Markov Bases for Logistic Regression Models: A Preliminary Report

Dane Wilburne

Illinois Institute of Technology

Joint work with:
Hisayuki Hara (Niigata University)

Loyola University
AMS Fall Central Section Meeting
Special Session on Algebraic Statistics and its Interactions with Combinatorics, Computation, and Network Science

Chicago, IL
October 4, 2015

Poisson Regression

$$x_{jkl} \sim Po(\lambda_{jkl}), \quad j = 0,1, \quad k = 0,1, \quad l = 1, \dots, L$$

$$\log(\lambda_{jkl}) = \alpha + \beta l + \gamma d_1^j + \delta d_1^k$$

$$d_1^j = \begin{cases} 1, & j = 1 \\ 0, & \text{otherwise} \end{cases}, \qquad d_1^k = \begin{cases} 1, & k = 1 \\ 0, & \text{otherwise} \end{cases}$$

	l			
(j,k)	1	2	...	L
$(0,0)$	$*$	$*$	...	$*$
$(0,1)$	$*$	$*$	...	$*$
$(1,0)$	$*$	$*$	...	$*$
$(1,1)$	$*$	$*$	...	$*$

Example

You own a fishing boat, and you would like to model how many fish are caught with l fisherman in fair/poor weather (d_1^j) on a weekend/weekday (d_1^k). Record data in a table such as the one on the left.

Poisson Regression

$$x_{jkl} \sim Po(\lambda_{jkl}), \quad j = 0, 1, \quad k = 0, 1, \quad l = 1, \ldots, L$$

$$\log(\lambda_{jkl}) = \alpha + \beta l + \gamma d_1^j + \delta d_1^k$$

$$d_1^j = \begin{cases} 1, & j = 1 \\ 0, & \text{otherwise} \end{cases}, \qquad d_1^k = \begin{cases} 1, & k = 1 \\ 0, & \text{otherwise} \end{cases}$$

Sufficient statistics:

- x_{+++}
- x_{1++}
- x_{+1+}
- $\sum_{l=1}^{L}(l \cdot x_{++l})$

Configuration matrix:

$$\mathcal{B} = \begin{pmatrix} 1 & \ldots & 1 & 1 & \ldots & 1 & 1 & \ldots & 1 & 1 & \ldots & 1 \\ 0 & \ldots & 0 & 0 & \ldots & 0 & 1 & \ldots & 1 & 1 & \ldots & 1 \\ 0 & \ldots & 0 & 1 & \ldots & 1 & 0 & \ldots & 0 & 1 & \ldots & 1 \\ 1 & \ldots & L & 1 & \ldots & L & 1 & \ldots & L & 1 & \ldots & L \end{pmatrix}$$

Markov bases for Poisson Regression

Proposition (Hara, W., 2015)

There exists a Markov basis for the Poisson regression model for $2 \times 2 \times L$ tables containing only degree two moves.

Example (The case $L = 3$)

The minimal Markov basis for the Poisson regression model for $2 \times 2 \times 3$ tables consists of 33 degree moves of degree two.

	l		
(j, k)	1	2	3
$(0, 0)$	1	0	0
$(0, 1)$	0	0	-1
$(1, 0)$	-1	0	0
$(1, 1)$	0	0	1

Generalization:

One way to generalize this model is to increase the number of dummy variables. In other words, we now consider the Possion regression model for $\underbrace{2 \times 2 \times \ldots \times 2}_{I \text{ times}} \times L$ tables.

Markov bases for Poisson Regression

Proposition (Hara, W., 2015)

There exists a Markov basis for the Poisson regression model for $2 \times 2 \times L$ tables containing only degree two moves.

Example (The case $L = 3$)

The minimal Markov basis for the Poisson regression model for $2 \times 2 \times 3$ tables consists of 33 degree moves of degree two.

(j, k)	l = 1	2	3
$(0,0)$	1	0	0
$(0,1)$	0	0	-1
$(1,0)$	-1	0	0
$(1,1)$	0	0	1

Proposition (Hara, W. 2015)

There exists a Markov basis for the Poisson regression model for $\underbrace{2 \times 2 \times \ldots \times 2}_{I \text{ times}} \times L$ tables containing only degree two moves.

Primitive Partition Identities

Definition

Fix a positive integer n. A partition identity is an identity of the form:

$$a_1 + a_2 + \ldots + a_k = b_1 + b_2 + \ldots + b_l, \quad (1)$$

where $0 < a_i, b_j \leq n$, $a_i, b_j \in \mathbb{Z}$. *The quantity* $k + l$ *is called the degree of the partition identity. We call the partition identity in* (1) *primitive if there is no proper subidentity*

$$a_{i_1} + a_{i_2} + \ldots + a_{i_r} = b_{j_1} + b_{j_2} + \ldots + b_{j_s}, \quad (2)$$

where $1 \leq r + s \leq k + l - 1$ *and we call it homogeneous if* $k = l$.

Theorem (Diaconis-Graham-Sturmfels)

The degree of any primitive partition identity satisfies $k + l \leq 2n - 1$.

Primitive Partition Identities

Example

- $1+2+3=3+3$ is a non-primitive partition identity with largest part $n=3$ and degree $3+2=5$;
- $1+3+3+3=5+5$ is a primitive partition identity with largest part $n=5$ and degree $4+2=6$.

- $I_{\mathcal{A}}$: toric ideal assoc. to $\mathcal{A} \in \mathbb{Z}^{d\times n}$
- $Gr(\mathcal{A})$: Graver basis of $\mathcal{A}$; $\langle Gr(\mathcal{A})\rangle = I_{\mathcal{A}}$

Observation

Let $\mathcal{A}$ be the $1\times n$ integer matrix $\mathcal{A} = (1 \;\; 2 \;\; \dots \;\; n)$. Then, the binomial $x_{a_1}x_{a_2}\cdots x_{a_k} - x_{b_1}x_{b_2}\cdots x_{b_l}$ is a primitive element of $I_{\mathcal{A}}$ iff $a_1+a_2+\ldots+a_k = b_1+b_2+\ldots+b_l$ is a ppi, i.e. there is a correspondence between ppi's and elements of $Gr(\mathcal{A})$.

Color-Homogeneous PPIs

Definition

A colored partition identity in the colors $1, \ldots, c$ is an identity of the form:

$$a_{1,1} + \ldots + a_{1,k_1} + a_{2_1} + \ldots + a_{2,k_2} + a_{c,1} + \ldots + a_{c,k_c} =$$
$$b_{1,1} + \ldots + b_{1,l_1} + b_{2_1} + \ldots + b_{2,l_2} + b_{c,1} + \ldots + b_{c,l_c}$$

where $1 \leq a_{p,j}, b_{p,j} \leq n_p$ are positive integers for all j, $1 \leq p \leq c$ and some positive integers $n_1, \ldots, n_c$. If $k_j = l_j$ for all $1 \leq j \leq c$, then it is called color-homogeneous. A chpi is primitive if there is no proper color-homogeneous subidentity

$$a_{-,i_1} + \ldots + a_{-,i_r} = b_{-,j_1} + \ldots + b_{-,j_s}$$

with $1 \leq r + s \leq k_1 + \ldots + k_c + l_1 + \ldots + l_c$. The degree is the number of summands $k_1 + \ldots + k_c + l_1 + \ldots + l_c$.

Color-Homogeneous PPIs

Example

Let $c = 2$ and let $n_1 = n_2 = 3$. Then, the following are examples of color-homogeneous partition identities in $c = 2$ colors with $n_1 = n_2 = 3$:

- $1 + 3 = 2 + 2$
- $1 + 3 = 3 + 1$
- $2 + 2 + 3 = 3 + 3 + 1$

Theorem (Petrović)

Let $n_P = \max\{n_i : 1 \le i \le c\}$ *and let* $n_Q = \max\{n_j : 1 \le j \le c\}, j \ne P\}$. *Then, any primitive color-homogeneous partition identity satisfies* $k_1 + \ldots + k_c + l_1 + \ldots + l_c \le n_P + n_Q - 2$.

Remark

Again, there is a correspondence between primitive color-homogeneous partition identitites and $Gr(\mathcal{A})$ *for some integer matrix* $\mathcal{A}$.

Moves of the Poisson model

(j,k)	l = 1	2	3
$(0,0)$	0	2	0
$(0,1)$	-3	0	1
$(1,0)$	0	0	-2
$(1,1)$	2	0	0

We record an entry in the $(l,(j,k))$ cell as l_{jk}, putting positive entries on the left and negative entries on the right, which gives:

$$2_{00}+2_{00}+3_{01}+1_{11}+1_{11}=1_{01}+1_{01}+1_{01}+3_{10}+3_{10}$$

Color according to j:

$$2_{00}+2_{00}+3_{01}+1_{11}+1_{11}=1_{01}+1_{01}+1_{01}+3_{10}+3_{10}$$

Color according to k:

$$2_{00}+2_{00}+3_{01}+1_{11}+1_{11}=1_{01}+1_{01}+1_{01}+3_{10}+3_{10}$$

Logistic regression

$$\log\left(\frac{p_{1jkl}}{1-p_{1jkl}}\right) = \alpha + \beta l + \gamma d_1^j + \delta d_1^k$$

$$j = 0,1; \quad k = 0,1 \quad l = 1,\ldots,L$$

$$d_1^j = \begin{cases} 1, & j = 1 \\ 0, & \text{otherwise} \end{cases}, \qquad d_1^k = \begin{cases} 1, & k = 1 \\ 0, & \text{otherwise} \end{cases}$$

Configuration matrix:

$$\Lambda(\mathcal{B}) = \begin{pmatrix} \mathcal{B} & 0 \\ I_{n\times n} & I_{n\times n} \end{pmatrix},$$

where $\mathcal{B}$ is the Poisson configuration (this is called the Lawrence lifting of $\mathcal{B}$).

Example

In the fishing example from earlier, we now think of modeling success/failure outcomes instead of counts (i.e. did you meet your quota for the day, or not?) where the variables have the same interpretation as before (# of fisherman, weather, weekend/weekday).

Markov bases for Logistic Regression

The idea is to exploit the connection to generalized CHPIs to prove bounds on the elements of the $Gr(\mathcal{A})$, since this corresponds to a Markov basis for $I_{\Lambda(\mathcal{A})}$.

Conjecture (Hara, W. 2015)

There exists a Markov basis for the logistic regression model for $2 \times 2 \times 2 \times L$ tables with maximum degree

$$d \leq \begin{cases} 6 & \text{if } L = 2 \\ 6L - 8 & \text{if } L \geq 3 \end{cases}.$$

Remark

This conjecture has been verified computationally for $L \leq 6$.