Fast Algorithm for Generalized Multinomial Models with Ranking Data

Jiaqi Gu (Joint work with Prof. Guosheng Yin)

June 13, 2019

Department of Statistics and Actuarial Science, HKU

Generalized multinomial models

Consider d basic cells $c_1, \ldots, c_d$, where c_i is assigned with cell probability p_i ($\sum_{i=1}^d p_i = 1$). Suppose cell c_i is chosen for a_i times ($i = 1, \ldots, d$), then the log-likelihood function is

$$\ell(\mathbf{p}) = \sum_{i=1}^{d} a_i \log p_i. \tag{1}$$

Completeness condition: Sets of basic cells for selection are always $\{c_1\}, \ldots, \{c_d\}$. (choose 1 from d)

If completeness condition is violated:

- Union of sets for selection includes only a fraction of basic cells. (choose 1 from $k < d$)
- Sets for selection consists of more than one basic cell. (choose $l > 1$ from d)

$\rightarrow$ Incomplete multinomial models.

Log-likelihood function

$$\ell(\mathbf{p}) = \sum_{j=1}^{n} \Big\{ \log(\sum_{c_i \in A^j} p_i) - \log(\sum_{c_i \in \mathcal{C}^j} p_i) \Big\}, \qquad (2)$$

where $\mathcal{C}^j$ is the union of sets for selection, A^j is the selected set in j-th record.

Examples:

- Placett–Luce model [3, 4];
- Bradley–Terry model [1];
- Contingency table model [2].

Markov chain based algorithm

Denote $W_i = \{j : c_i \in A^j\}$ and $L_i = \{j : c_i \in (\mathcal{C}^j \backslash A^j)\}$,
$q_j^+ = \sum_{c_i \in A^j} p_i$ and $q_j^* = \sum_{c_i \in \mathcal{C}^j} p_i$,

$$\ell(\mathbf{p}) = \sum_{j=1}^{n} \Big\{ \log(q_j^+) - \log(q_j^*) \Big\}.$$

Letting $\frac{\partial \ell(\mathbf{p})}{\partial p_i} = 0$, we have

$$\sum_{i' \neq i} p_{i'} \left[\sum_{j \in W_i \cap L_{i'}} \frac{p_i}{q_j^+ q_j^*} \right] = \sum_{i' \neq i} p_i \left[\sum_{j \in L_i \cap W_{i'}} \frac{p_{i'}}{q_j^+ q_j^*} \right], \qquad (i = 1, \ldots, d). \tag{3}$$

Markov chain based algorithm

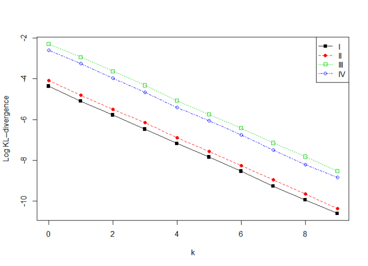
Algorithm 1 Markov chain based algorithm

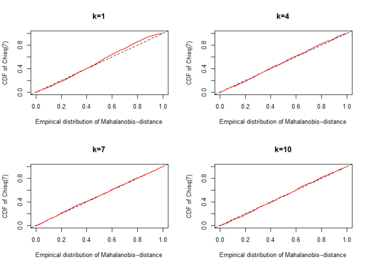
Input: Observations $\left\{(A^j, \mathcal{C}^j) : j = 1, \ldots, n\right\}$ and calculate $\{W_i, L_i\}$ for each c_i.
Initialize $\mathbf{p} = (1/d, \ldots, 1/d)^T$.
Initialize $\mathbf{\Sigma}(\mathbf{p}) = \mathbf{0}_{d \times d}$.
repeat
  for $i \in \{1, \ldots, d\}$ **do**
    for $i' \in \{1, \ldots, d\} \backslash \{i\}$ **do**
      Compute

$$\sigma_{ii'}(\mathbf{p}) \leftarrow \sum_{j \in L_i \cap W_{i'}} \frac{p_{i'}}{q_j^+ q_j^*}$$

    end for
  end for
  Compute $\sigma_{ii}(\mathbf{p})$ $(i = 1, \ldots, d)$ and then normalize $\mathbf{\Sigma}(\mathbf{p})$ so that $\forall i, \sum_{i'=1}^{d} \sigma_{ii'}(\mathbf{p}) = 1$.
  $\mathbf{p} \leftarrow T(\mathbf{p})$ under the transition matrix $\mathbf{\Sigma}(\mathbf{p})$.
until convergence.

Experiments: Convergence





Estimator obtained by our algorithm is close to the MLE, indicating our algorithm's convergence to the MLE.

Experiments: Convergence rate

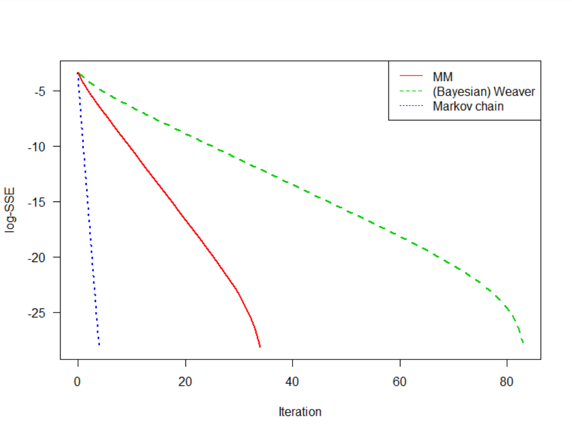
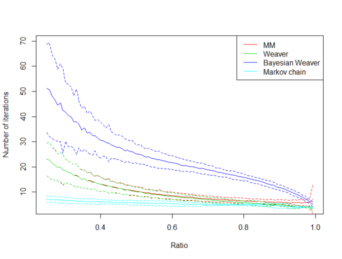


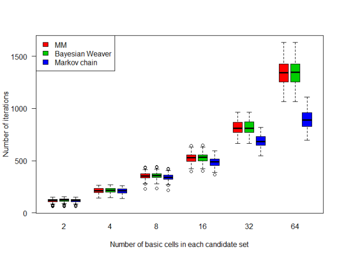
Figure 1: Path of iterations for three algorithms on sushi data

Experiments: Computational efficiency

- Choose 1 from $k < d$:



- Choose $l > 1$ from d:



Discussions

- Our algorithm obtain the MLE efficiently than existing methods.
- Further improvement. (Especially in situation "choose $l > 1$ from d")

Reference

- Ralph Allan Bradley and Milton E. Terry.
  Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons.
  Biometrika, 39(3/4):324–345, 1952.
- Tar Chen and Stephen E. Fienberg.
  The Analysis of Contingency Tables with Incompletely Classified Data.
  Biometrics, 32(1):133, 1976.
- R. Duncan Luce.
  Individual Choice Behavior: A Theoretical analysis.
  Wiley, 1959.
- R. L. Plackett.
  The Analysis of Permutations.
  Applied Statistics, 24(2):193–202, 1975.

End

Thank you for listening.