Notes for High-Dimensional Probability

Vishwak Srinivasan

Contents

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:*

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

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

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 $\mathrm{CONV}(\{z_i\}_{i=1}^m) = S$, 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 $i \in [k]$:

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

Step (i) holds true due to Lemma 1.1. Step (ii) follows from the fact that $Z_i, x \in S$ which implies that $\|Z_i - x\|_2 \leq \operatorname{diam}(S)$ 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:

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

□

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 $k \to \infty$ 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:*

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

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:

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

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

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

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

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:

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

$\square$

1.2 Quantities and Inequalities associated with RVs

- Expectation: $\mathbb{E}[X]$
- Variance: $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- 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: $\|X\|_{L^\infty} = \operatorname{ess\,sup}|X|$, where $\operatorname{ess\,sup}|X|$ denotes the supremum over all set with measure not 0. Also note that: $\operatorname{ess\,sup}|X| \le \sup|X|$.
- Covariance: $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$
- CDF: $F_X(t) = \Pr(X \le t)$, $t \in \mathbb{R}$

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

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

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

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

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

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

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

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

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

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

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:*

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

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:*

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

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:

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

1.3 Basic Limit Theorems

Theorem 1.3 (Strong Law of Large Numbers)**.** *Let $\{X_i\}_{i=1}^n$ be a sequence of identically and independently distributed random variables with mean μ. The quantity $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ satisfies:*

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

as $n \to \infty$.

Here $\xrightarrow{a.s.}$ denotes *almost sure convergence*, which is:

$$\Pr\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

There is a *weak law of large numbers*, which can be derived from Chebyshev's Inequality, for distributions with bounded variance. It is stated below:

Corollary 1.4 (Weak Law of Large Numbers)**.** *Let $\{X_i\}_{i=1}^n$ be a sequence of identically and independently distributed random variables with mean μ and variance $\sigma^2 < \infty$. The quantity $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ satisfies:*

$$\bar{X}_n \xrightarrow{p} \mu$$

where $\xrightarrow{p}$ denotes convergence in probability*, which is;*

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} \Pr\left(|\bar{X}_n - \mu| > \epsilon\right) = 0$$

Proof. First note $\mathbb{E}\left[\bar{X}_n\right] = \mu$, and hence $\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left(X_i\right) = \frac{\sigma^2}{n}$.

By Chebyshev's inequality, for any $\epsilon > 0$:

$$\Pr(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon} \Rightarrow \lim_{n\to\infty} \Pr(|\bar{X}_n - \mu| > \epsilon) = 0 \;\; (\because \text{Sandwich theorem})$$

□

Remark. This weak result is *weak* because $\xrightarrow{a.s.}$ implies $\xrightarrow{p.}$.

Next, we state a result that gives the asymptotic distribution of $\bar{X}_n$.

Theorem 1.4 (Central Limit Theorem)**.** *Let $\{X_i\}_{i=1}^n$ be a sequence of identically and independently distributed random variables with mean μ and variance $\sigma^2 < \infty$. Define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then:*

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1) \qquad \textit{as } n \to \infty$$

While this result states that the deviation between the sample mean and population mean is 0 in the limit, we can give some non asymptotic guarantees on the deviation as follows:

Lemma 1.5. *Let $\{X_i\}_{i=1}^n$ be a sequence of identically and independently distributed random variables with mean μ and variance $\sigma^2 < \infty$. We have that:*

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right|\right] = O\left(\frac{1}{\sqrt{n}}\right)$$

Proof. By Jensen's inequality:

$$\mathbb{E}\left[|Z|\right] \leq \sqrt{\mathbb{E}\left[Z^2\right]}$$

(Note that this also follows from the fact that $\|Z\|_{L^1} \leq \|Z\|_{L^2}$)

Therefore:

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right|\right] &\leq \sqrt{\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right]} \\
&= \sqrt{\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right)^2\right]} \\
&= \sqrt{\frac{1}{n^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}(X_i - \mu)\right)^2\right]} \\
&\overset{(i)}{=} \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left[(X_i - \mu)^2\right]} \\
&= \sqrt{\frac{\sigma^2}{n}} = O\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

where Step (i) follows from Lemma 1.1 for 1-D random variables. □

A special case of the Central Limit Theorem is to provide approximate distributions for binomial distributions. Recall that the binomial distribution $\mathrm{Bin}(n, p)$ is the sum of n independent Bernoulli distribution with parameter p. Therefore, we get that:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{n\bar{X}_n - n\mu}{\sigma\sqrt{n}} = \frac{B_{n,p} - np}{\sqrt{n}\sqrt{p(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ as } n \to \infty$$

where $X_i \sim \mathrm{Ber}(p), i \in [n]$ and $B_{n,p} \sim \mathrm{Bin}(n, p)$. This means that $B_{n,p} \xrightarrow{d} \mathcal{N}(np, np(1-p))$ as $n \to \infty$.

However, there is a better limit theorem in the regime where $p \to \infty, n \to \infty$ and $np = \lambda > 0$. This is the Poisson Limit Theorem:

Theorem 1.5 (Poisson Limit Theorem)**.** *Consider $\{X_i\}_{i=1}^n$ to be n independent Bernoulli variables with parameters p_i. Then, for $n \to \infty$, $\max\limits_{i\in[n]} p_i \to 0$ and $\sum\limits_{i=1}^{n} p_i = \lambda > 0$, we have that:*

$$\sum_{i=1}^{n} X_i \xrightarrow{d} \mathrm{Poi}(\lambda)$$

Remark. In the special case when all p_is are equal, we obtain the same result with $n \to \infty$, $p \to 0$ and $np = \lambda > 0$ as described informally earlier.

2 Concentration inequalities

2.1 Basic Gaussian Inequalities

Lemma 2.1 (Mill's inequalities)**.** *Let $g \sim \mathcal{N}(0,1)$. We have the following lower and upper bounds for the tail* $\Pr(g > t)$*,* $t > 0$ *as follows:*

$$\left(\frac{1}{t} - \frac{1}{t^3}\right)\frac{1}{\sqrt{2\pi}}e^{-t^2/2} \leq \Pr(g > t) \leq \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

Proof. First, the upper bound:

$$\begin{aligned}
\Pr(g > t) &= \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-x^2/2}dx \\
&= \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}}\int_t^\infty te^{-x^2/2}dx \\
&\leq \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}}\int_t^\infty xe^{-x^2/2}dx \\
&= \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}}\int_{t^2/2}^\infty ye^{-y}dy \qquad \left(\because y = \frac{x^2}{2}\right) \\
&= \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}}e^{-t^2/2}
\end{aligned}$$

Second, the lower bound:

$$\begin{aligned}
\Pr(g > t) &= \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-x^2/2}dx \\
&\geq \frac{1}{\sqrt{2\pi}}\int_t^\infty \left(1 - \frac{3}{x^4}\right)e^{-x^2/2}dx \qquad \left(\because 1 - \frac{3}{x^4} \leq 1 \;\forall\; x > 0\right) \\
&\geq \left(\frac{1}{t} - \frac{1}{t^3}\right)\frac{1}{\sqrt{2\pi}}e^{-t^2/2}
\end{aligned}$$

□

Remark. Note that one can get tail bounds for $\mathcal{N}(0, \sigma^2)$ by simply reparameterising the integrals as:

$$\left(\frac{\sigma}{t} - \frac{\sigma}{t^3}\right)\frac{1}{\sqrt{2\pi}}e^{-t^2/2} \leq \Pr(g > t) \leq \frac{\sigma}{t} \cdot \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

The Central Limit Theorem (Theorem 1.4) states that averages tend in distribution to a Gaussian. But what can be said about the distribution function itself? The following theorem gives this result:

Theorem 2.1 (Berry-Esseen CLT)**.** *Let $\{X_i\}_{i=1}^n$ be a sequence of identically and independently distributed random variables with mean μ and variance $\sigma^2 < \infty$. Define $\bar{X}_n = \frac{1}{n}\sum\limits_{i=1}^n X_i$ and $\bar{Z}_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$. Then:*

$$\left|\Pr(\bar{Z}_n > t) - \Pr(g > t))\right| \leq \frac{\rho}{\sqrt{n}}$$

where $\rho = \frac{\mathbb{E}\left[|X_i - \mu|^3\right]}{\sigma^3}$, $i \in [n]$ and $g \sim \mathcal{N}(0,1)$.

Remark. This theorem basically states that the error of approximation scales as $O\left(\frac{1}{\sqrt{n}}\right)$, which is bad, since we can't always leverage the normal approximation from the central limit theorem always.

2.1.1 Auxiliary Lemmata

Lemma 2.2. *Let $g \sim \mathcal{N}(0,1)$. For $t \geq 1$, we have that:*

$$\mathbb{E}\left[g^2 \mathbf{1}_{\{g>t\}}\right] = \frac{t}{\sqrt{2\pi}} e^{-t^2/2} + \Pr(g > t) \leq \left(t + \frac{1}{t}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[g^2 \mathbf{1}_{\{g>t\}}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \mathbf{1}_{\{x>t\}} e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} x^2 e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \left(\left(x \cdot e^{-x}\right)\Big|_{t^2/2}^{\infty} + \int_{t}^{\infty} e^{-x^2/2} dx \right) \qquad \left(\because \text{int. by parts with } f(x) = x, g(x) = xe^{-x^2/2}\right) \\
&= \frac{t}{\sqrt{2\pi}} e^{-t^2/2} + \Pr(g > t) \\
&\leq \frac{t}{\sqrt{2\pi}} e^{-t^2/2} + \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \\
&= \left(t + \frac{1}{t}\right) \frac{1}{\sqrt{2\pi}} e^{-t^2/2}
\end{aligned}$$

□

2.2 Hoeffding's inequality