Concentration inequalities and tail bounds

John Duchi

Outline

Motivation

- Often in this class, goal is to argue that sequence of random (vectors) $X_1, X_2, \ldots$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n} X_i \stackrel{p}{\to} \mathbb{E}[X].$$

- Law of large numbers: if $\mathbb{E}[\|X\|] < \infty$, then

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i \neq \mathbb{E}[X]\right) = 0.$$

Markov inequalities

Theorem (Markov's inequality)

Let X be a non-negative random variable. Then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Chebyshev inequalities

Theorem (Chebyshev's inequality)

Let X *be a real-valued random variable with* $\mathbb{E}[X^2] < \infty$*. Then*

$$\mathbb{P}(X - \mathbb{E}[X] \ge t) \le \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\mathrm{Var}(X)}{t^2}.$$

Example: i.i.d. sampling

Chernoff bounds

Moment generating function: for random variable X, the MGF is

$$M_X(\lambda) := \mathbb{E}[e^{\lambda X}]$$

Example: Normally distributed random variables

Chernoff bounds

Theorem (Chernoff bound)

For any random variable and $t \geq 0$,

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \inf_{\lambda \geq 0} M_{X - \mathbb{E}[X]}(\lambda) e^{-\lambda t} = \inf_{\lambda \geq 0} \mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] e^{-\lambda t}.$$

Sub-Gaussian random variables

Definition (Sub-Gaussianity)

A mean-zero random variable X is σ^2-*sub-Gaussian* if

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

Example: $X \sim \mathsf{N}(0, \sigma^2)$

Properties of sub-Gaussians

Proposition (sums of sub-Gaussians)

Let X_i be independent, mean-zero σ_i^2-sub-Gaussian. Then $\sum_{i=1}^n X_i$ is $\sum_{i=1}^n \sigma_i^2$-sub-Gaussian.

Concentration inequalities

Theorem

Let X be σ^2-sub-Gaussian. Then for $t \geq 0$,

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$$
$$\mathbb{P}(X - \mathbb{E}[X] \leq -t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

Concentration: convergence of an independent sum

Corollary

Let X_i be independent σ_i^2-sub-Gaussian. Then for $t \geq 0$,

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

Example: bounded random variables

Proposition

Let $X \in [a, b]$*, with* $\mathbb{E}[X] = 0$*. Then*

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2 (b-a)^2}{8}}.$$

Maxima of sub-Gaussian random variables (in probability)

$$\mathbb{E}\left[\max_{j \le n} X_j\right] \le \sqrt{2\sigma^2 \log n}$$

Maxima of sub-Gaussian random variables (in expectation)

$$\mathbb{P}\left(\max_{j \le n} X_j \ge \sqrt{2\sigma^2(\log n + t)}\right) \le e^{-t}.$$

Hoeffding's inequality

If X_i are bounded in $[a_i, b_i]$ then for $t \geq 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) \leq \exp\left(-\frac{2nt^2}{\frac{1}{n}\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \leq -t\right) \leq \exp\left(-\frac{2nt^2}{\frac{1}{n}\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Equivalent definitions of sub-Gaussianity

Theorem

The following are equivalent (up to constants)

i $\mathbb{E}[\exp(X^2/\sigma^2)] \le e$

ii $\mathbb{E}[|X|^k]^{1/k} \le \sigma\sqrt{k}$

iii $\mathbb{P}(|X| \ge t) \le \exp(-\frac{t^2}{2\sigma^2})$

If in addition X is mean-zero, then this is also equivalent to i–iii above

iv X *is σ^2-sub-Gaussian*

Sub-exponential random variables

Definition (Sub-exponential)

A mean-zero random variable X is (τ^2, b)-sub-Exponential if

$$\mathbb{E}\left[\exp\left(\lambda X\right)\right] \le \exp\left(\frac{\lambda^2 \tau^2}{2}\right) \quad \text{for } |\lambda| \le \frac{1}{b}.$$

Example: Exponential RV, density $p(x) = \beta e^{-\beta x}$ for $x \ge 0$

Sub-exponential random variables

Example: χ^2-random variable. Let $Z \sim \mathsf{N}(0, \sigma^2)$ and $X = Z^2$. Then

$$\mathbb{E}[e^{\lambda X}] = \frac{1}{[1 - 2\lambda\sigma^2]_+^{\frac{1}{2}}}.$$

Concentration of sub-exponentials

Theorem

Let X be (τ^2, b)-sub-exponential. Then

$$\mathbb{P}(X \geq \mathbb{E}[X]+t) \leq \begin{cases} e^{-\frac{t^2}{2\tau^2}} & \text{if } 0 \leq t \leq \frac{\tau^2}{b} \\ e^{-\frac{t}{2b}} & \text{if } t \geq \frac{\tau^2}{b} \end{cases} = \max\left\{e^{-\frac{t^2}{2\tau^2}}, e^{-\frac{t}{2b}}\right\}.$$

Sums of sub-exponential random variables

Let X_i be independent (τ_i^2, b_i)-sub-exponential random variables. Then $\sum_{i=1}^n X_i$ is $(\sum_{i=1}^n \tau_i^2, b_*)$-sub-exponential, where $b_* = \max_i b_i$

Corollary: If X_i satisfy above, then

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X_i]\right| \ge t\right) \le 2\exp\left(-\min\left\{\frac{nt^2}{2\frac{1}{n}\sum_{i=1}^n \tau_i^2}, \frac{nt}{2b_*}\right\}\right).$$

Bernstein conditions and sub-exponentials

Suppose X is mean-zero with

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

Then

$$\mathbb{E}[e^{\lambda X}] \le \exp\left(\frac{\lambda^2 \sigma^2}{2(1 - b|\lambda|)}\right)$$

Johnson-Lindenstrauss and high-dimensional embedding

Question: Let $u^1, \ldots, u^m \in \mathbb{R}^d$ be arbitrary. Can we find a mapping $F : \mathbb{R}^d \to \mathbb{R}^n$, $n \ll d$, such that

$$(1-\delta)\left\|u^i - u^j\right\|_2^2 \le \left\|F(u^i) - F(u^j)\right\|_2^2 \le (1+\delta)\left\|u^i - u^j\right\|_2^2$$

Theorem (Johnson-Lindenstrauss embedding)

For $n \gtrsim \frac{1}{\epsilon^2} \log m$ such a mapping exists.

Proof of Johnson-Lindenstrauss continued

$$\mathbb{P}\left(\left|\frac{\|Xu\|_2^2}{n\,\|u\|_2^2} - 1\right| \geq t\right) \leq 2\exp\left(-\frac{nt^2}{8}\right) \quad \text{for } t \in [0,1].$$

Reading and bibliography

1. S. Boucheron, O. Bousquet, and G. Lugosi. Concentration inequalities.
   In O. Bousquet, U. Luxburg, and G. Ratsch, editors, *Advanced Lectures in Machine Learning*, pages 208–240. Springer, 2004
2. V. Buldygin and Y. Kozachenko. *Metric Characterization of Random Variables and Random Processes*, volume 188 of *Translations of Mathematical Monographs*.
   American Mathematical Society, 2000
3. M. Ledoux. *The Concentration of Measure Phenomenon*.
   American Mathematical Society, 2001
4. S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: a Nonasymptotic Theory of Independence*.
   Oxford University Press, 2013