Partition Models As A Framework For Multiplicative Models For Categorical Data
Anna Klimova, Department Of Statistics, University Of Washington
Joint Work With Tamas Rudas, Adrian Dobra
June 4, 2009

1 Partition Models
A novel framework for the analysis of multiplicative models for categorical data.
A generalization of log-linear models, quasi-independence model, models with structural zeroes.
Lattice basis for MCMC testing is easy to find.

2 Partition: Definition And Examples
Let \( X = (X_1, X_2, \ldots, X_k) \) be a set of categorical variables cross-classified in a contingency table \( I \). A cell of the table \( i = (i_1, i_2, \ldots, i_k) \) has a cell probability \( p(i) \).

A partition \( q \) of a table \( I \) is a division of its cells into subsets \( S_q \) such that:
\[
S_q \neq \emptyset \quad \forall u, \quad S_{q_1} \cap S_{q_2} = \emptyset \quad \forall u \neq v; \quad \bigcup_u S_q = I.
\]

Examples:
- A marginal partition - all cells belonging to the same marginal cell form a subset;
- A partition induced by a cell - this cell forms a subset.

3 Partition Model: Matrix Representation
A partition model \( PM(Q) \) for a class \( Q \) of partitions of \( I \): \( \log\ p = A'V \).
Here \( A = (a_i) \) is a model matrix, \( a_q = I_q(j) \).
\( I_q \) is the indicator function of the subset \( S_q \). \( V \) is the vector of parameters.

\[
p \in PM(Q) \Leftrightarrow D_x \log p = 0.
\]

\( D_x \) is the matrix with rows that form a basis of \( \text{Null}(A) \).

4 Maximum Likelihood Estimation
Theorem: Let \( Q \) be a class of partitions and \( PM(Q) \) be a partition model. The sum of observed probabilities of the cells from the same subset of partition is equal to the sum of the MLIs of the probabilities \( p(i) \) under the \( PM(Q) \) model.
The degrees of freedom equal \( \text{dim Null}(A) \) (the number of rows in \( D_x \)).

5 Iterative Proportional Fitting
\( PM(Q) \) is a partition model.
Let \( S_q = \{S_q = S_1, \ldots, S_J \} \) be all the subsets of the partitions in \( Q \) and \( \{n(i)\}_{i \in I} \) be the set observed frequencies. Set \( m_0(i) = 1 \) and iterate cyclically over \( j \in J \). With \( d + 1 = J + j \)
\[
m_{i_{j+1}}(i) = m_i(i) \sum_{i \in S_j} n(i) \quad \text{if } i \in S_j,
\]
\[
m_{i_{j+1}}(i) = m_i(i) \quad \text{if } i \notin S_j.
\]

6 Markov Chains
- Rows of matrix \( D_x \) with integer entries is a lattice basis of matrix \( A \).
- They can be transformed into Markov moves for MCMC.
- Use moves to sample from the space of tables \( n' \) that have the same observed subset (e.g. marginal) totals and structural zeroes under the uniform distribution.
- Choose the test statistic. For example,
\[
\chi^2(n) = \sum_{i \in I} \frac{(n(i) - \hat{h}(i))^2}{\hat{h}(i)}.
\]
Here \( \hat{h} \) is the expected cell values under the partition model.
- For the observed table \( n \), MCMC estimate of the exact p-value is
\[
p(n) = \sum_{\{n' : \chi^2(n') \leq \chi^2(n)\}} p(n').
\]

7 Example: Rapallo, 2006
Consider a 2 x 3 contingency table with a structural zero in the cell (1, 2) (Rapallo, 2006) and frequencies:
\[
\begin{pmatrix}
\mu_{11} & \mu_{12} \\
\mu_{21} & \mu_{22} & \mu_{23}
\end{pmatrix}
\]

The class of partitions for the model of conditional independence with a structural zero:
\( q_0 \) is the trivial partition, \( q_1 \) is the partition induced by row marginal, \( q_2 \) is the partition induced by column marginal, \( q_3 \) is the partition induced by the cell (1, 2).

\[
A = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 1
\end{pmatrix}
\]
is the model matrix.

The model of independence: \( D_x \log p = \log \rho_{11} \rho_{22} \rho_{33} = 0 \),
where \( D_x = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \end{pmatrix} \).
A lattice basis consists of only one move:
\[
M_1 = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0
\end{pmatrix}
\]
Partition model leads to the same move as Rapallo obtained, but without special tools of computational algebra.

8 Example: Squirrel Monkeys Data: Data (Ploog, 1967)
The genital display data in six squirrel monkeys is summarized as a 5 x 6 table with 5 structural zeroes (a monkey does not display toward itself):

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>S</th>
<th>T</th>
<th>U</th>
<th>V</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>5</td>
<td>29</td>
<td>46</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>38</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>W</td>
<td>9</td>
<td>25</td>
<td>4</td>
<td>6</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

Fienberg (2007). The quasi-independence model. \( \chi^2 = 168.05, \text{d.f.} = 15, \text{p-value}=0 \).
\( \text{p-value} = 0.929 \pm 0.0006 (\text{one million sampled tables}) \).
\( \text{p-value} = 0.938 \pm 0.0006 \).

9 Squirrel Monkeys Data: Our Approach
Model: Partition model. Partitions induced by marginals and zero cells.

\[
A = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1
\end{pmatrix}
\]

Method: Exact testing with MCMC. Markov moves are sampled from the lattice basis. Five chains. One million sampled tables each. Burn-in 1000 iterations.
Result: \( \text{p-value} = 0.98 \pm 0.0046 \).

Convergence plot for exact p-values:

\[
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1
\end{pmatrix}
\]

10 What’s Next? Further Research Directions
- Applications of partition models: topological models etc.
- Sampling strategies from lattice basis
- Partition models and Neumann structure

Contact Information
email: klimova@u.washington.edu