# Math 690: Topics in Data Analysis and Computation Lecture notes for 10/31/2017

Scribed by Lihan Wang

Let's continue our topic on concentration of measure.

**Theorem.** *(Chernoff's Inequality) Let $y_1, \cdots, y_n$ be i.i.d. Bernoulli$(p)$ random variables (in other words $\mathbb{P}(y_i = 1) = p = 1 - \mathbb{P}(y_i = 0)$). Then for any $\lambda > 0$ we have the following concentration estimates of $S_n = \sum_{i=1}^{n} y_i$:*

$$\mathbb{P}(S_n < np - \lambda) \leq \exp(-\frac{\lambda^2}{2np})$$

$$\mathbb{P}(S_n > np + \lambda) \leq \exp(-\frac{\lambda^2}{2(np + \lambda/3)})$$

**Remark.** *The bound by Chernoff is non-asymptotic, which means that it holds for any $p$ and $n$. If $np$ is large, and $\lambda$ scales no greater than $np$, then the bound for the upper tail is generally $\exp(-\frac{c\lambda^2}{np})$, which is the normal limit (decays like Gaussian); if $\lambda >> np$, then the bound for the upper tail is essentially $\exp(-c\lambda)$, which is the Poisson limit (decays exponentially).*

**Remark.** *For the upper tail bound, if we directly apply Hoeffding's inequality, we get*

$$\mathbb{P}(S_n > np + \lambda) \leq \exp(-\frac{2\lambda^2}{n})$$

*This is not as sharp as Chernoff's inequality if $p << 1$.*

**Theorem.** *(Bernstein's Inequality) Suppose $X_1, \cdots, X_n$ are i.i.d random variables, $EX_i = 0, EX_i^2 = \sigma^2$, and $|X_i| \leq M$. Let $S_n = \sum_{i=1}^{n} X_i$, then $\forall \delta > 0$*

$$\mathbb{P}(|S_n| > \sqrt{n}\delta) \leq 2\exp(-\frac{3}{8}\min\{\frac{\delta^2}{\sigma^2}, \sqrt{n}\frac{\delta}{M}\})$$

**Remark.** *For $M = 1$, if we directly apply Hoeffding's inequality, we get*

$$\mathbb{P}(|S_n| > \sqrt{n}\delta) \leq \exp(-\frac{\delta^2}{2})$$

*This bound is not as sharp as Bernstein's inequality, where we get $\exp(-\frac{3\delta^2}{8\sigma^2})$ for $\sigma < \frac{\sqrt{3}}{2}$ and $\delta < \sqrt{n}\sigma^2$.*

Interested readers can refer to [**?**] for the proofs of these inequalities.

Now let's get back to the Johnson-Lindenstrauss Lemma. The proof given below is from [**?**].

**Theorem.** *(Johnson-Lindenstrauss Lemma) Let $X$ be a set of $n$ points in $\mathbb{R}^D$. Suppose $\varepsilon \in (0,1)$. If*

$$d > \frac{4}{\varepsilon^2/2 - \varepsilon^3/3} \log n,$$

*then for any projection $P$ from $\mathbb{R}^D$ to a random $d$-dimensional subspace, and $f(u) = \sqrt{\frac{D}{d}} Pu$, with high probability, we have*

$$(1-\varepsilon)||u-v||^2 \le ||f(u) - f(v)||^2 \le (1+\varepsilon)||u-v||^2$$

*for any $u, v \in X$.*

**Remark.** *This above lemma also implies the existence of such a projection $P$ and that it can be found in polynomial time: since for any projection $P$, the probability that it fails the condition is low, then as long as we pick many such projections as we want, the probability that none of these projections satisfy the condition will tend to zero.*

Now we can start to prove the Johnson-Lindenstrauss Lemma.

Proof: Fix $u, v \in X$, and let $w = u - v$. By linearity of $P$, we have

$$f(u) - f(v) = \sqrt{\frac{D}{d}} P(u-v) = \sqrt{\frac{D}{d}} Pw.$$

Now the remaining step is to estimate the operator norm of any such projection $P$, as we want

$$(1-\varepsilon) \le ||Pw||^2 \le (1+\varepsilon)$$

with high probability, for $w$ uniformly distributed on $S^{D-1}$.

Let $X_1, \cdots, X_D$ be $D$ i.i.d. $N(0,1)$ random variables, and $X = (X_1, \cdots, X_D)$, $w = \frac{X}{||X||}$. Then $w$ is uniformly distributed on $S^{D-1}$, and we are left to prove

$$||\sqrt{\frac{D}{d}} PX||^2 \in \big((1-\varepsilon)||X||^2, (1+\varepsilon)||X||^2\big)$$

with high probability. Without loss of generality (by some rotation), we can assume $P$ is the projection to the first $d$ coordinates. Thus

$$||\sqrt{\frac{D}{d}} PX||^2 = \frac{D}{d}(X_1^2 + \cdots + X_d^2).$$

Here we have a useful lemma:

**Lemma.** *Suppose $d, D, X$ are defined as above.*
*(i) If $0 < \beta < 1$, then*

$$\mathbb{P}(X_1^2 + \cdots + X_d^2 < \beta \frac{d}{D} ||X||^2) \le \beta^{\frac{d}{2}} \big(1 + \frac{(1-\beta)d}{D-d}\big)^{\frac{D-d}{2}} \le \exp\big(\frac{d}{2}(1 - \beta + \ln \beta)\big);$$

*(ii) If $\beta > 1$, then*

$$\mathbb{P}(X_1^2 + \cdots + X_d^2 > \beta \frac{d}{D} ||X||^2) \le \beta^{\frac{d}{2}} \big(1 + \frac{(1-\beta)d}{D-d}\big)^{\frac{D-d}{2}} \le \exp\big(\frac{d}{2}(1 - \beta + \ln \beta)\big);$$

Substitute $\beta$ with $1-\varepsilon$ and $1+\varepsilon$ in the corresponding part of the lemma, we have

$$\mathbb{P}\Big(||\sqrt{\frac{D}{d}}Pw||^2 < (1-\varepsilon)||w||^2\Big) \leq \exp\Big(\frac{d}{2}\big(1-(1-\varepsilon)+\ln(1-\varepsilon)\big)\Big) \leq \exp(-\frac{d\varepsilon^2}{4}).$$

Here we used the fact that $\ln(1-\varepsilon) \leq -\varepsilon - \frac{\varepsilon^2}{2}$. By assumption of the lemma, there exists $r > 0$ such that $d \geq \frac{8(1+r)}{\varepsilon^2}\ln n$, thus $\frac{d\varepsilon^2}{4} \geq 2(1+r)\ln n$, and

$$\mathbb{P}\Big(||\sqrt{\frac{D}{d}}Pw||^2 < (1-\varepsilon)||w||^2\Big) \leq \exp(-2(1+r)\ln n) = (\frac{1}{n})^{2+2r}.$$

Similarly we have the bound for the upper tail

$$\mathbb{P}\Big(||\sqrt{\frac{D}{d}}Pw||^2 > (1+\varepsilon)||w||^2\Big) \leq \exp(-2(1+r)\ln n) = (\frac{1}{n})^{2+2r}.$$

Now, for any pair $u, v \in X$, since there are $\frac{n(n-1)}{2}$ different pairs in $X$, by a naive union bound estimate, we have

$$\mathbb{P}\Big(\exists u, v \in X,\ s.t. ||f(u)-f(v)||^2 \notin \big((1-\varepsilon)||u-v||^2, (1+\varepsilon)||u-v||^2\big)\Big) \leq 2\frac{n(n-1)}{2}(\frac{1}{n})^{2+2r} = O(n^{-2r})$$

for large $n$. Hence we finish the proof.

**Remark.** *The Johnson-Lindenstrauss lemma shows that, loosely speaking, any set of size $n$ can be projected to a space with dimension $d = O(\ln n)$ without much distortion, and this can be achieved by a random projection with high probability.*

**Remark.** *Is there any "counter example" for Johnson-Lindenstrauss lemma?*

# 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.

[2] Tao, Terence. Topics in random matrix theory. Vol. 132. Providence, RI: American Mathematical Society, 2012.