On The Geometry Of The MLE In Log-Linear Model Analysis

Alessandro Rinaldo

Department of Statistics
Carnegie Mellon University

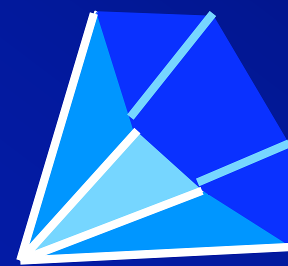
San Antonio, TX
January 13, 2006

It Is All About Geometry…

- Sample Space

Virtually all data-related quantities can be described by **polyhedra**.

Polytopes

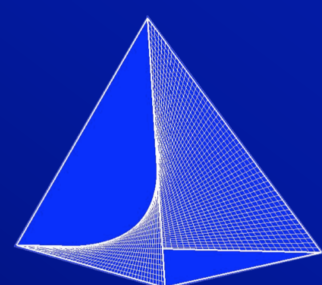


Polyhedral Cones

MLE

- Parameter Space

A log-linear model is a hyper-surface of points satisfying polynomial (binomial) equations.



Toric Varieties

Log-Linear Models

- Consider K random variable $(X_1,\ldots,X_K)$ each taking values on a finite set of labels $\mathcal{I}_j$, j=1,…,K.
- Log-linear modeling is concerned with the study of their joint distributions on the product set

$$\mathcal{I} = \otimes_{i=j}^{K} \mathcal{I}_j .$$

- Probabilistic dependence among the K variables is specified by a hypergraph Δ. (Focus is on hierarchical models.)

Generating classes of interactions $\quad \Delta \Longleftrightarrow A_\Delta \quad$ Design matrix

- The probability distributions over $\mathcal{I}$ are required to satisfy

$$\log p \in \mathcal{M}_\Delta \qquad \log p_\theta = A_\Delta^T \theta$$

log-linear subspace spanned by the rows of A_Δ

Example

X_1 **P**lacebo/**T**reatment

X_2 **D**isease/**H**ealthy

X_3 **F**emale/**M**ale

PDF	PHF
TDF	THF

PDM	PHM
TDM	THM

$$\mathcal{I} = \{T,P\} \otimes \{D,H\} \otimes \{F,M\}$$

$$\mathcal{I} = \{PDF, PDM, PHF, PHM, TDF, TDM, THF, THM\}$$

$$p \in \mathbb{R}^{\mathcal{I}} \cong \mathbb{R}^8$$

$\Delta = [1][2][3] \quad \longleftrightarrow \quad 1 \perp 2 \perp 3 \quad \longleftrightarrow$ Mutual independence

$\Delta = [13][23] \quad \longleftrightarrow \quad 1 \perp 2 \mid 3 \quad \longleftrightarrow$ Conditional independence

$$\Delta = [12][13][23]$$

Example

X_1 **P**lacebo/**T**reatment

X_2 **D**isease/**H**ealthy

X_3 **F**emale/**M**ale

PDF	PHF
TDF	THF

PDM	PHM
TDM	THM

$$\mathcal{I} = \{T,P\} \otimes \{D,H\} \otimes \{F,M\}$$

$$\mathcal{I} = \{PDF, PDM, PHF, PHM, TDF, TDM, THF, THM\}$$

$$p \in \mathbb{R}^{\mathcal{I}} \cong \mathbb{R}^8$$

$$\Delta = [12][13][23] \Longrightarrow A_\Delta = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Data are collected in a series of independent realizations over $\mathcal{I}$ that are cross-tabulated in a table of counts.

PDF	PHF
TDF	THF

PDM	PHM
TDM	THM

- Data are collected in a series of independent realizations over $\mathcal{I}$ that are cross-tabulated in a table of counts.

0	1
2	4

2	3
2	0

- A K-way contingency table $\mathsf{n} \in \mathbb{R}^{\mathcal{I}}$ is an integer-valued random function defined on $\mathcal{I}$.

- It is sufficient to study the cell mean vectors

$$\mathsf{m} = \mathbb{E}[\mathsf{n}]$$

under the constraint $\log \mathsf{m} = \mu \in \mathcal{M}$

- In fact

$$\log \mathsf{m} = \mu \in \mathcal{M}_{\Delta} \iff \log \mathsf{p} \in \mathcal{M}_{\Delta}$$

- The distributions of n form an exponential family with densities:

$$p_\theta(n) = \exp\{(A_\Delta n, \theta) - \phi(\theta)\}$$

Log-likelihood

$$\ell(\mu) = (\mathcal{P}_{\mathcal{M}_\Delta} n, \mu) + C_\mu + C_n$$

the margins $A_\Delta n = t$ is a set of sufficient statistics.

- The MLE's are the vectors $\hat{\mu} \in \mathcal{M}_\Delta$ such that:

$$\ell(\hat{\mu}) = \sup_{\mu \in \mathcal{M}_\Delta} \ell(\mu)$$

If the supremum is not attained then the MLE does not exist.

$$\left\{\hat{\mu} \in \mathcal{M}_\Delta : \ell(\hat{\mu}) = \sup_{\mu \in \mathcal{M}_\Delta} \ell(\mu)\right\} = \varnothing$$

Nonexistence Of The MLE

- Non-existence of the MLE is caused by “forbidden” patterns of zeros in the table.
- Virtually all existing statistical software are inadequate!
- It is quite likely that statistical analyses of large and sparse (i.e. containing many zeros) data-bases containing counts data are affected by this problem and hence are possibly incorrect!

0	1
2	4

2	3
2	**0**

[12][13][23]

Nonexistence Of The MLE

- Non-existence of the MLE is caused by “forbidden” patterns of zeros in the table.
- Virtually all existing statistical software are inadequate!
- It is quite likely that most statistical analyses of large and sparse (i.e. containing many zeros) data-bases containing counts data are affected by this problem and hence are possibly incorrect!
- Problem: what are the patterns of zero affecting the MLE?

0		
0	0	

	0	0
	0	

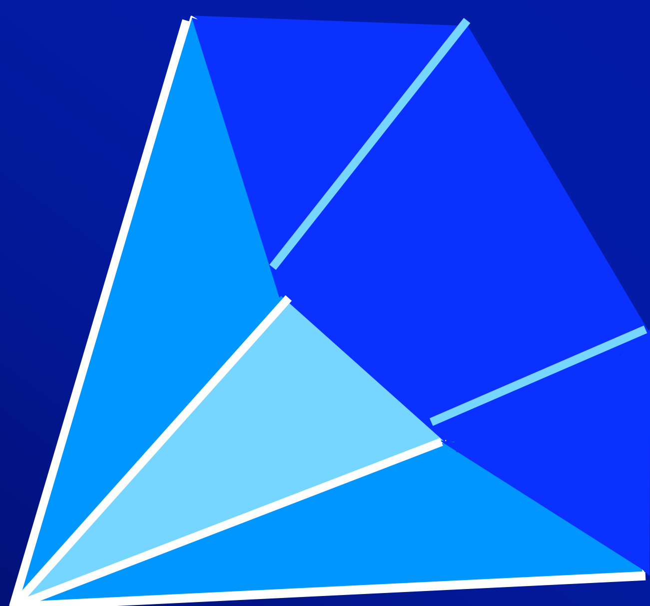
[12][13][23]

0		0
		0

?

Geometry Of The Sample Space: The Marginal Cone $C(A_\Delta)$

- The marginal cone is the convex hull of all the possible sufficient statistics (margins) that can be observed.



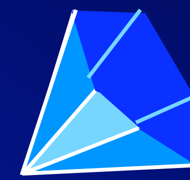
Pointed polyhedral cone generated by the columns of A_Δ

$$\dim C(A_\Delta) = \mathrm{rank}(A_\Delta)$$

- It is the most parsimonious representation of the sample space of all possible contingency tables according to the log-linear models defined by A_Δ

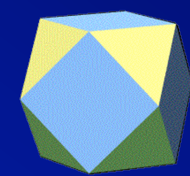
The convex hull of the sufficient statistics for general log-linear models are always polyhedra.

- Poisson sampling: marginal cone $C(A_\Delta)$



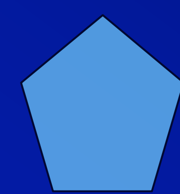
- Multinomial sampling: polytope $C(A_\Delta) \cap H$, with

$$H = \{x \in \mathbb{R}^{\mathcal{I}} : \sum_{i \in \mathcal{I}} x(i) = N\}$$

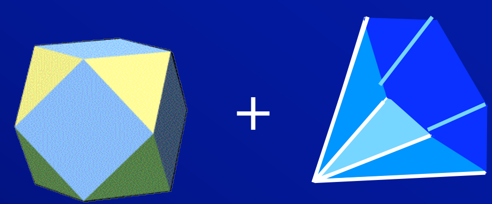


- Product-Multinomial sampling: polytope $C(A_\Delta) \bigcap_a H_a$ with

$$H_a = \{x \in \mathbb{R}^{\mathcal{I}} : \sum_{i \in \mathcal{I}} x(i) 1_a(i) = N_a\}$$



- Multinomial-Poisson sampling (Lang, 2004)



Existence Of The MLE

- The MLE exists uniquely if and only if $t \in \text{relint}\, C(A_\Delta)$

$\Updownarrow$

- The MLE does not exist if and only if $\text{supp}(n)$ is contained in a facial set of $C(A_\Delta)$.
  - The existence of the MLE is affected only by the position of the sampling zeros.
  - The patterns of zeros invalidating the MLE are the zeros of the facial set incidence vectors.

Eriksson, Fienberg, Rinaldo, Sullivant (2006)

Examples

0		
0	0	

	0	0
	0	

0		0
		0

[12][13][23]

nonexistent MLE

0	0
0	

0	

	0

	0
0	0

[12][23][34][14]

	0	0
0		0
0	0	

0	0	
	0	0
0	0	

0		0
0	0	
	0	0

[12][13][23]

existent MLE

- Finding patterns of forbidden zeros is an instance of the face enumeration problem, which is affected by a combinatorial explosion. Examples:

	Size	# facets
Δ=[12][13][23]	2x2x2	16
	3x3x3	207
	4x4x4	113,740

Eriksson, Fienberg, Rinaldo, Sullivant (2006) & Rinaldo(2006a)

Geometry Of The Parameter Space: Toric Varieties

- Let $\mathcal{L}_\Delta = \text{kernel}(A_\Delta) \cap \mathbb{Z}^{\mathcal{I}}$
  The mean functions are defined by binomial equations in the toric ideal

$$I_\Delta = \left\langle x^{z+} - x^{z-} : \forall z \in \mathcal{L}_\Delta \right\rangle$$

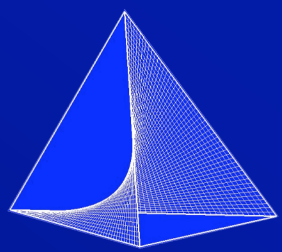
and belong to the toric variety

$$V_{\Delta,\geq 0} = V(I_\Delta) \cap \mathbb{R}^{\mathcal{I}}_{\geq 0}$$

Example. 2x2 table.

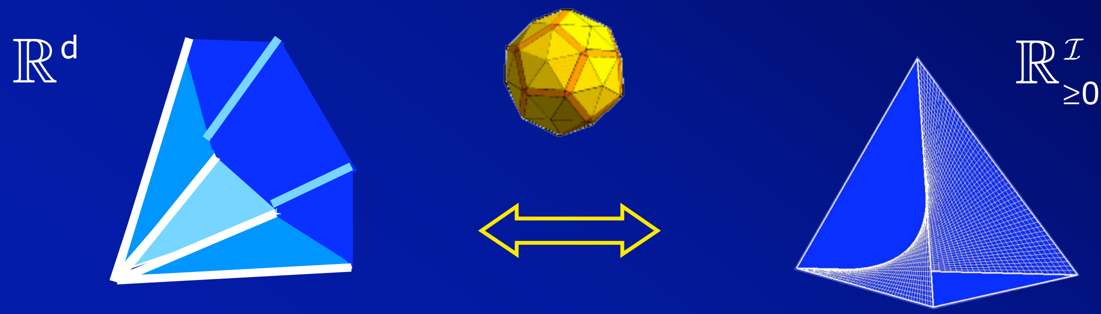
Model of independence

$$I_\Delta^{2x2} = \left\langle p_{11}p_{22} - p_{12}p_{21} \right\rangle$$



Segre Variety

surface of independence



- The variety $V_{\Delta,\geq 0}$ consists of all the mean vectors for the log-linear model Δ and their point-wise limits.
  It is homeomorphic to $C(A_\Delta)$ (Moment Map).

- For any $t \in C(A_\Delta)$ the *fiber* at t is the polytope:

$$P_{t,\Delta} = \{x \in \mathbb{R}^{\mathcal{I}}_{\geq 0} : A_\Delta x = t\}$$

  consisting of the set of all tables whose A_Δ-margins are t.

- For each $t \in C(A_\Delta)$, the intersection $P_{t,\Delta} \cap V_{\Delta,\geq 0}$ consists of a unique point in $\text{relint}(P_{t,\Delta})$.

The Geometry Of The MLE

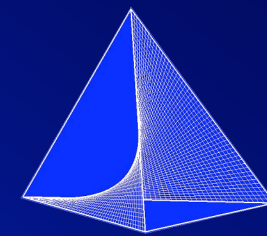
$$\hat{\mu} = \log \hat{m}$$

$$\ell(\hat{\mu}) = \sup_{\mu \in \mathcal{M}_\Delta} \ell(\mu)$$

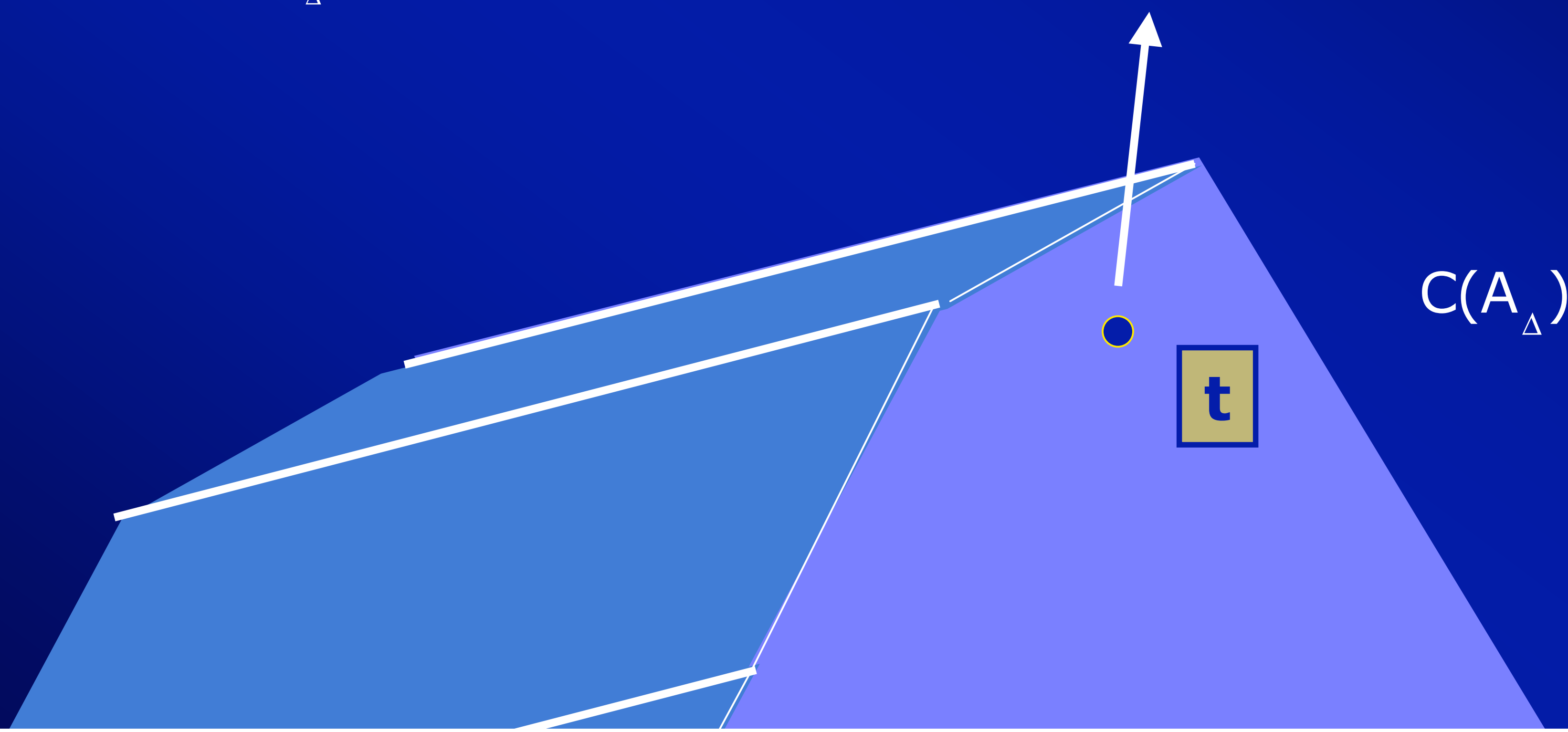
MLE

$\Longleftrightarrow$

$\cap$

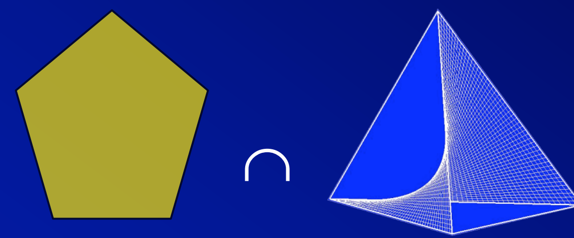


$$\{\hat{m}\} = P_{t,\Delta} \cap V_{\Delta,\geq 0}$$



The Boundary Cases

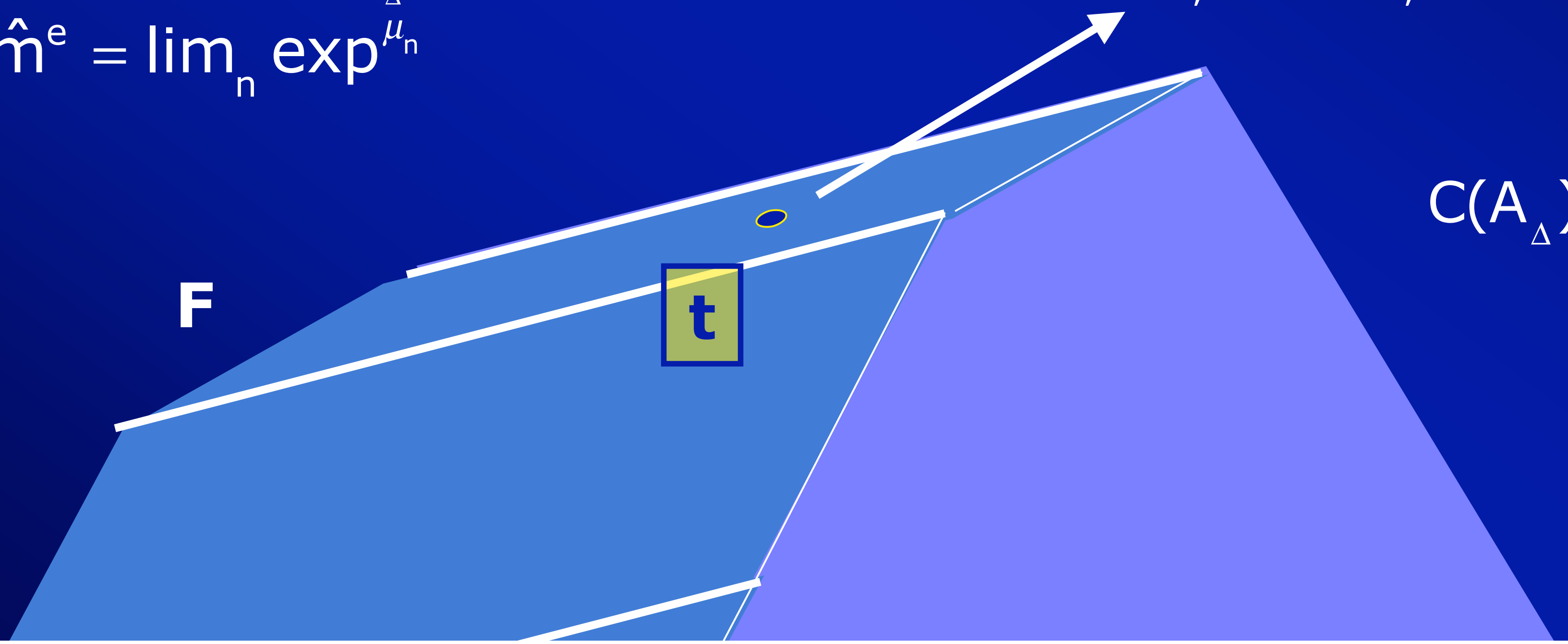
$$\hat{m}^e_F > 0$$
$$\hat{m}^e_{F^c} = 0$$



Extended MLE

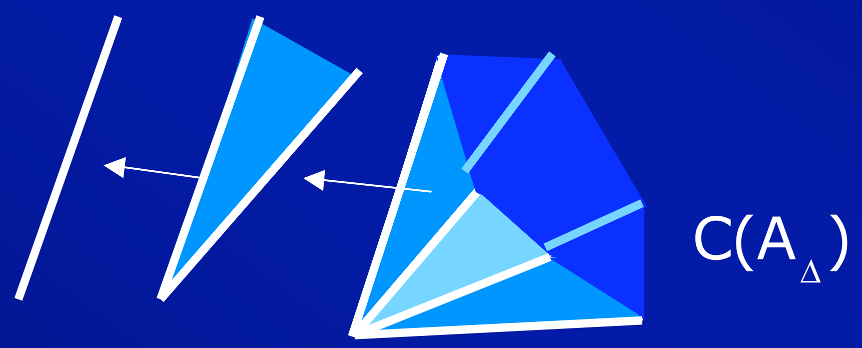
$$\lim_n \ell(\mu_n) = \sup_{\mu \in \mathcal{M}_\Delta} \ell(\mu) \quad \Longleftrightarrow \quad \{\hat{m}^e\} = P_{t,\Delta} \cap V_{\Delta,\geq 0}$$

$$\hat{m}^e = \lim_n \exp^{\mu_n}$$



More To Say About the Geometry Of Exponential Families…

- It is possible, using the same geometry, to construct extended exponential families.
  This is a larger set of family of distributions for the cell counts, each defined by the relative interior of a face of the marginal cone. These distributions can be "glued" together and are parametrized in a smoothly by the points inside toric variety.

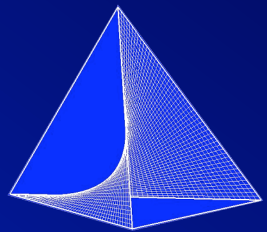


- This extended family has nice properties.
  (for example it is closed in r-I and total variation distance)

Rinaldo (2006b)

The Design Matrix A_Δ

← **MLE** →



Sample Space

A_Δ identifies:

- Marginal cone of sufficient statistics:

$$C(A_\Delta) = \text{cone}(A_\Delta)$$

- and the fiber of tables in the exact distribution:

$$P_{t,\Delta}\left\{x \geq 0, A_\Delta x = t\right\}$$

Parameter Space

A_Δ specifies the set of polynomial equations that encode the dependence among the variables.

All probability vectors satisfy binomial equations:

$$p^{u+} - p^{u-} = 0$$

all integer $u \in \text{kernel}(A)$