36-788: Topics in High Dimensional Statistics I **Fall 2015**

Lecture 4: Sept. 10

Lecturer: Alessandro Rinaldo *Scribes: Lingxue Zhu*

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

4.1 More on Hoeffding's Inequality

Last time, we went through Hoeffding's inequality:

Theorem 4.1 (Hoeffding) *Let $X_1, ..., X_n$ be independent r.v.'s such that $\mathbb{E}X_i = 0$, $a_i \le X_i \le b_i$ a.s.. Let $S_n = \sum_{i=1}^n X_i$, then for any $t > 0$,*

$$\mathbb{P}(|S_n| > t) \le 2 \exp\left\{ -\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}$$

Hoeffding's inequality is not always optimal. The proof uses the fact that

$$\Psi_{X_i - \mathbb{E}X_i}(\lambda) \le \frac{\lambda^2 (b_i - a_i)^2}{8} ,$$

which depends on the extreme values of X_i and thus accounts for the "worst" case. The following theorem sharpens the bound by taking into account where $\mathbb{E}X_i$ locates.

Theorem 4.2 (Berend-Kontorovich) *[BK13] Under the same assumptions as in Theorem 4.1, let*

$$\gamma_i = \frac{\mathbb{E}X_i - a_i}{b_i - a_i} ,$$

we have

$$\mathbb{P}(|S_n| > t) \le 2 \exp\left\{ -\frac{t^2}{4 \sum_{i=1}^n c_i (b_i - a_i)^2} \right\}$$

where

$$c_i = \begin{cases} 0, & \gamma_i = 0 \\ \frac{1 - 2\gamma_i}{4 \log\left(\frac{1-\gamma_i}{\gamma_i}\right)}, & 0 < \gamma_i < \frac{1}{2} \\ \frac{\gamma_i(1-\gamma_i)}{2}, & \frac{1}{2} \le \gamma_i \le 1 \end{cases}$$

Especially, when $\gamma_i = \frac{1}{2}$ (that is, $\mathbb{E}X_i = \frac{a_i + b_i}{2}$), it recovers Hoeffding's inequality. When $X_i \sim$ Bernoulli(p), the bound becomes $2\exp\left\{-\frac{t^2}{2np(1-p)}\right\}$, which is much better than Hoeffiding's when p is close to 0 or 1.

Example 1 (Rademacher distribution) *Let* $X_i = \begin{cases} 1, & \text{with probability } \frac{1}{2} \\ -1, & \text{with probability } \frac{1}{2} \end{cases}$, *and* $X = \sum_{i=1}^n \alpha_i X_i$, *then* $V(X) = ||\alpha||^2$, *and by Hoeffding's inequality,*

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

This is exactly the Chernoff bound we would get if $X \sim N(0, ||\alpha||^2)$. *Hoeffding's is optimal in this case.*

For bounded r.v. $a_i \leq X_i \leq b_i$, we have $V(X_i) \leq \frac{(b_i - a_i)^2}{4}$. We have seen that when $V(X_i)$ achieves this upper bound, Hoeffding's is optimal. But when $V(X_i)$ is much smaller, we will see that Bernstein's inequality becomes better. Before coming to that, we first introduce another class of random variables that is more general than sub-gaussian.

4.2 Sub-Exponential Random Variables

Definition 4.3 *A r.v. X is sub-exponential with parameters* $\nu, c > 0$, *if*

$$\Psi_{X-\mathbb{E}X}(\lambda) \leq \frac{\lambda^2 \nu^2}{2} \text{ for } \forall\ |\lambda| \in [0, \frac{1}{\mathrm{c}})$$

and we write $X \in S(\nu, c)$.

By definition, sub-gaussian r.v.'s are always sub-exponential.

Theorem 4.4 *If* $X \in S(\nu, c)$ *with* $\mathbb{E}X = 0$, *then for* $\forall\ \lambda \in [0, \frac{1}{c})$,

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

Proof: We know that

$$\mathbb{P}(X \geq t) \leq \exp\{-\lambda t + \Psi_X(\lambda)\} \leq \exp\left\{-\lambda t + \frac{\lambda^2 \nu^2}{2}\right\}$$

We want to minimize RHS over $\lambda \in [0, \frac{1}{c})$. Without this constraint, the minima is $\lambda^* = \frac{t}{\nu^2}$.

(i) If $\lambda^* \leq \frac{1}{c}$, i.e., $t \leq \frac{\nu^2}{c}$, then $\min_{\lambda \in [0, 1/c)} \text{RHS} = \mathrm{e}^{-\frac{t^2}{2\nu^2}}$.

(ii) If $\lambda^* > \frac{1}{c}$, i.e., $t < \frac{\nu^2}{c}$, then RHS is monotonically decreasing when $\lambda \in [0, \frac{1}{c})$, so $\min_{\lambda \in [0, 1/c)} \text{RHS} = \exp\left\{-\frac{\mathrm{t}}{\mathrm{c}} + \frac{\nu^2}{2\mathrm{c}^2}\right\} \leq \exp\left\{-\frac{\mathrm{t}}{2\mathrm{c}}\right\}$.

■

Example 2 (Chi-square distribution) *Let* $X \sim \mathcal{X}_1$, *then* $X \overset{d}{=} Z^2$ *where* $Z \sim N(0, 1)$, *and* $\mathbb{E}X = 1$. *We* *can show that* $X \in S(2,4)$*:*

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X-1)}\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{\lambda(z^2-1)}e^{-z^2/2}dz \\
&= \frac{e^{-\lambda}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp\left\{-\frac{z^2}{2}(1-2\lambda)\right\}dz \\
&= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}dy \quad \left(\text{Let y} = \sqrt{1-2\lambda}\text{z}, \text{ for } \lambda \in \left(0, \frac{1}{2}\right)\right) \\
&= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \le e^{2\lambda^2}, \text{ if} \lambda \in \left(0, \frac{1}{4}\right).
\end{aligned}$$

Properties of $S(\nu, c)$**:**

1. If $V(X) = \nu^2$ and $|X - \mathbb{E}X| \le c$ *a.e.*, then $X \in S(\sqrt{2}\nu, 2c)$.

   Proof:

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X-\mathbb{E}X)}\right] &= 1 + \frac{\lambda^2\nu^2}{2} + \sum_{n=3}^{\infty}\frac{\lambda^n}{n!}\mathbb{E}\left[(X-\mathbb{E}X)^n\right] \\
&\le 1 + \frac{\lambda^2\nu^2}{2} + \frac{\lambda^2\nu^2}{2}\sum_{n=3}^{\infty}(\lambda c)^{n-2} \quad \left(\text{ since } \mathbb{E}\left[(X-\mathbb{E}X)^n\right] \le \mathbb{E}\left[(X-\mathbb{E}X)^2 c^{n-2}\right]\right) \\
&= 1 + \frac{\lambda^2\nu^2}{2}\sum_{n=0}^{\infty}(\lambda c)^n \\
&= 1 + \frac{\lambda^2\nu^2}{2}\frac{1}{1-\lambda c} \quad \left(\text{if } \lambda < \frac{1}{c}\right) \\
&\le \exp\left\{\frac{\lambda^2\nu^2}{2(1-\lambda c)}\right\} \\
&\le \exp\{\lambda^2\nu^2\} \quad \left(\text{if } \lambda < \frac{1}{2c}\right)
\end{aligned}$$

■

2. If $X_i \in S(\nu_i, c_i)$ independently, $\mathbb{E}X_i = 0$, then $\sum_i X_i \in S\left(\sqrt{\sum_i \nu_i^2},\ \max_i c_i\right)$.

   Proof:

$$\Psi_{\sum_i X_i}(\lambda) = \sum_i \Psi_{X_i}(\lambda) \le \frac{\lambda^2\sum_i \nu_i^2}{2}, \text{ if } |\lambda| < \min_{\text{i}}\frac{1}{\text{c}_\text{i}}$$

■

   This implies that

$$\mathbb{P}\left(\frac{1}{n}\sum_i X_i \ge t\right) \le \begin{cases} e^{-\frac{nt^2}{2\nu^2}}, & 0 < t \le \frac{\nu^2}{c} \\ e^{-\frac{nt}{2c}}, & \frac{\nu^2}{c} < t \end{cases} \le \exp\left\{-n\cdot\min\left\{\frac{t^2}{2\nu^2}, \frac{t}{2\nu}\right\}\right\}$$

   where $\nu^2 = \frac{1}{n}\sum_i \nu_i^2$, $c = \max_i c_i$.

3. If $X \in S(\nu, c)$, then $\left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} \le cp$ for $p \ge 1$.

4.3 Bernstein's Inequality

Theorem 4.5 (Easy Bernstein) *Let X be such that $V(X) = \nu^2$, $|X - \mathbb{E}X| \leq c$ a.e., then*

$$\mathbb{P}(X - \mathbb{E}X \geq t) \leq \exp\left\{-\frac{t^2}{2(\nu^2 + ct)}\right\}$$

Note that when $\nu^2 \gg t$, the bound behaves like $\exp\left\{-\frac{t^2}{2\nu^2}\right\}$; when $\nu^2 \ll t$, it behaves like $\exp\left\{-\frac{t}{2c}\right\}$.

Proof: We saw in previous section that $X \in S(\sqrt{2}\nu, 2c)$, and

$$\mathbb{E}\left[e^{\lambda(X - \mathbb{E}X)}\right] \leq \exp\left\{\frac{\lambda^2\nu^2}{2(1-\lambda c)}\right\}, \quad \lambda \in [0, \frac{1}{c})$$

Picking $\lambda = \frac{t}{ct+\nu^2} < \frac{1}{c}$ gives the desired result. ■

Theorem 4.6 (Bernstein) *Let $X_1, ..., X_n$ be independent r.v.'s such that for $\nu, c > 0$, $q = 3, 4, ...$,*

(1) $\sum_{i=1}^n \mathbb{E}(X_i^2) \leq \nu^2$

(2) $\sum_{i=1}^n \mathbb{E}\left(|X_i|^q\right) \leq \frac{q!}{2}\nu^2 c^{q-2}$ *(in fact , it is enough to assume* $\sum_{i=1}^n \mathbb{E}\left((X_i)_+^q\right) \leq \frac{q!}{2}\nu^2 c^{q-2}$*)*

Then $\forall t > 0$,

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left\{-\frac{\nu^2}{c^2}h_1\left(\frac{ct}{\nu^2}\right)\right\} \leq \exp\left\{-\frac{t^2}{2\nu^2 + ct}\right\}$$

where $h_1(u) = 1 + u + \sqrt{1 + 2u}$, $u > 0$.

Remarks

- As a result, we have

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq \sqrt{2\nu^2 t} + ct\right) \leq e^{-t}$$

- If we add additional assumptions such that $|X_i - \mathbb{E}X_i| \leq B$ *a.e.*, $\sum_{i=1}^n \mathbb{E}(X_i^2) = \nu^2$, then

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left\{-\frac{t^2}{2\nu^2 + \frac{B}{3}t}\right\}$$

We defer the proof to next lecture.

References

[BK13] D. Berend and A. Kontorovich, "On the concentration of the missing mass", *Electron. Commun. Probab.*, 2013, 18 (3), 1-7.