10-755: Advanced Statistical Theory I **Fall 2017**

Lecture 5: September 18

Lecturer: Alessandro Rinaldo *Scribes: Boyan Duan*

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

5.1 Lipschitz functions of Gaussians

This section shows that concentration of normal distribution is good in some sense.
Recall: Let $Z \sim N_d(0, \sigma^2 I_d)$. If $d = 1$, $P(|Z| \geq t) \leq 2exp\{-\frac{t^2}{2\sigma^2}\}$.
The following thoerm is about multi-dimensional, Lipschitz case.

Theorem 5.1 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be L-Lipschitz, let $X = f(Z)$,then*

$$X - EX \in SG(\sigma^2 L^2)$$

and

$$P(|f(Z) - Ef(Z)| \geq t) \leq 2exp\{-\frac{t^2}{2L^2\sigma^2}\}$$

Note: L-Lipschitz: $|f(x) - f(y)| \leq L||x - y||$, where $|| \cdot ||$ is euclidean norm

Remark

1. *The above bound doesn't depend on d!*
2. *Many proofs exists. There's one in the reference book.*
   *One way is to use **Gaussian Isoperimetric Inequality** (stated as follows).*
   Let P be distribution of $N_d(0, I_d)$, and let $A \subseteq \mathbb{R}^d$.
   If H is a half space (definded as $\{x, < x, \nu > \leq 0$, for some unit norm $\nu\}$), and if $P(A) = P(H)$, then

   $$P(d(x, A) \geq \epsilon) \leq P(d(x, H) \geq \epsilon), \forall \epsilon > 0$$

   for $x \sim N(0, I)$. (Note: $d(x, A) = inf_{y \in A}||x - y||$)
   See book by Massart [PM2007] and book by Ledoux. [ML2005]

5.2 Maximum of Gaussian

In this section, we use the above theorem to show that maximum of Gaussian have the similar tail behavior as Gaussian.

Theorem 5.2 *Let $Y \sim N_d(0, \Sigma), \sigma^2 = max_i \Sigma_{ii}, X = max_i Y_i$ (or $X = max_i |Y_i|$), then*

$$P(|X - EX| \geq t) \leq 2exp\{-\frac{t^2}{2\sigma^2}\}$$

$$E[X] \sim \sqrt{2\sigma logd}, V[X] \leq \sigma$$

Proof: Let $Y = AZ, Z = N_d(0, I_d), \Sigma = AA^T$. Consider function $f : Z(\in \mathbb{R}^d) \to max_{i=1,\cdots,d}(AZ)_j$.
f is L-Lipschitz with $L = max_i \sqrt{\sum_{j=1}^d A_{ij}^2} = \sigma^2$. Then by Theorem 5.1, proof complete.
Now we show 1. f is L-Lipschitz; 2. $L = max_i \sqrt{\sum_{j=1}^d A_{ij}^2} = \sigma$.

1. $\forall Z, Z' \in \mathbb{R}^d$, for $i = 1, \cdots, d$, $|(AZ)_i - (AZ')_i| = |\sum_j A_{ij}(Z_j - Z'_j)| \le \sqrt{\sum_j A_{ij}^2}||Z - Z'|| \le L||Z - Z'||$
2. $\sum_{j=1}^d A_{ij}^2 = E[(\sum_j A_{ij} Z_j)^2] = V[\sum_j A_{ij} Z_j] = V[Y_i] = \sigma^2$

■

Remark *The theorem is stated for norm $||Y||_\infty$. It can be extended to $||Y||_p = (\sum_i |Y_i|^p)^{1/p}$, $p \ge 1$.*

Note that Theorem 5.1 can be extend to other distribution, but require a stronger condition on f.

Theorem 5.3 *Let $X_1, \cdots, X_d$ be independent taking value in $[0,1]^d$.*
Let $f : \mathbb{R}^d \to \mathbb{R}$ be L-Lipschitz and convex, then

$$P(|f(X) - Ef(X)| \ge t) \le 2exp\{-\frac{t^2}{2L^2}\}$$

For more, see Thm 3.3 in the book "High-dimensional statistics: A non-asymptotic viewpoint".

5.3 Covering and Packing

Background we often want to bound

$$max_{i\in I} X_i,$$

where I is finit or infinite, $X_i \in SG$ or SE, not necessarily independent.

Example 5.4

- $I = \{1, 2, \cdots\}, X_i \overset{iid}{\sim} P$, *then*

$$lim_{n\to\infty} P(max_{i=1,\cdots,n} X_i \le t) = lim_{n\to\infty}[P(X_i \le t)]^n = 0$$

- *If* $X_i = X$, *then for proper* t,

$$lim_{n\to\infty} P(max_i X_i \le t) = P(X \le t) > 0$$

Recall: Metric space $(\mathcal{X}, d)$

Example 5.5

- $(\mathbb{R}^d, ||\cdot||_p), p \ge 1$. *Especially,* $||x||_\infty = max_i |x_i|$
- L_p*-space (mostly infinite dimension):* $\mathcal{X}$ *is a set of functions on* $[0,1]$*, and* $d(f,g) = ||f - g||_p = (\int |f(x) - g(x)|^p dx)^{1/p}$. *Especially,* $||f - g||_\infty = sup_{x\in[0,1]}|f(x) - g(x)|$.

Definition 5.6 (Covering Number) *For $\delta > 0$, a δ-covering or a δ-net of $(\mathcal{X}, d)$ is a subset $\{\theta_1, \cdots, \theta_N\} \subset \mathcal{X}$ s.t. $\forall \theta \in \mathcal{X}, \exists \theta_i$ s.t. $d(\theta, \theta_i) \leq \delta$. The δ-covering number of $(\mathcal{X}, d)$ is the cardinality of a smallest cover, denoted as $N(\delta, \mathcal{X}, d)$.*

Remark

- *We assume that $N(\delta, \mathcal{X}, d) < \infty, \forall \delta > 0$, and $\delta <$diameter$(\mathcal{X}) = sup_{x,x' \in \mathcal{X}} d(x, x')$.*
- *Covering means $\mathcal{X} \subset \cup_{i=1}^N B(\theta_i, \delta)$, where $B(\theta_i, \delta) = \{x \in \mathcal{X} : d(x, \theta_i) \leq \delta\}$.*
- *$N(\delta, \mathcal{X}, d)$ decreases in δ, and will converge if $\delta \downarrow 0$.*
- *If $\mathcal{X}' \subset \mathcal{X}$, it is not true that $N(\delta, \mathcal{X}', d) \leq N(\delta, \mathcal{X}, d)$.*
- *The quantity $logN(\delta, \mathcal{X}, d)$ is also known as metric entropy.*

Example 5.7

- *$\mathcal{X} = [-1, 1], d(x, y) = |x - y|$, then $N(\delta, \mathcal{X}, d) \leq \frac{1}{\delta} + 1$.*
- *$\mathcal{X} = [-1, 1]^d, d(x, y) = |x - y|$,then $N(\delta, \mathcal{X}, d) \leq (\frac{1}{\delta} + 1)^d$.*
  Thus, $logN(\delta, \mathcal{X}, d) \asymp dlog(\frac{1}{\delta} + 1)$ (scale linearly in d).
- *(Infinite dimensional space) $\mathcal{F} = \{f : [0, 1]^d \to \mathbb{R}$, L-Lipschitz$\}$, then $logN(\delta, \mathcal{F}, || \cdot ||_\infty) \asymp (\frac{L}{\delta})^d$ (scale exponential in d, a reflect of size of space).*

Definition 5.8 (Packing Number) *A δ-packing of $(\mathcal{X}, d)$ is $\{\theta_1, \cdots, \theta_M\} \subseteq \mathcal{X}$ s.t. $d(\theta_i, \theta_j) > \delta, \forall i, j$. The δ-packing number of $(\mathcal{X}, d)$ is the cardinality of a largest packing.*

The following lemma bounds covering number between two packing number:

Lemma 5.9 $\forall \delta > 0, M(2\delta, \mathcal{X}, d) \leq N(\delta, \mathcal{X}, d) \leq M(\delta, \mathcal{X}, d)$

Covering number in euclidean space

Theorem 5.10 (Volumetric ratios) *Let $|| \cdot ||, || \cdot ||'$ be two norms in $\mathbb{R}^d$, with unit balls B and B' ($B = \{x \in \mathbb{R}^d, ||x|| \leq 1\}$). Then, the δ-covering number of B in $|| \cdot ||'$ satisfies*

$$(\frac{1}{\delta})^d \frac{vol(B)}{vol(B')} \leq N(\delta, B, || \cdot ||') \leq \frac{vol(\frac{2}{\delta}B + B')}{vol(B')},$$

where $vol(B) =$ volume of B; $\delta B = \{\delta x : x \in B\}(\delta > 0)$;
for $a, b > 0, aB + bB' = \{ax + by : x \in B, y \in B'\}$(Minkowski sum).

Proof: Use the fact that $V_d(\delta B) = \delta^d vol(B)$.

1. (Lower Bound) If $\{x_1, \cdots, x_N\}$ is a δ-covering of B in $|| \cdot ||'$,

   $$B \subset \cup_{i=1}^N (x_i + \delta B')$$

   in which $x_i + \delta B' = \{y : ||x - y||' \leq \delta\}$. Since volume is invariant to shift,

   $$vol(B) \leq Nvol(\delta B') = N\delta^d vol(B')$$

   .

   Therefore,

   $$N(\delta, B, || \cdot ||') \geq (\frac{1}{\delta})^d \frac{vol(B)}{vol(B')}$$

2. (Upper Bound) Let $\{x_1, \cdots, x_M\}$ be a maximum δ-packing of B in $||\cdot||'$, then $\{x_1, \cdots, x_M\}$ is also a δ-covering of B in $||\cdot||'$ (proof by contradiction).
   Now, the balls $x_i + \frac{\delta}{2}B', i = 1, \cdots, M$ are disjoint and

$$\cup_{i=1}^{M}(x_i + \frac{\delta}{2}B') \subset B + \frac{\delta}{2}B'.$$

   Take volume on both side,

$$M(\frac{\delta}{2})^d vol(B') \leq (\frac{\delta}{2})^d vol(\frac{2}{\delta}B + B')$$

   Therefore,

$$N \leq M \leq \frac{vol(\frac{2}{\delta}B + B')}{vol(B')}.$$

■

Remark

1. *If $||\cdot|| = ||\cdot||'$, then $dlog(\frac{1}{\delta}) \leq log[N(\delta, B, ||\cdot||)] \leq dlog(1 + \frac{2}{\delta}) \leq dlog(\frac{3}{\delta})$, $(\delta \leq 1)$.*
2. *If $B' \subseteq B$, B' is unit ball in $||\cdot||$, B is a euclidean unit ball, then*

$$N(\delta, B, ||\cdot||) \leq (1 + \frac{2}{\delta})^d \frac{vol(B)}{vol(B')}.$$

5.4 Discretization Argument

In this section, we will use covering number to bound $max_{i \in I} X_i$.

Definition 5.11 *A random vector X is sub-Gaussian(σ^2) if $\nu^T X \in SG(\sigma^2), \forall \nu \in S^{d-1}$. $(S^{d-1} = \{x \in \mathbb{R}^d, ||x|| = 1\})$*

Theorem 5.12 *Assume $X \in \mathbb{R}^d$ s.t. $X \in SG(\sigma^2)$ then,*

$$E[||X||] \leq 4\sigma\sqrt{d}$$

and

$$||X|| \leq 4\sigma\sqrt{d} + 2\sigma\sqrt{2log(\frac{1}{\delta})}$$

with prob $\geq 1 - \delta, \delta \in (0, 1)$.

References

[PM2007] P. Massart, "Concentration inequalities and model selection," *Vol.6. Berlin: Springer*, 2007.

[ML2005] M. Ledoux, "The concentration of measure phenomenon," *No. 89. American Mathematical Soc.*, 2005.