Study of diagonal-effect models as toric and mixture models

Fabio Rapallo

Department DISTA
University of Eastern Piedmont
Alessandria (Italy)

AMS Special Session on Advances in Algebraic Statistics,
Lexington, KY, March 27-28, 2010

Summary

- Toric models, Markov bases, and invariants
- Diagonal-effect models in torc form
- Diagonal-effect models in mixture form
- Concluding remarks

This work is joint with Cristiano Bocci (University of Siena, Italy) and Enrico Carlini (Politecnico di Torino, Italy).

Statistical models for contingency tables

A probability distribution for an $I \times J$ contingency table is a non-negative normalized matrix $P = (p_{i,j})$, i.e., is a point of the closed simplex

$$\Delta = \left\{ P = (p_{i,j}) \ : \ p_{i,j} \geq 0 \, , \ \sum_{i,j} p_{i,j} = 1 \right\} .$$

Definition

A statistical model is a subset of Δ.

Toric models

In a toric model, the raw probabilities of the cells are defined (apart from normalization) in parametric form as power products:

$$p_{i,j} = \prod_{h=1}^{s} \zeta_h^{A_{(i,j),h}}$$

where $\zeta_1, \ldots, \zeta_s$ are non-negative real parameters.

The structure of the toric model is encoded in an $IJ \times s$ non-negative integer matrix A.

Given an observed contingency table f (table of counts), it is easy to see that $A^t f$ is the sufficient statistic.

Invariants

Eliminating the ζ parameters from the system:

$$p_{i,j} = \prod_{h=1}^{s} \zeta_h^{A_{(i,j),h}}$$

one obtains the toric ideal

$$\mathcal{I}_A \subset \mathbb{R}[p] = \mathbb{R}[p_{1,1}, \ldots, p_{I,J}]$$

associated to the statistical model.

The ideal $\mathcal{I}_A$ is generated by pure binomials.

The polynomials in $\mathcal{I}_A$ are the invariants of the model.

Markov bases

A Markov basis for the statistical toric model defined by the matrix A is a finite set of tables $m_1, \ldots, m_\ell \in \mathbb{Z}^{IJ}$ that allows us to connect any two contingency tables f_1 and f_2 in the same fiber, i.e. such that $A^t(f_1) = A^t(f_2)$, with a path of elements of the fiber.

The path is therefore formed by tables of non-negative counts with constant image under A^t (the sufficient statistic).

Markov bases and invariants

The relation between the Markov basis and the toric ideal $\mathcal{I}_A$ is known.

Theorem

The set of moves $\{m_1, \ldots, m_\ell\}$ is a Markov basis if and only if the set $\{p^{m_1+} - p^{m_1-}, \ i = 1, \ldots, \ell\}$ generates the ideal $\mathcal{I}_A$.

Usually this theorem is used in its "if" part to deduce Markov bases from the computation of a system of generators of a toric ideal.

For our study, we will make use of this theorem in its "only if" implication.

The independence model

Given:

- I non-negative row parameters $\zeta_1^{(r)}, \ldots, \zeta_I^{(r)}$;
- J non-negative column parameters $\zeta_1^{(c)}, \ldots, \zeta_J^{(c)}$;

the independence model is parametric form is:

$$\mathcal{M} = \{p_{i,j} = \zeta_i^{(r)} \zeta_j^{(c)} \ , \ 1 \le i \le I, \ 1 \le j \le J\} \cap \Delta \, .$$

In implicit form, this translates into:

$$\mathcal{M}' = \{p_{i,j} p_{k,h} - p_{i,h} p_{k,j} = 0 \ , \ 1 \le i < k \le I, \ 1 \le j < h \le J\} \cap \Delta \, .$$

Remark

In the open simplex $\Delta_{>0}$, $\mathcal{M} = \mathcal{M}'$.

$\mathcal{M}$ and $\mathcal{M}'$ are in general different on the boundary $\Delta \setminus \Delta_{>0}$.

Diagonal-effect models in toric form

Let us consider $I \times I$ (square) tables.

Definition

The diagonal-effect model $\mathcal{M}_1$ is defined as the set of probability matrices $P \in \Delta$ such that:

$$p_{i,j} = \zeta_i^{(r)} \zeta_j^{(c)} \quad \text{for} \quad i \neq j$$

$$p_{i,j} = \zeta_i^{(r)} \zeta_j^{(c)} \zeta_i^{(\gamma)} \quad \text{for} \quad i = j$$

where $\zeta^{(r)}$, $\zeta^{(c)}$ and $\zeta^{(\gamma)}$ are non-negative vectors with length I.

Theorem

The invariants of the model $\mathcal{M}_1$ are the binomials

$$p_{i,j}p_{i',j'} - p_{i,j'}p_{i',j}$$

for i, i', j, j' all distinct, and

$$p_{i,i'}p_{i',i''}p_{i'',i} - p_{i,i''}p_{i'',i'}p_{i',i}$$

for i, i', i'' all distinct.

Sketch of the proof. In Aoki and Takemura (2005) it is shown that a minimal Markov basis for the model $\mathcal{M}_1$ is formed by:

(a) The basic degree 2 moves:

$$\begin{array}{ccc} & j & j' \\ i & +1 & -1 \\ i' & -1 & +1 \end{array}$$

with i, i', j, j' all distinct, for $I \geq 4$;

(b) The degree 3 moves of the form:

$$\begin{array}{cccc} & i & i' & i'' \\ i & 0 & +1 & -1 \\ i' & -1 & 0 & +1 \\ i'' & +1 & -1 & 0 \end{array}$$

with i, i', i'' all distinct, for $I \geq 3$.

Thus, it is enough to write such moves in polynomial form. $\square$

Diagonal-effect models in mixture form

Definition

The diagonal-effect model $\mathcal{M}_2$ is defined as the set of probability matrices P such that

$$P = \alpha c r^t + (1-\alpha) D$$

where r and c are non-negative vectors with length I and sum equal to one, $D = \mathrm{diag}(d_1, \ldots, d_I)$ is a non-negative diagonal matrix with sum equal to one, and $\alpha \in [0, 1]$.

Remark

While in the toric definition the normalization is applied once, in the mixture definition the normalization is applied twice.

Some geometry

Theorem

The models $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same invariants.

Proof. The proof is based on the properties of the elimination technique. $\square$

The non-negativity conditions imply that $\mathcal{M}_1 \neq \mathcal{M}_2$ and

Theorem

Neither $\mathcal{M}_2 \subset \mathcal{M}_1$ nor $\mathcal{M}_1 \subset \mathcal{M}_2$.

We see this with two examples.

First example

$$P = \begin{pmatrix} 0 & \frac{1}{I(I-1)} & \frac{1}{I(I-1)} & \cdots & \frac{1}{I(I-1)} \\ \frac{1}{I(I-1)} & 0 & \frac{1}{I(I-1)} & \cdots & \frac{1}{I(I-1)} \\ \frac{1}{I(I-1)} & \frac{1}{I(I-1)} & 0 & \cdots & \frac{1}{I(I-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{I(I-1)} & \frac{1}{I(I-1)} & \frac{1}{I(I-1)} & \cdots & 0 \end{pmatrix} .$$

$P \in \mathcal{M}_1$ by constructions, while it does not belong to $\mathcal{M}_2$ (easy to check).

Second example

$$P = \begin{pmatrix} \frac{1}{I} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{I} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{I} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{I} \end{pmatrix} .$$

$P \in \mathcal{M}_2$, by setting $\alpha = 0$ and $D = \mathrm{diag}(1/I, \dots, 1/I)$, while it does not belong to $\mathcal{M}_1$ (once again, easy to check).

Other results in the open simplex

Theorem

In the open simplex

$$\Delta_{>0} = \left\{ P = (p_{i,j}) \ : \ p_{i,j} > 0 \ , \ \sum_{i,j} p_{i,j} = 1 \right\}$$

the following inclusion holds:

$$\mathcal{M}_2 \subset \mathcal{M}_1 \ .$$

With some polynomial manipulations we are also able to characterize the set $\mathcal{M}_1 \setminus \mathcal{M}_2$.

Theorem

Let $P \in \mathcal{M}_1 \cap \Delta_{>0}$ be a strictly positive probability table given by the vectors $\zeta^{(r)}$, $\zeta^{(c)}$ and $\zeta^{(\gamma)}$. Define:

- $N_T = \sum_{i \neq j} \zeta_i^{(r)} \zeta_j^{(c)} + \sum_{i=j} \zeta_i^{(r)} \zeta_j^{(c)} \zeta_i^{(\gamma)}$
- $N = \sum_{i,j} \zeta_i^{(r)} \zeta_j^{(c)}$.

Then $P \in \mathcal{M}_1 \setminus \mathcal{M}_2$ if one of the following situations holds:

(i) $N_T < N$;

(ii) $N_T = N$ and there exists at least an index i such that $\zeta_i^{(\gamma)} \neq 1$;

(iii) $N_T > N$ and there exists at least an index i such that $\zeta_i^{(\gamma)} < 1$.

Maximum likelihood estimation

Theorem

For an independent sample of size n, the models $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same sufficient statistic.

Nevertheless the MLE can differ. Consider the table

$$\begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix} .$$

In this case, the normalized observed table belongs to $\mathcal{M}_1$ and to $\mathcal{M}_2$, and thus the MLE is equal to the observed table for both models.

On the other hand, consider the following observed table

$$\begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 1 & 2 & 2 \\ 2 & 2 & 1 & 2 \\ 2 & 2 & 2 & 1 \end{pmatrix} .$$

After normalization, this table belongs to $\mathcal{M}_1$ and thus under the toric model, the MLE is again equal to the observed table. However, the table does not belong to $\mathcal{M}_2$ and (with symmetry arguments), the MLE is the rank-one table

$$\frac{1}{16} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} .$$

Concluding remarks

- The models $\mathcal{M}_1$ (toric) and $\mathcal{M}_2$ (mixture) have the same invariants, but they are “essentially” different.
- Here “essentially” means: “not only on the boundary”.
- Therefore, we also have a different behavior in terms of maximum likelihood estimation.

Thank you!

The complete manuscript is available at
`http://arxiv.org/abs/0908.0232`.

(In press on the AMS Contemporary Mathematics volume
"Algebraic Methods in Statistics and Probability II")