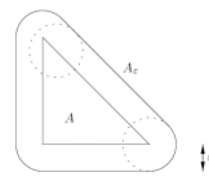
CS 798 - Convexity and Optimization, Winter 2017, Waterloo

Lecture 20: Measure concentration

We will discuss isoperimetric inequalities and then prove two concentration inequalities (one in the Gaussian setting, one in the sphere setting). The proofs will involve Brunn-Minkowski inequality and Prékopa-Leindler inequality.

Isoperimetric inequalities [Ball, lecture 8]

In L18, we have seen the classical isoperimetric inequality in $\mathbb{R}^n$.



from [Ball]

For each point $x \in \mathbb{R}^n$, let $d(x, A) := \min \{ \|x-y\|_2 \mid y \in A \}$ be the distance from x to A.

Let $A_\varepsilon := \{ x \in \mathbb{R}^n \mid d(x,A) \leq \varepsilon \}$ be the ε-neighborhood of A. Note that $A_\varepsilon = A + \varepsilon B_2^n$.

The Brunn-Minkowski inequality shows that if B is an Euclidean ball of the same volume as A, then we have $\mathrm{vol}(A_\varepsilon) \geq \mathrm{vol}(B_\varepsilon)$.

So, in a sense, the Euclidean balls are the least "expanding" sets in $\mathbb{R}^n$ under the Lebesgue measure.

We can ask the same question whenever we have a distance function d and a probability measure vol.

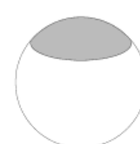
We will discuss two important examples in the following and see their probabilistic implications.

Sphere Let the domain be the sphere S^{n-1} in $\mathbb{R}^n$.

The distance function d is either the geodesic distance or the Euclidean distance inherited from $\mathbb{R}^n$.

The volume measure will be the rotational-invariant probability on S^{n-1}.

Lévy proved that the least expanding sets in this setting are the spherical caps:



[Ball]

If a subset A of the sphere has the same measure as a cap C of radius r,

then $\mathrm{vol}(A_\varepsilon) \geq \mathrm{vol}(C_\varepsilon)$, where C_ε is a cap of radius $r+\varepsilon$.

Gaussian The domain is $\mathbb{R}^n$, with the distance function d being the Euclidean distance.

The volume measure is the standard Gaussian probability measure on $\mathbb{R}^n$ with density $\gamma(x) = (2\pi)^{-\frac{n}{2}} e^{-\|x\|_2^2/2}$.

Borell proved that the least expanding sets in this setting are the half spaces:

In particular, if $A \subseteq \mathbb{R}^n$ and $\mathrm{vol}(A) = \frac{1}{2}$, then $\mathrm{vol}(A_\varepsilon) \geq \mathrm{vol}(H_\varepsilon)$ where H is the halfspace $\{x \in \mathbb{R}^n \mid x_1 \leq 0\}$.

Probabilistic implications

For the sphere setting, suppose A has the measure of a hemisphere H, which is normalized so that $\mathrm{vol}(H) = \frac{1}{2}$.

By Levy's isoperimetric theorem, we have $vol(A_\varepsilon) \geq vol(H_\varepsilon) \geq 1 - e^{-n\varepsilon^2}$ which we derived in L18.

So, for any set A with half the measure, almost the entire sphere lies within distance ε of A.

We can rephrase it in a functional setting that is closer to the concentration inequalities that we usually see.

Suppose $f: S^{n-1} \to \mathbb{R}$ is a function on the sphere that is 1-Lipschitz, i.e. $|f(x) - f(y)| \leq \|x-y\|_2$.

There exists at least one number M such that the sets $(f \leq M)$ and $(f \geq M)$ have measure at least half.

Since f is 1-Lipschitz, if a point x has distance at most ε from $A := (f \leq M)$, then $f(x) \leq M + \varepsilon$.

By the isoperimetric inequality, if we pick a random point on the sphere (e.g. using the sampling algorithm in L18),

then $\Pr_x(f(x) > M+\varepsilon) \leq 1 - vol(A_\varepsilon) \leq e^{-n\varepsilon^2}$.

This implies that with high probability the function value is close to the median value.

A similar conclusion can be drawn in the Gaussian setting: if $vol(A) = \frac{1}{2}$, then $vol(A_\varepsilon) \geq 1 - e^{-\varepsilon^2/2}$.

These are general conditions to guarantee measure concentration. and have various applications in different fields.

Next time, we will see one application in a combinatorial problem, discrepancy minimization.

Plan

We will not prove the isoperimetric inequalities, which are harder to prove.

Levy's theorem can be proved by a delicated symmetrization argument.

It is intuitive that the two theorems are closely related, as the Gaussian measure should concentrate on the sphere of radius $\sqrt{n}$, and in fact Borell's proof uses Levy's theorem.

Instead, we will directly prove these "approximate" isoperimetric / concentration inequalities, which follow from Brunn-Minkowski and Prékopa-Leindler inequalities elegantly.

Gaussian concentration inequality [Ball, lecture 8]

To motivate the statement, let us quickly review how some simple concentration inequalities are proved.

Let $X = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ where each X_i is +1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$, so that $Var[X] = 1$.

To show that X is concentrated around 0, the standard approach is to bound $E[e^{tX}]$ for $\lambda \in \mathbb{R}$,

where in this setting one can show that $E[e^{tX}] \leq e^{t^2/2}$ (see [Ball, lecture 7]).

Then, by Markov's inequality. $\Pr(X \geq t) = \Pr(e^{tX} \geq e^{t^2}) \leq E[e^{tX}]/e^{t^2} \leq e^{-t^2/2}$.

In the Gaussian setting, we will also bound $E[e^{tX}]$ where X is a Gaussian random variable.

Theorem Let $A \subseteq \mathbb{R}^n$ be measurable and let μ be the standard Gaussian measure on $\mathbb{R}^n$.

Then $\int e^{d(x,A)^2/4} \, d\mu \leq \frac{1}{\mu(A)}$, where the integral is $\mathbb{E}_{x \sim \mu}[d(x,A)^2/4]$.

proof The inequality follows from an application of Prékopa-Leindler inequality with $\lambda = \frac{1}{2}$ and with $f(x) = e^{d(x,A)^2/4} \gamma(x)$, $g(x) = \chi_A(x)\gamma(x)$ where $\chi_A(x) = 1$ if $x \in A$ and 0 o.w., and $m(x) = \gamma(x)$.

Then $\int_{\mathbb{R}^n} f(x) = \int e^{d(x,A)^2/4} d\mu$, $\int_{\mathbb{R}^n} g(x) = \mu(A)$ and $\int_{\mathbb{R}^n} m(x) = 1$, and so the inequality to prove can be written as $\left(\int_{\mathbb{R}^n} m(x)\right)^2 \geq \left(\int_{\mathbb{R}^n} f(x)\right)\left(\int_{\mathbb{R}^n} g(x)\right)$.

By the Prékopa-Leindler inequality, this will hold if $m((x+y)/2) \geq f(x)^{\frac{1}{2}} g(y)^{\frac{1}{2}}$ for all $x,y \in \mathbb{R}^n$.

So, to prove the inequality, it remains to check that $m((x+y)/2)^2 \geq f(x)g(y)$ for all $x,y \in \mathbb{R}^n$.

Recall that $\gamma(x) = (2\pi)^{-\frac{n}{2}} e^{-\|x\|_2^2/2}$.

If $y \notin A$, then $g(y) = 0$ and there is nothing to check.

Hence, we assume $y \in A$, and this implies that $d(x,A) \leq \|x-y\|_2$.

So, $f(x)g(y) = e^{d(x,A)^2/4} (2\pi)^{-\frac{n}{2}} e^{-\|x\|_2^2/2} (2\pi)^{-\frac{n}{2}} e^{-\|y\|_2^2/2}$

$$\leq (2\pi)^{-n} \exp\left(\|x-y\|_2^2/4 - \|x\|_2^2/2 - \|y\|_2^2/2\right) = (2\pi)^{-n} \exp\left(-\|x+y\|_2^2/4\right) = m((x+y)/2)^2.$$

Therefore, $m((x+y)/2)^2 \geq f(x)g(y)$ always holds and the inequality follows from Prékopa-Leindler inequality. $\square$

Corollary If $\mu(A) = \frac{1}{2}$, then $\mu(A_\varepsilon) \geq 1 - 2e^{-\varepsilon^2/4}$.

proof Every point $x \notin A_\varepsilon$ has $d(x,A) > \varepsilon$, and so $(1-\mu(A_\varepsilon)) e^{\varepsilon^2/4} \leq \int e^{d(x,A)^2/4} d\mu \leq \frac{1}{\mu(A)} = 2$.

Rearranging gives the corollary. $\square$

This is a slightly weaker bound (with the constant in the exponent being 4 instead of 2), but this makes no difference in most applications.

Concentration of the sphere [Barvinok, lecture 25]

It is possible to derive the following result using the previous result about Gaussians, but the proof is quite interesting and it can be extended to other strictly convex surfaces.

Theorem Let $S := S^{n-1}$. Let $X \subseteq S$ be a subset with $\mathrm{vol}(X)/\mathrm{vol}(S) \geq \frac{1}{2}$.

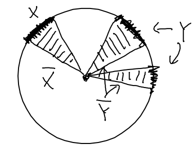
Then, for any $2 \geq \varepsilon > 0$, we have $\mathrm{vol}(X_\varepsilon)/\mathrm{vol}(S) \geq 1 - 2e^{-n\varepsilon^2/4}$.

Proof Let $Y := \{y \in S \mid d(y,X) \geq \varepsilon\}$ where $d(x,y) = \|x-y\|_2$ is the Euclidean distance in $\mathbb{R}^n$.

For any $x \in X$ and $y \in Y$, we have $\frac{1}{2}\|x+y\|^2 = \|x\|^2 + \|y\|^2 - \frac{1}{2}\|x-y\|^2$, and so $\|\frac{x+y}{2}\|^2 \le 1 - \frac{\varepsilon^2}{4}$ and thus $\|\frac{x+y}{2}\| \le 1 - \frac{\varepsilon^2}{8}$.

Consider $\bar{X} = \mathrm{conv}\{X \cup \{0\}\}$ and $\bar{Y} = \mathrm{conv}\{Y \cup \{0\}\}$ where $\bar{X}$ and $\bar{Y}$ are the cones with X and Y as surfaces.

Let $\bar{x} \in \bar{X}$ and $\bar{y} \in \bar{Y}$, so $\bar{x} = \alpha x$ for $0 \le \alpha \le 1$ and $x \in X$, and similary $\bar{y} = \beta y$ for $0 \le \beta \le 1$ and $y \in Y$.



Then, we also have $\|\frac{\bar{x}+\bar{y}}{2}\|_2 = \|\frac{\alpha x + \beta y}{2}\|_2 = \|\alpha(\frac{x+y}{2}) + (\beta-\alpha)\frac{y}{2}\|_2 \le \alpha(1-\frac{\varepsilon^2}{8}) + \frac{\beta-\alpha}{2} \le 1 - \frac{\varepsilon^2}{8}$.

Now, we apply the Brunn-Minkowski inequality to $\bar{X}$ and $\bar{Y}$.

The above calculations show that the Minkowski sum $(\bar{X}+\bar{Y})/2$ is contained in the ball $(1-\frac{\varepsilon^2}{8})B_2^n$.

Therefore,
$$(1-\tfrac{\varepsilon^2}{8})^n \,\mathrm{vol}(B_2^n) \ge \mathrm{vol}((\bar{X}+\bar{Y})/2)$$
$$\ge \mathrm{vol}(\bar{X})^{\frac{1}{2}}\,\mathrm{vol}(\bar{Y})^{\frac{1}{2}}$$ by the multiplicative Brunn-Minkowski inequality

Hence,
$$1 - \frac{\mathrm{vol}(X_\varepsilon)}{\mathrm{vol}(S)} = \frac{\mathrm{vol}(Y)}{\mathrm{vol}(S)} = \frac{\mathrm{vol}(\bar{Y})}{\mathrm{vol}(B_2^n)} \le (1-\frac{\varepsilon^2}{8})^{2n}\frac{\mathrm{vol}(B_2^n)}{\mathrm{vol}(\bar{X})} = (1-\frac{\varepsilon^2}{8})^{2n}\frac{\mathrm{vol}(S)}{\mathrm{vol}(X)} \le 2(1-\frac{\varepsilon^2}{8})^{2n} \le 2e^{-n\varepsilon^2/4}. \quad \square$$

It is not difficult to extend it to strictly convex surface; see [Barvinok].

References: [Ball, lecture 8].

- Barvinok, Math 710: Measure concentration. lecture 25, 2005.