

Notes for High-Dimensional Probability

Vishwak Srinivasan

Contents

1 Preliminaries	2
1.1 Example on approximate Caratheodory's Theorem	2
1.1.1 Auxiliary Lemmata	3
1.2 Quantities and Inequalities associated with RVs	4

1 Preliminaries

1.1 Example on approximate Caratheodory's Theorem

First, we begin by discussing Caratheodory's Theorem:

Theorem 1.1 (Caratheodory's Theorem). *Consider a convex set $S \subseteq \mathbb{R}^p$. Any point $x \in S$ can be represented as a convex combination of at most $p + 1$ distinct points from S .*

Remark. This result is a popular result in convex analysis, and is tight. The tight lower bound is achieved by a simplex in p dimensions, which corresponds to $p + 1$ vertices.

Now, we seek an approximation of the above theorem like so: given k points $\{x_i\}_{i=1}^k \subset S$, is it possible to approximate a point $x \in S$? We answer this in the affirmative below:

Theorem 1.2 (Approx. Caratheodory's Theorem). *Given $x \in S \subseteq \mathbb{R}^p$, where S is convex, there exists a set of k points $\{x_i\}_{i=1}^k \in S$, such that the following holds:*

$$\left\| x - \frac{1}{k} \sum_{i=1}^k x_i \right\|_2 \leq \frac{\text{diam}(S)}{\sqrt{k}}$$

where $\text{diam}(S) = \sup_{s, t \in S} \|s - t\|_2$.

Proof. By the fact that $x \in S$, we know that we can write x as a convex combination of a subset $\{z_i\}_{i=1}^m$ that satisfy $\text{CONV}(\{z_i\}_{i=1}^m) = S$, where $m \leq p + 1$. Let the coefficients be $\{\lambda_i\}_{i=1}^m$ where $\sum_{i=1}^m \lambda_i = 1$ and $\lambda_i \geq 0$ for all $i \in [m]$.

Consider a random variable Z that takes m different values from the set $\{z_i\}_{i=1}^m$ with probability λ_i . Note that $\mathbb{E}[Z] = x$, since $\mathbb{E}[Z] = \sum_{i=1}^m \Pr(Z = z_i) z_i = \sum_{i=1}^m \lambda_i z_i = x$.

We know that for any $x \in \mathbb{R}^p$ and independent random variables $\{Z_i\}_{i=1}^k$ that satisfy $\mathbb{E}[Z_i] = x$ for all $i \in [k]$:

$$\begin{aligned} \mathbb{E} \left[\left\| x - \frac{1}{k} \sum_{i=1}^k Z_i \right\|_2^2 \right] &= \mathbb{E} \left[\left\| \frac{1}{k} \sum_{i=1}^k (x - Z_i) \right\|_2^2 \right] \\ &= \frac{1}{k^2} \mathbb{E} \left[\left\| \sum_{i=1}^k (x - Z_i) \right\|_2^2 \right] \\ &\stackrel{(i)}{=} \frac{1}{k^2} \sum_{i=1}^k \mathbb{E} [\|x - Z_i\|_2^2] \\ &\stackrel{(ii)}{\leq} \frac{1}{k^2} \sum_{i=1}^k \text{diam}(S)^2 = \frac{\text{diam}(S)^2}{k} \end{aligned}$$

Step (i) holds true due to Lemma 1.1. Step (ii) follows from the fact that $Z_i, x \in S$ which implies that $\|Z_i - x\|_2 \leq \text{diam}(S)$ followed by the fact that $\mathbb{E}[c] = c$ for constant c .

Therefore, there exists a realization of $\{Z_i\}_{i=1}^k$, that satisfies:

$$\left\| x - \frac{1}{k} \sum_{i=1}^k Z_i \right\|_2 \leq \frac{\text{diam}(S)}{\sqrt{k}}$$

□

Remark. First note the dimension independence in the result. Secondly, in the special case where S consists of elements with bounded norms i.e., $\|x\|_2 \leq B$ for all $x \in S$, the diameter of the set is bounded by $2B$ by an application of the triangle inequality. Finally, note that if we have $k \rightarrow \infty$ samples from the set, then our approximation is going to be perfect.

The method used to prove Theorem 1.1 is called Maurey's Empirical Method.

1.1.1 Auxiliary Lemmata

Lemma 1.1. *Let $\{X_i\}_{i=1}^k$ be a set of independent zero-mean random variables. The following holds true:*

$$\mathbb{E} \left[\left\| \sum_{i=1}^k X_i \right\|_2^2 \right] = \sum_{i=1}^k \mathbb{E} [\|X_i\|_2^2]$$

Proof. First note that:

$$\begin{aligned} \left\| \sum_{i=1}^k X_i \right\|_2^2 &= \left\langle \sum_{i=1}^k X_i, \sum_{j=1}^k X_j \right\rangle \\ &= \sum_{i=1}^k \sum_{j=1}^k X_i^T X_j \\ &= \sum_{i=1}^k \|X_i\|_2^2 + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^k X_i^T X_j \end{aligned}$$

Taking expectations on both sides:

$$\begin{aligned} \mathbb{E} \left[\left\| \sum_{i=1}^k X_i \right\|_2^2 \right] &= \mathbb{E} \left[\sum_{i=1}^k \|X_i\|_2^2 \right] + 2 \mathbb{E} \left[\sum_{\substack{i,j=1 \\ i \neq j}}^k X_i^T X_j \right] \\ &= \sum_{i=1}^k \mathbb{E} [\|X_i\|_2^2] + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^k \mathbb{E} [X_i^T X_j] \end{aligned}$$

Since X_i s are independent, $\mathbb{E} [X_i^T X_j] = \mathbb{E} [X_i]^T \mathbb{E} [X_j] = 0$, and this completes the proof. \square

Lemma 1.2. *For all integers $m \in [1, n]$, we have the following series of inequalities:*

$$\left(\frac{n}{m} \right)^m \leq \binom{n}{m} \leq \sum_{k=0}^m \binom{n}{k} \leq \left(\frac{en}{m} \right)^m$$

Proof. First inequality:

$$\binom{n}{m} m^m = \frac{n!}{(n-m)! \cdot m!} m^m \geq \frac{n!}{(n-m)!} \geq n^m \Rightarrow \binom{n}{m} \geq \left(\frac{n}{m} \right)^m$$

Second inequality:

$$\binom{n}{m} \leq \binom{n}{m} + \sum_{k=0}^{m-1} \binom{n}{k} = \sum_{k=0}^m \binom{n}{k}$$

Third inequality:

$$\left(\frac{m}{n}\right)^m \sum_{k=0}^m \binom{n}{k} \leq \sum_{k=0}^m \binom{n}{k} \left(\frac{m}{n}\right)^k \leq \sum_{k=0}^n \binom{n}{k} \left(\frac{m}{n}\right)^k = \left(1 + \frac{m}{n}\right)^n \leq e^m \Rightarrow \sum_{k=0}^m \binom{n}{k} \leq \left(\frac{en}{m}\right)^m$$

□

1.2 Quantities and Inequalities associated with RVs

- Expectation: $\mathbb{E}[X]$
- Variance: $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- MGF: $M_X(t) = \mathbb{E}[e^{tX}]$, $t \in \mathbb{R}$
- p^{th} moment: $\mathbb{E}[X^p]$ and p^{th} absolute moment: $\mathbb{E}[|X|^p]$
- L^p norm: $\|X\|_{L^p} = \sqrt[p]{\mathbb{E}[|X|^p]}$
- L^∞ norm: $\|X\|_{L^\infty} = \text{ess sup } |X|$, where $\text{ess sup } |X|$ denotes the supremum over all set with measure not 0. Also note that: $\text{ess sup } |X| \leq \sup |X|$.
- Covariance: $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$
- CDF: $F_X(t) = \Pr(X \leq t)$, $t \in \mathbb{R}$

For a convex function f and any random variable X , we have by *Jensen's inequality* that:

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

Consequently, for a concave function f and any random variable X , we have:

$$f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)]$$

As a special case, consider $f(x) : x^{q/p}$ where $q > p$. Note that f is convex. Therefore:

$$(\mathbb{E}[|X|^p])^{q/p} \leq \mathbb{E}[|X|^q] \Rightarrow \|X\|_{L^p} \leq \|X\|_{L^q}$$

Another inequality is the *Cauchy-Schwarz inequality*, which states that for any two RVs X and Y :

$$\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]} \sqrt{\mathbb{E}[Y^2]} = \|X\|_{L^2} \|Y\|_{L^2}$$

We also have *Holder's inequality* which generalizes *Cauchy-Schwarz* to dual norms as:

$$\mathbb{E}[|XY|] \leq \|X\|_{L^p} \|Y\|_{L^q} \quad ; \quad \frac{1}{p} + \frac{1}{q} = 1$$

The following lemma characterizes the expectation as a quantity involving only tails:

Lemma 1.3. *Consider a non-negative random variable X . The expectation of this random variable can be written as:*

$$\mathbb{E}[X] = \int_0^\infty \Pr(X > t) dt$$

Proof. For any $x \geq 0$, we have that:

$$x = \int_0^\infty \mathbf{1}_{\{t < x\}} dt = \int_0^x 1 dt + \int_x^\infty 0 dt$$

Therefore:

$$\begin{aligned} X = \int_0^\infty \mathbf{1}_{\{t < X\}} dt &\Rightarrow \mathbb{E}[X] = \mathbb{E}\left[\int_0^\infty \mathbf{1}_{\{t < X\}} dt\right] \\ &= \int_0^\infty \int_{-\infty}^\infty \mathbf{1}_{\{t < x\}} \Pr(X = x) dx dt \\ &= \int_0^\infty \int_t^\infty \Pr(X = x) dx dt \\ &= \int_0^\infty \Pr(X > t) dt \end{aligned}$$

□

A simple generalization for real-valued random variables from the proof of Lemma 1.3 is as follows:

Corollary 1.1. *Consider a real valued random variable X . The expectation of this random variable can be written as:*

$$\mathbb{E}[X] = \int_0^\infty \Pr(X > t) dt - \int_{-\infty}^0 \Pr(X < t) dt$$

An application of Lemma 1.3 is to use it to bound the p^{th} absolute moments via tails:

Corollary 1.2. *For any random variable X :*

$$\mathbb{E}[|X|^p] = \int_0^\infty p t^{p-1} \Pr(|X| > t) dt$$

Classical inequalities: Markov and Chebyshev's:

Lemma 1.4 (Markov's Inequality). *Consider a non-negative random variable X . Then the tails of X can be bounded as:*

$$\Pr(X > t) \leq \frac{\mathbb{E}[X]}{t}$$

Proof. Note that:

$$\mathbb{E}[X] = \mathbb{E}[X \cdot \mathbf{1}_{\{X > t\}}] + \mathbb{E}[X \cdot \mathbf{1}_{\{X \leq t\}}] \geq \mathbb{E}[X \cdot \mathbf{1}_{\{X > t\}}] \geq t \mathbb{E}[\mathbf{1}_{\{X > t\}}] = t \Pr(X > t) \Rightarrow \Pr(X > t) \leq \frac{\mathbb{E}[X]}{t}$$

□

Corollary 1.3 (Chebyshev's Inequality). *Consider a random variable X . Then the probability of deviation from the expectation of X can be bounded as:*

$$\Pr(|X - \mathbb{E}[X]| > t) \leq \frac{\text{Var}(X)}{t^2}$$

Proof. Take $Y = |X - \mathbb{E}[X]|$ as the random variable as apply Markov's inequality:

$$\Pr(Y > t) = \Pr(Y^2 > t^2) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2}$$

□

Remark. Note that one can achieve better dependence on t by using higher moments - provided they exist:

$$\Pr(Y > t) = \Pr(Y^{2k} > t^{2k}) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^{2k}]}{t^{2k}}$$