Spring 2018: STA 6448
Advanced Probability and Inference II
Lecture 3

Yun Yang

- Concentration inequality

Bernstein condition

Directly verifying the sub-exponential property can be impractical in practice. We seek some alternative methods.

Definition

A random variable X with mean μ and variance σ^2 satisfies the Bernstein condition with parameter b if

$$\left|\mathbb{E}\left[(X-\mu)^k\right]\right| \le \frac{1}{2}\,k!\,\sigma^2\,b^{k-2} \qquad \text{for } k = 3, 4, \ldots$$

For example, a bounded random variable X satisfies the Bernstein condition with parameter b if

$$\left|X-\mu\right| \le b \quad a.s.$$

Bernstein condition

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X-\mu)}\right] &= 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \lambda^k \, \frac{\mathbb{E}\left[(X-\mu)^k\right]}{k!} \\
&\le 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2} \sum_{k=3}^{\infty} (|\lambda|^k \, b)^{k-2} \\
&= 1 + \frac{\lambda^2\sigma^2/2}{1 - b\,|\lambda|} \le \exp\left(\frac{\lambda^2\sigma^2/2}{1 - b\,|\lambda|}\right).
\end{aligned}$$

Therefore, X is sub-exponential with parameters $(2\,\sigma^2, 2b)$.

Bernstein type bound

Theorem

If random variable X with mean μ and variance σ^2 satisfies the Bernstein condition with parameter b, then

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

Moreover,

$$\mathbb{P}(|X-\mu| \geq t) \leq 2\,\exp\left(-\frac{t^2}{2(\sigma^2+bt)}\right) \qquad \textit{for all } t > 0.$$

For bounded random variables, Bennet's inequality can be used to provide sharper control on the tails.

Proof: Take $\lambda = \frac{t}{\sigma^2+bt} \in (0,\, b^{-1})$ in the Chernoff bound.

Sum of independent sub-exponential variables

Suppose X_k is sub-exponential with parameters (ν_k^2, b_k) for $k = 1, \ldots, n$, and they are independent. Let $\mu_k = \mathbb{E}[X_k]$.

$$\mathbb{E}\left[e^{\lambda \sum_{k=1}^n (X_k - \mu_k)}\right] = \prod_{k=1}^n \mathbb{E}\left[e^{\lambda (X_k - \mu_k)}\right]$$
$$\leq \exp\left(\frac{\lambda^2}{2} \sum_{k=1}^n \nu_k^2\right) \quad \text{for all } |\lambda| \leq \left(\max_k b_k\right)^{-1}.$$

Property

$\sum_{k=1}^n X_k$ has mean $\sum_{k=1}^n \mu_k$, and is sub-exponential with parameters $\left(\sum_{k=1}^n \nu_k^2,\ \max_k b_k\right)$.

Sum of independent sub-exponential variables

Let $\nu^2 = n^{-1} \sum_{k=1}^n \nu_k^2$ and $b = \max_k b_k$.

Theorem (Sub-exponential concentration inequality)

$$\mathbb{P}\Big(\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu_k) \geq t\Big) \leq \begin{cases} \exp\left(-\frac{nt^2}{2\nu^2}\right) & \textit{if } 0 \leq t \leq \frac{\nu^2}{b}, \\ \exp\left(-\frac{nt}{2b}\right) & \textit{if } t > \frac{\nu^2}{b}. \end{cases}$$

Or equivalently (up to constants),

$$\mathbb{P}\Big(\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu_k) \geq \sqrt{\frac{2\nu^2}{n}x} + \frac{2b}{n}x\Big) \leq e^{-x} \quad \textit{for all } x > 0.$$

In particular, a similar inequality holds when each X_i satisfies the Bernstein condition. This leads to the sharper version of the concentration inequality for sum of Bernoulli variables.

Example: χ^2-variables

Consider chi-squared random variable $Y \sim \chi_n^2$ with n degrees of freedom. We can write

$$Y = \sum_{k=1}^{n} Z_k^2, \qquad Z_k \overset{iid}{\sim} \mathcal{N}(0,1).$$

Z_k^2 is sub-exponential with parameters $(2^2, 4)$. Therefore, Y is sub-exponential with parameters $(4n,\, 4)$ ($\nu^2 = 4$ and $b = 4$), and

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n} Z_k^2 - 1\Big| \geq t\Big) \leq 2e^{-nt^2/8} \quad \text{for all } t \in (0,1),$$

$$\mathbb{P}\Big(\frac{1}{n}\sum_{k=1}^{n} Z_k^2 - 1 \geq t\Big) \leq e^{-nt/8} \quad \text{for all } t \geq 1.$$

Application: Johnson-Lindenstrauss embedding

Theorem

For m points $x_1, \ldots, x_m$ in $\mathbb{R}^d$, there is a projection $F : \mathbb{R}^d \to \mathbb{R}^n$ that preserves distances in the sense that, for each pair (x_i, x_j),

$$(1-\delta)\,\|x_i - x_j\|_2^2 \leq \|F(x_i) - F(x_j)\|_2^2 \leq (1+\delta)\,\|x_i - x_j\|_2^2,$$

provided that $n > (16/\delta^2)\log m$.

Johnson-Lindenstrauss: proof

We consider a random projection:

$$F(u) = \frac{1}{\sqrt{n}}\, X\, u,$$

where $X \in \mathbb{R}^{n \times d}$ has independent $\mathcal{N}(0, 1)$ entries.

Let $X_i \in \mathbb{R}^d$ denote the ith row of X. Then $\langle X_i,\, u/\|u\|_2 \rangle$ follows $\mathcal{N}(0, 1)$, and

$$Y = \frac{\|X\, u\|_2^2}{\|u\|_2^2} = \sum_{i=1}^{n} \langle X_i,\, u/\|u\|_2 \rangle^2 \sim \chi_n^2.$$

Johnson-Lindenstrauss: proof

Therefore, the chi-squared concentration inequality leads to

$$\mathbb{P}\Big(\Big|\frac{1}{n}\,\frac{\|X\,u\|_2^2}{\|u\|_2^2} - 1\Big| \geq \delta\Big) \leq 2e^{-n\delta^2/8} \quad \text{for all } \delta \in (0,1)$$

$$\Leftrightarrow\ \mathbb{P}\Big(\frac{\|F(u)\|_2^2}{\|u\|_2^2} \notin [1-\delta,\ 1+\delta]\Big) \leq 2e^{-n\delta^2/8} \quad \text{for all } \delta \in (0,1)$$

There are at most $\binom{m}{2}$ distinct pair of points, we apply a union bound

$$\mathbb{P}\Big(\exists i \neq j \text{ s.t. } \frac{\|F(x_i) - F(x_j)\|_2^2}{\|x_i - x_j\|_2^2} \notin [1-\delta,\ 1+\delta]\Big) \leq 2\binom{m}{2}e^{-n\delta^2/8}.$$

For any $\varepsilon \in (0,1)$, this probability will be below ε if $n > (16/\delta^2)\ \log(m/\varepsilon)$.

Concentration for martingale difference sequence

Example

For some function $f : \mathbb{R}^n \to \mathbb{R}$ and independent variables $\{X_k\}_{k=1}^n$, want to understand the deviation of $f(X_1, \ldots, X_n)$ from its mean $\mathbb{E}[f(X_1, \ldots, X_n)]$.

Apply the telescoping identity

$$f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] = \sum_{k=1}^{n} (Y_k - Y_{k-1}),$$

where $Y_k = \mathbb{E}[f(X_1, \ldots, X_n) \,|\, X_1, \ldots, X_k]$ for $k = 0, 1, \ldots, n$.