36-755: Advanced Statistical Theory 1 **Fall 2016**

Lecture 1: August 31

Lecturer: Alessandro Rinaldo *Scribes: Jaehyeok Shin*

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

1.1 Tail bounds and Concentration inequality

Let $X_1, \dots, X_n \overset{iid}{\sim} (\mu, \sigma^2)$. We know

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n \xrightarrow{p} \mu \Longleftrightarrow \bar{X}_n = \mu + o_p(1)$$

$$\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) \Longrightarrow Z \sim N(0,1) \Longleftrightarrow \bar{X}_n = \mu + O_p\left(\frac{1}{\sqrt{n}}\right)$$

By the central limit theorem,

$$\lim_{n \leftarrow \infty} \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) > t\right) = \mathbb{P}(Z > t) \leq \frac{1}{2}e^{-t^2/2} \quad \text{for } t > 0,$$

If n is large enough,

$$\lim_{n \leftarrow \infty} \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) > t\right) \leq C_1 \exp\left\{-t^2 C_2\right\}, \quad C_1, C_2 > 0$$

In fact, this is true for a large class of random variables for all finite n!

1.2 Concentration phenomenon

Let P be uniform distribution over $B_d = \left\{x \in \mathbb{R}^d : \|x\| \leq 1\right\}$. Then, if $X \sim P, \epsilon > 0$,

$$\mathbb{P}\left(\|X\| < 1 - \epsilon\right) = \frac{v_d(1-\epsilon)^d}{v_d} \leq e^{-\epsilon d} \longrightarrow 0 \ \text{ as } \ d \longrightarrow \infty$$

where $v_d := Vol(B_d)$. Similar phenomenon happens for the Normal distribution.

Back to concentration

Let $(X_1, \dots, X_n)$ be independent. Let $Z = f(X_1, \dots, X_n)$. If f does not depend too much on its individual coordinates, Z concentrates well around its mean $\mathbb{E}[[Z]$. We will focus

$$f(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

1.3 Markov Inequality

If $X \geq 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}, \quad t > 0$$
$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\sigma^2}{t^2}, \quad \sigma^2 = \mathbb{V}[X]$$

To bound $\mathbb{P}(|X-\mu| \geq t)$, $\mu = \mathbb{E}[X]$, we could observe

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\mathbb{E}[X-\mu|^k]}{t^k}, \quad k = 1, 2, \dots$$
$$\implies \mathbb{P}(|X - \mu| \geq t) \leq \min_{k=1,2,\dots} \frac{\mathbb{E}[X-\mu|^k]}{t^k}$$

This is a good bound but we need to know all moments of X which requires strong and unrealistic assumptions on X.

1.4 Chernoff Bound

Let $\psi_X(\lambda) := \log\left(\mathbb{E}[e^{\lambda(X-\mu)}]\right)$. Assume $\psi_X(\lambda)$ exists for $\forall\lambda \in [0, b)$ $, 0 < b \leq \infty$. Then, for $t > 0$, $0 \leq \lambda < b$,

$$\begin{aligned}
\mathbb{P}(X - \mu \geq t) &\leq \mathbb{P}\left(e^{\lambda(x-\mu)} \geq e^{\lambda t}\right) \\
&\leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda(x-\mu)}\right], \quad \text{(by the Markov inequality)} \\
&= \exp\{\psi_X(\lambda) - \lambda t\}
\end{aligned}$$

which implies $\mathbb{P}(X - \mu \geq t) \leq \exp\{-\psi_X^*(t)\}$ where $\psi_X^*(t) = \sup_{\lambda\in[0,b)} \{\lambda t - \psi_X(\lambda)\}$

Example

Let $X \sim N(\mu, \sigma^2)$. We know $\mathbb{E}[e^{\lambda X}] = \exp\left\{\mu\lambda + \frac{\sigma^2\lambda^2}{2}\right\}, \quad \forall\lambda \in \mathbb{R}$. So,

$$\sup_{\lambda\geq 0}\left\{\lambda t - \log \mathbb{E}\left[e^{\lambda(X-\mu)}\right]\right\} = \sup_{\lambda\geq 0}\left\{\lambda t - \frac{\sigma^2\lambda^2}{2}\right\} = \frac{t^2}{2\sigma^2}$$

By using Chernoff, $t > 0$,

$$\mathbb{P}\left(|X - \mu| \geq t\right) \leq 2\exp\left\{-\frac{t^2}{2\sigma^2}\right\}$$

How good is this bound?

$$\sup_{t\geq 0}\left\{\mathbb{P}(Z \geq t)\exp\left\{\frac{t^2}{2}\right\}\right\} = \frac{1}{2}$$

In Normal case, Chernoff method gave a bound

$$C_1 \exp\left\{-t^2 C_2\right\}, \quad C_1, C_2 > 0 \quad \text{(Gaussian-like tail behavior)}$$

1.5 Sub-Gaussian Random Variables

Definition 1.1 *A random variable X with finite $\mu = \mathbb{E}[X]$ is said to be sub-gaussian with parameter σ^2, $X \in SG(\sigma^2)$, $\sigma > 0$ if*

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \le \exp\left\{\frac{\lambda^2\sigma^2}{2}\right\}, \quad \forall \lambda \in \mathbb{R}$$

Note that if X is sub-gaussian, $-X$ is sub-gaussian.

Result

If $X \in SG(\sigma^2)$

$$\mathbb{P}\left(|X-\mu| \ge t\right) \le 2\exp\left\{-\frac{t^2}{2\sigma^2}\right\}, \quad \forall t > 0$$

Properties of $SG(\sigma^2)$

1. $X \in SG(\sigma^2) \Longrightarrow \mathbb{V}[X] \le \sigma^2$. (by Taylor expansion of $\mathbb{E}\left[e^{\lambda(X-\mu)}\right]$)

2. $a \le X - \mu \le b$ a.s. $\Longrightarrow X \in SG\left(\left(\frac{b-a}{2}\right)^2\right)$

   Proof: Without loss of generality, assume $\mu = 0$. We need to show

   $$\psi_X(\lambda) \le \frac{(b-a)^2\lambda^2}{8}$$

   First, note that $\mathbb{V}[X] \le \left(\frac{b-a}{2}\right)^2$ since $\left|X - \frac{a+b}{2}\right| \le \frac{b-a}{2}$ a.s. For any random variable X such that $a \le X \le b$ a.s., let Z be a random variable such that $\frac{dP_Z}{dP_X}(z) = e^{\lambda z}e^{-\psi_X(\lambda)}$. Then $a \le Z \le b$ a.s. and $\mathbb{V}[Z] = \psi_X''(\lambda)$. So $\psi_X''(\lambda) \le \left(\frac{b-a}{2}\right)^2$. Now,

   $$\begin{aligned}
   \psi_X(0) &= \log 1 = 0 \\
   \psi_X'(0) &= \mathbb{E}[X] = 0 \\
   \psi_X(\lambda) &= \int_0^\lambda \psi_X'(\lambda')d\lambda' = \int_0^\lambda \int_0^{\lambda'} \psi_X''(\lambda'')d\lambda''d\lambda' \\
   &\le \frac{\lambda^2}{2}\frac{(b-a)^2}{4} = \frac{\lambda^2(b-a)^2}{8}
   \end{aligned}$$

   ■

3. $X \in SG(\sigma^2) \Longrightarrow \alpha X \in SG(\alpha^2\sigma^2), \quad \alpha \in \mathbb{R}$

4. $X \in SG(\sigma^2)$ and $Y \in SG(\tau^2) \Longrightarrow X + Y \in SG\left((\sigma+\tau)^2\right)$ $(\sigma, \tau > 0)$.
   If $X \perp\!\!\!\perp Y$, $X + Y \in SG(\sigma^2 + \tau^2)$

   Proof: of the first case

Without loss of generality, assume $\mathbb{E}[X] = \mathbb{E}[X] = 0$

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X+Y)}\right] &= \mathbb{E}e^{\lambda X}e^{\lambda Y} \\
&\leq_{H\ddot{o}lder} \left(\mathbb{E}e^{\lambda pX}\right)^{1/p}\left(\mathbb{E}e^{\lambda qY}\right)^{1/q}, \quad \text{where } \frac{1}{p}+\frac{1}{q}=1 \\
&\leq \exp\left\{\frac{\lambda^2\sigma^2p^2}{2}\frac{1}{p}+\frac{\lambda^2\tau^2q^2}{2}\frac{1}{q}\right\} \\
&= \exp\left\{\frac{\lambda^2}{2}(p\sigma^2+q\tau^2)\right\} \\
&= \exp\left\{\frac{\lambda^2}{2}(\sigma+\tau)^2\right\} \quad \text{by setting } \ p=\frac{\tau}{\sigma}+1
\end{aligned}$$

∎

1.6 Hoeffding Inequality

Let $X_1, \ldots, X_n$ be independent and $X_i \in SG(\sigma_i^2), \quad i = 1, \ldots, n$ Then,

$$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mu_i) \geq t\right) \leq \exp\left\{-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2}\right\}, \quad t > 0$$

where $\mu_i = \mathbb{E}[X_i], \quad i = 1, \ldots, n$ In particular, if $\sigma_i = \sigma, \quad \forall i$

$$\mathbb{P}\left(|\bar{X}_n - \bar{\mu}| \geq t\right) \leq 2\exp\left\{-\frac{t^2 n}{2\sigma^2}\right\}$$

where $\bar{\mu} = \frac{1}{n}\sum_{i=1}^n \mu_i$

Example $X_i \sim Bernoulli(p_i), \quad i = 1, \ldots, n$

By Hoeffding,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n p_i\right| \geq t\right) \leq 2\exp\{-2t^2 n\}$$

$\implies$ With probability at least $1 - \delta, \quad \delta \in (0, 1)$

$$\left|\bar{X}_n - \bar{p}_n\right| \leq \sqrt{\frac{1}{2n}\log\frac{2}{\delta}} \qquad (\text{set } 2\exp\{-2t^2n\} = \delta \text{ and solve it for } t)$$

Remark

- Sharper results can be obtained by using Chernoff method.
- In this case, there are multiplicative bounds.

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq (1+\epsilon)\mu\right) \leq e^{-\epsilon^2\mu/3}$$
$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq (1-\epsilon)\mu\right) \leq e^{-\epsilon^2\mu/2}$$

where $\mu = \sum_{i=1}^n p_i, \quad \epsilon \in (0, 1)$. They are useful if $p_i \longrightarrow 0$

Hoeffding vs Multiplicative bounds ($p_i = p, \quad \forall i = 1, \dots, n$)

$$\mathbb{P}\left(p - \bar{X}_n \geq t\right) \leq e^{-2t^2 n} \iff \mathbb{P}\left(p - \bar{X}_n \geq \sqrt{\frac{1}{2n} \log \frac{1}{\delta}}\right) \leq \delta \qquad \text{(Hoeffding)}$$

$$\mathbb{P}\left(p - \bar{X}_n \geq \epsilon p\right) \leq e^{-n\epsilon^2/2} \iff \mathbb{P}\left(p - \bar{X}_n \geq \sqrt{\frac{2p}{n} \log \frac{1}{\delta}}\right) \leq \delta \qquad \text{(Multiplicative)}$$

Multiplicative bound is better if $p \leq \frac{1}{4}$, and much better if $p_i \longrightarrow 0$ as $n \longrightarrow \infty$