Feynman Liang
STAT 210B
Lecture #3
1/24/2017

1 Review

Commit to memory: *Sub-gaussian* mean $\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2 \sigma^2}{2}}$

- λ is *tilting* the distribution, picking out a moment
- σ is a *scale parameter*, like a variance

$\implies P(|X - \mu| > t) \leq e^{-\frac{t^2}{2\sigma^2}}$

Sub-exponential example: For χ^2

$$\mathbb{E}e^{\lambda X} = \frac{e^{-\lambda}}{\sqrt{1-2}} \leq e^{2\lambda^2} \leq e^{2\lambda^2}, \quad |\lambda| < \frac{1}{2} \tag{1.1}$$

Equivalent definitions of sub-Gaussian:

(a) $\exists \sigma$ st $\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2\sigma^2}{2}}$

(b) $\exists c_0 > 0$ st $\mathbb{E}e^{\lambda X} < \infty, \quad |\lambda| \leq c_0$

(c) $\exists c_1, c_2 > 0$ st $P(|X| > t) \leq c_1 e^{-c_2 t}, \quad t > 0$

(d) $\sup_{k \geq 0} \left(\frac{\mathbb{E}|X|_k}{k!}\right)^{1/k}$ is finite

2 Tail bounds on sub-Exponential

Let X be zero-mean (subtract $\mathbb{E}X$ off otherwise).

$$P(X \geq t) \leq \exp\left(-\lambda t + \frac{\lambda^2 \nu}{2}\right), \quad \lambda < \frac{1}{\alpha} \tag{2.1}$$

Find infimum over $0 \leq \lambda < \frac{1}{\alpha}$.

Unconstrianed optimum

$$\lambda^* = \frac{t}{\nu^2} \stackrel{?}{=} \frac{1}{\alpha} \tag{2.2}$$

$$\tag{2.3}$$

If $t < \frac{\nu^2}{\alpha}$ (small deviations), then

$$\implies P(X \geq t) \leq \exp\left(-\frac{t^2}{2\nu^2}\right) \tag{2.4}$$

"Gaussian-like."

Otherwise $t > \frac{\nu^2}{\alpha}$ (large deviations),

$$\implies \phi(\lambda) = -\lambda t + \frac{\lambda^2 \nu}{2} \tag{2.5}$$

$$\frac{d\phi}{d\lambda} = -t + \frac{\nu^2}{2} \leq -\frac{\nu^2}{\alpha} + \lambda \nu^2 \tag{2.6}$$

$$= \nu^2 \left(\nu - \frac{1}{\alpha} \right) \leq 0 \tag{2.7}$$

$$\implies \lambda^* = \frac{1}{\alpha} \tag{2.8}$$

$$\implies P(X \geq t) \leq \exp\left(-\frac{t}{2\alpha} \right) \tag{2.9}$$

"Exponential tail."

In summary, there are two regimes depending on how far out (t)

$$P(X - \mu \geq t) \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}}, & 0 \leq t \leq \frac{\nu^2}{\alpha} \\ e^{-\frac{t}{2\alpha}}, & t > \frac{\nu}{\alpha} \end{cases} \tag{2.10}$$

3 Bernstein's Condition

TODO: Fig 4.1 TODO: Fig 4.2

Definition 3.1. A RV X satisfies *Bernstein's condition* if

- $\sigma^2 = \mathbb{E}X^2 - \mu^2$ exists
- $\exists b > 0$ such that $\mathbb{E}(X-\mu)^k \leq \frac{1}{2}k!\sigma^2 b^{k-2}$ for all $k = 3, 4, \cdots$

Remark 3.2. This holds, e.g., for bounded RVs, improves on Hoeffding

Remark 3.3. Bounded $\implies$ Sub-Exponential

Theorem 3.4. *Bernstein's condition implies sub-exponential*

Proof.

$$\mathbb{E}e^{\lambda(X-\mu)} = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \lambda^k \frac{E(X-\mu)^k}{k!} \tag{3.1}$$

$$\leq 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2} \sum_{k=3}^{\infty} \lambda^{k-2} b^{k-2} \tag{3.2}$$

$$= 1 + \frac{\lambda^2\sigma^2}{2} \left(\frac{1}{1 - |\lambda| b} \right), \quad |\lambda| < \frac{1}{b} \tag{3.3}$$

$$\leq e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}} \leq e^{\frac{\lambda^2(\sqrt{2}\sigma)^2}{2}} \implies SE(\sqrt{2}\sigma, 2b) \tag{3.4}$$

$\square$

Remark 3.5. $\mathbb{E}e^{\lambda(X-\mu)} \leq e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}}$ can be used to give more tighter bounds than the sub-exponential (special case when $\lambda = \frac{1}{2b}$)

Aside: This is a convexity bound, follows from the general framework of conjugate duality (*Fenchenl transform*)

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k f(t) \tag{3.5}$$

$$f^*(\lambda) := \sup_{\lambda} (\lambda t - f(t)) \tag{3.6}$$

$$|x| < 1 \tag{3.7}$$

$$1 + t \leq e^t \tag{3.8}$$

TODO: Fig 4.3

The following is worth memorizing:

Proposition 3.6 (Bernstein's inequality). *From*

$$\mathbb{E}e^{\lambda(X-\mu)} \leq e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}} \tag{3.9}$$

we get

$$P(|X-\mu| \geq t) \leq 2\exp\left(\lambda t - \frac{\lambda^2\sigma^2}{2(1-b|\lambda|)}\right) \tag{3.10}$$

$$= 2e^{-\frac{t^2}{2(bt+\sigma^2)}} \tag{3.11}$$

where the last line is obtained by setting $\lambda = \frac{t}{bt+\sigma^2}$.

Whenever you have bounded RVs, instead of reaching for Hoeffding, consider Bernstein.

Proposition 3.7 (Sum of sub-exponentials). *If $X_i \sim SE(\nu_i, \alpha_i)$, then $\sum_i X_i \sim SE(\sqrt{\sum_i \nu_i^2}, \max_i \alpha_i)$*

Proof. Let X_i be independent with mean μ_i and sub-exponential parameters ν_i, α_i.

$$\mathbb{E}e^{\lambda \sum_{i=1}^n (X_i-\mu_i)} = \prod_{i=1}^n \mathbb{E}e^{\lambda(X_i-\mu_i)} \tag{3.12}$$

$$\leq \prod_{i=1}^n e^{\lambda^2\nu_i^2/2} \tag{3.13}$$

$$= e^{\lambda^2 \sum_i \nu_i^2/2} \tag{3.14}$$

This holds for $\lambda \leq \frac{1}{\alpha_i}$ for all i, so $\lambda \leq \frac{1}{\max_i \alpha_i}$. □

Example 3.8. Let $Y = \sum_{k=1}^{n} Z_k^2$ where $Z_k \sim N(0,1)$. $Y \sim \chi_n^2$ and since it's a sum of sub-exponentials, $Y \sim SE(2\sqrt{n}, 4)$.

Know $Z_K^2 \sim SE(2,4)$, so

$$P\left(\left|\frac{1}{n}\left(\sum_{k=1}^{n} Z_k^2 - 1\right)\right| \geq t\right) \leq 2e^{-\frac{nt^2}{8}} \tag{3.15}$$

Example 3.9 (Johnson-Lindenstrauss)**.** $N > 2$ points in $\mathbb{R}^d$, say $u^1, \ldots, u^N$.

Project onto $m \ll d$ dimensions, $F : \mathbb{R}^d \to \mathbb{R}^m$.

Want to preserve metric approximately (i.e. with "fudge-factor" δ)

$$1 - \delta \leq \frac{\|F(u^i) - F(u^j)\|_2^2}{\|u^i - u^j\|_2^2} \leq 1 + \delta \tag{3.16}$$

Can do this with high probability if

$$m \gtrsim \frac{1}{\delta^2} \log N \tag{3.17}$$