Convex Geometry and Concentration of Measure in High Dimension

December 1, 2015

Joseph Zappa

UNC Chapel Hill

December 1st, 2015

Overview

Questions

- How does convexity influence the properties of certain sets?
- How close is a function of random variables to its mean?
- Can convexity be used to create powerful inequalities relating to functions of random variables?
- Can concentration inequalities yield useful properties about convex sets?

Convex Geometry

Overview

- Concerned with the study of compact, convex subsets of $\mathbb{R}^n$
- Key Idea: All convex shapes behave a bit like the Euclidean ball
- Will focus on finding subspaces with desirable properties with high dimension

Examples

- The cube
- The n-dimensional regular solid simplex
- The cross-polytope
- The Euclidean ball

A Notion of Distance

Definition

Let K, L be convex bodies then $d(K, L) =$ the smallest d such that there exists a linear image of $\tilde{L}$ s.t. $\tilde{L} \subset K \subset d\tilde{L}$

Important Properties

- This is not a proper metric, if $K = L$ then $d(K, L) = 1$
- The distance is multiplicative, to get an additive one must take $\log(d)$
- Under this distance a polytope must have exponentially many facets to be close to the Euclidean ball

An Interesting Result

Theorem

Let K be a (symmetric) polytope in $\mathbb{R}^n$ with $d(K, B_2^n) = d$. Then K must have at least $\exp(\frac{n}{2d^2})$ facets. Moreover $\forall n \geq 1, \exists$ a polytope, L, with 4^n facets such that $d(L, B_2^n) \leq 2$.

Proof Concept

Look at spherical caps and bound the area of each cap. Then we can bound the area of any symmetric convex polytope inscribed in the Euclidean ball.

Spherical Caps

Fix a unit vector v and $\epsilon \in [0, 1)$, then the set $C(\theta, v) = \{\theta \in S^{n-1} : \langle \theta, v \rangle \geq \epsilon\}$ is called a spherical cap. For $0 \leq \epsilon \leq 1$, the cap $C(\epsilon, u)$ on S^{n-1} has measure at most $e^{-n\epsilon/2}$

John's Theorem

Theorem

Each Convex body K contains an unique ellipsoid of maximal volume. This is B_2^n iff

1. $B_2^n \subset K$
2. $\exists m$ s.t.
   1. there are Euclidean unit vectors $\{u_i\}_i^m$ on the boundary of K
   2. $\{c_i\}_i^m$, $c_i \in \mathbb{R}^+$ such that $\sum_{i=1}^m c_i u_i = 0$ and $\sum_{i=1}^m c_i \langle x, u_i \rangle^2 = |x|^2 \; \forall x \in \mathbb{R}$.

- Condition (1) guarantees that $\{u_i\}$ don't all lie on one side of the sphere
- The second condition insures that all of the $\{u_i\}$ do not lie too close to a proper subspace of $\mathbb{R}^n$
- This means that $\{u_i\}$ behaves like an orthonormal basis of K

The Brunn-Minkowski Inequality

Theorem

If A and B are non-empty compact subsets of $\mathbb{R}^n$ and $\lambda \in (0,1)$ then

$$\text{Vol}((1-\lambda)A + \lambda B) \geq (1-\lambda)\text{Vol}(A)^{1/n} + \lambda \text{Vol}(B)^{1/n}$$

Corollary

If A and B are compact then for fixed $\lambda \in (0,1)$

$$\text{Vol}((1-\lambda)A + \lambda B) \geq \text{Vol}(A)^{1-\lambda} + \text{Vol}(B)^{\lambda}$$

The Corollary is Useful

1. We no longer need A and B to be non-empty
2. The inequality is dimension neutral

The Concentration Phenomenon

The Idea

If $X_1, ..., X_n$ are independent (or weakly dependent) random variables, then the random variable $f(X_1, ..., X_n)$ is close to its mean provided f is not too "sensitive" to its coordinates.

Lemma

Let X be any random variable. Then

$$\mathbb{V}[f(X)] \leq \frac{1}{4}(\sup f - \inf f)^2$$

$$\mathbb{V}[f(X)] \leq \mathbb{E}[(f(X) - \inf f)^2]$$

Tensorization

Definition

Define the new variables

$$\mathbb{V}_i[f(X_1, ..., X_n)] := \mathbb{V}[f(x_1, ..., x_{i-1}, X_i, x_{i+1}, ..., x_n)]$$

Theorem

If $X_1, ..., X_n$ are independent then

$$\mathbb{V}[f(X_1, ..., X_n)] \leq \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{V}_i[f(X_1, ..., X_n)]\right]$$

Things to Note

- If f is linear this holds with equality $\Rightarrow$ gives an exact bound on the LLN
- Generally difficult to calculate $\mathbb{V}_i f$ so produce bounds similar to Chernoff type inequalities

Two More Useful Inequalities

Definition

$$D_i f := \sup_z f(x_1, ..., x_{i-1}, z, x_{i+1}, ..., x_n) - \inf_z f(x_1, ..., x_{i-1}, z, x_{i+1}, ..., x_n)$$

$$D_i^- f := f(x_1, ..., x_n) - \inf_z f(x_1, ..., x_{i-1}, z, x_{i+1}, ..., x_n)$$

Corollary

If $X_1, ..., X_n$ are independent then

$$\mathbb{V}[f(X_1, ..., X_n)] \leq \frac{1}{4}\mathbb{E}\left[\sum_{i=1}^{n}(D_i f(X_1, ..., X_n))^2\right]$$

$$\mathbb{V}[f(X_1, ..., X_n)] \leq \mathbb{E}\left[\sum_{i=1}^{n}(D_1^- f(X_1, ..., X_n))^2\right]$$

Concentration on Convex Spaces

Theorem

Let $X_1, ..., X_n$ be iid Bernoulli random variables with $p = 1/2$ (also known as the Rademacher Distribution) then $\forall \epsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \epsilon\right) \leq 2\exp\left(-\frac{n\epsilon^2}{2(1+\epsilon/3)}\right)$$

- The above is the simplest Bernstein inequality. Proofs of these types of inequalities typically revolve around leveraging the Markov inequality and applying it to $\exp\left(\lambda \sum_{i=1}^{n} X_i\right)$.
- Examples include Chernoff bounds, Azuma, and Hoeffding inequalities

ϵ-Neighborhoods of Sets

Definition

- Let $A \subset \mathbb{R}^n$ be compact then $\forall x \in \mathbb{R}$ define the distance $d(x, A) = \min\{|x - y| : y \in A\}$.
- $\forall \epsilon > 0$ the set addition $A + \epsilon B_2$ are exactly the points whose distance is at most ϵ from A we will denote this as A_ϵ and it is the ϵ-neighborhood of the set A.

- We will compare volumes of sets and their ϵ-neighborhoods.
- The Brunn-Minkowski inequality tells us that the sphere has the smallest surface area for a convex body of a given volume. This implies that if B is a Euclidean ball and Vol$(B) =$ Vol(A) then Vol$(A_\epsilon) \geq$ Vol(B_ϵ), $\forall \epsilon > 0$.

Example

The Setting

- Let $\Omega = S^{n-1} \subset \mathbb{R}^n$ with the Euclidean distance in $\mathbb{R}^n$
- Let $\mu = \sigma_{n-1}$ be the rotation invarient measure on S^{n-1}
- Then the sets with smallest surface area of a given volume are spherical caps
- Look at convex sets of volume $\mu(A) = 1/2$

The Result

- Note that A has the measure of a hemisphere H
- $\forall \epsilon$ we have $\mu(A_\epsilon) \geq \mu(H_\epsilon)$
- But we have an upper bound on the area of spherical caps and H_ϵ^c is a spherical cap
- So $\mu(H_\epsilon^c) \approx e^{-n\epsilon^2/2} \Rightarrow \mu(A_\epsilon) \geq 1 - \exp(-n\epsilon^2/2)$

Another Example

The Setting

Suppose $f : S^{n-1} \to \mathbb{R}$ is 1-Lipschitz and define the same measure and metric as in the prior example.

The Result

- Lipschitz condition guarantees differentiability
- There exists $M = \text{med}(f)$ such that $\sigma(f \leq M) \geq 1/2$ and $\sigma(f \geq M) \geq 1/2$
- $d(x, \{f \leq M\}) \leq \epsilon$ then $f(x) \leq M + \epsilon$

$$\Rightarrow \sigma(f > M + \epsilon) \leq \exp(-n\epsilon^2/2)$$

$$\sigma(f < M - \epsilon) \leq \exp(-n\epsilon^2/2)$$

$$\Rightarrow \sigma(|f - M| > \epsilon) \leq 2\exp(-n\epsilon^2/2)$$

- So f is almost equal to M on almost all of the sphere.

One Final Example

The Setting

- Let $\Omega = \mathbb{R}^n$ with the Euclidean distance metric and $\mu = N_n(0,1)$ or $\gamma(x) = (2\pi)^{-n/2}\exp(-|x|^2/2)$
- Half-spaces are the sets with smallest surface area of a given volume (e.g. $H = \{x \in \mathbb{R}^n : x_1 \leq 0\}$)

The Result

- $\mu(H_\epsilon^c) = \frac{1}{\sqrt{2\pi}}\int_\epsilon^\infty \exp(-t^2/2)dt \leq e^{-\epsilon^2/2}$
- $\mu(A_\epsilon) \geq 1 - \exp(-\epsilon^2/2)$
- It appears to be weaker than previous example due to dimension invariance but the Gaussian measure is not dimension invariant and in fact the Gaussian concentrates on a spherical shell of thickness approximately 1 and radius $\sqrt{n}$

Dvoretzky's Theorem

Theorem

$\forall \epsilon > 0$ and $n \in \mathbb{N}$, $\exists c > 0$ s.t. every symmetric convex body of dimension n has a slice of dimension

$$k \geq \frac{c\epsilon^2}{\log(1+\epsilon^{-1})} \log(n)$$

which is at most $1 + \epsilon$ away from the k-dimensional Euclidean ball.

Proof Idea

- Prove for cube and build cube around convex set
- For cube need to estimate

$$\frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \max |x_i| d\mu(x)$$

- Use previous example to show median is not too far from the maximum and thus simplify the integral

Conclusion

- We have shown that assuming convexity gives us powerful tools for bounding the volume of sets and for solving the isoperimetric problem
- If we allow certain assumptions about independence or boundedness we can achieve strong inequalities describing how concentrated a function of random variables is about its mean. To achieve better bounds we need more complex machinery such as Markov semi-groups
- Because the Gaussian concentrates on a sphere we are able to deduce more advanced concentration inequalities
- Concentration of measure can be used to prove Dvoretzky type theorems about almost ellipsoidal slices of large dimension