Sections of convex bodies, statistical estimation and (in)stability

Luis Rademacher
(joint work with Navin Goyal)
Minkowski’s uniqueness theorem (injectivity)

- Theorem: If two centrally symmetric convex bodies have the same central sectional \((n - 1)\)-dimensional volumes, then they are equal.
- Stability?
- Upper bounds on error can be found in Groemer’s book (inverse spherical Radon transform). Not useful for algorithms as they need too high precision.
- Focus here on distinguishing between polynomial and exponential dependence on dimension.
Distance

\[ d(K, L) = \frac{\text{vol}(K \setminus L)}{\text{vol } K} \]

\[ \text{vol } K \geq \text{vol } L \]

- \(d(K, L)\) = total variation distance (statistical distance) between uniform distributions on \(K\) and \(L\), \(f_K\) and \(f_L\), respectively:

\[ d(K,L) = \frac{1}{2} \int |f_K(x) - f_L(x)|dx \]
Stability of reconstruction

• Relative sectional volumes for $\theta \in S^{n-1}$:

$$A_K(\theta) := \frac{vol_{n-1}(K \cap \theta^\perp)}{vol(K)}$$

• Can determine volume from it, so it is enough to determine centrally symmetric convex bodies (!)

• **Corollary** [Goyal R.]
  There exists universal constants $0 < c < 1$, $c' > 0$ such that for all $n$ large enough:
  There exist two centrally symmetric convex bodies $K, L \subseteq R^n$ such that

$$\sup_\theta |A_K(\theta) - A_L(\theta)| < c^{\sqrt{n}}$$

but $d_{TV}(K, L) > c'$.

• Relative v/s absolute not really important.
It is a corollary of

**Theorem** [Goyal R ‘09]: There exists a distribution $D$ on $n$-dimensional convex bodies and $c > 1$ such that if

- $f$ is a function that takes $q$ points and outputs a convex body and,
- when given $q$ random points $X_1, \ldots, X_q$ from a convex body $K$ according to $D$ it satisfies

$$P \left( d_{TV} \left( K, f(X_1, \ldots, X_q) \right) < 1/8 \right) \geq \frac{1}{2}. $$

Then $q > c^{\sqrt{n}}$. 
Idea: from random points to sections

• Approximate area of section by volume of narrow band around it.
• Given $m$ random points from $K$, approximate relative volume of band by fraction of points in it.
• Concentration bound implies additive error for relative volume of band to within $\varepsilon$ with $O\left(\frac{1}{\varepsilon^2}\right)$ points for a single band.
• How to get all bands simultaneously?
• Want uniform convergence over all bands.
Connection between samples and relative sectional areas

- Use Vapnik Chervonenkis theory, “uniform convergence”.
- If \( X = R^n \), \( H = \) set of all halfspaces then VC-dimension \( d \) of \((X, H)\) is \( n + 1 \).
- **Theorem** (Vapnik and Chervonenkis) (restatement of [Anthony and Bartlett, thm 4.3]):
  Suppose that \( H \) is a family of subsets of a set \( X \) and that \((X, H)\) has finite VC-dimension \( d \). Let \( D \) be a probability distribution on \( X \). For any \( 0 < \epsilon < 1 \) and \( m > d \) a positive integer we have

\[
P_{X_1, \ldots, X_m \sim_D} \left( \left( \forall A \in H \right) \left| P_D(A) - \frac{\# \{ i : X_i \in A \}}{m} \right| < \epsilon \right) \geq 1 - 4m^{d+1} e^{-\epsilon^2 m/8}
\]
Connection between samples and relative sectional areas

- **Corollary:**
  Let $H$ be the set of all bands of the form
  \[\{x \in \mathbb{R}^n : v \cdot x \in [a, b]\}\]
  for any $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$
  and $v \in \mathbb{R}^n$. Let $D$ be a probability distribution on $\mathbb{R}^n$.
  For $0 < \varepsilon < 1$ and $m > 10n$ a positive integer we have:

  \[
P_{X_1,\ldots,X_m \sim D} \left( (\forall A \in H) \left| P_D(A) - \frac{\#\{i : X_i \in A\}}{m} \right| < \varepsilon \right) \geq 1 - 4m^{10n} e^{-\varepsilon^2 m/8}
  \]

- **Proof:** VC-dimension of bands is less than $10n$. 

Connection between samples and relative sectional areas

Given a centrally symmetric convex body $K \subseteq \mathbb{R}^n$ and $\theta \in S^{n-1}$, let $A_K(\theta)$ denote the $(n - 1)$-dimensional area of the section orthogonal to $\theta$ relative to $\text{vol}(K)$:

$$A_K(\theta) := \frac{\text{vol}_{n-1}(K \cap \theta^\perp)}{\text{vol} K}.$$  

Given $\delta > 0$, let $A_{K,\delta}$ be the following $\delta/2$-neighborhood approximation to $A_K$:

$$A_{K,\delta}(\theta) := \frac{1}{\delta} \frac{\text{vol}(K \cap \{x : |\theta \cdot x| \leq \delta/2\})}{\text{vol} K}.$$  

Given a random sample $X_1, \ldots, X_m$ from $K$, let $\tilde{A}_{K,\delta}$ be the following sample approximation to $A_{K,\delta}$ (and $A_K$):

$$\tilde{A}_{K,\delta}(\theta) := \frac{1}{\delta} \frac{\#\{i : |\theta \cdot X_i| \leq \delta/2\}}{m}.$$
Connection between samples and relative sectional areas

For any direction $\theta$, among all hyperplane sections of $K$ perpendicular to $\theta$, the one with maximal volume is the section containing the origin. This implies $A_{K,\delta}(\theta) \leq A_K(\theta)$ for all $\delta > 0$. If $K$ is isotropic (covariance = identity), then it contains the ball of radius 1. This implies $K \cap (\theta^\perp + \delta \theta) \supseteq (1 - \delta) K \cap \theta^\perp + \delta \theta$ and $\text{vol}_{n-1}(K \cap (\theta^\perp + \delta \theta)) \geq (1 - \delta)^{n-1} \text{vol}_{n-1}(K \cap \theta^\perp)$. Thus,

$$A_{K,\delta}(\theta) \geq \left(1 - \frac{\delta}{2}\right)^{n-1} A_K(\theta) \geq \left(1 - \frac{(n-1)\delta}{2}\right) A_K(\theta).$$

If $\delta \leq 2\epsilon/(n - 1)$ we have

$$(1 - \epsilon) A_K(\theta) \leq A_{K,\delta}(\theta) \leq A_K(\theta).$$

Using that $B_n \subseteq K \subseteq c\sqrt{n}B_n$, we also have $c/\sqrt{n} \leq A_K(\theta) \leq 2n$. This implies the previous multiplicative bound on $A_{K,\delta}$ is also additive: If we let $\epsilon' = \epsilon/(2n)$ (so $\delta \leq \frac{\epsilon'}{n(n-1)}$) we have

$$A_K(\theta) - \epsilon' \leq A_{K,\delta}(\theta) \leq A_K(\theta).$$

(1)
Connection between samples and relative sectional areas

• **Corollary:**
  Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric isotropic convex body. Let $X_1, \ldots, X_m$ be iid and uniform in $K$. Let $0 < \varepsilon'' < 1$. Then for $\delta = \varepsilon''/(2n^2)$ and $m > 10n$ we have:
  \[
P \left( (\forall \theta \in S^{n-1}) | \tilde{A}_{K,\delta}(\theta) - A_K(\theta) | < \varepsilon'' \right) \\
  \geq 1 - 4m^{10n} e^{-\varepsilon''^4 m/(16n^2)}
  \]

• In particular, for $m = \text{poly} \left( n, \frac{1}{\varepsilon''}, \log \frac{1}{\delta} \right)$ the rhs is at least $1 - \delta$. 
Proof of the initial claim

• Corollary [Goyal R.]
There exists universal constants $0 < c < 1$, $c' > 0$ such that for all $n$ large enough:
There exist two centrally symmetric convex bodies $K, L \subseteq R^n$ such that
\[
\sup_{\theta} |A_K(\theta) - A_L(\theta)| < c\sqrt{n}
\]
but $d_{TV}(K, L) > c'$.

• Proof: Statistical lower bound in [G R’09] gives finite family of convex bodies such that any pair is at TV distance at least $c'$ but to be estimated need more than $c_2^{\sqrt{n}}$ random points. At the same time VC corollary implies $c_2^{\sqrt{n}}$ points are enough to estimate $A_K(\theta)$ to within additive $c\sqrt{n}$ for some $c > 0$. If the corollary were false, one could pinpoint the input body. This is a contradiction.
VC-dimension

- $X$: any set
- $H$: any family of subsets of $X$
- VC-dim of $(X, H)$ is maximum $\#S$ over $S \subseteq X$ such that $\{S \cap A : A \in H\} = 2^S$.
- VC-dimension of halfspaces in $\mathbb{R}^n$ is $n + 1$ by Radon’s theorem
  “Any set of $n + 2$ points in $\mathbb{R}^n$ can be partitioned into two subsets so that their convex hulls intersect”.
Instability of volume

• Exact relative sectional areas determine volume (based on inverse spherical Radon transform).

• [Eldan] For some $\epsilon > 0$, need $2^{n^\epsilon}$ random points to determine volume within factor $2^{n^\epsilon}$ with probability $2^{-n^\epsilon}$.

⇒ As before then, volume needs relative sectional areas to within very high precision.
Instability of volume

- **Corollary:** There exists a universal constant $\epsilon > 0$ such that for all $n$ large enough: There exist two centrally symmetric convex bodies $K, L \subseteq R^n$ such that

$$\sup_{\theta} |A_K(\theta) - A_L(\theta)| < 2^{-n\epsilon}$$

but $\text{vol}(K)/\text{vol}(L) > 2^{n\epsilon}$
Message

• Instability of a functional (like volume) from random points implies instability from relative sectional areas.

• Argument is robust: instability from any other family with VC-dimension at most $poly(n)$.

• Also for Gaussian measure:
Similar result for Gaussian measure

• [Klivans O’Donnell Servedio] Need $2^{\Omega(\sqrt{n})}$ samples for reconstruction under Gaussian distribution (getting +,- labeled examples). (Algorithm with matching complexity.)
Questions

• Motivation: efficient (algorithmic) estimation of polytopes with \( \text{poly}(n) \) facets (or vertices) from random points.

• (Q1) Injectivity of polytopes with \( \text{poly}(n) \) (say, \( "n^2" \)) facets (or vertices) from first "100" moment tensors.
  • [Gravin Lasserre Pasechnik Robins ‘11] Can recover a \( n \)-polytope with \( v \) vertices from \( O(nv) \) moments along \( n \) random directions. But efficient estimation of higher order moments unlikely from samples.
  • [Frieze Jerrum Kannan ‘96] Can estimate parallelepiped from first 4 moment tensors

• (Q2) Stability of reconstruction of polytopes with \( \text{poly}(n) \) facets (or vertices) from central sectional areas. Algorithmic?