MATH 690: Topics in Probablity Theory	October 26, 2017

Concentration of Measure

Lecturer: Xiuyuan Cheng — *Scribe: Xiangying Huang*

1 Introduction

It is known from the central limit theorem (CLT) that given $X_1, X_2, \ldots, X_n i.i.d$ with $\mathbb{E}X_i = 0$ and $Var(X_i) = \sigma^2$, we have

$$\frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}(0,1) \text{ in distribution.}$$

Write $S_n = \sum_{i=1}^n X_i$, $S_n \sim O(\sqrt{n})$.

Remark 1. *Notice that if $X_1, X_2, \ldots, X_n$ are not mutually independent, then S_n does not necessarily scale like $O(\sqrt{n})$. For example, when $X_1 = X_2 = \cdots = X_n$, $S_n \sim O(n)$.*

While CLT ensures the convergence of S_n in distribution, it does not give information on the rate of convergence, which is the main topic of this lecture. In terms of S_n, we want a result of the form

$$\mathbb{P}(|S_n - \mathbb{E}S_n| > \alpha) \leq ?$$

where $\alpha \sim O(\sqrt{n})$.

More generally, given a function $F(X_1, X_2, \ldots, X_n)$, we want to bound the probability

$$\mathbb{P}(|F(X_1, X_2, \ldots, X_n) - \mathbb{E}F(X_1, X_2, \ldots, X_n)| > \alpha) \leq ?$$

To motivate the study of concentration of measure, we list three applications of it:

1. **Johnson-Lindenstrauss Lemma**

   Lemma 1. *Johnson-Lindenstrauss Lemma.*

   Given a set X of n points in $\mathbb{R}^D$, $0 < \varepsilon < 1$, and a number $d > \dfrac{8\ln n}{\varepsilon^2}$, there exists a linear map $f : \mathbb{R}^D \to \mathbb{R}^d$ such that

   $$(1-\varepsilon)\|u-v\| \leq \|f(u) - f(v)\| \leq (1+\varepsilon)\|u-v\|$$

   for all $u, v \in X$.

   Remark 2. *$f \cdot u = \sqrt{\frac{D}{d}} P \cdot u$, where P is a projection to a d-dimensional subspace and $u \in \mathbb{R}^D$.*

   See [1] for a proof.

2. **Fast randomized SVD**

   Let G be a $n \times k$ matrix where $k << n$, $Gij \sim \mathcal{N}(0,1)$ *i.i.d* and G is well-conditioned. Let u, v be different columns of the normalized matrix $\frac{G}{\sqrt{n}}$. We have

(a) $\mathbb{E}\left\|u\right\|^2 = 1$

(b) $\mathbb{E}u^T v = 0$

Hence, $\frac{G}{\sqrt{n}}$ has "almost" orthonormal columns and singular values close to 1.

3. **Spectral norm of Wigner matrices**

2 Preliminaries

Lemma 2. *Markov Inequality*

Let X be a non-negative random variable with $\mathbb{E}X < \infty$. Then, for $\alpha > 0$,

$$\mathbb{P}(X > \alpha) \leq \frac{\mathbb{E}X}{\alpha}$$

Proof. $\mathbb{E}X = \mathbb{E}[X(1_{X>\alpha} + 1_{X\leq\alpha})] \geq \mathbb{E}[X(1_{X>\alpha}] \geq \mathbb{E}[\alpha 1_{X>\alpha}] = \alpha\mathbb{P}(X > \alpha)$ □

Corollary 1. *Chebyshev's Inequality.*

$$\mathbb{P}(|X - EX| > \alpha) \leq \frac{Var(X)}{\alpha^2}$$

3 Key Idea

Our goal is control the large deviation of S_n. Let us first consider the most basic case, where $X_1, X_2, \ldots, X_n$ are *i.i.d*. (Constraints on identical distribution or independence can be relaxed.)

For $t > 0$,

$$\mathbb{P}(S_n > \alpha) = \mathbb{P}(e^{tS_n} > e^{t\alpha}) \leq \frac{E(exp(tS_n))}{e^{t\alpha}}.$$

Hence it suffices to give an upper bound of $E(exp(tS_n))$. By *i.i.d* property of $\{X_i\}$,

$$\begin{aligned} \mathbb{E}[exp(tS_n)] &= \mathbb{E}[exp(t\sum_{i=1}^{n} X_i)] \\ &= \prod_{i=1}^{n} \mathbb{E}e^{tX_i} \\ &= (\mathbb{E}e^{tX_1})^n \end{aligned}$$

If we assume $\mathbb{E}e^{tX_1} \leq e^{ct^2}$ for some constant $c > 0$ (which holds under rather general conditions), the upper bound is then obtained:

$$\mathbb{P}(S_n > \alpha) \leq \frac{E(exp(tS_n))}{e^{t\alpha}} \leq \frac{e^{ct^2 n}}{e^{\alpha t}} = e^{cnt^2 - \alpha t}$$

Since the above inequality holds for any $t > 0$, we can pick $t^* = \frac{\alpha}{2cn}$ so that it minimizes the expression on the right hand side of the inequality. Plugging in t^* gives

$$\mathbb{P}(S_n > \alpha) \leq e^{\frac{-\alpha}{4cn}}$$

Notice that we pick $\alpha \sim O(\sqrt{n})$ so that the above bound gives meaningful results.

4 Hoeffding's Inequality

The idea in Section 3 can be applied in the proof of similar results. One example is **Hoeffding's inequality**.

Theorem 1. *Hoeffding's Inequality.*

Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with $a_i \leq X_i \leq b_i$. Write $S_n = \sum_{i=1}^n X_i$. Then, for $\alpha > 0$,

$$P(|S_n - \mathbb{E}S_n| > \alpha) \leq 2exp(\frac{-2\alpha^2}{\sum_{i=1}^n (b_i - a_i)^2})$$

In order to prove the inequality, let us first introduce Hoeffding's lemma.

Lemma 3. *Hoeffding's lemma.*

Suppose X is a random variable such that $a \leq X \leq b$ for some constant a, b. For any $t > 0$, we have

$$\mathbb{E}e^{tX} \leq exp(\frac{t^2(b-a)^2}{8}).$$

Proof of Hoeffding's Inequality:

Proof. Without loss of generality we can consider the case where each X_i has $\mathbb{E}X_i = 0$. Otherwise we just replace X_i with $\bar{X}_i = X_i - \mathbb{E}X_i$ in later analysis. Notice that if $a_i \leq X_i \leq b_i$, then $a_i - \mathbb{E}X_i \leq \bar{X}_i \leq b_i - \mathbb{E}X_i$, and $(b_i - \mathbb{E}X_i) - (a_i - \mathbb{E}X_i) = b_i - a_i$. Hence the result remains the same.

With the above being said, from now on we work with X_i that has $\mathbb{E}X_i = 0$.

$$\begin{aligned}
\mathbb{P}(S_n > \alpha) &= \mathbb{P}(e^{tS_n} > e^{t\alpha}) \\
&\leq \frac{\mathbb{E}e^{tS_n}}{e^{t\alpha}} \\
&= \frac{\prod_{i=1}^n \mathbb{E}e^{tX_i}}{e^{t\alpha}} \\
&\leq e^{\frac{t^2 \sum_{i=1}^n (b_i - a_i)^2}{8}} e^{-t\alpha}
\end{aligned}$$

By picking $t = \frac{4\alpha}{\sum_{i=1}^n (b_i - a_i)^2}$ we can minimize the right hand side of the inequality, which gives

$$\mathbb{P}(S_n > \alpha) \leq exp(\frac{-2\alpha^2}{\sum_{i=1}^n (b_i - a_i)^2}).$$

Following the same steps, we have

$$\mathbb{P}(S_n < -\alpha) = P(-S_n > \alpha) \leq exp(\frac{-2\alpha^2}{\sum_{i=1}^n (b_i - a_i)^2}).$$

Therefore,

$$P(|S_n - \mathbb{E}S_n| > \alpha) \leq 2exp(\frac{-2\alpha^2}{\sum_{i=1}^n (b_i - a_i)^2}).$$

$\square$

Proof of Hoeffding's lemma:

Proof. By convexity, for any $x \in (0,1)$ we have

$$e^{tX} \leq xe^{tb} + (1-x)e^{ta}.$$

Hence,

$$e^{tX} \leq \frac{X-a}{b-a}e^{tb} + \frac{b-X}{b-a}e^{ta}.$$

Then

$$\mathbb{E}e^{tX} \leq \frac{\mathbb{E}X-a}{b-a}e^{tb} + \frac{b-\mathbb{E}X}{b-a}e^{ta} = \frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{t}a.$$

Let $h = t(b-a)$, $p = \dfrac{-a}{b-a}$ and $L(h) = -hp + \ln(1-p+pe^h)$. Then $\frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{t}a = e^{L(h)}$.

Taking derivative of $L(h)$,

$L(0) = L^{'}(0) = 0$ and

$$L^{''}(h) = \frac{(1-p)pe^h}{(1-p+pe^h)^2} \leq \frac{(1-p)pe^h}{4(1-p)pe^h} = \frac{1}{4}$$

for all h.

By Taylor's expansion,

$$L(h) \leq \frac{1}{8}h^2 = \frac{1}{8}t^2(b-a)^2$$

Hence, $\mathbb{E}e^{tX} \leq e^{\frac{1}{8}t^2(b-a)^2}$ □

References

[1] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of a theorem of Johnson and Lindenstrauss." Random Structures & Algorithms 22.1 (2003): 60-65.