Summary Part 3: Applications of Concentration Inequalities

Tianpei Xie

Jan. 26th., 2023

Contents

1 Applications **2**
1.1 U-Statistics 2
1.2 Jackknife Estimation and Boostrapping 2
1.3 Kernel-Density Estimation 2
1.4 Random Graph 3
1.5 Minimum Weight Spanning Tree 3
1.6 Rademacher Complexity 3
1.7 Dimensionality Reduction 3

2 Self-Bounding Functions **3**
2.1 Definitions, Variance Bounds and Concentration 3
2.2 Configuration Function 4
2.3 VC-Dimension and Growth Function 5
2.4 Longest Increasing Subsequence 5
2.5 Weakly Self-Bounding Functions 5

3 Random Matrices **5**
3.1 Definitions 5
3.2 Concentration Inequalities of Random Vectors 5
3.3 Concentration of Norm of Gaussian Vectors 5
3.4 Spectral Distribution of Hermitian Matrix: Semi-Circular Law 5
3.5 Largest Eigenvalue of Hermitian Random Matrix 5

4 Empirical Process **5**
4.1 Definition 5
4.2 Uniform Law of Large Numbers 5
4.3 Suprema of Gaussian Process 5
4.4 Covering Number, Packing Number and Metric Entropy 5
4.5 Chaining 5
4.6 VC-Dimension 5
4.7 Variance Bounds 5

1 Applications

1.1 U-Statistics

1.2 Jackknife Estimation and Boostrapping

1.3 Kernel-Density Estimation

- **Example** (***Kernel Density Estimation***)
  Let $Z_1, \ldots, Z_n$ be i.i.d. samples drawn according to some (unknown) density ϕ on the real line. The density is estimated by the kernel estimate

$$\phi_n(z) = \frac{1}{n\,h_n} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{h_n}\right),$$

  where $h_n > 0$ is a *smoothing parameter*, and K is a nonnegative function with $\int K(z) = 1$. The performance of the estimate is typically measured by ***the*** L_1 ***error***:

$$X(n) := f(Z_1, \ldots, Z_n) = \int |\phi(z) - \phi_n(z)|\, dz.$$

  It is easy to see that

$$\begin{aligned} \left|f(z_1, \ldots, z_n) - f_i(z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_n)\right| &\le \frac{1}{nh_n} \int \left|K\left(\frac{z - z_i}{h_n}\right) - K\left(\frac{z - z_i'}{h_n}\right)\right| dz \\ &\le \frac{2}{n}, \end{aligned}$$

  so without further work we obtain

$$\mathrm{Var}\left(X(n)\right) \le \frac{1}{n}$$

  It is known that for every ϕ, $\sqrt{n}\mathbb{E}\left[X(n)\right] \to \infty$, which implies, by *Chebyshev's inequality*, that for every $\epsilon > 0$

$$\mathbb{P}\left\{\left|\frac{X(n)}{\mathbb{E}\left[X(n)\right]} - 1\right| > \epsilon\right\} = \mathbb{P}\left\{\left|X(n) - \mathbb{E}\left[X(n)\right]\right| > \epsilon\mathbb{E}\left[X(n)\right]\right\} \le \frac{\mathrm{Var}(X(n))}{\epsilon^2(\mathbb{E}\left[X(n)\right])^2} \to 0$$

  as $n \to \infty$. That is, $\frac{X(n)}{\mathbb{E}[X(n)]} \to 1$ ***in probability***, or in other words, $X(n)$ is ***relatively stable***. This means that ***the random*** L_1***-error*** essentially *behaves like* ***its expected value***.

  By bounded difference inequality, we have

$$\mathbb{P}\left\{\left|X(n) - \mathbb{E}\left[X(n)\right]\right| \ge t\right\} \le 2\exp\left(-\frac{nt^2}{2}\right) \quad \blacksquare$$

1.4 Random Graph

1.5 Minimum Weight Spanning Tree

1.6 Rademacher Complexity

1.7 Dimensionality Reduction

2 Self-Bounding Functions

2.1 Definitions, Variance Bounds and Concentration

- Another simple property which is satisfied for many important examples is the so-called *self-bounding property*.

  Definition (***Self-Bounding Property***)
  A ***nonnegative*** *function* $f : \mathcal{X}^n \to [0, \infty)$ has the ***self-bounding property*** if *there exist* functions $f_i : \mathcal{X}^{n-1} \to \mathbb{R}$ such that for all $z_1, \ldots, z_n \in \mathcal{X}$ and all $i = 1, \ldots, n$,

  $$0 \le f(z_1, \ldots, z_n) - f_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n) \le 1 \tag{1}$$

  and also

  $$\sum_{i=1}^{n} \left(f(z_1, \ldots, z_n) - f_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n)\right) \le f(z_1, \ldots, z_n). \tag{2}$$

- **Remark** Clearly if f has the ***self-bounding property***,

  $$\sum_{i=1}^{n} \left(f(z_1, \ldots, z_n) - f_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n)\right)^2 \le f(z_1, \ldots, z_n) \tag{3}$$

  Taking expectation on both sides, we have the following inequality

- **Corollary 2.1** *[Boucheron et al., 2013]*
  *If f has the **self-bounding property**, then*

  $$Var(f(Z)) \le \mathbb{E}\left[f(Z)\right].$$

- **Remark** (***Relative Stability***) [Boucheron et al., 2013]
  A sequence of nonnegative random variables $(Z_n)_{n \in \mathbb{N}}$ is said to be ***relatively stable*** if

  $$\frac{Z_n}{\mathbb{E}\left[Z_n\right]} \xrightarrow{\mathbb{P}} 1.$$

  This property guarantees that ***the random fluctuations*** *of Z_n around its* ***expectation*** *are* ***of negligible size*** *when compared to the expectation*, and therefore ***most information about the size of Z_n is given by*** $\mathbb{E}\left[Z_n\right]$.

Bounding the variance of Z_n ***by its expected value*** *implies, in many cases,* ***the relative stability*** *of* $(Z_n)_{n\in\mathbb{N}}$. If Z_n has the ***self-bounding property***, then, by *Chebyshev's inequality*, for all $\epsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{Z_n}{\mathbb{E}\left[Z_n\right]} - 1\right| > \epsilon\right\} \le \frac{\mathrm{Var}(Z_n)}{\epsilon^2(\mathbb{E}\left[Z_n\right])^2} \le \frac{1}{\epsilon^2\mathbb{E}\left[Z_n\right]}.$$

Thus, for relative stability, it suffices to have $\mathbb{E}\left[Z_n\right] \to \infty$.

2.2 Configuration Function

- An important class of functions satisfying *the self-bounding property* consists of the so-called ***configuration functions***.

  Definition (***Configuration Function***)
  Assume that we have a property Π ***defined over the union of finite products*** of a set $\mathcal{X}$, that is, a sequence of sets

  $$\Pi_1 \subset \mathcal{X},\ \Pi_2 \subset \mathcal{X}\times\mathcal{X},\ \ \ldots,\ \ \Pi_n \subset \mathcal{X}^n.$$

  We say that $(z_1, \ldots, z_m) \in \mathcal{X}^m$ ***satisfies the property*** Π if $(z_1, \ldots, z_m) \in \Pi_m$.

  We assume that Π is ***hereditary*** in the sense that if $(z_1, \ldots, z_m)$ satisfies Π then so does ***any sub-sequence*** $\{z_{i_1}, \ldots, z_{i_k}\}$ of $(z_1, \ldots, z_m)$.

  The function f that maps any vector $z = (z_1, \ldots, z_n)$ to ***the size*** *of a* ***largest sub-sequence*** *satisfying* Π is ***the configuration function*** *associated with property* Π.

- **Corollary 2.2** *[Boucheron et al., 2013]*
  Let f be a ***configuration function****, and let $X = f(Z_1, \ldots, Z_n)$, where $Z_1, \ldots, Z_n$ are* ***independent*** *random variables. Then*

  $$Var(f(Z)) \le \mathbb{E}\left[f(Z)\right].$$

2.3 VC-Dimension and Growth Function

2.4 Longest Increasing Subsequence

2.5 Weakly Self-Bounding Functions

3 Random Matrices

3.1 Definitions

3.2 Concentration Inequalities of Random Vectors

3.3 Concentration of Norm of Gaussian Vectors

3.4 Spectral Distribution of Hermitian Matrix: Semi-Circular Law

3.5 Largest Eigenvalue of Hermitian Random Matrix

4 Empirical Process

4.1 Definition

4.2 Uniform Law of Large Numbers

4.3 Suprema of Gaussian Process

4.4 Covering Number, Packing Number and Metric Entropy

4.5 Chaining

4.6 VC-Dimension

4.7 Variance Bounds

References

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.