Why do we care about inequalities in (algebraic) statistics?

Piotr Zwiernik
Mittag-Leffler Institute, Stockholm
(j.w. Jim Q. Smith)

WOGAS 3,
Warwick, 2 April 2011

Outline of the talk

- Algebraic statistical models and constrained multinomial models.
- How inequalities affect inference for the tripod tree model?
- Extensions to more general models.

Algebraic statistical model

- X a random variable with m values, $p = (p_1, \dots, p_m)$
- probability simplex $\Delta = \{p \in \mathbb{R}^m : \sum_i p_i = 1, p_i \geq 0\}$
- ASM: $p : \Theta \to \Delta$, $\mathcal{M} = p(\Theta)$, p polynomial map
- ASM given by polynomial equations and inequalities

- X, Y binary and $X \perp\!\!\!\perp Y$: $p : [0,1]^2 \to \Delta$

$$(\theta_x, \theta_y) \quad \mapsto \quad \big((1-\theta_x)(1-\theta_y),\ (1-\theta_x)\theta_y,\ \theta_x(1-\theta_y),\ \theta_x\theta_y\big)$$

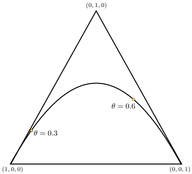
- implicit equation: $p_{00}p_{11} - p_{01}p_{10} = 0$

ASM as a constrained multinomial model

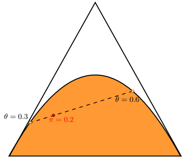
- interpret Δ as a set of parameters for the multinomial model.
- $\mathcal{M} = p(\Theta) \subseteq \Delta$ gives a *constrained multinomial model*

Example:

$\mathrm{Bin}(2, \theta)$
no inequalities
$\mathrm{codim} = 1$



$\mathrm{MixBin}(2, \theta)$
inequalities
$\mathrm{codim} = 0$

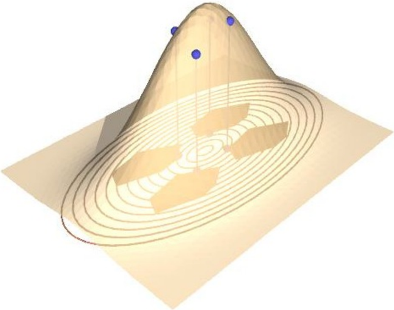


The likelihood function

- $X \in \{1, \ldots, m\}$, $X^{(1)}, \ldots, X^{(N)}$ an *iid* sample from $\mathcal{M}$
- likelihood: $L(\theta; x) \propto \prod_{i=1}^{m} p_i(\theta)^{x_i}$
- constrained: $L(p; x) \propto \prod_{i=1}^{m} p_i^{x_i}$ such that $p \in \mathcal{M} = p(\Theta)$
- unconstrained: $L(p; x) \propto \prod_{i=1}^{m} p_i^{x_i}$ such that $p \in \Delta$

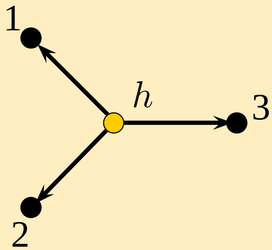
The constrained likelihood

- typically MLE does not have a closed-form solution
- the MLE will often lie on the boundary of the parameters space
  - no usual asymptotics for the LR statistic, MLE etc.
  - posterior heavily depends on prior specification



The tripod tree model

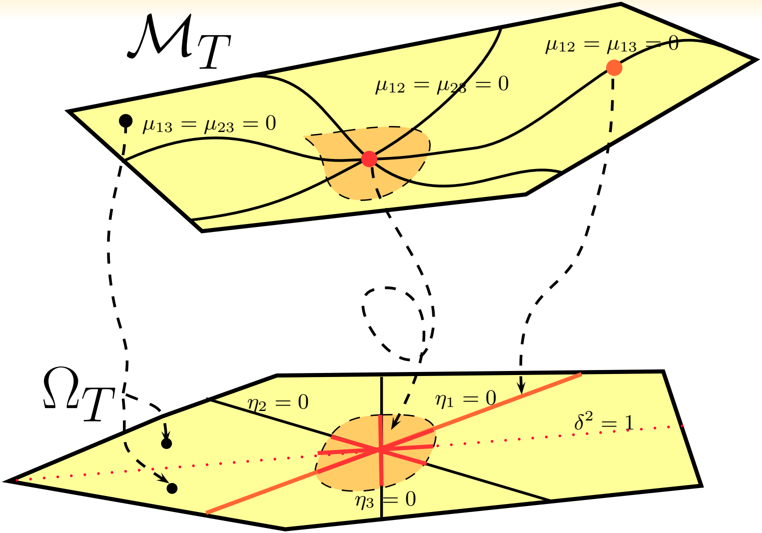
- $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3 | H$ or
- a Bayesian network a tripod tree
- seven free parameters
- the codimension is zero



Alternative parametrization

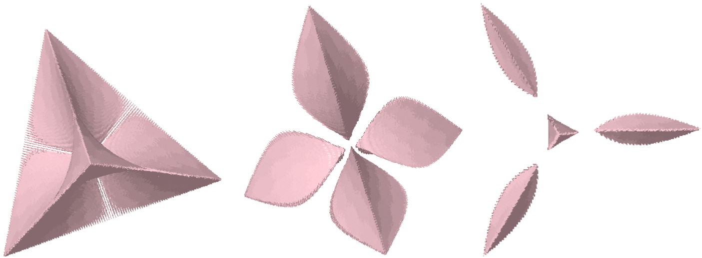
$$\mathcal{M}_T : \begin{aligned} k_{12} &= \frac{1}{4}(1-\delta^2)\eta_1\eta_2, \\ k_{13} &= \frac{1}{4}(1-\delta^2)\eta_1\eta_3, \\ k_{23} &= \frac{1}{4}(1-\delta^2)\eta_2\eta_3, \\ k_{123} &= \frac{1}{4}(1-\delta^2)\delta\eta_1\eta_2\eta_3 \end{aligned}$$

Application: Identifiability



The model structure

- The model covers 8% of the probability simplex.
- Picture: $\mu_1 = \mu_2 = \mu_3 = \frac{1}{2}$ and $k_{123} = 0, 0.005, 0.02$



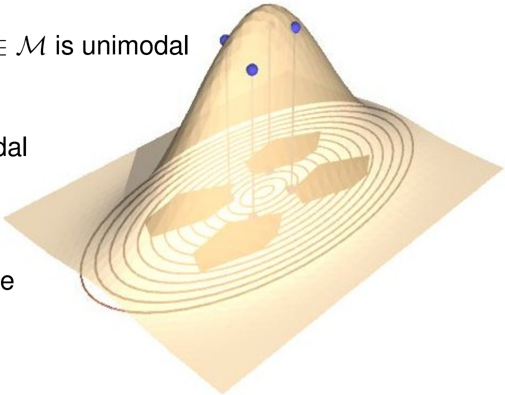
The constrained likelihood for the tripod

Three possible scenarios:

(i) $\hat{p} \in \mathcal{M}_T$ and then $\ell(p)$ for $p \in \mathcal{M}$ is unimodal and $\ell(\theta)$ bimodal

(ii) $\hat{p} \notin \mathcal{M}_T$ and $\ell(p)$ is multimodal with one global maximum but many local maxima.

(iii) $\hat{p} \notin \mathcal{M}_T$ and $\ell(p)$ has multiple global maxima.



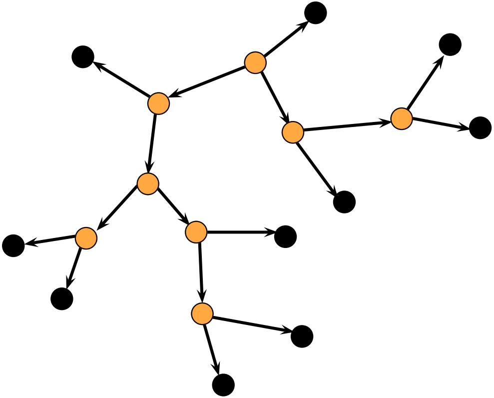
Implications for inference

$$\begin{bmatrix} x_{000} & x_{001} & x_{100} & x_{101} \\ x_{010} & x_{011} & x_{110} & x_{111} \end{bmatrix} = \begin{bmatrix} 2069 & 16 & 2242 & 331 \\ 2678 & 863 & 442 & 1359 \end{bmatrix}.$$

- only $k_{12}k_{13}k_{23} \geq 0$ does not hold
- the EM algorithm gives 4 different maxima of $L(\theta)$

	$\theta_1^{(r)}$	$\theta_{1\|0}^{(1)}$	$\theta_{1\|1}^{(1)}$	$\theta_{1\|0}^{(2)}$	$\theta_{1\|1}^{(2)}$	$\theta_{1\|0}^{(3)}$	$\theta_{1\|1}^{(3)}$
1	0.4658	0.3371	0.5524	1.0000	0.0000	0.4159	0.0745
2	0.5342	0.5524	0.3371	0.0000	1.0000	0.0745	0.4159
3	0.4771	0.0000	0.9167	0.6369	0.4216	0.1468	0.3775
4	0.5229	0.9167	0.0000	0.4216	0.6369	0.3775	0.1468

Does it generalize?



Main theorems

Theorem[Z.,JQ Smith]

- There exist explicit formulae for the MLEs if $\hat{p} \in \mathcal{M}_T$.
- Generically the map is $2^{\text{int}(V)}$ $-$ to $-$ 1,
  - for some point the preimage is a manifold with corners, otherwise it is a singular subset of the parameter space.
- All the equations and inequalities defining the model can be listed.

Models with equality constraints

- Can we determine from $\hat{p}$ whether the MLE lies on the boundary?
- Can we perform a full Bayesian analysis?
- Can we test these models?

- For the Bayesian perspective see the results of the Dutch group
  - Utrecht: Herbert Hoijtink, Irene Klugkist, Olav Laudy, Bernet Kato et al.
  - Amsterdam: Ruud Wetzels, Raoul P.P.P. Grasman, Eric-Jan Wagenmakers

Thank you!

The bibliography

- DAVIS-STOBER, C., ***Analysis of multinomial models under inequality constraints: Applications to measurement theory***, ***Journal of Mathematical Psychology***, **2009.**
- P. ZWIERNIK AND J. Q. SMITH, ***Tree-cumulants and the geometry of binary tree models***, **to appear in** ***Bernoulli***.
- P. ZWIERNIK AND J. Q. SMITH, ***The geometry of tree models and some consequences for inference***, **on arXiv.**