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

1 Sub-Gaussian RVs and tail bounds

Recall *Markov's inequality*:

$$P(X \geq t) \leq \frac{\mathbb{E}X}{t} \tag{1.1}$$

$$P(X \geq t) = P(X^2 \geq t^2) \leq \frac{\operatorname{Var} X}{t^2} \tag{1.2}$$

$$P(X \geq t) = P(X^k \geq t^k) \leq \frac{(X - \mathbb{E}X)^k}{t^k} \tag{1.3}$$

Can get tighter with *Chernoff bound*:

$$P(X \geq t) = P(e^{\lambda X} \geq e^{\lambda t}) \leq \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda t}}, \lambda > 0 \tag{1.4}$$

$$\log P(X \geq t) \leq \inf_{\lambda > 0} \left(\log \mathbb{E}e^{\lambda X} - \lambda t \right) \tag{1.5}$$

Some useful facts:

- Exponential Taylor series: $e^X = \sum_{k=0}^{\infty} \frac{X^k}{k!}$
- Gaussian MGF: For $X \sim N(\mu, \sigma^2)$, $\phi(t) := \mathbb{E}e^{tX} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
- Cauchy-Schwarz: $(\mathbb{E}XY)^2 \leq \mathbb{E}X^2 + \mathbb{E}Y^2$

If $X \sim N(\mu, \sigma^2)$, Chernoff bound becomes

$$\inf_{\lambda > 0} \left(\log \mathbb{E}e^{\lambda X} - \lambda t \right) = -\frac{t^2}{2\sigma^2} \tag{1.6}$$

$$\implies P(|X - \mu| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}} \tag{1.7}$$

where the 2 comes from a *union bound* $P(A \cup B) \leq P(A) + P(B)$.

Definition 1.1. A RV X is *sub-Gaussian* if $\exists \sigma > 0$ such that $\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2 \sigma^2}{2}}$

Any sub-Gaussian RV has a Chernoff tail bound.

Example 1.2. Let X be a *Rademacher* RV, i.e. $X \in \{-1, 1\}$ with equal probability. Then

$$\mathbb{E}e^{\lambda X} = \frac{1}{2}(e^{-\lambda} + e^{\lambda}) \tag{1.8}$$
$$= \frac{1}{2}\left(\sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) \tag{1.9}$$
$$= \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} = 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{(k+1)(k+2)\cdots(2k)k!} \tag{1.10}$$
$$\leq 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{2^k k!} = \sum_{k=0}^{\infty} \frac{(\lambda^2/2)^k}{k!} \tag{1.11}$$
$$= e^{\lambda^2/2} \tag{1.12}$$

So X is sub-Gaussian.

Proposition 1.3. *Let $X \in [a, b]$ be a bounded random variable. Then $\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2(a-b)^2}{8}}$ (Hoeffding Lemma). Hence,* any bounded RV is sub-Gaussian

Proof. Let Z be such that $\mathbb{E}Z = 0$, $\psi(\lambda) = \log \mathbb{E}e^{\lambda Z}$. By mean-value theorem, $\exists \theta \in (0, \lambda)$ such that

$$\psi(\lambda) = \psi(0) + \lambda\psi'(0) + \frac{\lambda^2}{2}\psi''(\theta) \tag{1.13}$$

is exact. But since Z is zero mean and $\psi(0) = 0$

$$\phi'(\lambda)|_{\lambda=0} = \frac{\mathbb{E}Ze^{\lambda Z}}{\mathbb{E}e^{\lambda Z}} = \frac{\cancelto{0}{\mathbb{E}Z}}{1} \tag{1.14}$$
$$\implies \psi(\lambda) = \cancelto{0}{\psi(0)} + \lambda\cancelto{0}{\psi'(0)} + \frac{\lambda^2}{2}\psi''(\theta) \tag{1.15}$$

$Z_\lambda \in [a, b]$ is a RV with density $f(x) = e^{-\psi(\lambda)}e^{\lambda x} = \frac{e^{\lambda x}}{\mathbb{E}e^{\lambda Z}}$ (wrt P).
$\psi''(\lambda) = \text{Var } Z$, because

$$\psi''(\lambda) = \frac{\mathbb{E}z^2 e^{\lambda z}}{\mathbb{E}e^{\lambda Z}} - \frac{(\mathbb{E}Ze^{\lambda Z})^2}{(\mathbb{E}e^{\lambda Z})^2} = \mathbb{E}Z_\lambda^2 - (\mathbb{E}Z_\lambda)^2 \tag{1.16}$$

Hence

$$\text{Var } Z_\lambda = \text{Var}\left(Z_\lambda - \frac{a+b}{2}\right) \leq \frac{(b-a)^2}{4} \tag{1.17}$$
$$\psi(\lambda) \leq \frac{\lambda^2}{2}\frac{(b-a)^2}{4} = \frac{\lambda^2(b-a)^2}{8} \tag{1.18}$$

We denote this as $X \sim SG\left(\frac{b-a}{2}\right)$ □

Proof. (This time using symmetrization) Let Z' be a copy of Z. $\mathbb{E}Z' = 0$ so

$$\mathbb{E}_Z e^{\lambda Z} = \mathbb{E}_Z e^{\lambda Z - \lambda \mathbb{E}_{Z'} Z'} \tag{1.19}$$

$$(\text{Jensen}) \implies \leq \mathbb{E}_Z \mathbb{E}_{Z'} e^{\lambda(Z-Z')} \tag{1.20}$$

Let $\varepsilon \sim \text{Rad}$, so $\text{dist}(Z - Z') = \text{dist}(\varepsilon(Z - Z'))$ by symmetry and

$$\mathbb{E}_Z \mathbb{E}_{Z'} e^{\lambda(Z-Z')} = \mathbb{E}_Z \mathbb{E}_{Z'} \mathbb{E}_\varepsilon e^{\lambda \varepsilon (Z-Z')} \leq \mathbb{E}_Z \mathbb{E}_{Z'} e^{\frac{\lambda^2 (Z-Z')}{2}} \tag{1.21}$$

where we have used the Rademacher MGF bound $\mathbb{E}e^{\lambda X} \leq e^{\lambda^2/2}$.

Since Z and Z' are bounded, $|Z - Z'| \leq b - a$ hence altogether

$$\mathbb{E}e^{\lambda Z} \leq \mathbb{E}_Z \mathbb{E}_{Z'} e^{\frac{\lambda^2 (Z-Z')^2}{2}} \leq e^{\frac{\lambda^2 (b-a)^2}{2}} \tag{1.22}$$

Therefore $Z \sim SG(b - a)$. □

Example 1.4. Let $X_i \overset{\text{i.i.d.}}{\sim} P$, $X_i \sim SG(\sigma)$, $Z = n^{-1} \sum_{i=0}^n X_i$. Then Z is sub-Gaussian, which we can show by multiplying the MGFs

$$\mathbb{E}e^{\lambda Z} = \mathbb{E}e^{\lambda n^{-1} \sum_{i=0}^n X_i} \tag{1.23}$$

$$= \mathbb{E} \prod_{i=0}^n e^{\lambda n^{-1} X_i} \tag{1.24}$$

$$X_i \text{ indep.} \implies = \prod_{i=0}^n \mathbb{E}e^{\lambda n^{-1} X_i} \tag{1.25}$$

$$\leq \prod_{i=0}^n e^{\frac{\lambda^2 \sigma^2}{2n^2}} \tag{1.26}$$

$$= e^{\frac{\lambda^2 \sigma^2}{2n}} \tag{1.27}$$

So $Z \sim SG(\frac{\sigma}{\sqrt{n}})$ and hence $P(|Z - \mu| \geq t) \leq 2e^{-\frac{nt^2}{2\sigma^2}}$.

This holds for sum of i.i.d. sub-Gaussians, so since binomials are sums of Bernoullis (which are bounded), we can conclude binomials are sub-Gaussian.

These concenration inequalities also give rise to the $n \sim \log \frac{1}{\varepsilon}$ seen in papers because

$$\varepsilon = P(|Z - \mu| \geq t) \leq 2e^{-\frac{nt^2}{2\sigma^2}} \tag{1.28}$$

$$\implies n \geq \frac{2\sigma^2}{t^2} \log \frac{2}{\varepsilon} \tag{1.29}$$

Proposition 1.5 (Equivalent characterizations of sub-Gaussian)**.** *The following are equivalent:*

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

(b) $\exists c$ *and* $Z \sim N(0, \tau^2)$ *such that* $P(|X| \geq s) \leq cP(|Z \geq s)$, $\forall s \geq 0$

(c) $\exists \theta \geq 0$ *such that* $\mathbb{E}X^{2k} \leq \frac{(2k)!}{2^k k!} \theta^{2k}$

(d) $\mathbb{E}e^{\frac{\lambda X^2}{2\sigma^2}} \leq \frac{1}{\sqrt{1-\lambda}}$ *where* $\lambda \in (0, 1)$ *(MGF of* χ^2*)*

2 Sub-Exponential RVs

All sub-Gaussians are sub-Exponentials.

χ^2 are not sub-Gaussian, but the left tail is Gaussian-like while the right-tail decays more slowly (like an exponential).

Useful facts:

- Gamma density: $X \sim \Gamma(\alpha, \beta)$ has density $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
  - A polynomial $x^{\alpha-1}$ (entropy term) trading off with an exponential $e^{-\beta x}$ (energy term)
- $\mathbb{E}X = \frac{\alpha}{\beta}$
- $\chi^2 \sim \Gamma(1/2, 1/2)$
- TODO: Calculate MGF of Gamma

Example 2.1. $Z \sim N(0,1)$, $X = Z^2$ is χ^2 distributed.

The MGF

$$\mathbb{E}e^{\lambda(X-1)} = \frac{1}{\sqrt{2\pi}} \int e^{\lambda(z^2-1)} e^{-\frac{z^2}{2}} dz = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \tag{2.1}$$

exists only for $0 \le \lambda < \frac{1}{2}$. This constraint on the support of the MGF is what defines sub-Exponential RVs.

Definition 2.2. A RV X with mean μ is *sub-Exponential* with parameters μ and α ($SE(\nu, \alpha)$) if $\exists$ non-negative (ν, α) such that

$$\mathbb{E}e^{\lambda(X-\mu)} \le e^{\frac{\nu^2\lambda^2}{2}}, \quad |\lambda| < \frac{1}{\alpha} \tag{2.2}$$

Remark 2.3. $SG(\sigma) \implies SE(\sigma, 0)$

χ^2 RVs are sub-Exponential:

$$\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \le e^{2\lambda^2} = e^{\frac{4\lambda^2}{2}}, \quad |\lambda| < \frac{1}{4} \tag{2.3}$$

So $\chi^2 \sim SE(2, 4)$. TODO: PROVE THIS

Next class: will consider

$$\inf_{\frac{1}{\alpha} > \lambda > 0} \left(\frac{\nu^2\lambda^2}{2} - \lambda t \right) \tag{2.4}$$

In general, sub-Exponentials begin with e^{-t^2} behaviour until a break-point and continue as e^{-t} after. TODO: Fig 3.1