Recent Bayesian and NON-BAYESIAN Methods for Analyzing Sparse Classifications

Yves Thibaudeau

U.S. Census Bureau

yves.thibaudeau@census.gov

August 7 2023

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.

Purpose of the Talk

1. Review 3 analytical methods to handle likelihood zeros -sparse tables- in multinomial and Poisson regression for modeling prevalence of categories.
2. Emphasize the advantages of *orthogonalization* -a less often used method.

Outline

:

- Context
- Example: Poisson/multinomial regression in presence of likelihood zeros
- EMLE
- Orthogonalization
- The LASSO
- Discussion

Context

Methods for log-linear (Poisson or multinomial product) modeling *sparse* contingency tables arising in

- Social Sciences - Demography: Classification of U.S. population by voting status (Abowd Bell Brown Hawes Hegeness Keller Mule Schafer Spence Warren Yi 2020)
- Record Linkage - Entity Resolution (Latent-Class Modeling)
- Biomedical Sciences

Example

x_{111}	x_{121}
x_{211}	x_{221}

Table: $k = 1$

x_{112}	x_{122}
x_{212}	x_{222}

Table: $k = 2$

Example (cont'd)

Poisson Regression: No 2nd Order Interaction:

$$\xi_{j,k,l} = \log\left(\mathrm{E}\left[x_{jkl}\right]\right) = u + u_{1(j)} + u_{2(k)} + u_{3(l)} + u_{12(j,k)} + u_{13(j,l)} + u_{23(k,l)}$$

Example (cont'd)

$$\boldsymbol{\xi} = \boldsymbol{\Delta}\boldsymbol{\mu}$$

$\boldsymbol{\Delta}$ is a Poisson design matrix including an intercept and characterizes a multinomial:

$$\boldsymbol{\Delta} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{\mu} = -2 \begin{bmatrix} -u/2 \\ u_{1(1)} \\ u_{2(1)} \\ u_{3(1)} \\ u_{12(1,1)} \\ u_{13(1,1)} \\ u_{23(1,1)} \end{bmatrix}$$

Example (cont'd)

Hypothetical Observations (Haberman 1974) - **likelihood zeros in bold**

0	2
1	3

Table: $k = 1$

4	6
5	**0**

Table: $k = 2$

Example (cont'd)

In the linear algebra of the model we have:

$$\boldsymbol{x} = \begin{bmatrix} \mathbf{0} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \mathbf{0} \end{bmatrix}$$

Likelihood zeros in bold -Friedlander (2016). No *sampling zero*.

Iterate Newton-Ralphson:

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

$$\hat{x}_{1,1,1},\ \hat{x}_{2,2,2} \to 0$$

Multinomial Model

$$\hat{\pi}_{1,1,1},\ \hat{\pi}_{2,2,2} \to 0$$

$$2\mu_1 - \mu_5 - \mu_6 - \mu_7 \to -\infty$$

MLE does not exist over a one-dimensional subspace of the parameter space $\Re^7$. But we have a perfect fit on the *fascia* (NOT in general).

Extended Maximum Likelihood (Fienberg Rinaldo 2012): Remove the rows associated to the likelihood zeros:

$$\boldsymbol{\Delta}_{emle} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Substitute *structural zeros* for the estimate of likelihood zero cells in $\hat{\boldsymbol{x}}$ by postulate.

Example (cont'd)

Exposed Subspace in $\mathcal{CS}\{\boldsymbol{\Delta}\}$: $E = \mathcal{CS}\{\ \boldsymbol{\Delta}_E\ \}$

$$\boldsymbol{\Delta}_E \quad = \quad \begin{bmatrix} \mathbf{1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \mathbf{1} \end{bmatrix} \quad = \quad \boldsymbol{\Delta} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -0.5 \\ -0.5 \\ -0.5 \end{bmatrix}$$

Orthogonalization

Orthogonalization (suggestion from Fienberg Rinaldo 2012): Remove the exposed subspace from the column space of the design matrix:

$$\boldsymbol{\Delta}_{orth} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In general $\boldsymbol{\Delta}_{orth}$ has no intercept and does not characterize a multinomial.

LASSO

LASSO (Least Absolute Shrinkage and Selection Operator, Tibshrani 1996). "Flagship" regularization method.

$$\min_{\boldsymbol{\mu}} \left[\, \mathcal{L}\left(\boldsymbol{x}, \boldsymbol{\Delta}\boldsymbol{\mu}\right) + \lambda \|\boldsymbol{\mu}\|_1 \,\right], \quad \lambda > 0$$

LASSO effectively removes columns from $\boldsymbol{\Delta}$ as $\lambda \uparrow$ to obtain a new Poisson design matrix $\boldsymbol{\Delta}_\lambda$ by setting components of $\boldsymbol{\mu}$ equal to zero. $\| \, \|_1$ is the L-1 norm (sum of absolute values).

Example (cont'd)

D.F.	Data							
	0	1	2	3	4	5	6	0

	EMLE							
6	0^*	1	2	3	4	5	6	0^*

	LASSO $\lambda = .05,\ .125,\ 2$							
7	0.044	1.255	2.255	2.643	4.256	4.643	5.643	0.256
6	0.175	1.575	2.374	2.374	4.374	4.374	5.175	0.575
1	2.625	2.625	2.625	2.625	2.625	2.625	2.625	2.625

	Orthogonalization							
6	.572	1.58	2.58	2.41	4.58	4.41	5.41	1.74

Conclusion

Given a nominal Poisson design matrix $\mathbf{\Delta}$ leads to likelihood zeros:

$\mathbf{\Delta}_{emle}$ EMLE: Exact solution. Correct degrees of freedom. Multinomial. May leave no residual d.f. for testing.

$\mathbf{\Delta}_{\lambda}$ LASSO: Flexible choice of λ allows for calibrating variance vs. bias. Multinomial. Testing the submodel possible after adjusment of d.f.'s. May involve ancillary information through remaining dimensions of the exposed subspace.

$\mathbf{\Delta}_{orth}$ Orthogonalization: Correct degrees of freedom. Residual d.f.'s allow for testing the submodel. Salvages entire estimable subspace. No ancillary information. Not multinomial in general.

Latent-Class Models: regularization - Schafer (Cvam 2020-24). Weinberg Thibaudeau (simulations of orthogonalization 2024).

References

- Abowd, J., Bell, W., Brown, J., Hawes, M., Hegeness, M., Keller, A., Mule, V., Schafer, J., Spence, M., Warren, M., Yi, M. (2020) Determination of the 2020 U.S. Citizen Voting Age Population (CVAP) Using Administrative Records and Statistical Methodology. U.S. Census Bureau, CES 20-33.
- Bishop, Y., Fienberg, S., Holland, P. (1974). Discrete Multivariate Analysis Theory and Applications. Springer.
- Fienberg, S., Rinaldo, A. (2012) "Maximum Likelihood Estimation in Log-Linear Models". Ann. Stats 40, 996-1023
- Friedlander, M. (2016). "Fitting log-linear models in sparse contingency tables using the eMLEloglin R package". arXiv:1611.07505v2.
- Schafer, J. (2020). Cvam: Coarsened-variable modeling. R package.
- Tibshrani, R. (1996). "Regression Shrinkage and Selection via the Lasso". J. R. Statist. Soc. B, 58, 1, 267-288.

The End