The Cone of Sufficient Statistics

Outline

1. Polyhedral Geometry
2. Discrete Exponential Families
3. Gaussian Exponential Families

① Polyhedral Geometry

Def: a) A set $S \subseteq \mathbb{R}^d$ is **convex** if for all $p, q \in S$ and all $\lambda \in [0,1]$, $\lambda p + (1-\lambda) q \in S$.

b) The **convex hull** $\mathrm{conv}(S)$ of a set $S \subseteq \mathbb{R}^d$ is the smallest convex set that contains S.

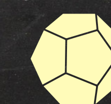
Note: $\mathrm{conv}(S)$ = intersection of all convex sets that contain S.

can be $\emptyset$ or $\mathbb{R}^d$

Def: a) For $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$, $H_{a,b} := \{p \in \mathbb{R}^d \mid a^T p \leq b\}$ is a **half-space**.

b) A set $P \subseteq \mathbb{R}^d$ is a **polyhedron** if it is the intersection of finitely many half-spaces.

c) A **polytope** is a bounded polyhedron.



d) The **dimension** of a polyhedron is the dimension of the smallest affine space containing it.

Example: $C_d := \{p \in \mathbb{R}^d \mid 0 \leq p_i \leq 1 \text{ for all } i\}$ is **d-dimensional hypercube**.

$\Delta_{d-1} := \{p \in \mathbb{R}^d \mid p_i \geq 0, \sum_{i=1}^{d} p_i = 1\}$ is **(d-1)-dimensional simplex**.

note: equality constraint $a^T p = b$ is intersection of 2 half-spaces: $a^T p \leq b$ & $a^T p \geq b$

Def: a) A **face** of a polyhedron $P \subseteq \mathbb{R}^d$ is a set of the form $\{p \in \mathbb{R}^d \mid a^T p = b\} \cap P$ where the half-space $H_{a,b}$ contains P.

b) A **vertex** of P is a zero-dimensional face of P.

c) An **edge** of P is a one-dimensional face of P.

d) A **facet** of P is a $(\dim(P)-1)$-dimensional face of P.

Note:
- A face of a polyhedron is a polyhedron.
- $\emptyset$ and P are faces of P $\quad (\dim(\emptyset) = -1)$.
- Every polyhedron has finitely many faces.

Def: Let $P \subseteq \mathbb{R}^d$ and $Q \subseteq \mathbb{R}^e$ be polyhedra.

a) P and Q are **affinely isomorphic** if there are affine transformations $\phi: \mathbb{R}^d \to \mathbb{R}^e$ and $\psi: \mathbb{R}^e \to \mathbb{R}^d$ such that $\phi(P) = Q$ and $\psi(Q) = P$.

b) P and Q are **combinatorially equivalent** if there is a bijection between the faces of P and the faces of Q which preserves the incidences between faces.

Note:
- aff. isom. $\Rightarrow$ comb. equiv.

- P comb. equiv. to Δ_{k-1} $\Rightarrow$ P aff. isom. to Δ_{k-1}.

- P comb. equiv. to C_k $\not\Rightarrow$ P aff. isom. to C_k.

Thm: a) The convex hull of any finite set is a polytope.

b) Every polytope is the convex hull of its set of vertices.

How to do this for polyhedra?

Def: A polyhedral cone is a set of the form $C=\{p\in\mathbb{R}^d \mid Ap\le 0\}$ for $A\in\mathbb{R}^{k\times d}$.

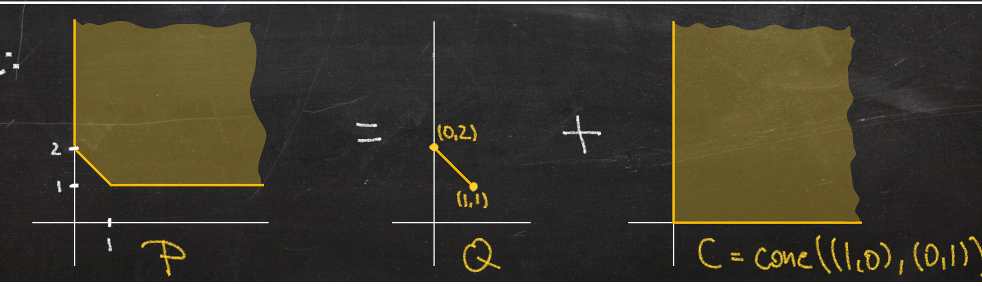
- Equivalently, a polyhedral cone is a polyhedron where all defining half-spaces pass through the origin.
- A polyhedral cone can have at most one vertex (the origin).

Def: a) A polyhedral cone is pointed if it has the origin as a vertex.
b) The cone generated by $V=\{v_1,\dots,v_n\}\subseteq\mathbb{R}^d$ is
$\text{cone}(V):=\{\lambda_1 v_1+\dots+\lambda_n v_n \mid \lambda_1,\dots,\lambda_n\ge 0\}$.
c) The Minkowski sum of two sets $S,T\subseteq\mathbb{R}^d$ is $S+T:=\{p+q \mid p\in S, q\in T\}$.

Note: a) A polyhedral cone is pointed if and only if it does not contain any lines.
b) For a pointed polyhedral cone C, there is a unique minimal set of vectors V (up to scaling) such that $\text{cone}(V)=C$. They are called the extreme rays of C.
c) $S,T\subseteq\mathbb{R}^d$ convex $\Rightarrow$ $S+T$ convex

Thm (Minkowski-Weyl)
a) For two finite sets $V_1,V_2\subseteq\mathbb{R}^d$, $\text{conv}(V_1)+\text{cone}(V_2)$ is a polyhedron.
b) Every polyhedron is of this form.
c) Let $P=C+Q$ where P is a polyhedron, C a polyhedral cone, Q a polytope.
(i) Then C is uniquely determined by P, called recession cone of P.
(ii) If P contains no lines, Q can be taken to be the convex hull of the set of vertices of P.

Example:



② Discrete Exponential Families

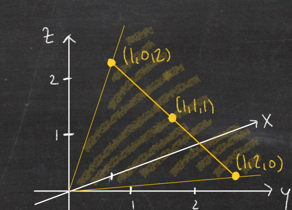
Recall: The discrete exponential families are the log-affine models:
$$M_{A,h}=\{p\in\Delta_{r-1} \mid \log(p)\in\log(h)+\text{rowspan}(A)\}$$
where $A\in\mathbb{Z}^{k\times r}$, $\mathbb{1}\in\text{rowspan}(A)$, $h\in\mathbb{R}^r_{>0}$.

- Given i.i.d. samples $X^{(1)},\dots,X^{(n)}$, the vector of counts $u\in\mathbb{N}^r$ is given by $u_j=|\{i \mid X^{(i)}=j\}|$.
- The vector of sufficient statistics is Au.
- The set of all vectors of sufficient statistics is the affine semigroup $\mathbb{N}A:=\{Au \mid u\in\mathbb{N}^r\}$.
- The cone of sufficient statistics is the polyhedral cone $\text{cone}(A):=\{Au \mid u\in\mathbb{R}^r_{\ge 0}\}=\text{cone}(a_1,\dots,a_r)$ where $A=[a_1\ a_2\ \cdots\ a_r]$

Thm: Given a vector of counts $u\in\mathbb{N}^r$, the maximum likelihood estimate exists in the model $M_{A,h}$ (and is unique) if and only if Au lies in the relative interior of $\text{cone}(A)$.

In that case, the MLE is the unique $p\in M_{A,h}$ such that $Au=nAp$ where $n:=\sum u_i$.

Example: $A=\begin{bmatrix}1&1&1\\2&0&1\\0&2&1\end{bmatrix}$



$\dim(\text{cone}(A))=2$
$\Rightarrow \text{int}(\text{cone}(A))=\emptyset$ but
$\text{relint}(\text{cone}(A))\neq\emptyset$

③ Gaussian Exponential Families

Notation:
- $\mathbb{S}_m$ = space of $m \times m$ symmetric matrices
- $PD_m \subseteq \mathbb{S}_m$: cone of positive definite matrices
- $PSD_m \subseteq \mathbb{S}_m$: cone of positive semidefinite matrices

The Gaussian exponential families are exactly the models of the form $\mathcal{M}_L := \{(\mu, \Sigma) \mid (\Sigma\mu, \Sigma^{-1}) \in L \cap (\mathbb{R}^m \times PD_m)\}$ where $L \subseteq \mathbb{R}^m \times \mathbb{S}_m$ is a linear subspace.

mean in $\mathbb{R}^m$ — covariance matrix in PD_m

(see Lian's lecture notes)

Assume: $L \subseteq \{0\} \times \mathbb{S}_m$, i.e. centered Gaussian exponential families.
$\Rightarrow$ ignore mean and write $L \subseteq \mathbb{S}_m$ & $\mathcal{M}_L = \{\Sigma \mid \Sigma^{-1} \in L \cap PD_m\}$.

- Given i.i.d. samples $X^{(1)}, \dots, X^{(n)}$, the sample covariance matrix is $S = \sum_{i=1}^{n} X^{(i)} (X^{(i)})^T \in PSD_m$.
- The sufficient statistics is $\pi_L(S)$ where $\pi_L : \mathbb{S}_m \longrightarrow L$ is the orthogonal projection onto L.
  w.r.t. trace inner product $(A, B) \mapsto tr(AB)$ for $A, B \in \mathbb{S}_m$
- The cone of sufficient statistics is $\pi_L(PSD_m)$.
  usually not polyhedral 

Thm: Given $S \in PSD_m$, the maximum likelihood estimate exists in the model $\mathcal{M}_L$ (and is unique) if and only if $\pi_L(S)$ lies in the relative interior of $\pi_L(PSD_m)$. In that case, the MLE is the unique $\hat{S} \in \mathcal{M}_L$ such that $\pi_L(S) = \pi_L(\hat{S})$.

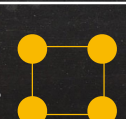
Gaussian graphical models

Def: The Gaussian graphical model associated to a graph $G = (V, E)$ is
$\mathcal{M}_G := \mathcal{M}_{L(G)}$ where $L(G) := \{K \in \mathbb{S}_m \mid K_{ij} = 0$ if $i \neq j$ and $ij \notin E\}$. ← linear space

$$\Rightarrow \pi_G := \pi_{L(G)} : \mathbb{S}_m \longrightarrow \mathbb{R}^V \oplus \mathbb{R}^E \qquad (m := |V|)$$
$$S \longmapsto (s_{ii})_{i \in V} \oplus (s_{ij})_{ij \in E}$$

Understanding the cone of sufficient statistics $\pi_G(PSD_m)$ is equivalent to:

Positive Semidefinite Matrix Completion Problem:
Given a graph G and a partially observed symmetric matrix S°, determine whether or not there is an $S \in PSD_m$ such that $\pi_G(S) = S^\circ$.

Example: $G =$ 
$$S^\circ = \begin{bmatrix} 1 & 2 & x & -2 \\ 2 & 1 & 2 & y \\ x & 2 & 1 & 2 \\ -2 & y & 2 & 1 \end{bmatrix}$$
Can we find x and y such that S° becomes PSD?

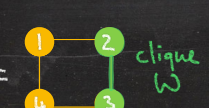
Graph Basics

Def: Let $G = (V, E)$ be a graph.

a) Given a subset $W \subseteq V$, the induced subgraph is the graph $G_W = (W, E_W)$ where $E_W := \{ij \mid i, j \in W$ & $ij \in E\}$.

b) A clique is a subset $W \subseteq V$ such that G_W is a complete graph.

c) Given $S \in \mathbb{S}_m$ and $W \subseteq V$, $S_W \in \mathbb{S}_{|W|}$ denotes the submatrix formed by the rows and columns of S that are indexed by W.

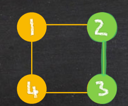
Example:
$G =$ 

$$S = \begin{bmatrix} 1 & 2 & x & -2 \\ 2 & 1 & 2 & y \\ x & 2 & 1 & 2 \\ -2 & y & 2 & 1 \end{bmatrix} \quad S_W$$

Back to: Positive Semidefinite Matrix Completion Problem

Prop: Let $G=(V,E)$ be a graph with a clique $W \subseteq V$ and let $S^\circ = \Pi_G(S)$ for some $S \in \mathcal{S}_m$.
If there is an $S' \in PSD_m$ such that $\Pi_G(S') = S^\circ$, then $S_W \in PSD_{|W|}$.

Example:

$G=$ 

clique W

$$S^\circ = \begin{bmatrix} 1 & 2 & x & -2 \\ 2 & 1 & 2 & y \\ x & 2 & 1 & 2 \\ -2 & y & 2 & 1 \end{bmatrix}$$

$S_W \notin PSD_2$

S° cannot be completed to a PSD matrix!

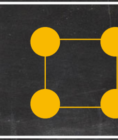
Def: $S \in \mathcal{S}_m$ satisfies the clique condition with respect to a graph G if $\det(S_W) \geq 0$ for all cliques W of G. $\leftarrow$ equivalently, $S_W \in PSD_{|W|}$ $\forall$ cliques W

By Prop., every S° that can be completed to a PSD matrix satisfies the clique condition.

When is this an "if and only if"?

Def: A graph $G=(V,E)$ is chordal if on every cycle $(v_0, v_1, \ldots, v_k = v_0)$ in G of length $k \geq 4$, there is a pair of vertices v_i, v_j with $i-j \not\equiv -1, 0, 1 \mod k$ such that $v_i v_j \in E$.

Equivalently, a graph is chordal if every induced subgraph that is a cycle is a 3-cycle.



not chordal

Def: a) A graph $G=(V,E)$ has a reducible decomposition into induced subgraphs $G_1=(V_1,E_1)$ and $G_2=(V_2,E_2)$ if
① $G_1 = G_{V_1}$ and $G_2 = G_{V_2}$, ② $G = G_1 \cup G_2$, and
③ $G_{V_1 \cap V_2}$ is complete.

b) A graph with a reducible decomposition is reducible.

c) A graph is decomposable if it is complete or reducible into decomposable subgraphs G_1 and G_2.

Thm (Dirac): A graph is chordal if and only if it is decomposable.

Example:

chordal graph

decomposition into maximal cliques

Thm: If G is a chordal graph, then $S \in \mathcal{S}_m$ satisfies the clique condition with respect to G if and only if there is an $S' \in PSD_m$ such that $\Pi_G(S) = \Pi_G(S')$.

As soon as G is not chordal, there are other conditions that must be satisfied to guarantee that there is a PSD matrix completion!

Example:

$G=$ 

$$S^\circ = \begin{bmatrix} 1 & \alpha & x & -\alpha \\ \alpha & 1 & \alpha & y \\ x & \alpha & 1 & \alpha \\ -\alpha & y & \alpha & 1 \end{bmatrix}$$

For which α can we find x and y such that S° becomes PSD?

① clique condition:

$\det \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \geq 0 \iff \alpha \in [-1, 1]$

$\iff$ all 1x1 and 2x2 principal minors are non-negative

② 3x3 principal minors:

$0 \leq \det \begin{bmatrix} 1 & \alpha & x \\ \alpha & 1 & \alpha \\ x & \alpha & 1 \end{bmatrix} = -(x-1)(x+1-2\alpha^2)$

$\iff x \in [2\alpha^2 - 1, 1]$

$0 \leq \det \begin{bmatrix} 1 & x & -\alpha \\ x & 1 & \alpha \\ -\alpha & \alpha & 1 \end{bmatrix} = -(x+1)(x-1+2\alpha^2)$

$\iff x \in [-1, 1-2\alpha^2]$

$2\alpha^2 - 1 \leq x \leq 1 - 2\alpha^2 \Rightarrow \alpha^2 \leq \frac{1}{2}$

Def: The maximum likelihood threshold of a linear space $L \subseteq \mathbb{S}_m$, denoted $\mathrm{mlt}(L)$, is the smallest N such that for all $n \geq N$ and generic samples $X^{(1)}, \dots, X^{(n)}$, the MLE exists in the model $\mathcal{M}_L$.

equivalently: for generic $S \in PSD_m$ of rank $\min\{m,n\}$, $\pi_L(S) \in \mathrm{relint}(\pi_L(PSD_m))$.

or: for generic $S \in PSD_m$ of rank $\min\{m,n\}$, there is $S' \in PD_m$ such that $\pi_L(S) = \pi_L(S')$.

Thm: If G is a chordal graph, then $\mathrm{mlt}(L(G))$ is the size of the largest clique in G.

Thm: If G is a planar graph, then $\mathrm{mlt}(L(G)) \leq 4$.

Def: The generic completion rank of a linear space $L \subseteq \mathbb{S}_m$, denoted $\mathrm{gcr}(L)$, is the smallest r such that for generic $S \in PD_m$ there is an $S' \in \mathbb{S}_m$ of rank $\leq r$ such that $\pi_L(S) = \pi_L(S')$.

Thm: $\mathrm{mlt}(L) \leq \mathrm{gcr}(L)$.

The End