36-755: Advanced Statistics **Fall 2016**

Lecture 3: September 7

Lecturer: Alessandro Rinaldo *Scribes: Kayla Frisoli*

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

3.1 Recap

Sub-gaussian random variables are a large class of variables that haev good concentration properties. They have tail behavior like gaussian random variables and they concentrate well around their means.

We proved **hoeffding's inequality**, which will be (and should be) used often with sub-gaussian random variables, and we note that this bound is not always the sharpest. **Chernoff** and multiplicative bounds can be much better if p $\to 0$.

These results are often useful to get confidence intervals for the mean. Last time we saw that we can bound the distance from the mean using Bernoulli variables. Since we only know the tail behavior, and don't know the actual distribution, we consider these variables distribution free.

Exercise:

$x_1 \cdots x_n \overset{iid}{\sim} \text{Bern}(p)$

$$\mathbb{P}\left(|\overline{x_n} - p| \geq \sqrt{\frac{1}{2}\ln\left(\frac{1}{\delta}\right)}\right) \leq \delta$$

3.2 Sub-gaussianity

In general, if $x \in SG(\sigma^2)$, we expect

$$\mathbb{P}\left(\overline{x_n} - \mu = \mathcal{O}\left(\sqrt{\frac{\ln n}{n}}\right)\right) \geq 1 - \frac{1}{n^c} \qquad \text{for any constant, c}$$

3.2.1 Equivallent characterizations of being $SG(\sigma^2)$

1. $\mathbb{E}[e^{\lambda x}] \leq e^{\frac{\lambda^2\sigma^2}{2}}$ $\qquad \lambda \in \mathbb{R},\ \mathbb{E}[x] = 0$
2. $\exists c > 0$ s.t. $\mathbb{P}(|x| \geq t) \leq \mathbb{P}(|z| \geq t)$ $\qquad t > 0,$

3. $\exists \theta > 0$ s.t. $[x^{2k}] \leq \frac{(2k)!}{2^k k!}\theta^{2k}$

4. if $p \geq 1$ $\mathbb{E}[|x|^p] \leq (e\sigma^2)^{\frac{p}{2}} \cdot p \cdot \Gamma(\frac{p}{2})$ where $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$
$a > 0$
$\Gamma(p) = (p-1)!$ if $p \in \mathbb{N}$

Proof:
Assume sub-gaussianity (item 1)

$$
\begin{aligned}
\mathbb{E}[x] &= \int_0^\infty \mathbb{P}\left(x \geq t\right) dt \\
\mathbb{E}[|x|^p] &= \int_0^\infty \mathbb{P}\left(|x|^p \geq t\right) dt \\
&= \int_0^\infty \mathbb{P}\left(|x| \geq t^{\frac{1}{p}}\right) dt \\
&\text{now let's use the fact that we have sub-gaussianity to bound this} \\
&\leq 2\int_0^\infty e^{\frac{-t^{2/p}}{2\sigma^2}} dt \\
&\text{call the inside u, use chain rule} \\
&= (2\sigma^2)^{p/2} \int_0^\infty e^{-\mu} \cdot \mu^{p/2-1} d\mu \qquad \text{making substitution } \mu = \frac{t^{2/p}}{2\sigma^2} \\
&= (2\sigma^2)^{p/2} \cdot p \cdot \Gamma(p/2)
\end{aligned}
$$

■

Notice that if $X \sim N(0, \sigma^2)$ then $\mathbb{E}[|x|^p] \leq \sqrt{\frac{2p}{\pi}} \cdot \sigma^p \cdot \Gamma(\frac{p+1}{2})$

3.2.2 Back to Hoeffding

$$
\begin{aligned}
a \leq X \leq b \quad &\text{a.e. and } \mathbb{E}[x] = 0 \\
V[x] \leq \frac{(b-a)^2}{4} &\text{(not dependent on distribution)}
\end{aligned}
$$

Hoeffding assumes the worst case scenario and achieves the bound by:

$$
\varepsilon = \begin{cases} -1 & p = \frac{1}{2} \\ 1 & p = \frac{1}{2} \end{cases}
$$

In fact, if $\varepsilon_1 \cdots \varepsilon_n \overset{iid}{\sim}$ Rademacher and $X = \sum_{i=1}^n \alpha_i \varepsilon_i$ is a linear combination of the random variables, then

$$\mathbb{P}\left(|x| \geq t\right) \leq 2exp\left(\frac{-t^2}{2\|\alpha\|^2}\right) \qquad \sum \alpha_i^2 = V[x] = \sigma^2$$

$$= 2exp\left(\frac{-t^2}{2\sigma^2}\right) \qquad \text{same bound as we'd expect if X were normal}$$

specifically, same bound (up to a constant) as if $X \sim N(0, \sigma^2)$

Finding: Hoeffding is sharp if $V[x]$ is maximal, given $a \leq X \leq b$. But, if $V[x] \ll \frac{(b-a)^2}{4}$ than we can do much better than Hoeffding. With Hoeffding, we assume the worst case scenario of variance.

3.3 Sub-Exponential Variables (a larger class than SG)

We can make claims about variables that have thicker tails than gaussian. These variables are called sub-exponential.

For large t, the gaussian tail is of order e^{-t^2}. For laplace(1), the tail behavior is different. Specifically, the tail behavor is $\mathbb{P}(|X| \geq t) \leq e^{-t}$

Definition: A random variable $X \in SE(\nu, \alpha)$ if $\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\lambda^2 \nu^2}{2}}$ where $\nu > 0, \alpha \geq 0, |\lambda| < \frac{1}{\alpha}, \mu = \mathbb{E}[X]$

Sub-exponential variables only differ from sub-gaussian variables in the range of λ. If $X \in SG(\sigma^2)$ then $X \in SE(\sigma, 0)$ where $\frac{1}{0} = \infty$. This means that the moment generating function of sub-exponential variables may not be defined everywhere. As long as it holds for $|\lambda| < \frac{1}{\alpha}$ we're good.

Exercise:

$Z \sim N(0,1)$, $X = Z^2 \sim \chi_1^2$

Take $\lambda < \frac{1}{2}$, $\mathbb{E}[X] = \mu = 1$

$$\mathbb{E}[e^{\lambda(X-1)}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda(Z^2-1) - \frac{Z^2}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\lambda} \int_{-\infty}^{\infty} e^{\frac{Z^2}{2}(1-2\lambda)} dz$$

$$\text{setting } Y = \sqrt{1-2\lambda} \text{ and } dz = \frac{dy}{\sqrt{1-2y}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\lambda} \int_{-\infty}^{\infty} \frac{e^{\frac{-Y^2}{2}}}{\sqrt{1-2\lambda}} dy$$

$$= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-Y^2}{2}} dy \right)$$

$$= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \qquad \text{because the integral of the standard normal pdf is 1}$$

$$\leq e^{\frac{4\lambda^2}{2}} \qquad \text{if } |\lambda| < \frac{1}{4}$$

$$\implies X \in SE(\nu = 2,\ \alpha = 4)$$

Theorem 3.1 *Sub-exponential tail bounds*

Let $X \in SE(\nu, \alpha)$, then

$$\mathbb{P}\left(|x-\mu| \geq t\right) \leq \begin{cases} 2e^{\frac{-t^2}{2\nu^2}} & 0 \leq t \leq \frac{\nu^2}{\alpha} \\ 2e^{\frac{-t}{2\alpha}} & t > \frac{\nu^2}{\alpha} \end{cases}$$

Finding: When t is small, we recover gaussian behavior. When t is large, we recover exponential behavior.

Proof:
Assume $\mathbb{E}[x] = 0, t \geq 0$

$$\begin{aligned} \mathbb{P}[x \geq t] &\leq e^{-\lambda t}\ \mathbb{E}[e^{\lambda X}] \\ &\leq e^{-\lambda t}\ e^{\frac{\lambda^2 \nu^2}{2}} \\ &= g(\lambda, t) \qquad \lambda \in [0, 1/\alpha) \end{aligned}$$

We found a bound, with an extra parameter, and call it g.

$$g^*(t) = \inf_{\lambda \in [0,1/\alpha)} g(\lambda, t) \qquad t \geq 0$$

The unconstrained minimum of $g(\lambda, t)$ as a function of λ is at $\lambda^* = t/\nu^2$. If ($\lambda^* < 1/\alpha \iff t < \nu^2/\alpha$) then λ^* is also the constrained minimum and $g^*(t) = \frac{-t^2}{2\nu^2}$.

If $\lambda^* \geq 1/\alpha$, notice that $g(\lambda, t)$ is $\searrow$ in $(0, \lambda^*)$. It is a continuous function so constraining happens at the boundary. By continuity the constrained minimum occurs at the boundary $\lambda^* = 1/\alpha$.

So, $g^*(t) = g(\lambda^*, t) = \frac{-t}{\alpha} + \frac{\nu^2}{2\alpha^2} \leq \frac{-t}{2\alpha}$ because $\frac{\nu^2}{\alpha} \leq t$. ■

3.4 Bernstein Condition

Let X be such that $\mathbb{E}[X] = \mu,\ v[X] = \sigma^2$.

X satisfies the bernstein condition with parameter $b > 0$ if k=3, 4, 5, ...

$\mathbb{E}[\ |X-\mu|^k] \leq \frac{1}{2}\ k!\ \sigma^2\ b^{k-2}$

This gives us a bound on the centered moments of a random variable. Note that all bounded variables satisfy this condition. We'll see that the bernstein condition can be tighter than Hoeffding and often more applicable.

Theorem 3.2 *If X satisfies the Bernstein condition with parameter b, then*

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

$X \in SE(\sqrt{2}\sigma, 2b)$

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

Under the bernstein condition, we see that it is sub-exponential and can use the bounds from before (Theorem 3.1), but we also get this new bound.

Proof: Given that all of the moments exist, and are well defined we get that

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

$$\leq 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2}\sum_{k=3}^{\infty}(|\lambda|\ b)^{k-2} \qquad \sum_{k=0}^{\infty} X^k = \frac{1}{1-X} for|X| < 1$$

for $|\lambda| < 1/b$ we can evaluate the geometric series

$$\leq 1 + \frac{\lambda^2\sigma^2}{2}\frac{1}{1-b|\lambda|} \qquad [1 + x \leq e^x]\ x \in \mathbb{R}$$

Note: all of these proofs are about **bounds**!!

$$\leq e^{\frac{\lambda^2\sigma^2}{2}\frac{1}{1-b|\lambda|}}$$

If we further restrict $|\lambda| < 1/2b,$ we get

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\lambda^2\sigma^2} = e^{\frac{\lambda^2(\sqrt{2}\sigma)^2}{2}}$$

$$\implies X \in SE(\nu = \sqrt{2}\sigma, 2b)$$

To get our final result, we set $\lambda = \frac{2}{bt+\sigma^2}$ in the chernoff argument.

If $\sigma^2 \ll t$ the bound is e^{-t}. If $\sigma^2 \gg t$ the bound is e^{-t^2}.

Hoeffding doesn't take into consideration the variance, so the top bound can sometimes be better than the bottom. We use the variance term as a quantity that can get us better results.

Note: This bernstein condition proof isn't as in depth as it could be. Another useful version is where $\sigma^2 = V[x]$ and $|X - \mu| \leq c$ a.e.

$$\mathbb{P}(|X-\mu| \geq t) \leq 2e^{\frac{-t^2}{2(\sigma^2 + \frac{c}{3}t)}} \qquad t \geq 0$$

■

If we have a composition of sub-exponential variables $X_1 \cdots X_n$ which are independent and $\in SE(\nu_i, \alpha_i)$ it is easy to see that $\sum_{i=1}^n X_i$ is also sub-exponential where $\alpha^* = max(\alpha_i)$ and $\nu^* = \sqrt{\sum_{i=1}^n \nu_i^2}$.

If they are not independent, then $\nu^* = \sum_{i=1}^n \nu_i^2$ which isn't as good, but still sub-exponential.

Upper tail bound:

$$\mathbb{P}\left(\frac{\sum(x_i-\mu_i)}{n} \geq t\right) \leq \begin{cases} 2e^{\frac{-nt^2}{2(\nu_*^2/n)}} & t < \frac{\nu_*^2}{\alpha_*} \\ 2e^{\frac{-nt}{2\alpha_*}} & t \geq \frac{\nu_*^2}{\alpha_*} \end{cases}$$

Exercise:

$Z \overset{iid}{\sim} N(0,1)$, $\quad Y = \sum_{i=1}^n Z_i^2 \sim \chi_n^2$

$Z_i^2 \in SE(2,4),\quad Y \in SE(2\sqrt{n}, 4)$

$\mathbb{P}(|Y/n - 1| \geq t) \leq 2e^{-nt^2/8}$ $\qquad t \in (0,1)$

probability an average exceeds a value

Note on homework 1: $|X| = \frac{\sum(x_i - \mu_i)}{n}$ in $\mathbb{P}(|X| \geq t) \leq c_1 e^{-c_2 n t^\alpha}$

3.5 Next topic

- One more concentration inequality
- Maxima

Useful links

http://www.stat.cmu.edu/~arinaldo/36788/subgaussians.pdf
http://www.stat.berkeley.edu/~bartlett/courses/2013spring-stat210b/notes/4notes.pdf
https://en.wikipedia.org/wiki/Hoeffding%27s_inequality
https://en.wikipedia.org/wiki/Chernoff_bound
http://www.cs.cornell.edu/~sridharan/concentration.pdf
https://en.wikipedia.org/wiki/Concentration_inequality