Spring 2018: STA 6448
Advanced Probability and Inference II
Lecture 2

Yun Yang

- Concentration inequality

Why concentration inequalities?

Often we would like to obtain bounds on tail probability like $\mathbb{P}(X \geq t)$ for some random variable X. For example, central limit theorem tells us

$$\lim_{n\to\infty} \mathbb{P}(\bar{X}_n \geq \mu + \sigma\, n^{-1/2}\, \varepsilon) = 1 - \Phi(\varepsilon) \leq \frac{1}{2}\, e^{-\varepsilon^2/2}.$$

This tells us the asymptotic limit for each fixed ε, but what if ε is also changing with n? For example, what is

$$\mathbb{P}(\bar{X}_n \geq \mu + t)?$$

Do we still have

$$\mathbb{P}(\bar{X}_n \geq \mu + t) \leq C_1\, e^{-C_2\, n\, t^2} \text{ for all } t > 0?$$

We need to exploit information about the random variables.

Some universal bounds

Theorem (Berry-Esseen central limit theorem)

$$\sup_{\varepsilon \in \mathbb{R}} \left| \mathbb{P}(\bar{X}_n \le \mu + \sigma\, n^{-1/2}\, \varepsilon) - \Phi(\varepsilon) \right| \le \frac{\mathbb{E}|X_1 - \mu|^3}{\sqrt{n}\, \sigma^3}.$$

As an implication, we get

$$\mathbb{P}(\bar{X}_n \ge \mu + \varepsilon) \le \frac{C_1}{\sqrt{n}} + \frac{1}{2}\, e^{-n\varepsilon^2/2}.$$

The approximation error is of order $n^{-1/2}$, which ruins the desired exponential decay for $\varepsilon \gtrsim \sqrt{\log n / n}$.

Moment bounds

$$\text{(Markov):} \qquad \mathbb{P}(\bar{X}_n \ge \mu + t) \le \frac{C_1}{C_2 + t},$$

$$\text{(Chebyshev):} \qquad \mathbb{P}(\bar{X}_n \ge \mu + t) \le \frac{C_1}{n\, t^2}.$$

From Markov to Chernoff

If random variable X has a central moment of order k, then

$$\mathbb{P}(|X-\mu| \geq t) \leq \frac{\mathbb{E}|X-\mu|^k}{t^k}.$$

Assume a even stronger assumption that X has a moment generating function near zero: for some $b > 0$,

$$\phi(\lambda) = \mathbb{E}[e^{\lambda(X-\mu)}] \text{ exists for all } \lambda \leq b.$$

Chernoff bound

$$\mathbb{P}(X-\mu \geq t) \leq \inf_{\lambda \in [0,b]} \frac{\mathbb{E}[e^{\lambda(X-\mu)}]}{e^{\lambda t}}.$$

Gaussian tail bounds

Let $X \sim \mathcal{N}(0, \sigma^2)$. We know that

$$\mathbb{E}[e^{\lambda X}] = e^{\mu\lambda + \sigma^2\lambda^2/2}, \quad \forall \lambda \geq 0.$$

By using Chernoff bound and optimizing over $\lambda \in [0, \infty)$, we obtain

$$\mathbb{P}(X - \mu \geq t) \leq e^{-\frac{t^2}{2\sigma^2}}, \quad \text{for all } t \geq 0.$$

This bound is sharp up to a constant!

$$\text{For standard normal } Z, \quad \sup_{t \geq 0} \left\{ e^{\frac{t^2}{2}} \, \mathbb{P}(Z \geq t) \right\} = \frac{1}{2}.$$

Sub-Gaussian random variables

Definition

A random variable X with mean $\mu = \mathbb{E}[X]$ is said to be sub-Gaussian with parameter σ^2, if

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq \exp\left\{\frac{\lambda^2\sigma^2}{2}\right\}, \quad \text{for all } \lambda \in \mathbb{R}.$$

Similar to the derivation of the Gaussian tail bound, we have the upper deviation inequality:

$$\mathbb{P}(X \geq \mu + t) \leq e^{-\frac{t^2}{2\sigma^2}}, \quad \text{for all } t \geq 0.$$

By the symmetry of the definition, $-X$ is also sub-Gaussian with parameter σ^2, which implies the lower deviation inequality:

$$\mathbb{P}(X \leq \mu - t) \leq e^{-\frac{t^2}{2\sigma^2}}, \quad \text{for all } t \geq 0.$$

Sub-Gaussian random variables

Theorem (Sub-Gaussian concentration inequality)

If X is sub-Gaussian with parameter σ^2, then

$$\mathbb{P}(|X-\mu| \geq t) \leq 2\, e^{-\frac{t^2}{2\sigma^2}}, \quad \textit{for all } t \geq 0.$$

Start from the above, we have

$$\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}|X-\mu|^2 &= 2\int_0^\infty t\, \mathbb{P}(|X-\mu| \geq t)\, dt \\
&\leq 4\int_0^\infty t\, e^{-\frac{t^2}{2\sigma^2}}\, dt = 4\sigma^2.
\end{aligned}$$

Example: Rademacher variables

A Rademacher random variable ε takes values $\{-1, +1\}$ equally likely.

Claim

ε is sub-Gaussian with parameter $\sigma^2 = 1^2$.

Proof: Apply Taylor expansions to show

$$\mathbb{E}[e^{\lambda\varepsilon}] \leq e^{\lambda^2/2}.$$

Example: bounded random variables

Let X be a random variable satisfying

$$\mathbb{P}(a \leq X \leq b) = 1.$$

Define

$$A(\lambda) = \log \mathbb{E}[e^{\lambda X}] = \log \left(\int e^{\lambda x} \, \mathbb{P}(dx) \right).$$

Then A is the log-normalization of the exponential family random variable X_λ with reference measure $\mathbb{P}$ and sufficient statistic x. Therefore,

$$A'(\lambda) = \mathbb{E}[X_\lambda] \qquad \text{and} \qquad A''(\lambda) = \mathrm{Var}(X_\lambda).$$

Since X_λ is supported on $[a, b]$, $\mathrm{Var}(X_\lambda) \leq (b-a)^2/4$ (why?). Therefore, a Taylor expansion of A at $\lambda = 0$ gives

$$A(\lambda) \leq \lambda \, \mathbb{E}[X] + \frac{\lambda^2}{2} \, \frac{(b-a)^2}{4}.$$

Example: bounded random variables

Let X be a random variable satisfying

$$\mathbb{P}(a \leq X \leq b) = 1.$$

Property

X is sub-Gaussian with parameter $\sigma^2 = (b-a)^2/4$.

Let $b = 1$ and $a = -1$ leads to the claim of Rademacher variables.

Hoeffding Bound

Property

X_1 and X_2 are independent sub-Gaussian variables with parameter σ_1^2 and σ_2^2, then $X_1 + X_2$ is sub-Gaussian with parameter $\sigma_1^2 + \sigma_2^2$.

Theorem (Hoeffding bound)

For $X_1, \ldots, X_n$ independent, $\mathbb{E}[X_i] = \mu_i$, X_i sub-Gaussian with parameter σ_i^2, then for all $t > 0$,

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

For example, $X_i \in [a, b]$, we have $\sigma_i^2 = (b-a)^2/4$, and

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_i) \geq t \right) \leq \exp\left(- \frac{2nt^2}{(b-a)^2} \right).$$

Example: Sum of Bernoulli random variables

$X_i \sim$Bernoulli(p_i), $i = 1, \ldots, n$. Let $\mu = \sum_{i=1}^n p_i$. By Hoeffding,

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

Sharper bounds can be obtained by using the Chernoff bound:

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

Sub-exponential random variables

Definition

A random variable X with mean $\mu = \mathbb{E}[X]$ is said to be sub-exponential if there are nonnegative parameters $(\nu^2,\, b)$ such that

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq \exp\left\{\frac{\lambda^2\,\nu^2}{2}\right\}, \quad \text{for all } |\lambda| \leq \frac{1}{b}.$$

Sub-Gaussian random variables are sub-exponential with $\nu^2 = \sigma^2$ and $b = 0$.

Example (Sub-exponential but not sub-Gaussian)

Let $Z \sim \mathcal{N}(0, 1)$, and consider $X = Z^2$. For $\lambda \leq 1/4$,

$$\mathbb{E}\left[e^{\lambda(X-1)}\right] = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq e^{2\lambda^2}.$$

Therefore, X is sub-exponential with parameters $(2^2, 4)$.

Sub-exponential random variables

Theorem (Sub-exponential concentration inequality)

If X is sub-exponential with parameters (ν^2, b), then

$$\mathbb{P}(X \geq \mu + t) \leq \begin{cases} \exp\left(-\frac{t^2}{2\nu^2}\right) & \text{if } 0 \leq t \leq \frac{\nu^2}{b}, \\ \exp\left(-\frac{t}{2b}\right) & \text{if } t > \frac{\nu^2}{b}. \end{cases}$$

Another useful version,

$$\mathbb{P}(X \geq \mu + \sqrt{2\nu^2 x} + 2bx) \leq e^{-x} \quad \text{for all } x > 0.$$

We can also obtain a two sided concentration inequality with an additional factor of two.

When t is small, the bound is sub-Gaussian and when t is large, the bound has exponential decay.

Proof: Apply the Chernoff bound.