PARAMETER REDUNDANCY AND THE EXISTENCE OF MAXIMUM LIKELIHOOD ESTIMATES IN LOG-LINEAR MODELS

Serveh Sharifi Far, Michail Papathomas* and Ruth King

*University of Edinburgh and *University of St Andrews*

Supplementary Material

S1 Details of paper examples

Example 1. The initial log-linear model $\log \boldsymbol{\mu}_{27\times1} = A_{27\times19}\boldsymbol{\theta}_{19\times1}$, in the matrix form is as follows. Note that the μ indices correspond to cell counts 1 to 27 respectively.

$$
\begin{bmatrix}
\log\mu_{000}\\ \log\mu_{100}\\ \log\mu_{200}\\ \log\mu_{010}\\ \log\mu_{110}\\ \log\mu_{210}\\ \log\mu_{020}\\ \log\mu_{120}\\ \log\mu_{220}\\ \log\mu_{001}\\ \log\mu_{101}\\ \log\mu_{201}\\ \log\mu_{011}\\ \log\mu_{111}\\ \log\mu_{211}\\ \log\mu_{021}\\ \log\mu_{121}\\ \log\mu_{221}\\ \log\mu_{002}\\ \log\mu_{102}\\ \log\mu_{202}\\ \log\mu_{012}\\ \log\mu_{112}\\ \log\mu_{212}\\ \log\mu_{022}\\ \log\mu_{122}\\ \log\mu_{222}
\end{bmatrix}
=
\begin{bmatrix}
1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
1&0&1&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&0&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0\\
1&0&1&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&0&0&1&0&0&0&0&0&0&0&0&0&1&0&0&0\\
1&0&1&0&0&1&0&0&0&0&0&0&0&0&0&0&1&0&0\\
1&0&0&1&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0\\
1&1&0&1&0&1&0&1&0&0&0&1&0&0&0&1&0&0&0\\
1&0&1&1&0&1&0&0&1&0&0&1&0&0&0&0&1&0&0\\
1&0&0&0&1&1&0&0&0&0&0&0&1&0&0&0&0&0&0\\
1&1&0&0&1&1&0&0&0&1&0&0&1&0&0&1&0&0&0\\
1&0&1&0&1&1&0&0&0&0&1&0&1&0&0&0&1&0&0\\
1&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&1&0\\
1&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1\\
1&0&0&1&0&0&1&0&0&0&0&0&0&1&0&0&0&0&0\\
1&1&0&1&0&0&1&1&0&0&0&0&0&1&0&0&0&1&0\\
1&0&1&1&0&0&1&0&1&0&0&0&0&1&0&0&0&0&1\\
1&0&0&0&1&0&1&0&0&0&0&0&0&0&1&0&0&0&0\\
1&1&0&0&1&0&1&0&0&1&0&0&0&0&1&0&0&1&0\\
1&0&1&0&1&0&1&0&0&0&1&0&0&0&1&0&0&0&1
\end{bmatrix}
\begin{bmatrix}
\theta\\ \theta_1^X\\ \theta_2^X\\ \theta_1^Y\\ \theta_2^Y\\ \theta_1^Z\\ \theta_2^Z\\ \theta_{11}^{XY}\\ \theta_{21}^{XY}\\ \theta_{12}^{XY}\\ \theta_{22}^{XY}\\ \theta_{11}^{YZ}\\ \theta_{21}^{YZ}\\ \theta_{12}^{YZ}\\ \theta_{22}^{YZ}\\ \theta_{11}^{XZ}\\ \theta_{21}^{XZ}\\ \theta_{12}^{XZ}\\ \theta_{22}^{XZ}
\end{bmatrix},
$$

The reduced model $\log \boldsymbol{\mu}'_{21\times 1} = A'_{21\times 18}\boldsymbol{\theta}'_{18\times 1}$ in the matrix form is,

$$
\begin{bmatrix}
\log\mu_{200}\\ \log\mu_{010}\\ \log\mu_{110}\\ \log\mu_{210}\\ \log\mu_{020}\\ \log\mu_{120}\\ \log\mu_{220}\\ \log\mu_{001}\\ \log\mu_{101}\\ \log\mu_{201}\\ \log\mu_{011}\\ \log\mu_{111}\\ \log\mu_{021}\\ \log\mu_{121}\\ \log\mu_{202}\\ \log\mu_{012}\\ \log\mu_{112}\\ \log\mu_{212}\\ \log\mu_{022}\\ \log\mu_{122}\\ \log\mu_{222}
\end{bmatrix}
=
\begin{bmatrix}
0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&1&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&1&0&0&0&0&1&0&0&0&0&0&0&0&0&0\\
0&1&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&0&1&0&0&0&0&0&0&0&0&0&1&0&0&0\\
0&1&0&0&1&0&0&0&0&0&0&0&0&0&0&1&0&0\\
0&0&1&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0\\
1&0&1&0&1&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&0&0&1&1&0&0&0&0&0&0&1&0&0&0&0&0&0\\
1&0&0&1&1&0&0&0&1&0&0&1&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&1&0&0&1&0&0&0&0&0&0&1&0&0&0&0&0\\
1&0&1&0&0&1&1&0&0&0&0&0&1&0&0&0&1&0\\
0&1&1&0&0&1&0&1&0&0&0&0&1&0&0&0&0&1\\
0&0&0&1&0&1&0&0&0&0&0&0&0&1&0&0&0&0\\
1&0&0&1&0&1&0&0&1&0&0&0&0&1&0&0&1&0\\
0&1&0&1&0&1&0&0&0&1&0&0&0&1&0&0&0&1
\end{bmatrix}
\begin{bmatrix}
\theta^{X}_{1}\\ \theta+\theta^{X}_{2}\\ \theta+\theta^{Y}_{1}\\ \theta+\theta^{Y}_{2}\\ \theta+\theta^{Z}_{1}\\ \theta^{Z}_{2}\\ \theta^{XY}_{11}\\ -\theta+\theta^{XY}_{21}\\ \theta^{XY}_{12}\\ -\theta+\theta^{XY}_{22}\\ -\theta+\theta^{YZ}_{11}\\ -\theta+\theta^{YZ}_{21}\\ \theta^{YZ}_{12}\\ \theta^{YZ}_{22}\\ \theta^{XZ}_{11}\\ -\theta+\theta^{XZ}_{21}\\ \theta^{XZ}_{12}\\ \theta^{XZ}_{22}.
\end{bmatrix}.
$$

Example 2. The vector of 59 estimable parameters obtained by parameter redundancy for the $3^5 \times 2^1$ contingency table is,

$$
\begin{aligned}
\boldsymbol{\theta}'^{\mathsf{T}} =&(\theta, \theta^{A}_{1}, \theta^{A}_{2}, \theta^{B}_{1}, \theta^{B}_{2}, \theta^{C}_{1}, \theta^{C}_{2}, \theta^{D}_{1}, \theta^{D}_{2}, \theta^{E}_{1}, \theta^{E}_{2}, \theta^{F}_{1}, \theta^{AB}_{11}, \theta^{AB}_{21}, \theta^{AB}_{12}, \theta^{AB}_{22}, \theta^{AC}_{11}, \theta^{AC}_{21},\\
&\theta^{AC}_{12}, \theta^{AC}_{22}, \theta^{AD}_{11}, \theta^{AD}_{21}, \theta^{AD}_{12}, \theta^{AE}_{11}, \theta^{AE}_{21}, \theta^{AE}_{12}, \theta^{AF}_{11}, \theta^{AF}_{21}, \theta^{BC}_{11}, \theta^{BC}_{21}, \theta^{BC}_{12}, \theta^{BC}_{22},\\
&\theta^{BD}_{11}, \theta^{BD}_{21}, \theta^{BD}_{12}, \theta^{BD}_{22}, \theta^{BE}_{11}, \theta^{BE}_{21}, \theta^{BE}_{12}, \theta^{BF}_{11}, \theta^{BF}_{21}, \theta^{BE}_{22}, \theta^{CD}_{11}, \theta^{CD}_{21}, \theta^{CD}_{12}, \theta^{CD}_{22},\\
&\theta^{CE}_{11}, \theta^{CE}_{21}, \theta^{CE}_{12}, \theta^{CE}_{22}, \theta^{CF}_{11}, \theta^{CF}_{21}, \theta^{DE}_{11}, \theta^{DE}_{21}, \theta^{DE}_{12}, \theta^{DF}_{11}, \theta^{DF}_{21}, \theta^{EF}_{11}, \theta^{EF}_{21}).
\end{aligned}
$$

Example 3. Model (4.11) can be written as,

$$\begin{bmatrix} \log\mu_{000} \\ \log\mu_{100} \\ \log\mu_{010} \\ \log\mu_{110} \\ \log\mu_{001} \\ \log\mu_{101} \\ \log\mu_{011} \\ \log\mu_{111} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \theta \\ \theta^X \\ \theta^Y \\ \theta^{XY} \\ \theta^Z \\ \theta^{XZ} \\ \theta^{YZ} \end{bmatrix}.$$

The derivative matrix for contingency table in Table 2(a) is,

$$D = \left[\begin{array}{c|cccccccc} & \mu_{000} & \mu_{100} & \mu_{010} & \mu_{110} & \mu_{001} & \mu_{101} & \mu_{011} & \mu_{111} \\ \hline \theta & 0 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & 0 \\ \theta^X & 0 & y_2 & 0 & y_4 & 0 & y_6 & 0 & 0 \\ \theta^Y & 0 & 0 & y_3 & y_4 & 0 & 0 & y_7 & 0 \\ \theta^{XY} & 0 & 0 & 0 & y_4 & 0 & 0 & 0 & 0 \\ \theta^Z & 0 & 0 & 0 & 0 & y_5 & y_6 & y_7 & 0 \\ \theta^{XZ} & 0 & 0 & 0 & 0 & 0 & y_6 & 0 & 0 \\ \theta^{YZ} & 0 & 0 & 0 & 0 & 0 & 0 & y_7 & 0 \end{array}\right].$$

Example 4. The derivative matrix for contingency table in Table 2(b) is,

$$D = \left[\begin{array}{c|cccccccc} & \mu_{000} & \mu_{100} & \mu_{010} & \mu_{110} & \mu_{001} & \mu_{101} & \mu_{011} & \mu_{111} \\ \hline \theta & 0 & y_2 & y_3 & 0 & y_5 & y_6 & y_7 & y_8 \\ \theta^X & 0 & y_2 & 0 & 0 & 0 & y_6 & 0 & y_8 \\ \theta^Y & 0 & 0 & y_3 & 0 & 0 & 0 & y_7 & y_8 \\ \theta^{XY} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_8 \\ \theta^Z & 0 & 0 & 0 & 0 & y_5 & y_6 & y_7 & y_8 \\ \theta^{XZ} & 0 & 0 & 0 & 0 & 0 & y_6 & 0 & y_8 \\ \theta^{YZ} & 0 & 0 & 0 & 0 & 0 & 0 & y_7 & y_8 \end{array}\right].$$

Example S1. It is known that for a log-linear model fitted to a contingency table with all positive $y_{\mathbf{i}}$, the log-likelihood function is strictly concave and the maximum likelihood estimates exist for all the model parameters. Consider fitting a saturated Poisson log-linear model to an l^m $(m \geqslant 1, l \geqslant 2)$

contingency table. The derivative matrices for a 2^1 and a 2^2 table are,

$$D_1 = \left[\begin{array}{c|cc} & \mu_0 & \mu_1 \\ \hline \theta & y_1 & y_2 \\ \theta^X & 0 & y_2 \end{array}\right], \qquad D_2 = \left[\begin{array}{c|cccc} & \mu_{00} & \mu_{10} & \mu_{01} & \mu_{11} \\ \hline \theta & y_1 & y_2 & y_3 & y_4 \\ \theta^X & 0 & y_2 & 0 & y_4 \\ \theta^Y & 0 & 0 & y_3 & y_4 \\ \theta^{XY} & 0 & 0 & 0 & y_4 \end{array}\right].$$

Even for larger tables, we can always arrange an ordering of cell means and corresponding parameters that produces an upper triangular D matrix in which the main diagonal elements are the cell counts, as shown in D_1 and D_2 above (and also in the proof of Theorem 1). So when $y_{\mathbf{i}} > 0, \forall \mathbf{i} \in L$, the D matrix is always full rank, as expected, and all of the model parameters are estimable.

S2 Proof of Theorem 1

To prove Theorem 1, we use the induction method for two variables in two steps. First, the statement is proven to be true for an l^1 table for all integers $l \geqslant 2$. Then we show that if the statement is assumed to be true for an l^m table, it is also true for l^{m+1} for all integers $l \geqslant 2$ (Earl, 2017). For simplicity, instead of $y_{\mathbf{i}}$ and 0 in the derivative matrix we write 1 and 0. This helps relate the derivative matrix of m variables and the one with $m+1$ variables. Recall that a zero cell turns a corresponding column to zero in the derivative matrix. To clarify the notation, without loss of generality, assume the contingency table has m variables and each of them has l levels.

We set $D_r(\boldsymbol{\theta}_r) = \dfrac{d\boldsymbol{\mu}_r}{d\boldsymbol{\theta}_r}$, in which $\boldsymbol{\mu}_r$ and $\boldsymbol{\theta}_r$ are the set of cell means and parameters added to the model because of adding the rth variable to the table. Then we define $D_r = D_{\underline{r}}(\underline{\boldsymbol{\theta}_r}) = \dfrac{d\underline{\boldsymbol{\mu}_r}}{d\underline{\boldsymbol{\theta}_r}}$, as the derivative matrix for $\underline{\boldsymbol{\mu}_r} = \boldsymbol{\mu}_1 \cup \boldsymbol{\mu}_2 \cup \cdots \cup \boldsymbol{\mu}_r$ and $\underline{\boldsymbol{\theta}_r} = \boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2 \cup \cdots \cup \boldsymbol{\theta}_r$, which are union of sets of cell means and model parameters for having variables 1 to r. Accordingly, $D_r(\underline{\boldsymbol{\theta}_r}) = \dfrac{d\boldsymbol{\mu}_r}{d\underline{\boldsymbol{\theta}_r}}$. In the tables and matrices, the $y_{\mathbf{i}}$'s are ordered according to (1.2) in the main paper.

Before we derive the derivative matrix and nonestimable parameters for a general case of $m = k$, we start with a simple table and gradually discover the pattern in the structure of the derivative matrices. For a 2^1 table, $\boldsymbol{\alpha}$ and the nonestimable parameters in presence of zero cell counts are shown here. Since only one cell count is zero, the deficiency is one and there is one $\boldsymbol{\alpha}$ vector for each case.

$$m = 1, \quad D_1 = D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) = D_1(\boldsymbol{\theta}_1) = \left[\begin{array}{c|cc} & \mu_0 & \mu_1 \\ \hline \theta & 1 & 1 \\ \theta^X & 0 & 1 \end{array}\right],$$

$$\boldsymbol{\theta}_1 = (\theta, \theta^X), \quad \boldsymbol{\mu}_1 = (\mu_0, \mu_1).$$

zero cell	$\boldsymbol{\alpha}$ vector	nonestimable parameters
$y_0 = 0$	$\boldsymbol{\alpha}_{11} = (1, -1)$	$\boldsymbol{\gamma}_1 = \{\theta, \theta^X\}$
$y_1 = 0$	$\boldsymbol{\alpha}_{12} = (0, 1)$	$\boldsymbol{\gamma}_2 = \{\theta^X\}$

Those $\boldsymbol{\alpha}$ vectors are actually $\boldsymbol{\alpha}_{11} = (\alpha, -\alpha)$ and $\boldsymbol{\alpha}_{12} = (0, \alpha)$, where α

could be any non-zero number but for simplification the value 1 is used.

For the model corresponding to a 2^2 table, the derivative matrix, and nonestimable parameters for setting each cell count to zero are,

$$m = 2, \quad D_2 = D_{\underline{2}}(\underline{\boldsymbol{\theta}_2}) = \left[\begin{array}{c|cc:cc} & \mu_{00} & \mu_{10} & \mu_{01} & \mu_{11} \\ \hline \theta & 1 & 1 & 1 & 1 \\ \theta^X & 0 & 1 & 0 & 1 \\ \hdashline \theta^Y & 0 & 0 & 1 & 1 \\ \theta^{XY} & 0 & 0 & 0 & 1 \end{array}\right]$$

$$= \begin{bmatrix} D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) & D_2(\boldsymbol{\theta}_1) \\ \mathbf{0} & D_2(\boldsymbol{\theta}_2) \end{bmatrix} = \begin{bmatrix} D_1 & D_2(\boldsymbol{\theta}_1) \\ \mathbf{0} & D_1 \end{bmatrix},$$

$$\boldsymbol{\theta}_1 = (\theta, \theta^X), \qquad \boldsymbol{\theta}_2 = (\theta^Y, \theta^{XY}), \qquad \underline{\boldsymbol{\theta}_2} = (\theta, \theta^X, \theta^Y, \theta^{XY}),$$

$$\boldsymbol{\mu}_1 = (\mu_{00}, \mu_{10}), \qquad \boldsymbol{\mu}_2 = (\mu_{01}, \mu_{11}), \qquad \underline{\boldsymbol{\mu}_2} = (\mu_{00}, \mu_{10}, \mu_{01}, \mu_{11}).$$

zero cell	$\boldsymbol{\alpha}$ vector	nonestimable parameters
$y_{00} = y_1 = 0$	$\boldsymbol{\alpha}_{21} = (1, -1, -1, 1) = (\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{11})$	$\boldsymbol{\gamma}_1 = \{\theta, \theta^X, \theta^Y, \theta^{XY}\}$
$y_{10} = y_2 = 0$	$\boldsymbol{\alpha}_{22} = (0, 1, 0, -1) = (\boldsymbol{\alpha}_{12}, \boldsymbol{\alpha}_{12})$	$\boldsymbol{\gamma}_2 = \{\theta^X, \theta^{XY}\}$
$y_{01} = y_3 = 0$	$\boldsymbol{\alpha}_{23} = (0, 0, 1, -1) = (\mathbf{0}, \boldsymbol{\alpha}_{11})$	$\boldsymbol{\gamma}_3 = \{\theta^Y, \theta^{XY}\}$
$y_{11} = y_4 = 0$	$\boldsymbol{\alpha}_{24} = (0, 0, 0, 1) = (\mathbf{0}, \boldsymbol{\alpha}_{12})$	$\boldsymbol{\gamma}_4 = \{\theta^{XY}\}$

The expression $\boldsymbol{\alpha}_{21} = (\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{11})$ is true in terms of places of zero and non-zero elements which indicate estimable and nonestimable parameters. The pattern in the derivative matrices and $\boldsymbol{\alpha}$ vectors holds for increasing m and any l, as used in the proof below.

Proof. *Step one:* We prove that the statement is true for l^1 for all integers $l \geqslant 2$. Assume the only variable in the model is X with $[l] = \{0, 1, ..., l-1\}$ levels, therefore the saturated model includes l parameters. The derivative matrix for this model is,

$$D_1 = D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) = \left[\begin{array}{c|cccccc} & \mu_0 & \mu_1 & \mu_2 & \mu_3 & & \mu_{l-1} \\ \hline \theta & 1 & 1 & 1 & 1 & \ldots & 1 \\ \theta_1^X & 0 & 1 & 0 & 0 & \ldots & 0 \\ \theta_2^X & 0 & 0 & 1 & 0 & \ldots & 0 \\ \theta_3^X & 0 & 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta_{l-1}^X & 0 & 0 & 0 & 0 & \ldots & 1 \end{array}\right].$$

For this model, we show the $\boldsymbol{\alpha}$ vectors and the nonestimable parameters in the presence of zero cell counts. Since only one cell count is zero, the deficiency is one and there is one $\boldsymbol{\alpha}$ for each case.

zero cell	$\boldsymbol{\alpha}$ vector	nonestimable parameters
$y_0 = 0$	$\boldsymbol{\alpha}_{11} = (1, -1, -1, -1, \ldots, 1)$	all parameters
$y_1 = 0$	$\boldsymbol{\alpha}_{12} = (0, 1, 0, 0, \ldots, 0)$	θ_1^X
$y_2 = 0$	$\boldsymbol{\alpha}_{13} = (0, 0, 1, 0, \ldots, 0)$	θ_2^X
$y_3 = 0$	$\boldsymbol{\alpha}_{14} = (0, 0, 0, 1, \ldots, 0)$	θ_3^X
$\vdots$	$\vdots$	$\vdots$
$y_{l-1} = 0$	$\boldsymbol{\alpha}_{1l} = (0, 0, 0, \ldots, 0, 1)$	θ_{l-1}^X

According to the $\boldsymbol{\alpha}$ vectors, the theorem statement is true for this model. We can fix the number of variables at $m = 2$ and show that the statement is still true for this model with any number of levels. Assume the variables in this model are X and Y with $[l] = \{0, 1, ..., l-1\}$ levels, the derivative matrix for the model for this l^2 table is, $D_2 = D_{\underline{2}}(\underline{\boldsymbol{\theta}_2})$,

$$
= \left[\begin{array}{c|ccccc:ccccc:c:ccccc}
 & \multicolumn{5}{c}{Y=0} & \multicolumn{5}{c}{Y=1} & & \multicolumn{5}{c}{Y=l-1} \\
 & \mu_{00} & \mu_{10} & \mu_{20} & \dots & \mu_{l-10} & \mu_{01} & \mu_{11} & \mu_{21} & \dots & \mu_{l-11} & \dots & \mu_{0l-1} & \mu_{1l-1} & \mu_{2l-1} & \dots & \mu_{l-1l-1} \\
\hline
\theta & 1 & 1 & 1 & \dots & 1 & 1 & 1 & 1 & \dots & 1 & \dots & 1 & 1 & 1 & \dots & 1 \\
\theta_1^X & 0 & 1 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\
\theta_2^X & 0 & 0 & 1 & \dots & 0 & 0 & 0 & 1 & \dots & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\theta_{l-1}^X & 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & 0 & \dots & 1 \\
\hdashline
\theta_1^Y & 0 & 0 & 0 & \dots & 0 & 1 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\
\theta_{11}^{XY} & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
\theta_{21}^{XY} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & \dots & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\theta_{l-11}^{XY} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\
\hdashline
\vdots & & & \vdots & & & & & \vdots & & & \vdots & & & \vdots & & \\
\hdashline
\theta_{l-1}^Y & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & 1 & \dots & 1 \\
\theta_{1l-1}^{XY} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\
\theta_{2l-1}^{XY} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\theta_{l-1l-1}^{XY} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & \dots & 1
\end{array}\right]
$$

$$
= \begin{bmatrix} D_1 & D_1 & \dots & D_1 \\ \mathbf{0} & D_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & D_1 \end{bmatrix}.
$$

The derivative matrix is upper triangular and all elements on the main diagonal are 1. Let $y_{\mathbf{i}(0)}$ be a cell count such that its index ends with zero and $\boldsymbol{\gamma}_{\mathbf{i}}$ is the set including corresponding nonestimable parameters. We can order cells from 1 to l^m according to (1.2). Thus, in the case of having one zero cell count, the nonestimable parameters and unique $\boldsymbol{\alpha}$ vectors are as follows which satisfy the theorem's statement.

zero cell	$\boldsymbol{\alpha}$ vector	nonestimable parameters
$y_{\mathbf{i}(0)} = y_1 = 0$	$\boldsymbol{\alpha}_{21} = \overbrace{(\boldsymbol{\alpha}_{11}, \ldots, \boldsymbol{\alpha}_{11})}^{\#l}$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_1 = \{\text{all parameters}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(0)} = y_l = 0$	$\boldsymbol{\alpha}_{2l} = (\boldsymbol{\alpha}_{1l}, \ldots, \boldsymbol{\alpha}_{1l})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_l = \{\theta^{X}_{l-1}, \theta^{XY}_{l-11}, \ldots, \theta^{XY}_{l-1l-1}\}$
$y_{\mathbf{i}(1)} = y_{l+1} = 0$	$\boldsymbol{\alpha}_{2(l+1)} = (\mathbf{0}, \boldsymbol{\alpha}_{11}, \mathbf{0}, \ldots, \mathbf{0})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l+1} = \{\theta^{Y}_{1}, \theta^{XY}_{11}, \ldots, \theta^{XY}_{l-11}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(1)} = y_{l\times 2} = 0$	$\boldsymbol{\alpha}_{2(l\times 2)} = (\mathbf{0}, \boldsymbol{\alpha}_{1l}, \mathbf{0}, \ldots, \mathbf{0})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l\times 2} = \{\theta^{XY}_{l-11}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^2-l+1} = 0$	$\boldsymbol{\alpha}_{2(l^2-l+1)} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{11})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^2-l+1} = \{\theta^{Y}_{l-1}, \theta^{XY}_{1l-1}, \ldots, \theta^{XY}_{l-1l-11}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^2} = 0$	$\boldsymbol{\alpha}_{2l^2} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{1l})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^2} = \{\theta^{XY}_{l-1l-1}\}$

Step two: The statement is assumed to be true for l^m when $m = k$, we will show it is also true when $m = k + 1$. For $m = k$ when any of the cell counts is zero, the corresponding parameter to that cell and given that, all other parameters with a higher order interaction of the variables are assumed to be nonestimable. The derivative matrix is,

$$D_k = D_{\underline{k}}(\underline{\boldsymbol{\theta}_k}) = \begin{bmatrix} D_{\underline{k-1}}(\underline{\boldsymbol{\theta}_{k-1}}) & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ 0 & D_k(\boldsymbol{\theta}_k) \end{bmatrix} = \begin{bmatrix} D_{k-1} & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_k(\boldsymbol{\theta}_k) \end{bmatrix},$$

in which,

$$D_k(\boldsymbol{\theta}_k) = \begin{bmatrix} D_{k-1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & D_{k-1} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & D_{k-1} \end{bmatrix}_{l-1\times l-1}.$$

Derivative matrices are upper triangular and all elements on their main diagonals are 1. Say $y_{\mathbf{i}(0)}$ is a cell count such that its index ends with zero.

$\boldsymbol{\gamma}_{\mathbf{i}}$ is the set including the corresponding parameter to that cell and given that, all other parameters associated with a higher order interaction of the variables. The order of setting cell counts to zero here is the same order used in forming the derivative matrix. Thus, the nonestimable parameters must be as follows (same for $\boldsymbol{\alpha}$ vectors, because of the repetitive pattern in models and the point that in each case there is only one $\boldsymbol{\alpha}$ vector).

zero cell	$\boldsymbol{\alpha}$ vector	nonestimable parameters
$y_{\mathbf{i}(0)} = y_1 = 0$	$\boldsymbol{\alpha}_{k1} = \overbrace{(\boldsymbol{\alpha}_{k-1(1)}, \dots, \boldsymbol{\alpha}_{k-1(1)})}^{\#l}$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_1 = \{\text{all parameters}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(0)} = y_{l^{k-1}} = 0$	$\boldsymbol{\alpha}_{kl^{k-1}} = (\boldsymbol{\alpha}_{k-1(l^{k-1})}, \dots, \boldsymbol{\alpha}_{k-1(l^{k-1})})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^{k-1}}$
$y_{\mathbf{i}(1)} = y_{l^{k-1}+1} = 0$	$\boldsymbol{\alpha}_{k(l^{k-1}+1)} = (\mathbf{0}, \boldsymbol{\alpha}_{k-1(1)}, \mathbf{0}, \dots, \mathbf{0})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^{k-1}+1}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(1)} = y_{l^{k-1}\times 2} = 0$	$\boldsymbol{\alpha}_{k(l^{k-1}\times 2)} = (\mathbf{0}, \boldsymbol{\alpha}_{k-1(l^{k-1})}, \mathbf{0}, \dots, \mathbf{0})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^{k-1}\times 2}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{(l^{k-1}\times l-1)+1} = 0$	$\boldsymbol{\alpha}_{k((l^{k-1}\times l-1)+1)} = (\mathbf{0}, \mathbf{0}, \dots, \boldsymbol{\alpha}_{k-1(1)})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{(l^{k-1}\times l-1)+1}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^k} = 0$	$\boldsymbol{\alpha}_{kl^k} = (\mathbf{0}, \mathbf{0}, \dots, \boldsymbol{\alpha}_{k-1(l^{k-1})})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^k} = \{$only the highest order parameter$\}$

Now the theorem statement must be proven for $m = k + 1$. We have,

$$D_{k+1} = D_{\underline{k+1}}(\underline{\boldsymbol{\theta}_{k+1}}) = \begin{bmatrix} D_{\underline{k}}(\underline{\boldsymbol{\theta}_k}) & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ 0 & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix} = \begin{bmatrix} D_k & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ 0 & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix},$$

in which,

$$D_{k+1}(\boldsymbol{\theta}_{k+1}) = \begin{bmatrix} D_k & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & D_k & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & D_k \end{bmatrix}_{l-1\times l-1}.$$

So the nonestimable parameters are expected to be,

zero cell	nonestimable parameters
$y_{\mathbf{i}(0)} = y_1 = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_1 = \{\text{all parameters}\}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(0)} = y_{l^k} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^k}$
$y_{\mathbf{i}(1)} = y_{l^k+1} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^k+1}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(1)} = y_{l^k\times 2} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^k\times 2}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{(l^k\times l-1)+1} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{(l^k\times l-1)+1}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^{k+1}} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^{k+1}} = \{\text{only the highest order parameter}\}$

To prove that these are nonestimable parameters, we need to obtain the corresponding $\boldsymbol{\alpha}$ vectors. According to the repetitive pattern of $\boldsymbol{\alpha}$ vectors, that was observed when constructing the derivative matrices by increasing the number of variables in the table, they are made of vectors of the previous step. Therefore the unique $\boldsymbol{\alpha}$ vectors are,

zero cell	$\boldsymbol{\alpha}$ vector
$y_{\mathbf{i}(0)} = y_1 = 0$	$\boldsymbol{\alpha}_{k+1(1)} = \overbrace{(\boldsymbol{\alpha}_{k1}, \ldots, \boldsymbol{\alpha}_{k1})}^{\#l}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(0)} = y_{l^k} = 0$	$\boldsymbol{\alpha}_{k+1(l^k)} = (\boldsymbol{\alpha}_{kl^k}, \ldots, \boldsymbol{\alpha}_{kl^k})$
$y_{\mathbf{i}(1)} = y_{l^k+1} = 0$	$\boldsymbol{\alpha}_{k+1(l^k+1)} = (\mathbf{0}, \boldsymbol{\alpha}_{k1}, \mathbf{0}, \ldots, \mathbf{0})$
$\vdots$	$\vdots$
$y_{\mathbf{i}(1)} = y_{l^k\times 2} = 0$	$\boldsymbol{\alpha}_{k+1(l^k\times 2)} = (\mathbf{0}, \boldsymbol{\alpha}_{kl^k}, \mathbf{0}, \ldots, \mathbf{0})$
$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{(l^k\times l-1)+1} = 0$	$\boldsymbol{\alpha}_{k+1((l^k\times l-1)+1)} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{k1})$
$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^{k+1}} = 0$	$\boldsymbol{\alpha}_{k+1l^{k+1}} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{kl^k})$

For the first $\frac{1}{l}$ proportion of the cases in the previous table, having a zero cell count makes $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{ki}, \ldots, \boldsymbol{\alpha}_{ki})$. Since the theorem is assumed to be true for $m = k$, the first $\boldsymbol{\alpha}_{ki}$ makes the corresponding parameter to that cell and given that, all other parameters with a higher order interaction of variables be nonestimable for the last smaller model ($m = k$). Repeating $\boldsymbol{\alpha}_{ki}$, $l-1$ times in the $\boldsymbol{\alpha}$ vector makes some other parameters of the new model to be nonestimable, which are the same previous parameters corresponding to all levels of the new variable. Hence, the corresponding parameter to that cell and given that, all other parameters with a higher order interaction of the variables are nonestimable.

For the rest of the $\frac{1}{l}$ parts of the cases, having a zero cell count makes an $\boldsymbol{\alpha}_{ki}$ appear in the vector. This $\boldsymbol{\alpha}_{ki}$ makes the corresponding parameter to that cell and given that, all other parameters with a higher order interaction of the variables be nonestimable for the last smaller model, but as it appeared after one or more vectors of zeroes here, those parameters will have the higher levels of the new variable in their superscript and subscript. Hence, the corresponding parameter to that cell and given that, all other parameters with a higher order interaction of the variables are nonestimable. Therefore the statement is true for $m = k + 1$. □

References

Earl, R. (2017). *Towards Higher Mathematics: A Companion* . Cambridge University Press.