

# Orthogonal Log-Linear Models for Tables in Presence of Zero Cells

Yves Thibaudeau

U.S. Census Bureau

## Abstract

Fienberg and Rinaldo (2012) propose the "extended maximum likelihood estimator" (EMLE) to recuperate estimable parameters when zero cells in a contingency table preclude the estimation of a complete MLE for a given log-linear model. The EMLE provides positive estimates for all the estimable cells and postulates structural zeros in place of the likelihood zeros. That is the zero cells impeding the existence of an intact MLE. We propose extending the method of Fienberg to involve cell predictions for the likelihood zeros instead of postulating structural zeros while also preserving the parametric structure recuperated through the EMLE. Our approach leads to a principled method to identify surviving parameterizations and to define orthogonalized designs staying clear of estimation sinkholes. We include "R" code bringing to life all the algebraic constructs in the paper.

## Disclaimer

*Any views expressed on statistical methodology, technical or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.*

## 1 Introduction

Fienberg and Rinaldo (2012) discuss the concept of "likelihood zeros". Those are typically patterned zero cells in the contingency tables whose presence as a group lead to a divergent Newton-Raphson maximization procedure when attempting to maximize the likelihood. Of course not all zero cells are likelihood zeros meaning covering or uncovering some of the zero cells will have no effect on estimability. Such cells are "sampling zeros".

The same authors propose general methods to identify in a contingency table the likelihood zeros impeding maximization of the likelihood. Friedlander (2016) designs an "R" package to dig out all the likelihood zero given a log-linear model. In the situation of the paper the likelihood zeros are trivial to identify and/or have been identified by authors previously.

Sharifi Far et al. (2021) algebraically explore the poisson likelihood to motivate a parametric model approach to explain and transcend the EMLE of Fienberg and Rinaldo but stop short of expanding the linear algebra characterizing the parametric "sinkhole" generated by likelihood zeros.

We show how to capture a comprehensive representation of the linear subspace

associated to the likelihood zeros. It is over this subspace that the parameters to be estimated diverge to infinity. In the case of a multinomial distribution this analysis can be carried in the space of log-odd ratios of the cell probabilities and is easily recasted as a poisson regression in this case with unbounded parameters.

We will derive general algebraic results throughout and illustrate them on a simple example as we expand the methodology. In the next sections we present a demonstration example of a multinomial log-linear model.

In the last section we introduce an application that has already been the object of analysis and we implement our approach along with glmnet regularization.

## 2 Multinomial Models in the Log-Odds Space

### 2.1 Multinomial Design Matrices

For simplicity we limit our presentation to binary response variables. The methodology fully carries to situations involving multicategorical variables. Without loss of generality we consider the following product multinomial likelihood subject to log-linear constraints:

$$L_X(\mathbf{x}; \boldsymbol{\pi}) = \prod_{m_1=1, \dots, m_O=1}^2 \pi_{m_1, \dots, m_O}^{x_{m_1, \dots, m_O}} \quad \boldsymbol{\pi} \in \boldsymbol{\Omega} \quad (1)$$

where  $\pi_{m_1, \dots, m_O}^{x_{m_1, \dots, m_O}}$  is a probability vector and  $m_1 = \dots = m_O = 2$ .  $\boldsymbol{\Omega}$  is a multinomial hierarchical log-linear model on  $O$  binary categorical variables.

Let  $O = 3$  and assume a "no second-order interaction model" on a  $2 \times 2 \times 2$  contingency table. Then

$$\mathbf{x} = [x_{1,1,1} \quad x_{2,1,1} \quad \dots \quad x_{2,2,2}] \quad (2)$$

is a realization of the random vector of counts

$$\mathbf{X} = [X_{1,1,1} \quad X_{2,1,1} \quad \dots \quad X_{2,2,2}] \quad (3)$$

The ANOVA-style representation (Bishop, Fienberg, Holland, 1974, p. 37) for  $\boldsymbol{\Omega}$  in (1) is:

$$\log(\pi_{j,k,l}) = u + u_{1(j)} + u_{2(k)} + u_{3(l)} + u_{12(j,k)} + u_{13(j,l)} + u_{23(k,l)} \quad (4)$$

$$j, k, l = 1, 2$$

where

$$\sum_{j=1}^2 u_{1(j)} = \sum_{k=1}^2 u_{2(k)} = \sum_{l=1}^2 u_{3(l)} = 0 \quad (5)$$

$$\sum_{j=1}^2 u_{12(j,k)} = \sum_{j=1}^2 u_{13(j,l)} = \sum_{k=1}^2 u_{23(k,l)} = 0 \quad (6)$$

$$\sum_{k=1}^2 u_{11(j,k)} = \sum_{l=1}^2 u_{13(j,l)} = \sum_{l=1}^2 u_{23(k,l)} = 0 \quad (7)$$

where the index not involved in each summation is set at 1 and 2 in turn. and

$$\sum_{j,k,l=1}^2 \pi_{j,k,l} = 1 \quad (8)$$

Note the constraints (4) - (7) are linear in  $\log(\pi_{j,k,l})$  and in the model parameters, whereas the constraint (8) is not. So (4) - (8) does not describe a linear model. Formulating an equivalent model in the space of the log odd ratios reduces the number of parameters and lead to linearity. To do so we designate a reference category, which we also call the baseline. Let

$$\psi_{j,k,l} = \log\left(\frac{\pi_{j,k,l}}{\pi_{1,1,1}}\right) \quad (9)$$

The baseline implicitly defined in (9) is category  $(j, k, l) = (1, 1, 1)$ . In general the log-odd ratio evaluated at the baseline is 0. In this case we have  $\psi_{1,1,1} = 0$ . Let the vector of log-odd ratios be

$$\boldsymbol{\psi} = [\psi_{1,\dots,1} \quad \dots \quad \psi_{2,\dots,2}]^t \quad (10)$$

In general for a log-linear hierarchical multinomial model we have:

$$\boldsymbol{\psi} = \mathbf{D}\mathbf{u} \quad (11)$$

where

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{u} = -2 \begin{bmatrix} u_{1(1)} \\ u_{2(1)} \\ u_{3(1)} \\ u_{12(1,1)} \\ u_{13(1,1)} \\ u_{23(1,1)} \end{bmatrix} \quad (12)$$

In general, for hierarchical log-linear models matrices such as  $\mathbf{D}$  in (11) can be derived directly from the constraints of the model through toric algebra (Geiger, Meek, Sturmfels, 2006). The problem here is  $\mathbf{D}$  does not comprehensively characterize the multinomial distribution of  $\mathbf{X}$  in the sense  $\mathcal{CS}\{\mathbf{D}\}$ , the

column space of  $\mathbf{D}$  in (12), does not. For example let  $\mathbf{D}^*$  be the matrix of log odd ratios computed using category (2, 1, 1) for baseline instead of (1, 1, 1). Then  $\mathcal{CS}\{\mathbf{D}^*\} \neq \mathcal{CS}\{\mathbf{D}\}$ .

## 2.2 Poisson Design Matrices

To comprehensively characterize a multinomial log-linear model as a general linear model we represent multinomial log-linear modeling in the space of log-odd ratios as poisson design matrices. Let  $T = \sum_{j,k,l=1}^2 X_{j,k,l}$  be the total of the random counts. Under a multinomial model we have assumed a realization of  $T$  is observed. Let  $T \sim \text{Poisson}(\exp(\lambda))$ , where  $\lambda = \log(\mathbb{E}[T])$  and let

$$\xi_{j,k,l} = \log(\mathbb{E}[X_{j,k,l}]) \quad (13)$$

$$= \log(\mathbb{E}[T_s] \pi_{j,k,l}) \quad (14)$$

$$= \log(\pi_{j,k,l}) + \lambda \quad (15)$$

Keeping the same baseline as in (9) and from (15) we have

$$\xi_{j,k,l} - \xi_{1,1,1} = \phi_{j,k,l} \quad (16)$$

So

$$\xi_{j,k,l} = \xi_{1,1,1} + \phi_{j,k,l} \quad (17)$$

and

$$\boldsymbol{\xi} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \xi_{1,1,1} + \boldsymbol{\phi} \quad (18)$$

where

$$\boldsymbol{\xi}_s = [\xi_{1,1,\dots,1} \quad \dots \quad \xi_{2,2,\dots,2}]^t \quad (19)$$

We have

$$\boldsymbol{\xi} = \boldsymbol{\Delta} \boldsymbol{\mu} \quad (20)$$

where

$$\boldsymbol{\Delta} = [\mathbf{1} \quad | \quad \mathbf{D}] \quad (21)$$

and

$$\boldsymbol{\mu} = [\xi_{1,1,1} \quad | \quad \mathbf{u}^t]^t \quad (22)$$

$\boldsymbol{\Delta}$  is a poisson design matrix with an intercept (R glm documentation). Now let

$$\boldsymbol{\Delta}^* = [\mathbf{1} \quad | \quad \mathbf{D}^*] \quad (23)$$

then  $\mathcal{CS}\{\boldsymbol{\Delta}^*\} = \mathcal{CS}\{\boldsymbol{\Delta}\}$ .

In general changing the baseline when computing the log-odd ratios will not

alter the column space of the poisson design matrix with an intercept. The column space of a poisson design matrix comprehensively characterizes a specific multinomial likelihood in the sense it contains the column spaces of all the associated multinomial design matrices. There is no ambiguity when reconstructing the multinomial log-linear model from a poisson design matrix with an intercept.

This no longer holds in general if the poisson design matrix does not include an intercept. Then there may not be a single log-linear multinomial model for all the baselines. We can no longer assume the poisson design matrix unambiguously characterize a multinomial.

### 3 Estimation

#### 3.1 Likelihood Maximization

If the MLE exists the R function glm will identify it through the Newton Raphson procedure given an appropriate starting point. For the example in the paper we show using the default starting point leads to valid results.

#### 3.2 Likelihood Zeros

It is well known (Haberman 1973) the MLE does not exist when  $x_{1,1,1} = x_{2,2,2} = 0$ . Then suppose we observe

$$\mathbf{x} = [0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 0] \quad (24)$$

. Attempting to Using the Newton-Raphson procedure to attempt maximize the likelihood associated to the design matrix  $\Delta$  fails since

$$\hat{x}_{1,1,1}, \hat{x}_{2,2,2} \rightarrow 0 \quad (25)$$

while

$$\hat{x}_{j,k,l} \rightarrow x_{j,k,l} \quad (j, k, l) \neq (1, 1, 1), (2, 2, 2) \quad (26)$$

as the procedure is iterated. Furthermore estimates of associated log-odd ratios are unbounded:

$$\log \left( \frac{\hat{\pi}_{1,1,1}}{1 - \hat{\pi}_{1,1,1}} \right), \log \left( \frac{\hat{\pi}_{2,2,2}}{1 - \hat{\pi}_{2,2,2}} \right) \rightarrow -\infty \quad (27)$$

as Newton-Raphson is iterated. This result aligns on that reported by Fienberg and Rinaldo.  $x_{111}, x_{222}$  are patterned likelihood zeros and while the model loses only one degree of freedom the table itself loses two degrees of freedom in the failed attempt to maximize the likelihood. At the same time the surviving estimates,  $\hat{x}_{2,1,1}, \dots, \hat{x}_{2,2,1}$  perfectly fit the surviving table  $x_{2,1,1}, \dots, x_{2,2,1}$  and we cannot assess the fit of the model. As such  $\hat{\mathbf{x}}$  is the EMLE.

### 3.3 Surviving Parameterization

The EMLE is defined at the cell level. To embed the parameter space of the EMLE in a poisson regression we first identify a *surviving parameterization*  $\mathbf{R}\boldsymbol{\mu}$ , which is any non-singular parameterization for the row space spanned by the rows of  $\boldsymbol{\Delta}$  associated to the *fascia*. The fascia comprises the elements of  $\mathbf{x}$  other than the likelihood zeros. In this case a valid choice for  $\mathbf{R}$  is simply the 6 middle rows of  $\boldsymbol{\Delta}$ . We obtain

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad (28)$$

$\mathbf{R}$  leads us to elicit the *exposed subspace*. To do this let  $\mathbf{R}^\perp$  be of full row rank and let the row space of  $\mathbf{R}^\perp$  span the orthogonal complement of the row space of  $\mathbf{R}$ .

### 3.4 Exposed Subspace

Consider the following construct

$$[\mathbf{C} \mid \mathbf{C}^\top] = \begin{bmatrix} \mathbf{R} \\ -\frac{\mathbf{R}}{\mathbf{R}^\perp} \end{bmatrix}^{-1} \quad (29)$$

In general  $\mathbf{C}$  is a right inverse for  $\mathbf{R}$ .  $\mathbf{C}^\top$  is any full column rank matrix whose columns spans the orthogonal complement of the column space of  $\mathbf{C}$ . We The dimensions of  $\mathbf{C}$  as the same as those of  $\mathbf{R}^t$ . Then the column space of  $\mathbf{DC}^\top$  is the exposed space. The proof will be available in a full version of the paper. In the meanwhile we give a heuristic and intuitive definition for the exposed space:

*The exposed space is the dimensionally largest linear subspace such that the parameterization for any linear combination of vectors in the exposed space is not estimable.*

We have

$$\boldsymbol{\Xi} = \mathbf{DC}^\top = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1]^t \quad (30)$$

(30) should not come as a surprise. The exposed space is unidimensional and matches the likelihood zeros. Rather than integrating structural zeros in the EMLE it may possible to identify the orthogonal complement of the exposed subspace in the column space of the design matrix  $\boldsymbol{\Delta}$  and parametrize a dimensionally reduced design matrix whose columns span the estimable subspace.

Table 1: Estimates for the Demonstration Example

$\mathbf{x}$	0	1	2	3	4	4	6	0
EMLE	0	1	2	3	4	5	6	0
Orth Proj	.572	1.58	2.58	2.41	4.58	4.41	5.41	1.74

Orthogonality insures no ancillary residuals from the exposed space are ported to the orthogonal subspace and any new model constraint does not superseed existing ones.

### 3.5 Orthogonal Design Matrix

The projection of the orthogonal complement of the exposed space  $\Xi$  onto the column space of  $\Delta$  can be expressed as the column space of the following matrix

$$\mathbf{P} = [\Delta^\top \mid \Xi]^\top \quad (31)$$

In the form of an executable statement in "R", using the package "pracma", (31) can be written as

```
P <- null( t( cbind( null( t( Delta ) ), Xi ) ) )
```

Since the column space  $\mathbf{P}$  is orthogonal to  $\Xi$ , the intercept is not in the column space of  $\mathbf{P}$ . So, while  $\mathbf{P}$  is a poisson design matrix, it does not characterize a multinomial. Table 1 displays the estimates for the EMLE along with the orthogonal projection obtained using  $\mathbf{P}$  as poisson design matrix. the total of the poisson estimates is not the same as the total of the observations and the data cannot be interpreted as independent multinomial observations with a fixed total in general.

## 4 Discussion

In the case of our demonstration example the likelihood zeros and the number of degrees of freedom were previously identified. It is fair to say half the work had already been done. In a more complexe situation identifying the likelihood zeros and the number of lost degrees of freedom can be challenging. Wang, Rauhyand, Massam (2019) report when the number of dimensions exceed 16 the linear programming methods proposed by Fienberg and Rinaldo become difficult to implement. These authors propose a more nuanced approach. We hope our presentation will help further the development of strictly likelihood-based estimation methods based on log-linear models in high dimension and/or in presence of sparse tables. A by-product could be, in the case of multinomial log-linear models, to offer alternatives to regularization methods -see Friedman, Hastie, Tibshirani (2008), Tibshirani (1996)- which have become so popular.

## References

- [1] BISHOP, Y., FIENBERG, S., HOLLAND, P. (1974). “Discrete Multivariate Analysis Theory and Applications”. Springer.
- [2] FIENBERG, S., RINALDO, A. (2012) “Maximum Likelihood Estimation in Log-Linear Models”. *Ann. Stats* 40, 996-1023
- [3] Friedman, J., Hastie, T. and Tibshirani, R. (2008). “Regularization Paths for Generalized Linear Models via Coordinate Descent”. *Journal of Statistical Software*, Vol. 33(1), 1-22
- [4] FRIEDLANDER, M. (2016). “Fitting log-linear models in sparse contingency tables using the eMLEloglin R package”. *arXiv:1611.07505v2*
- [5] GEIGER, D., MEEK, C., STURMSFELS, B. (2006). “On The Toric Algebra of Graphical Models”. *Annals of Statistics*, 34, 3, 1463-1492
- [6] HABERMAN, S. (1974) “Log-Linear Models for Frequency Data”. *Annals of Statistics*, 1, 4 (617-632)
- [7] Sharifi Far, S., Papathomas, M., King, R. (2021). “Parameter Redundancy and the Existence of Maximum Likelihood Estimates in Log-Linear Models”. *Statistica Sinica* 31 (2021), 1125-1143.
- [8] Tibshirani, R. (1996). “Regression Shrinkage and Selection via the Lasso”. *J. R. Statist. Soc. B*, 58, No. 1.
- [9] WANG, N., RAUHYAND, J. AND MASSAM, H. (2019). “Approximating faces of marginal polytopes in discrete hierarchical models”. *The Annals of Statistics*, 47(3), 1203–1233.



## A R Code

```
library(pracma)

Delta <- matrix(nrow=8, ncol=7)
Delta[,1] <- c(1,1,1,1,1,1,1)
Delta[,2] <- c(0,1,0,1,0,1,0)
Delta[,3] <- c(0,0,1,1,0,0,1)
Delta[,4] <- c(0,0,0,0,1,1,1)
Delta[,5] <- c(0,1,1,0,0,1,1)
Delta[,6] <- c(0,1,0,1,1,0,1)
Delta[,7] <- c(0,0,1,1,1,1,0)

DeltaFrame <- matrix(nrow=8, ncol=8)
DeltaFrame[,1:7] <- Delta
DeltaFrame[,8] <- c(0,1,2,3,4,5,6,0)
colnames(DeltaFrame) <- c('D0','D1','D2','D3','D4','D5','D6','OBS')
DeltaFrame <- as.data.frame(DeltaFrame)

DeltaFrameOutput4 <- glm( formula = OBS ~ D0 + D1 + D2 + D3 + D4 + D5 + D6 -1,
data = DeltaFrame, family = poisson, maxit = 1000, epsilon = .00001 )
DeltaFrameOutput8 <- glm( formula = OBS ~ D0 + D1 + D2 + D3 + D4 + D5 + D6 -1,
data = DeltaFrame, family = poisson, maxit = 1000, epsilon = .000000001 )

R <- matrix(nrow=6, ncol=7)
R[, ] <- Delta[2:7, ]
Rorth <- null(R)
# Rorth <- c(-2,0,0,0,1,1,1)

RoverRorth <- rbind(R, t(Rorth))
CandCtop <- solve(RoverRorth)
Ctop <- matrix(nrow = 8, ncol = 1)
Ctop <- CandCtop[,7]
Xi <- Delta %*% Ctop

P <- null( t( cbind( null( t(Delta) ), Xi ) ) )

PFrame <- matrix(nrow=8, ncol=7)
PFrame[,1:6] <- P
PFrame[,7] <- c(0,1,2,3,4,5,6,0)
colnames(PFrame) <- c('SD1','SD2','SD3','SD4','SD5','SD6','OBS')
PFrame <- as.data.frame(PFrame)

PFrameOutput4 <- glm( formula = OBS ~ SD1 + SD2 + SD3 + SD4 + SD5 + SD6 -1,
data = PFrame, family = poisson, maxit = 1000, epsilon = .00001 )
PFrameOutput8 <- glm( formula = OBS ~ SD1 + SD2 + SD3 + SD4 + SD5 + SD6 -1,
```

```
data = PFrame, family = poisson, maxit = 1000, epsilon = .000000001 )
```

```
-----  
  
> PFrameOutput4$fitted.values  
      1      2      3      4      5      6      7      8  
0.5721164 1.5878898 2.5878898 2.4121110 4.5878898 4.4121103 5.4121103 1.7478961  
  
> PFrameOutput8$fitted.values  
      1      2      3      4      5      6      7      8  
0.5721165 1.5878897 2.5878897 2.4121103 4.5878897 4.4121103 5.4121103 1.7478959
```