36-788: Topics in High Dimensional Statistics I — Fall 2015

Lecture 2: September 3

Lecturer: Alessandro Rinaldo — *Scribes: Jisu KIM*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

This lecture's notes illustrate some uses of various LaTeX macros. Take a look at this and imitate.

2.1 Concentration

Let $(\mathscr{X}, d)$ be a metric space and P be a probability on $(\mathscr{X}, d)$, Let $A \subset \mathscr{X}$ be such that $P(A) \geq \frac{1}{2}$. Let $A_r = \{x \in \mathscr{X} : \; d(x, A) < r\}$ as in Figure 2.1, where $d(x, A) = \inf_{y \in A} d(x, y)$. A way to look at concentration of measure is to let r grow and look at $P(A_r^c)$.

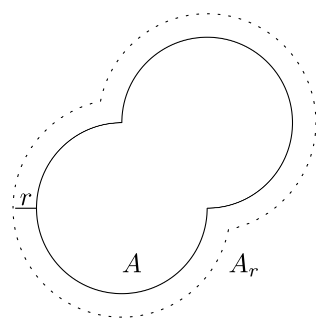

Figure 2.1: A and $A_r = \{x \in \mathscr{X} : \; d(x, A) < r\}$

Definition 2.1 *The concentration function $\alpha = \alpha(\mathscr{X}, d, P) : \; \mathbb{R}_+ \to [0, 1]$ is defined as* [1]

$$\alpha(r) = \sup_{A \subset \mathscr{X} \; s.t. \; P(A) \geq \frac{1}{2}} P(A_r^c).$$

Question: How fast does $\alpha(r)$ decrease?

Definition 2.2 *A space $(\mathscr{X}, d)$ with prob P has normal concentration if* [2]

$$\alpha(r) \leq C e^{-cr^2}.$$

[1] [L2005], p.3
[2] [L2005], p.4

Gaussian satisfies normal concentration.

Example. 1) Let $\mathscr{X} = \mathbb{S}^n = \{x \in \mathbb{R}^{n+1},\ \|x\| = 1\}$(sphere) with $d(x,y) = arc\cos < x, y >$(angular distance), and P is uniform. Then [3]

$$\alpha(r) \leq e^{-(n-1)r^2/2}.$$

So when n is large, $P(A_r) \approx 1$ for small r. And P-almost of $\mathbb{S}^n$ are within distance $\frac{1}{\sqrt{n}}$ from A.

Example. 2) Let $\mathscr{X} = \{0,1\}^n$(unit cube) with $d(x,y) = \frac{1}{n}\left|\{i:\ x_i \neq y_i\}\right|$(hamming distance), and P is uniform. Then [4]

$$\alpha(r) \leq e^{-2nr^2}.$$

Example. 3) Let $\mathscr{X}$ =unit ball in $\mathbb{R}^n$ with euclidean distance, and P is uniform. Then [5]

$$\alpha(r) \leq e^{-cnr^2}$$

Example. 4) Many many product spaces [6]

Example. 5) Let $\mathscr{X}$ =set of all permutations on n elements with $d(\sigma,\tau) = \frac{1}{n}\left|\{i:\ \sigma_i \neq \tau_i\}\right|$, and P is uniform. Then it has normal concentration, with c depends on n. [7]

Example. 6) Let P a probability on $(\mathbb{R}^n, \mathcal{B}_n)$ with density of the form $e^{-U(x)}$, where $U(x) + U(y) - 2U\left(\frac{x+y}{2}\right) \geq \frac{c\|x-y\|^2}{4}$, with $\|\cdot\|$: Euclidean form. (log-concave density. $N_n(0,I)$ satisfies this with $c = 1$) Then [8]

$$\alpha(r) \leq 2e^{-cr^2/4}.$$

Bound doesn't depend on n: This is how people go to infinite dimension. [9]

2.2 Connection to 1 Liptschitz function

Let $(\mathscr{X}, d)$ be a metric space and P be a probability on $(\mathscr{X}, d)$.

Definition 2.3 *$f : \mathscr{X} \to \mathbb{R}$ is called 1-Lipschitz if*

$$|f(x) - f(y)| \leq d(x,y).$$

Definition 2.4 *Let m_f be median of f if* [10]

$$P\left(f(X) \leq m_f\right) \geq \frac{1}{2} \text{ and } P\left(f(X) \geq m_f\right) \geq \frac{1}{2}.$$

• Concentration of 1-Liptschitz function $\Longrightarrow$ Concentration of probability

Pick $A \subset \mathscr{X}$ and let

$$f(x) = d(x, A).$$

[3] Theorem 2.3 in [L2005], vii, p.1-2, p.26, Theorem 14.3 in [B2005], p.3-4, p.59-60, Theorem 1 in [BN2009], p.5
[4] Theorem 2.11 in [L2005], p.3, p.31, Corollary 4.4 in [B2005], p.17
[5] Proposition 2.9 in [L2005], p.30
[6] Chapter 4. Concentration in product spaces in [L2005], p.67-90
[7] Theorem 8.10 in [L2005], p.159
[8] Theorem 2.15 in [L2005], p.36
[9] Theorem 7.1 in [L2005], p.133-134
[10] [L2005], p.5, [B2005], p.3

Then f is 1-Lipschitz (Exercise!). Now assume $P(A) \geq \frac{1}{2}$ and $X \sim P$. Then

$$Pr(f(X) = 0) = P(A) \geq \frac{1}{2}.$$

This implies 0 is a median of $f(x) = d(x, A)$. So

$$P(A_r^c) = Pr\left(f(X) - m_f \geq r\right) \leq \alpha(r),$$

where $\alpha(r)$ is a concentration function.

• Concentration of probability $\Longrightarrow$ Concentration of 1-Liptschitz function

Let f be a 1-Lipschitz function and let

$$A = \left\{x \in \mathscr{X} : \ f(x) \leq m_f\right\}.$$

Then $P(A) \geq \frac{1}{2}$, and

$$\forall r > 0, \ A_r \subset \left\{x \in \mathscr{X} : \ f(x) < m_f + r\right\} (\because \text{exercise}).$$

Therefore,

$$Pr\left(f(X) - m_f \geq r\right) \leq P\left(A_r^c\right) \leq \alpha(r).$$

Theorem 2.5 *A Borel space $(\mathscr{X}, d)$ with probability P has concentration function α iff for every 1-Liptschitz function f and for every $r > 0$,* [11]

$$P\left(f(X) \geq m_f + r\right) \leq \alpha(r).$$

Apply this to $-f$, we have

$$P\left(|f(X) - m_f| \geq r\right) \leq 2\alpha(r).$$

2.2.1 Extensions

- You can always replace m_f with $E\left[f(X)\right]$ with different constants C and c in α. [12]
- $E\left[f(X)\right] - m_f$ is small
- If there is a normal concentration, then [13]

  $$V\left[f(X)\right] \leq \frac{C}{c},$$

  and there exists $K(C)$ such that for any $q \geq 1$, [14]

  $$\left(E\left|f(X) - E\left[f(X)\right]\right|^q\right)^{\frac{1}{q}} \leq K\sqrt{\frac{q}{c}}.$$

  (very similar behavior as Gaussian)
- Gaussian measure in $\mathbb{R}^n$. Let $X = (X_1, \cdots, X_n) \sim N_n(0, I)$ and f is 1-Lipschitz. Then [15]

  $$P\left(|f(X) - E[f(X)]| > r\right) \leq 2e^{-\frac{r^2}{2}}.$$

  for all n! (dimension free).

[11] Proposition 1.3 in [L2005], p. 7
[12] Proposition 1.7 in [L2005], p. 9-10
[13] Proposition 1.9 in [L2005], p. 11-12
[14] Proposition 1.10 in [L2005], p. 12-13
[15] Corollary 2.6 in [L2005], p. 2, p. 28-29, Theorem 14.6 in [B2005], p.61, Proposition 1 and Theorem 10 in [BN2009], p.9-10, p.32-33, p.40-42

- Moreover, there exists g 1-Lipschitz and $Z \sim N(0,1)$ such that

$$f(\underbrace{X}_{\text{high dim}}) \stackrel{d}{=} g(\underbrace{Z}_{\text{1 dim}}).$$

End of interesting materials.

2.3 Chernoff inequality

2.3.1 Jensen inequality

Let $f : \mathbb{R} \to \mathbb{R}$ be convex on $-\infty \le a < b \le +\infty$ and X random variable supported on subset of (a, b), then

$$f(E[X]) \le E[f(X)].$$

($=$ holds if $X = c$ a.s. for some c)

if f is concave, reverse inequality holds.

2.3.2 Markov's inequality

Let $f : \mathbb{R}_+ \to \mathbb{R}$ be non-decreasing, and X be a random variable, then

$$P(|X| > r) \le \frac{E[f(|X|)]}{f(r)}.$$

2.3.3 Chernoff-bounds

Let $X_1, \cdots, X_n$ be independent random variables, and let $Z = f(X_1, \cdots, X_n)$. We are interested in bounding $P(Z - E[Z] > r)$, $P(Z - E[Z] < -r)$, $P(|Z - E[Z]| > r)$. We would like bounds that

1. are analytically simple,
2. apply to general random variables,
3. are sharp.

Theorem 2.6

$$P(Z \ge x) \le \exp\{\psi_Z^*(x)\},$$

with

$$\psi_Z^*(x) = \sup_{\lambda > 0}\{\lambda x - \psi_Z(\lambda)\}.$$

Example. Let $X = (X_1, \cdots, X_n) \stackrel{iid}{\sim} P$, with $E[X_1] = \mu$ and $V[X_1] = \sigma^2$. Let $f(X) = \bar{X}_n$. Then

$$P\left(\left|\bar{X}_n - \mu\right| > r\right) \le \frac{\sigma^2}{nr^2}$$

by Chebyshev. But we should be able to do better: by Central Limit Theorem,

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \rightsquigarrow N\left(0, \sigma^2\right),$$

so

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}}{\sigma^2}\left(\bar{X}_n - \mu\right) > r\right) \to 1 - \Phi(r) \leq e^{-r^2/2},$$

where $\Phi(r)$ is cdf of $N(0,1)$. So as $n \to \infty$,

$$P\left(|X_n - \mu| > r\right) \leq 2e^{-\frac{nr^2}{2\sigma^2}}.$$

Proof:

Step 1

for $x \in \mathbb{R}$ and $\forall \lambda > 0$,

$$\begin{aligned} P\left(Z \geq x\right) &= P\left(e^{\lambda Z} \geq e^{\lambda x}\right) \\ &\leq \frac{E\left[e^{\lambda Z}\right]}{e^{\lambda x}} \text{ (markov inequality)} \\ &= \exp\left\{\psi_Z(\lambda) - \lambda x\right\}, \ \left(\psi_Z(\lambda) = \log\left(\underbrace{E\left[e^{\lambda Z}\right]}_{mgf}\right)\right). \end{aligned}$$

Step 2

Minimize the RHS with respect to $\lambda > 0$, and then we obtain

$$P\left(Z \geq x\right) \leq \exp\left[\psi_Z^*(x)\right],$$

with

$$\psi_Z^*(x) = \sup_{\lambda > 0}\left\{\lambda x - \psi_Z(\lambda)\right\}.$$

■

Remarks

- we need to know ψ_Z.
- $\psi_Z(0) = 0$ implies $\psi_Z^* \geq 0$.
- ψ_Z is finite on $(0, b)$ where $b \leq \infty$.
- ψ_Z is convex.
- ψ_Z is infinite time differentiable.
- If $E[Z] = 0$, then $\psi_Z^{'}(0) = \psi_Z(0) = 0$.
- How do you get $\psi_Z^*(x)$?

- $$\psi_Z^*(x) = x\lambda_x - \psi_Z(\lambda_x),$$

  where $\psi_Z^{'}(\lambda_x) = x$. In particular,

  $$\lambda_x = \left(\psi_Z^{'}\right)^{-1}(x).$$

  (Since ψ_Z is strictly convex, $\psi_Z^{'}$ is strictly increasing.)

Example. 1) Normal : $Z \sim N(0, \sigma^2)$

$\psi_Z(\lambda) = \frac{\lambda^2\sigma^2}{2}$. Then $\lambda_x = \frac{x}{\sigma^2}$ and

$$\psi_Z^*(x) = x\lambda_x - \psi_Z(\lambda_x) = \frac{x^2}{2\sigma^2}.$$

Hence

$$P\left(Z \geq x\right) \leq e^{-\frac{x^2}{2\sigma^2}}.$$

This result is not optimal: missing a constant $\frac{1}{2}$.

Example. 2) Poisson : $X \sim Poisson(\nu)$, $\nu > 0$

Let $Z = X - \nu$. Then $E\left[e^{\lambda Z}\right] = e^{\nu(e^\lambda - \lambda - 1)}$ and

$$\lambda_x = \log\left(1 + \frac{x}{\nu}\right),\ x > 0.$$

And

$$\psi_Z^*(x) = \nu h\left(\frac{x}{\nu}\right),$$

where

$$h(\mu) = (1+\mu)\log(1+\mu) - \mu,\ \mu \geq -1.$$

similarly if $x \leq \nu$,

$$\psi_{-Z}^*(x) = \nu h\left(-\frac{x}{\nu}\right).$$

Reference

[B2005] Measure Concentration, Lecture Notes for Math 710, by Alexander Barvinok, 2005

[BN2009] Concentration of measure, lecture notes by N. Berestycki and R. Nickl

[L2005] The Concentration of Measure Phenomenon, by M. Ledoux, 2005, AMS.