - 1 \right| \geq \delta \right] \leq 2e^{-\frac{m\delta^2}{8}} \quad \text{for } \delta \in (0,1)$$

and for $u \neq 0$ and $u \in \mathbb{R}^d$.

- There $\binom{N}{2} = \frac{N(N-1)}{2}$ pairs of data points and by union bound:

$$P\left[\frac{\|F(u_i - u_j)\|_2^2}{\|u_i - u_j\|_2^2} \notin [1-\delta, 1+\delta] \right] \leq 2\binom{N}{2} e^{-\frac{m\delta^2}{8}}$$ (17

$$\leq 2 \cdot \frac{N(N-1)}{2} e^{-\frac{m\delta^2}{8}}$$

$$\leq N^2 e^{-m\frac{\delta^2}{8}} \leq \varepsilon^2$$

when
$$\varepsilon^2 \geq N^2 e^{-\frac{m\delta^2}{8}}$$

$$\log \varepsilon^2 \geq 2 \log N - m\frac{\delta^2}{8}$$

$$\Rightarrow \quad m\frac{\delta^2}{8} \geq 2 \log N - \log \varepsilon^2 = 2 \log (N/\varepsilon)$$

$$\therefore m \geq \frac{16}{\delta^2} \log (N/\varepsilon) = \Omega\left(\frac{\log N}{\delta^2} \right)$$

---

20) Properties of SE Variables

. Let X be SE variable with $E(X) = 0$. Then the following are equivalent.

1) $\exists (\nu, b) : E[e^{\lambda X}] \leq e^{\frac{\nu^2 \lambda^2}{2}}$ for $|\lambda| \leq \frac{1}{b}$

2) $\exists c_0 > 0 : E[e^{\lambda X}] < \infty \quad \forall |\lambda| \leq c_0$

3) $\exists c_1, c_2 > 0 : P[|X| > t] \leq c_1 e^{-c_2 t}, \quad t > 0$

4) $$\gamma = \sup_{k \geq 2} \left[\frac{E(X^k)}{k!} \right]^{1/k} \text{ is finite}$$

Lipschitz functions of Gaussian Variables:

- Exhibit attractive form of dimension-free Concentration
- $f: \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz w.r.to 2-norm:

$$\| f(x) - f(y) \|_2 \leq L \| x - y \|_2 \quad ; x, y \in \mathbb{R}^n \to (1)$$

- claim: This function is Sub Gaussian with parameter atmost L.

Thm 2.26: Let $x \in \mathbb{R}^n$ with $x \sim N(0, I_n)$ and $f: \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz as in (1) above. Then $Y = f(x) - E(f(x))$ is SG with parameter atmost L, and

$$P\left[|f(x) - E[f(x)]| \geq t \right] \leq 2 e^{-\frac{t^2}{2L}} \quad \text{for } t \geq 0 \quad \to (2)$$

Note: It is remarkable that L-Lipschitz fn. of standard Gaussian Vector, regardless of the dimension n of x, exhibits Concentration like a scalar Gaussian Variable with variance L^2.

Proof: Prove a weaker result for both Lipschitz + differentiable (since Lipschitz functions are differentiable almost everywhere) functions. We need a Lemma

Lemma 2.27: Let $f: \mathbb{R}^n \to \mathbb{R}$ be $C^1(\mathbb{R}^n)$ and ϕ be a Convex fn. $\phi: \mathbb{R} \to \mathbb{R}$.

$$\mathbb{E}\left[\phi\left[\underbrace{f(x)-\mathbb{E}(f(x))}\right]\right] \le \mathbb{E}\left[\phi\left[\frac{\pi}{2}\langle \nabla f(x), Y\rangle\right]\right]^{2} \rightarrow (3)$$

for $x, y \sim N(0, I_n)$ and are independent.

Proof of Theorem:

$\boxed{\text{Lemma 2.27}}$

. Let $\lambda \in \mathbb{R}$ be fixed, apply (3) to the convex fn $t \to \exp(\lambda t)$:

$\boxed{x, y \text{ are iid}}$

$$\mathbb{E}\left[\exp\left(\lambda\left[\underbrace{f(x)-\mathbb{E}(f(x))}\right]\right)\right]$$

$$\le \mathbb{E}\left[\exp\left(\frac{\lambda\pi}{2}\sum_{k=1}^{n} Y_k \frac{\partial f}{\partial x_k}\right)\right]$$

$$= \mathbb{E}_x\left[\mathbb{E}_y\left[\exp\left(\underbrace{\frac{\lambda\pi}{2}\sum_k Y_k \frac{\partial f}{\partial x_k}}\right)\right]\right]$$

$$= \mathbb{E}_x\left[\prod_{k=1}^{n} \mathbb{E}_{Y_k}\left[\exp\left(\underbrace{\frac{\lambda\pi}{2} Y_k \frac{\partial f}{\partial x_k}}\right)\right]\right]$$

But $Y_k \sim N(0, 1)$ set $\lambda = \left(\frac{\lambda\pi}{2}\frac{\partial f}{\partial x_k}\right)$, $\mu = 0$

$Y \sim N(\mu, \sigma^2)$

$\mathbb{E}[e^{\lambda Y}] = e^{\mu\lambda + \frac{\sigma^2\lambda^2}{2}}$

$$\Rightarrow \mathbb{E}_{Y_k}\left[\exp\left(\frac{\lambda\pi}{2} Y_k \frac{\partial f}{\partial x_k}\right)\right] = \exp\left[\frac{\lambda^2\pi^2}{8}\left(\frac{\partial f}{\partial x_k}\right)^2\right]$$

$$\therefore \mathbb{E}\left[\exp\left[\underbrace{\lambda\left(f(x)-\mathbb{E}(f(x))\right)}\right]\right]$$

$$\le \mathbb{E}\left[e^{\frac{\lambda^2\pi^2}{8}\|\nabla f(x_k)\|_2^2}\right]$$

$$\le e^{\frac{1}{8}\lambda^2\pi^2 L^2} \qquad \left[\because \|\nabla f(x)\|_2^2 \le L^2\right]$$

$P[|x-\mu| \ge t] \le 2e^{\frac{-t^2}{2\sigma^2}}$

SG. with $\sigma > 0$

$\mathbb{E}[e^{\lambda(x-\mu)}] \le e^{\frac{\sigma^2\lambda^2}{2}}$

$\Rightarrow \left[f(x) - \mathbb{E}(f(x))\right]$ is SG with parameter $\frac{\pi L}{2}$

$$\Rightarrow P\left[\underbrace{|f(x)-\mathbb{E}(f(x))|} \ge t\right] \le 2\exp\left[-\frac{2t^2}{\pi^2 L}\right]$$

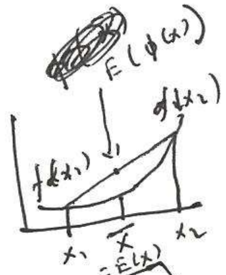
(HOEFFDING)

Proof of lemma 2.27:

- Based on interpolation method exploiting rotation invariance of Gaussian
- For $\theta \in [0, \pi/2]$, define $Z(\theta) \in \mathbb{R}^n$ with components: $x_k, y_k \sim N(0,1)$ iid

$$Z_k(\theta) = x_k \sin\theta + y_k \cos\theta \qquad 1 \le k \le n$$

- Let $Z_k(\theta), Z_k'(\theta)$ for all $\theta \in [0, \pi/2]$ be a jointly Gaussian pair with zero mean and covariance $= I_2$



- ϕ is convex: $\Rightarrow$

$\boxed{E[\phi(x)] \le \phi(E(x))}$

$$E_x\left[\phi\left(\underbrace{f(x) - E_y f(y)}\right)\right] \le E_{x,y}\left[\phi\left(\underbrace{f(x) - f(y)}\right)\right] \quad \to (2.41)$$

- $Z_k(0) = y_k, \quad Z_k(\pi/2) = x_k$ for $1 \le k \le n$

$$\Rightarrow f(x) - f(y) = \int_0^{\pi/2} \frac{d}{d\theta} f(Z(\theta))\, d\theta$$

$$= \int_0^{\pi/2} \langle \nabla f(Z(\theta)), Z'(\theta) \rangle\, d\theta \quad \to (2.42)$$

element-wise derivative

- using (2.42) in (2.41)

$$E_x\left[\phi\left(\boxed{f(x) - E_y(f(y))}\right)\right]$$

$$\le E_{x,y}\left[\phi\left(\int_0^{\pi/2} \langle \nabla f(Z(\theta)), Z'(\theta) \rangle\, d\theta\right.\right.$$

$$= E_{x,y}\left[\phi\left(\frac{1}{(\pi/2)} \int_0^{\pi/2} \frac{\pi}{2} \langle \nabla f(Z(\theta), Z'(\theta)) \rangle\, d\theta\right.\right.$$

$$\le \frac{1}{(\pi/2)} \int_0^{\pi/2} E_{x,y}\left[\phi\left(\frac{\pi}{2} \langle \nabla f(Z(\theta), Z'(\theta)) \rangle\right)\right.\, d\theta \quad \to (2.43)$$

- Since $(z_k(\theta), z'_k(\theta)) \sim N(0, I_2)$ for all θ, the expectation does not depend on θ.
- Hence

$$\frac{1}{(\pi/2)} \int_0^{\pi/2} E_{x,y}\left[\phi\left(\frac{\pi}{2} < \nabla f(z(\theta)), z'(\theta) >\right)\right] d\theta$$

$$= E\left[\phi\left(\frac{\pi}{2} < \nabla f(\tilde{x}), \tilde{y} >\right)\right]$$

where $\tilde{x}, \tilde{y}$ are i.i.d $N(0,1)$. Hence the claim.

<u>Note:</u> A similar concentration result holds good for non-Gaussian dist - uniform distribution on S^n and strictly log-concave distribution. <u>Refer chapter 3</u>:

- But for dim. free concentration bounds for Lipschitz fn we need additional conditions (eg: convexity) for arbitary s.G. distributions.

.x.

<u>Applications</u>:

S.E.

<u>Example 2.28</u>: $Z \in \mathbb{R}^n$, $Z_k \sim$ i.i.d $N(0,1)$

$$Y = \sum Z_k^2 \sim \chi^2_n$$

- To get tail ~~bound~~: Z_k^2 is SE and use independence as in the example above.

Lipschitz

- <u>Alternate route</u>: via Lipschitz functions of Gaussian r.v.
- Let $V = \frac{\sqrt{Y}}{\sqrt{n}} \Rightarrow V = \frac{\|Z\|_2}{\sqrt{n}}$
- Euclidean norm is <u>1-Lipschitz fn</u>.

· By Thm 2.26:

$$P[\underbrace{V \geq E(V) + \delta}] \leq \exp\left[-\frac{n\delta^2}{2}\right]$$

· By concavity of $\sqrt{x}$ fn. + Jensen inequalities:

$$E(V) \leq \sqrt{E(V^2)}$$

$$= \frac{1}{n}\sum E(Z_k^2)$$

$$= 1 \quad (\because E(Z_k^2)=1)$$

$\boxed{V \geq E(V) + \delta = (1+\delta)}$

$\boxed{\begin{aligned} Var(V) &= E(V^2) - E^2(V) \geq 0 \\ E(V^2) &\geq E^2(V) \\ E(V) &\leq [E(V^2)]^{1/2} \end{aligned}}$

· From $V = \frac{\sqrt{Y}}{\sqrt{n}}$, putting all pieces together:

$$P\left[\frac{Y}{n} \geq (1+\delta)^2\right] \leq \exp\left[-\frac{n\delta^2}{2}\right], \forall \delta \geq 0$$

· $(1+\delta)^2 = 1 + 2\delta + \delta^2 < 1 + 3\delta \quad \forall \delta \in [0,1]$

$$\Rightarrow P[Y \geq \underbrace{n(1+t)}] \leq \exp\left(-\frac{nt^2}{18}\right)$$

where $t = 3\delta$ $\quad \forall t \in [0,3]$

$\rightarrow (2.44)$

Note: Compare this (2.44) with the one in earlier example 2.11 using SE argument.

$\boxed{\text{Ex } 2.11}$

$$P\left[\left|\frac{1}{n}\left(\sum Z_k^2 - 1\right)\right| \geq t\right] \leq 2e^{-\frac{nt^2}{8}} \quad \forall t \in (0,1)$$

?