Stat210A: Theoretical Statistics **Lecture Date: November 18, 2014**

Concentration Inequalities

Lecturer: Michael I. Jordan *Scribe: Jonathan Levy*

We have finished Keener at this point and are now moving onto high dimensional statistics: Concentration Bounds. For references, refer to Boucheron, Lugosi and Messart.

Assume X is nonneg. with finite mean, then Markov states: P(X≥t)≤ $\frac{EX}{t}$ for t > 0.This follows by noting $X/t \geq 1_{(x \geq t)}$ and then taking expectations on both sides.

Chebyshev then follows: $P(\mid X - \mu \mid \geq t) \leq \frac{E(X-\mu)^2}{t^2} = \frac{varX}{t^2}$

$P(\mid X - \mu \mid \geq t) \leq \frac{E|X-\mu|^*}{t^*}$

Assume $Ee^{\lambda(X-\mu)}$ exists for $\mid \lambda \mid \leq b$
then $\lambda \in [0, b] : P((X - \mu) \geq t) = P(Ee^{\lambda|X-\mu|} \geq e^{\lambda t}) \leq \frac{Ee^{\lambda(X-\mu)}}{e^{\lambda t}}$

Now minimize over λ so $log(P(X - \mu \geq t) \leq - \underset{\lambda \in [0,b]}{sup} \{\lambda t - log(Ee^{\lambda(X-\mu)})\}$

Example: $X \sim N(\mu, \sigma^2)$

Mgf: $Ee^{\lambda X} = e^{\mu\lambda + \frac{\lambda^2\sigma^2}{2}}$ for all $\lambda \in \mathbb{R} \Longrightarrow \underset{\lambda}{sup}\{\lambda t - \frac{\lambda^2\sigma^2}{2}\}$

differentiating wrt λ and setting to 0 the inside we get:$\lambda = t/\sigma^2$ so our sup $= \frac{t^2}{\sigma^2} - \frac{t^2}{2\sigma^2} = \frac{t^2}{2\sigma^2}$

So $P((X - \mu) \geq t) \leq e^{\frac{-t^2}{2\sigma^2}}$ and thus Gaussian has tails decaying at rate $-t^2$

Definition: A RV X with mean μ is called sub-gaussian if there exists $\sigma > 0$ such that $Ee^{\lambda(X-\mu)} \leq e^{\frac{\lambda^2\sigma^2}{2}} \lambda \in \mathbb{R}$

$\sigma =$ **sub gaussian parameter**.

The symmetry of the definition $\Longrightarrow$ X is subgaussian iff -X is subgaussian $\Longrightarrow P(|X - \mu|) \geq t) \leq 2e^{\frac{-t^2}{2\sigma^2}}$

**Are there other vars other than gaussian that have subgaussian tails?

Example: Rademacher $\{-1, 1\}$ equiprob.

$$\begin{aligned}
Ee^{\lambda\varepsilon} &= \tfrac{1}{2}e^{-\lambda} + \tfrac{1}{2}e^{\lambda} \\
&= \tfrac{1}{2}\left(\sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!} + \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right) \\
&= \sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!} \\
&\leq 1 + \sum_{k=0}^{\infty}\frac{\lambda^{2k}}{2^k k!} \\
&= e^{\frac{\lambda^2}{2}}
\end{aligned}$$

Example: Bounded RV's

Let X be zero mean and supported on [a,b]

Let $X^{'}$ be an independent copy of X

$$\begin{aligned}
Ee^{\lambda X} &= E_X\left(e^{\lambda X - E_{x'}X^{'}}\right) \leq E_{X,X'}\left(e^{\lambda(X-X^{'})}\right) \\
&= E_{X,X'}\left(E_\varepsilon e^{\lambda\varepsilon(X-X^{'})}\right) \leq E_{X,X'}\left(e^{\lambda^2(X-X^{'})^2/2}\right) \\
&\leq e^{\lambda^2(b-a)^2/2} \\
\sigma &= \tfrac{b-a}{2} \ \textbf{sub-gaussian parameter}
\end{aligned}$$

X_1 and X_2 indep. subgaussian w params σ_1^2 and $\sigma_2^2 \implies$
$X_1 + X_2$ is subgaussian with parameters $\sigma_1^2 + \sigma_2^2$

Prop 2.1 Hoeffding Bound

X_i index., X_i has mean μ_i and subgaussian parameter σ_i

$\implies$ for all $t \geq 0$, $P\left(\sum_{i=1}^n (X_i - \mu_i) \geq t\right) \leq exp\left(\dfrac{t^2}{-2\sum_{i=1}^n \sigma_i^2}\right)$

$\implies$ for all $X_i \in [a,b] \implies P\left(\sum_{i=1}^n (X_i - \mu_i) \geq t\right) \leq exp\left(\dfrac{-2t^2}{n(b-a)^2}\right)$

Equivalent characterizations of subgaussian

(1) Mgf condition

(2) $\exists c \geq 0$ and for a Z $\sim$ N(0,τ^2), $P(\|X\| \geq s) \leq cP(\|Z\| \geq s)$ for all s ≥ 0

That finishes up subgaussian stuff.

Subexponential Bounds

Definition An RV X with mean μ is **sub exponential** if there are non-neg parameters (ν, b) such that $Ee^{\lambda(X-\mu)} \leq e^{\frac{\nu^2\lambda^2}{2}}$ for all $|\lambda| \leq \frac{1}{b}$