Fast and Stable Maximum Likelihood Estimation for Incomplete Multinomial Models

Chenyang Zhang, Guosheng Yin

Department of Statistics and Actuarial Science,
The University of Hong Kong

June 13, 2019

What is Incomplete Multinomial Model?

- A toy example: Incompelte contingency table

	Young	Middle	Senior
Female	p_1	p_2	p_3
Male	p_4	p_5	p_6

- Sample 1:

	Young	Middle	Senior
Female	21	24	18
Male	20	25	12

- Sample 2:

Female	18
Male	22

- Sample 3:

Young	Middle	Senior
10	20	10

- Sample 4:

	Young
Female	53
Male	47

What is Incomplete Multinomial Model (Cont'd)

Multinomial model: the sample space Ω is partitioned into K disjoint subspaces. Incomplete cases:

(a) a subset of categories rather than a unique category is reported (partial classification).

(b) the set of possible outcomes contains only part of all categories (truncated outcomes).

$$L(\boldsymbol{p}|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\Delta}) \propto \prod_{k=1}^{K} p_k^{a_k} \prod_{j=1}^{q} \tilde{p}_j^{b_j} = \prod_{k=1}^{K} p_k^{a_k} \prod_{j=1}^{q} (\boldsymbol{\delta}_j^\intercal \boldsymbol{p})^{b_j}.$$

- $\boldsymbol{p} = (p_1, \ldots, p_K)^\intercal$: parameters of the incomplete multinomial model.
- $\boldsymbol{a} = (a_1, \ldots, a_K)^\intercal$: counts of fully classified observations.
- $\boldsymbol{b} = (b_1, \ldots, b_q)^\intercal$: counts of incomplete observations. Positive terms for partial classification, and negative terms for truncated outcomes.
- $\boldsymbol{\Delta} = \{\Delta_{kj}\}_{K \times q} = [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_q]$: indicator matrix.

What is Incomplete Multinomial Model (Cont'd)

$$
\begin{aligned}
\mathrm{L}(\boldsymbol{p}) \propto\ & p_1^{21} p_2^{24} p_3^{18} p_4^{20} p_5^{25} p_6^{12} \\
& \times (p_1 + p_2 + p_3)^{18} (p_4 + p_5 + p_6)^{22} \\
& \times (p_1 + p_4)^{10} (p_2 + p_5)^{20} (p_3 + p_6)^{10} \\
& \times \left(\frac{p_1}{p_1 + p_4} \right)^{53} \left(\frac{p_4}{p_1 + p_4} \right)^{47}.
\end{aligned}
$$

$$
\boldsymbol{a}^\top = \left[\overset{p_1}{21 + 53}, \ \overset{p_2}{24}, \ \overset{p_3}{18}, \ \overset{p_4}{20 + 47}, \ \overset{p_5}{25}, \ \overset{p_6}{12} \right],
$$

$$
\boldsymbol{b}^\top = \left[\overset{1}{18}, \ \overset{2}{22}, \ \overset{3}{10 - 53 - 47}, \ \overset{4}{20}, \ \overset{5}{10} \right],
$$

$$
\boldsymbol{\Delta}^\top = \begin{array}{c} j \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\begin{array}{c}
\begin{array}{cccccc} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 \end{array} \\
\left[\begin{array}{cccccc}
1 & 1 & 1 & & & \\
 & & & 1 & 1 & 1 \\
1 & & & 1 & & \\
 & 1 & & & 1 & \\
 & & 1 & & & 1
\end{array} \right]
\end{array}.
$$

Optimality condition

Let $s = \sum_{k=1}^{K} a_k + \sum_{j=1}^{q} b_j$, $Q^+ = \{j \mid b_j > 0, j = 1, \ldots, q\}$ and $Q^- = \{j \mid b_j < 0, j = 1, \ldots, q\}$ be the sets of indices of positive and negative elements in $\boldsymbol{b}$ respectively.

$$\ell(\boldsymbol{p}|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\Delta}) = \sum_{k=1}^{K} a_k \log p_k + \sum_{j=1}^{q} b_j \log \boldsymbol{\delta}_j^{\mathsf{T}} \boldsymbol{p} - s \left(\sum_{k=1}^{K} p_k - 1 \right).$$

Optimality condition: $\nabla \ell(\boldsymbol{p}) = 0$,

$$\frac{\partial \ell}{\partial p_k} = \frac{a_k}{p_k} + \sum_{j \in Q^+} \frac{|b_j| \Delta_{kj}}{\boldsymbol{\delta}_j^{\mathsf{T}} \boldsymbol{p}} - \sum_{j \in Q^-} \frac{|b_j| \Delta_{kj}}{\boldsymbol{\delta}_j^{\mathsf{T}} \boldsymbol{p}} - s = 0,$$

which is equivalent to

$$a_k + \left(\sum_{j \in Q^+} \frac{|b_j| \Delta_{kj}}{\boldsymbol{\delta}_j^{\mathsf{T}} \boldsymbol{p}} - \sum_{j \in Q^-} \frac{|b_j| \Delta_{kj}}{\boldsymbol{\delta}_j^{\mathsf{T}} \boldsymbol{p}} - s \right) p_k = 0.$$

Stable Weaver Algorithm

Algorithm 1 Stable Weaver Algorithm

Input: Observations $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\Delta})$

Initialize: $\boldsymbol{p}^{(0)} = (1/K, \ldots, 1/K)^\intercal$, $s = \mathbf{1}^\intercal \boldsymbol{a} + \mathbf{1}^\intercal \boldsymbol{b}$

repeat

$\boldsymbol{\tau} = \boldsymbol{b}/\boldsymbol{\Delta}^\intercal \boldsymbol{p}^{(t)}$ (element-wise division)

$\boldsymbol{\tau}^+ = \max(\boldsymbol{\tau}, \mathbf{0})$, $\boldsymbol{\tau}^- = \min(\boldsymbol{\tau}, \mathbf{0})$

$\boldsymbol{p}^{(t+1)} = \left[\boldsymbol{a} + (\boldsymbol{\Delta}\boldsymbol{\tau}^+) \circ \boldsymbol{p}^{(t)}\right] / (s\mathbf{1} - \boldsymbol{\Delta}\boldsymbol{\tau}^-)$

($\circ$ represents element-wise product)

$\boldsymbol{p}^{(t+1)} = \boldsymbol{p}^{(t+1)} / \text{sum}(\boldsymbol{p}^{(t+1)})$

until convergence

- The weaver algorithm updates the parameter by $\boldsymbol{p} = \boldsymbol{a}/(s\mathbf{1} - \boldsymbol{\Delta}\boldsymbol{\tau})$.
- Bayesian weaver is time-consuming due to the inner–outer iteration structure and the selection of the thickening parameter is difficult.

Application

- Contingency tables with merged and truncated cells.
- Polytomous response data with underlying categories. For example, the phenotype expressions on blood types.
- Interval censored time-to-event data with truncation in survival analysis.
- Include several well-known ranking models as special cases, such as the Bradley–Terry, Plackett–Luce models and their variants.

Results on Real Datasets

Algorithm		NASCAR (w/o ties)	NASCAR (w/ ties)	HKJC1416 (w/o ties)	HKJC1416 (w/ ties)
Stable Weaver	Iteration	22	459	40.4K	27.2K
	Time (s)	<0.01	0.03	38.46	86.40
Bayesian Weaver	Iteration	128K	263K	>1M	>1M
	Time (s)	25.27	50.12	>5000	>5000
MM	Iteration	22	–	40.4K	–
	Time (s)	<0.01	–	375.79	–
Trust Region*	Iteration	1937	5048	636†	649†
	Time (s)	74.31	125.68	1139.14	1835.37
ILSR	Iteration	12	–	4056	–
	Time (s)	0.06	–	1166.97	–
Self Consistency	Iteration	36798	11282	–‡	–
	Time (s)	11.61	2.08	–	–

* The number of iterations for the trust region constrained algorithm refers to the number of the objective function evaluations.

† We use the approximated Hessian matrix when fitting the trust region constrained algorithm to the HKJC1416 data because its calculation is too time-consuming.

‡ For the HKJC1416 data, the self-consistency approach converges to a wrong solution.

Results on Real Datasets (Cont'd)

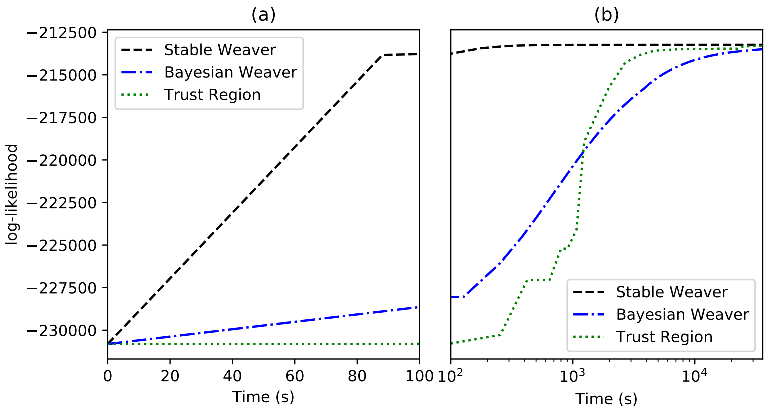

Figure 1: Convergence plot of the stable weaver algorithm compared with existing methods on the dataset HKJC9916 against running time (a) $t \in [0, 100]$ and (b) $t \in [100, 36000]$ (s).

Thanks for listening.