36-755: Advanced Statistics Theory I **Fall 2017**

Lecture 3: September 11

Lecturer: Alessandro Rinaldo *Scribes: Yufei Yi*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Review

Definition 3.1 *(Sub-exponential) A random variable X with mean $\mu = \mathbb{E}[X]$ is sub-exponential if there are non-negative parameters (ν, α) such that*

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

Remark : 1). SE can be used to characterize variables with a thicker tail than SG.
2). SG $\subset$ SE.

3.1 Tail Bound of SE Variable

Theorem 3.2 *Suppose that X is sub-exponential with parameters (ν, α). Then*

$$\mathbb{P}[X \ge \mu + t] \le \begin{cases} e^{-\frac{t^2}{2\nu^2}} & \text{if } 0 \le t \le \frac{\nu^2}{\alpha} \\ e^{-\frac{t}{2\alpha}} & \text{if } t > \frac{\nu^2}{\alpha} \end{cases}$$

or we can write it as another form:

$$\mathbb{P}[X \ge \mu + t] \le \exp\{-\frac{1}{2}\min\{\frac{t}{\alpha}, \frac{t^2}{\nu^2}\}\}$$

Proof: WLOG, let $\mathbb{E}X = 0$.

$$\begin{aligned}\mathbb{P}(X \ge t) &\le exp\{-\lambda t + \frac{\lambda^2\nu^2}{2}\}, \quad \forall 0 \le \lambda < \frac{1}{\alpha} \qquad\qquad Markov \\ &= \exp\left\{\inf_{\lambda\in[0,\frac{1}{\alpha})}\{-\lambda t + \frac{\lambda^2\nu^2}{2}\}\right\}\end{aligned}$$

Let $g(t,\lambda) = -\lambda t + \frac{\lambda^2\nu^2}{2}$. The unconstrained minimum of $g(t,\lambda)$ is at $\lambda^\star = \frac{t}{\nu^2}$. So,

- If $\frac{t}{\nu^2} \le \frac{1}{\alpha} \Rightarrow \quad t \le \frac{\nu^2}{\alpha}$, $\quad \inf_{\lambda\in[0,\frac{1}{\alpha})} g(t,\lambda) = g(t,\lambda^\star) = -\frac{t^2}{2\nu^2}$.
- If $\frac{t}{\nu^2} > \frac{1}{\alpha} \Rightarrow \quad t > \frac{\nu^2}{\alpha}$, $\quad \inf_{\lambda\in[0,\frac{1}{\alpha})} g(t,\lambda) = g(t,\frac{1}{\alpha}) = -\frac{t}{\alpha} + \frac{\nu^2}{2\alpha^2} \le -\frac{t}{2\alpha}$. By $g(t,\lambda)$ is decreasing in $[0,\frac{1}{\alpha})$

■

3.2 A Sufficient Condition for SE

Definition 3.3 *(Bernstein's Condition) Given a random variable X with mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = \mathbb{E}[X^2] - \mu^2$. The Bernstein's condition with parameter b holds if*

$$|\mathbb{E}(X-\mu)^k| \leq \frac{k!}{2}\sigma^2 b^{k-2} \qquad \textit{for } k = 3, 4, ... \tag{3.1}$$

Remark If $|X - \mathbb{E}X| \leq b$ a.e., X satisfies BC with parameter b.

Besides explicitly computing or bounding the moment-generating function, Bernstein's ondition is another way to verify SE property.

Theorem 3.4 *(Sufficiency) For any random variable satisfying the Bernstein's condition (3.1), it is sub-exponential with parameters σ^2 and b. Explicitly:*

$$\mathbb{E}[e^{\lambda(X-\mathbb{E}X)}] \leq \exp\{\frac{\lambda^2\sigma^2}{2(1-b|\lambda|)}\} \qquad \forall|\lambda| < \frac{1}{b}$$

Proof:

$$\begin{aligned}
\mathbb{E}[e^{\lambda(X-\mathbb{E}X)}] &\leq 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty}\frac{|\lambda|^k\mathbb{E}|X-\mathbb{E}X|^k}{k!} && \text{Taylor expansion} \\
&\leq 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2}\sum_{k=3}^{\infty}(|\lambda|b)^{k-2} && \text{Bernstein's condition} \\
&\leq 1 + \frac{\lambda^2\sigma^2}{2}\frac{1}{1-|\lambda|b} && \text{for } |\lambda| < \frac{1}{b}\text{, then use the sum of geometric series} \\
&\leq \exp\{\frac{\lambda^2\sigma^2}{2(1-|\lambda|b)}\} && 1 + x \leq e^x
\end{aligned}$$

■

Remark The parameters of SE depend on σ and b, but we don't have a fixed pair of parameters here. $(\sqrt{2}\sigma, 2b)$ or $(\frac{\sqrt{3}}{2}\sigma, 3b)$ are both ok.

Theorem 3.5 *(Bernstein-type Bound) For any random variable satisfying the Bernstein's condition (3.1), we have the concentration inequality*

$$\mathbb{P}[|X-\mu| \geq t] \leq 2e^{-\frac{t^2}{2(\sigma^2+bt)}} \qquad \forall t > 0$$

Proof: Use Theorem 3.4, and set $\lambda = \frac{t}{bt+\sigma^2}$ in the Chernoff bound. Then simplify it. ■

3.3 Hoeffding's Inequality v.s. Bernstein's Inequality

Assume $|X| < b$ $a.e.$ and $Var[X] = \sigma^2$. WLOG assume $\mathbb{E}[X] = 0$. We can apply both Hoeffding-type bound and Bernstein-type bound to obtain

$$\mathbb{P}[|X| \geq t] \leq \begin{cases} 2\exp\{-\frac{t^2}{2b^2}\} & \text{Hoeffding} \\ 2\exp\{-\frac{t^2}{2(\sigma^2+bt)}\} & \text{Bernstein} \end{cases}$$

Bernstein is always sharper than Hoeffding because $\sigma^2 = \mathbb{E}[X^2] \leq b^2$. It is substantially better if $\sigma^2 \ll b^2$. Notice that Hoeffding always assumes the worst variance by using the range to bound the 2nd moment. However, if we don't know any information about the variance, we might only be able to use Hoeffding.

Remark Even if sharper than Hoeffding, Bernstein is not the sharpest for a bounded random variable. Bennett's inequality can be used to provide sharper control on the tails. (See reference)

3.4 Composition Property of SE Variables

Lemma 3.6 *If $X_i \in SE(\nu_i, \alpha_i)$, then*

$$\sum_{i=1}^{n}[X_i - \mathbb{E}X_i] \sim \begin{cases} SE[\sum_{i=1}^{n} \nu_i^2, \max_{i=1,\ldots,n} \alpha_i] & \text{if } X_i s \text{ independent} \\ SE[(\sum_{i=1}^{n} \nu_i)^2, \max_{i=1,\ldots,n} \alpha_i] & \text{if } X_i s \text{ not independent} \end{cases}$$

Proof: Check the definition of SE on $\sum_{i=1}^{n}[X_i - \mathbb{E}X_i]$. ■

Lemma 3.7 *Assume X_is are independent, sub-exponential with parameters (ν_i, α_i), then we have this sub-exponential tail bound:*

$$\mathbb{P}\left[\left|\frac{\sum_{i=1}^{n}[X_i - \mathbb{E}X_i]}{n}\right| \geq t\right] \leq \begin{cases} 2e^{-\frac{nt^2}{2(\nu_\star^2/n)}} & \text{if } 0 \leq nt \leq \frac{\nu_\star^2}{\alpha_\star} \\ 2e^{-\frac{nt}{2\alpha_\star}} & \text{if } nt \geq \frac{\nu_\star^2}{\alpha_\star} \end{cases}$$

where $\nu_\star = \sqrt{\sum_{i=1}^{n} \nu_i^2}$, $\alpha_\star = \max_{i=1,\ldots,n} \alpha_i$.

Example 3.8 *By using the result in Lemma 3.7, for i.i.d $Z_i \sim \mathcal{N}(0,1)$, we can obtain the tail bound for a chi-square variable:*

$$\mathbb{P}\Big[\frac{1}{n}\Big|\sum_{i=1}^{n} Z_i^2 - n\Big| \geq t\Big] \leq e^{-\frac{nt^2}{8}}, \quad \forall t \in (0,1)$$

When we need to bound a chi-square variable, we usually use the inequality in **Corollary 3.10** because it's sharper than the inequality in **Example 3.8**. It's derived from **Lemma 3.9**.

Lemma 3.9 *(Laurent-Massart) [LM2000] Let $a_1, ..., a_n$ be nonnegative, and set*

$$|a|_\infty = \sup_{i=1,\ldots,n} |a_i|, \qquad |a|_2^2 = \sum_{i=1}^{n} a_i^2$$

For i.i.d $Z_i \sim \mathcal{N}(0,1)$, let

$$X = \sum_{i=1}^{n} a_i(Z_i^2 - 1)$$

Then the following inequalities hold for any positive t:

$$\mathbb{P}(X \geq 2|a|_2^2\sqrt{t} + 2|a|_\infty t) \leq e^{-t}$$
$$\mathbb{P}(X \leq -2|a|_2\sqrt{t}) \leq e^{-t}$$

Corollary 3.10 *(A Sharper Bound for a Chi-square Variable) Let X be chi-square with n degrees of greedom. For any positive t,*

$$\mathbb{P}(X - n \geq 2\sqrt{nt} + 2t) \leq e^{-t}$$
$$\mathbb{P}(X - n \leq -2\sqrt{nt}) \leq e^{-t}$$

3.5 Application : Maxima

Theorem 3.11 *$X_1, ..., X_n$ are independent, $\mathbb{E}X_i = 0$ and $\log \mathbb{E}[e^{\lambda X_i}] \leq \psi_{X_i}(\lambda)$ for all $i \in [1, n]$. ψ_{X_i} is convex and $|\lambda| < \frac{1}{b}, b < \infty$. Then we have:*

$$\mathbb{E}[\max X_i] \leq \inf_{\lambda \in [0,1/b)} \frac{\log n + \psi_{X_i}(\lambda)}{\lambda} \tag{3.2}$$

Proof: Arbitrarily choose $\lambda \in [0, 1/b)$.

$$\begin{aligned} \exp\{\lambda \mathbb{E}[\max X_i]\} &\leq \mathbb{E}[e^{\lambda \max X_i}] && \textit{Jensen} \\ &\leq \sum_{i=1}^{n} \mathbb{E}[e^{\lambda X_i}] \\ &\leq n e^{\psi_{X_i}(\lambda)} \end{aligned}$$

$\Rightarrow \mathbb{E}[\max X_i] \leq \frac{\log n + \psi_{X_i}(\lambda)}{\lambda}$ for all $\lambda \in [0, 1/b) \Rightarrow \mathbb{E}[\max X_i] \leq \inf_{\lambda \in [0,1/b)} \frac{\log n + \psi_{X_i}(\lambda)}{\lambda}$ ■

Corollary 3.12 *We can apply the inequality (3.2) to SG and SE variable and get:*

$$\mathbb{E}[\max X_i] \leq \sqrt{2\sigma^2 \log n} \qquad \text{if } X_i s \in SG(\sigma^2)$$
$$\mathbb{E}[\max X_i] \leq \sqrt{2\nu^2 \log n} + b \log n \qquad \text{if } X_i s \in SE(\nu^2, b)$$

Proof:

- For $X_i s \in SG(\sigma^2)$, set $\psi_{X_i}(\lambda) = \frac{\lambda^2 \sigma^2}{2}$ and compute the inf directly.
- For $X_i s \in SE(\nu^2, b)$, use the facts:

$$\inf_{\lambda \in [0,1/b)} \frac{\log n + \psi_{X_i}(\lambda)}{\lambda} = \psi_{X_i}^{\star -1}(\log n) \quad \text{and}$$
$$\psi_{X_i}^{\star -1}(x) = \sqrt{2\nu^2 x} + bx \quad \text{if } X_i \in SE(\nu^2, b) \ [BLM2013]$$

where

$$\psi_{X_i}^{\star}(x) = \sup_{\lambda \in [0, \frac{1}{b})} \{\lambda x - \psi_{X_i}(\lambda)\} \quad \text{is the Cramr transformation } [PM2003]$$
$$\psi_{X_i}^{\star -1}(x) = \inf\{t \geq 0 : \psi_{X_i}^{\star}(t) > x\} \quad \text{is the generalized inverse of } \psi_{X_i}^{\star}(x)$$

■

Reference

[LM2000] B.Laurent and P.Massart "Adaptive estimation of a quadratic functional by model selectionration Inequalities and Model Selection," *Annals of Statistics, 28(5), 1302-1338.*

[PM2003] P. Massart, "Concentration Inequalities and Model Selection," *Vol.6. Berlin: Springer, 2007. pp.16*

[BLM2013] S. Boucheron, G. Lugosi and P. Massart, "Concentration inequalities: A nonasymptotic theory of independence," *Oxford university press, 2013. pp.28*