Marginal Models for categorical data

Application to conditional independence and graphical models

Wicher Bergsma[1] Marcel Croon[2] Jacques Hagenaars[2]
Tamas Rudas[3]

[1]London School of Economics and Political Science

[2]Tilburg University

[3]ELTE, Hungary

www.cmm.st

Cambridge University, February 2012

Outline

Marginal models: for what types of data?

Interest lies in ‘population averaged’ quantities, but through design data are dependent (clustered). For correct inference, marginal modeling is needed.

Marginal models: for what types of data?

Interest lies in 'population averaged' quantities, but through design data are dependent (clustered). For correct inference, marginal modeling is needed.

Dependencies in the data arise in many situations:

- Comparing marginal distributions of two characteristics measured on the same respondents, e.g., preference prime minister and party preference.

Marginal models: for what types of data?

Interest lies in 'population averaged' quantities, but through design data are dependent (clustered). For correct inference, marginal modeling is needed.

Dependencies in the data arise in many situations:

- Comparing marginal distributions of two characteristics measured on the same respondents, e.g., preference prime minister and party preference.
- Respondents are clustered, e.g., husbands and wives, but interest lies in overall population differences men and women.

Marginal models: for what types of data?

Interest lies in ‘population averaged’ quantities, but through design data are dependent (clustered). For correct inference, marginal modeling is needed.

Dependencies in the data arise in many situations:

- Comparing marginal distributions of two characteristics measured on the same respondents, e.g., preference prime minister and party preference.
- Respondents are clustered, e.g., husbands and wives, but interest lies in overall population differences men and women.
- Panel studies (repeated measurements): are there overall changes in the population?

Marginal models: for what types of data?

Interest lies in ‘population averaged’ quantities, but through design data are dependent (clustered). For correct inference, marginal modeling is needed.

Dependencies in the data arise in many situations:

- Comparing marginal distributions of two characteristics measured on the same respondents, e.g., preference prime minister and party preference.
- Respondents are clustered, e.g., husbands and wives, but interest lies in overall population differences men and women.
- Panel studies (repeated measurements): are there overall changes in the population?
- Trend studies: comparing changes in *two* variables over time.

This talk: two types of marginal models

1. Conditional independence models for certain non-iid data.

This talk: two types of marginal models

1. Conditional independence models for certain non-iid data.
2. Intersections of conditional independencies

This talk: two types of marginal models

1. Conditional independence models for certain non-iid data.
2. Intersections of conditional independencies

Problems: degrees of freedom and minimal specification, smoothness

This talk: two types of marginal models

1. Conditional independence models for certain non-iid data.
2. Intersections of conditional independencies

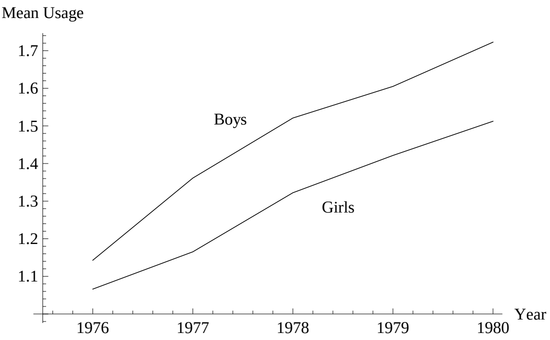
Problems: degrees of freedom and minimal specification, smoothness

Focus on categorical data

Example: longitudinal data

	Age (A)									
	13	%	14	%	15	%	16	%	17	%
Boys' Marijuana use (B)										
1. Never	106	89.1	89	74.8	78	65.5	73	61.3	65	54.6
2. Once a month	9	7.6	17	14.3	20	16.8	20	16.8	22	18.5
3. More than once a month	4	3.4	13	10.9	21	17.6	26	21.8	32	26.9
Girls' Marijuana use (G)										
1. Never	114	94.2	106	87.6	91	75.2	85	70.2	75	62.0
2. Once a month	6	5.0	10	8.3	21	17.4	21	17.4	30	24.8
3. More than once a month	1	.8	5	4.1	9	7.4	15	12.4	16	13.2

Growth curves marijuna use



Variables of interest: age (A), marijuana use (M), gender (G). Longitudinal study, i.e., same boys and girls at each point in time, so data in Table AMG not iid.

Fitting and testing: maximum likelihood

How to test a model such as $M \perp\!\!\!\perp G | A$ (at all ages, marijuana use same for boys and girls)?

Fitting and testing: maximum likelihood

How to test a model such as $M \perp\!\!\!\perp G | A$ (at all ages, marijuana use same for boys and girls)?

Model induces constraints on multinomial probabilities in full table $GM_1M_2M_3M_4M_5$.

Fitting and testing: maximum likelihood

How to test a model such as $M \perp\!\!\!\perp G | A$ (at all ages, marijuana use same for boys and girls)?

Model induces constraints on multinomial probabilities in full table $GM_1M_2M_3M_4M_5$.

Maximize kernel of multinomial log-likelihood

$$L = \sum p_i \log \pi_i - \sum \pi_i$$

subject to the constraint

$$B' \log(A'\pi) = 0$$

Use scoring type Lagrange multiplier method, algorithm of B. (1997), based on Aitchison and Silvey (1959) and Lang and Agresti (1994).

Problems

For many models: no problems at all with fitting and testing.

For some models we encountered problems…

Outline

How many degrees of freedom (df)?

For $i = 1, \ldots, K$ and $j = 1, \ldots, K$:

$$\pi^{ABCD}_{i\,j\,++} = \pi^{ABCD}_{+i\,j\,+} = \pi^{ABCD}_{++i\,j}$$

How many degrees of freedom (df)?

For $i = 1, \ldots, K$ and $j = 1, \ldots, K$:

$$\pi^{ABCD}_{i\,j\,++} = \pi^{ABCD}_{+i\,j\,+} = \pi^{ABCD}_{++i\,j} \qquad (K^2 - 1) \text{ restrictions per eq.} \qquad (1)$$

How many degrees of freedom (df)?

For $i = 1, \ldots, K$ and $j = 1, \ldots, K$:

$$\pi^{ABCD}_{i\,j\,++} = \pi^{ABCD}_{+i\,j\,+} = \pi^{ABCD}_{++i\,j} \qquad (K^2 - 1) \text{ restrictions per eq.} \qquad (1)$$

Naive calculation: $\text{df} = 2(K^2 - 1)$

How many degrees of freedom (df)?

For $i = 1, \ldots, K$ and $j = 1, \ldots, K$:

$$\pi^{ABCD}_{i\,j\,++} = \pi^{ABCD}_{+i\,j\,+} = \pi^{ABCD}_{++i\,j} \qquad (K^2 - 1) \text{ restrictions per eq.} \tag{1}$$

Naive calculation: $\mathrm{df} = 2(K^2 - 1)$
Wrong: $K - 1$ of the restrictions are not needed.

How many degrees of freedom (df)?

For $i = 1, \ldots, K$ and $j = 1, \ldots, K$:

$$\pi^{ABCD}_{i\,j\,++} = \pi^{ABCD}_{+i\,j\,+} = \pi^{ABCD}_{++i\,j} \qquad (K^2 - 1) \text{ restrictions per eq.} \qquad (1)$$

Naive calculation: $\text{df} = 2(K^2 - 1)$
Wrong: $K - 1$ of the restrictions are not needed.

Solution using marginal loglinear parameterizations (B. & Rudas, 2002); (1) equivalent to

$$\lambda^{AB}_{i\,j} = \lambda^{BC}_{i\,j} = \lambda^{CD}_{i\,j} \qquad (K-1)^2 \text{ restrictions per eq.}$$
$$\lambda^{A}_{i} = \lambda^{B}_{i} = \lambda^{C}_{i} = \lambda^{D}_{i} \qquad (K-1) \text{ restrictions per eq.}$$

These form restrictions on a *parameterization*, so this is a minimal specification; hence $\text{df} = 2(K-1)^2 + 3(K-1)$

How many degrees of freedom (df)? Ex. 2

$$\pi_{j\ i}^{B|A} = \pi_{j\ i}^{C|B} = \pi_{j\ i}^{D|C}$$

How many degrees of freedom (df)? Ex. 2

$$\pi_{j\ i}^{B|A} = \pi_{j\ i}^{C|B} = \pi_{j\ i}^{D|C} \qquad K(K-1) \text{ restrictions per eq.} \tag{2}$$

How many degrees of freedom (df)? Ex. 2

$$\pi_{j\ i}^{B|A} = \pi_{j\ i}^{C|B} = \pi_{j\ i}^{D|C} \qquad K(K-1) \text{ restrictions per eq.} \tag{2}$$

Naive calculation: $\mathrm{df} = 2K(K-1)$

How many degrees of freedom (df)? Ex. 2

$$\pi_{j\,i}^{B|A} = \pi_{j\,i}^{C|B} = \pi_{j\,i}^{D|C} \qquad K(K-1) \text{ restrictions per eq.} \qquad (2)$$

Naive calculation: $\text{df} = 2K(K-1)$
Correct. But how do we know?

How many degrees of freedom (df)? Ex. 2

$$\pi_{j\,i}^{B|A} = \pi_{j\,i}^{C|B} = \pi_{j\,i}^{D|C} \qquad K(K-1) \text{ restrictions per eq.} \tag{2}$$

Naive calculation: $\text{df} = 2K(K-1)$
Correct. But how do we know?

Using marginal loglinear parameters (2) equivalent to

$$\lambda_{i\,j}^{AB} = \lambda_{i\,j}^{BC} = \lambda_{i\,j}^{CD} \qquad (K-1)^2 \text{ restrictions per eq.}$$

$$\lambda_{*\,j}^{AB} = \lambda_{*\,j}^{BC} = \lambda_{*\,j}^{CD} \qquad (K-1) \text{ restrictions per eq.}$$

How many degrees of freedom (df)? Ex. 2

$$\pi_{j\,i}^{B|A} = \pi_{j\,i}^{C|B} = \pi_{j\,i}^{D|C} \qquad K(K-1) \text{ restrictions per eq.} \tag{2}$$

Naive calculation: $\text{df} = 2K(K-1)$
Correct. But how do we know?

Using marginal loglinear parameters (2) equivalent to

$$\lambda_{i\,j}^{AB} = \lambda_{i\,j}^{BC} = \lambda_{i\,j}^{CD} \qquad (K-1)^2 \text{ restrictions per eq.}$$

$$\lambda_{*\,j}^{AB} = \lambda_{*\,j}^{BC} = \lambda_{*\,j}^{CD} \qquad (K-1) \text{ restrictions per eq.}$$

Again, restrictions on a *parameterization*, so this is minimal specification; $\text{df} = 2(K-1)^2 + 2(K-1) = 2K(K-1)$

Cause of problems

Same loglinear 'effect' is restricted in two different marginal tables.

First example revisited:

$$\pi^{ABCD}_{i\,j\,++} = \pi^{ABCD}_{+i\,j\,+} = \pi^{ABCD}_{++i\,j}$$

Naive marginal loglinear specification:

$$\lambda^{AB}_{i\,j} = \lambda^{BC}_{i\,j} = \lambda^{CD}_{i\,j}$$

$$\lambda^{AB}_{i\,*} = \lambda^{BC}_{i\,*} = \lambda^{CD}_{i\,*}$$

$$\lambda^{AB}_{*\,j} = \lambda^{BC}_{*\,j} = \lambda^{CD}_{*\,j}$$

The loglinear B-effect is restricted in both tables AB and BC, and the C-effect in BC and CD $\Rightarrow$ problems!

A not-so-obvious example

Panel study drugs use of youth with 5 waves (ages 13 to 17)
Response variables: alcohol and marijuana use

Marginal tables of interest: transitions from time t to $t+1$ for both alcohol and marijuana usage.
Artificial table IPRS:
I - item (marijuana or alcohol)
P - period (age 13-14, 14-15, 15-16, or 16-17)
R, S - Response at 1st and 2nd measurement

Conditional independence models for marginal table IPRS:
$IP \perp\!\!\!\perp RS$: all turnover tables identical, whatever I, P
$P \perp\!\!\!\perp RS|I$: for both alcohol and marijuana (I), turnover tables same for all periods
$I \perp\!\!\!\perp RS|P$: for any period, turnover table alcohol same as for marijuana

Solution to the not-so-obvious example

Probabilities in table IPRS formed by sums of probabilities in the original multinomial table $M_1 M_2 M_3 M_4 M_5 A_1 A_2 A_3 A_4 A_5$ ($3^{10} = 59,049$ cells).

Solution is to formulate models such as $P \perp\!\!\!\perp RS | I$ in terms of restrictions on a marginal loglinear parameterization for original table.

Specification model Model $IP \perp\!\!\!\perp RS$

T_k: measurement at time point k. Model involves these restrictions on marginals of multinomial table:

$$\pi^{I\,T_1\,T_2}_{1\,i\,\,j} = \pi^{I\,T_2\,T_3}_{1\,i\,\,j} = \pi^{I\,T_3\,T_4}_{1\,i\,\,j} = \pi^{I\,T_4\,T_5}_{1\,i\,\,j} = \pi^{I\,T_1\,T_2}_{2\,i\,\,j} = \pi^{I\,T_2\,T_3}_{2\,i\,\,j} = \pi^{I\,T_3\,T_4}_{2\,i\,\,j} = \pi^{I\,T_4\,T_5}_{2\,i\,\,j},$$

A minimal specification is obtained by first, imposing equality of the conditional marginal association parameters:

$$\lambda^{T_1\,T_2|I}_{i\,\,j\,\,1} = \lambda^{T_2\,T_3|I}_{i\,\,j\,\,1} = \lambda^{T_3\,T_4|I}_{i\,\,j\,\,1} = \lambda^{T_4\,T_5|I}_{i\,\,j\,\,1} = \lambda^{T_1\,T_2|I}_{i\,\,j\,\,2} = \lambda^{T_2\,T_3|I}_{i\,\,j\,\,2} = \lambda^{T_3\,T_4|I}_{i\,\,j\,\,2} = \lambda^{T_4\,T}_{i\,\,j}$$

and second, by constraining the following univariate marginals:

$$\lambda^{T_1|I}_{i\,1} = \lambda^{T_2|I}_{i\,1} = \lambda^{T_3|I}_{i\,1} = \lambda^{T_4|I}_{i\,1} = \lambda^{T_5|I}_{i\,1} = \lambda^{T_1|I}_{i\,2} = \lambda^{T_2|I}_{i\,2} = \lambda^{T_3|I}_{i\,2} = \lambda^{T_4|I}_{i\,2} = \lambda$$

The number of independent constraints in this minimal specification is $(K-1)(7K+2)$.

Specification model Model $P \perp\!\!\!\perp RS | I$

Model assumes that the turnover tables are different for the two items but are identical over time for each item. A minimal specification:

$$\lambda_{i\ j\ 1}^{T_1 T_2 | I} = \lambda_{i\ j\ 1}^{T_2 T_3 | I} = \lambda_{i\ j\ 1}^{T_3 T_4 | I} = \lambda_{i\ j\ 1}^{T_4 T_5 | I} \, ,$$
$$\lambda_{i\ j\ 2}^{T_1 T_2 | I} = \lambda_{i\ j\ 2}^{T_2 T_3 | I} = \lambda_{i\ j\ 2}^{T_3 T_4 | I} = \lambda_{i\ j\ 2}^{T_4 T_5 | I} \, ,$$

and

$$\lambda_{i\ 1}^{T_1 | I} = \lambda_{i\ 1}^{T_2 | I} = \lambda_{i\ 1}^{T_3 | I} = \lambda_{i\ 1}^{T_4 | I} = \lambda_{i\ 1}^{T_5 | I} \, ,$$
$$\lambda_{i\ 2}^{T_1 | I} = \lambda_{i\ 2}^{T_2 | I} = \lambda_{i\ 2}^{T_3 | I} = \lambda_{i\ 2}^{T_4 | I} = \lambda_{i\ 2}^{T_5 | I} \, .$$

The number of independent constraints in this minimal specification is $(K-1)(6K+2)$.

Specification model Model $I \perp\!\!\!\perp RS|P$

Model assumes that the turnover tables change over period, but are the same for both items at each period. Minimal specification:

$$\begin{aligned}
\lambda_{i\ j\ 1}^{T_1 T_2|I} &= \lambda_{i\ j\ 2}^{T_1 T_2|I} \\
\lambda_{i\ j\ 1}^{T_2 T_3|I} &= \lambda_{i\ j\ 2}^{T_2 T_3|I} \\
\lambda_{i\ j\ 1}^{T_3 T_4|I} &= \lambda_{i\ j\ 2}^{T_3 T_4|I} \\
\lambda_{i\ j\ 1}^{T_4 T_5|I} &= \lambda_{i\ j\ 2}^{T_4 T_5|I},
\end{aligned}$$

and:

$$\begin{aligned}
\lambda_{i\ 1}^{T_1|I} &= \lambda_{i\ 2}^{T_1|I} \\
\lambda_{i\ 1}^{T_2|I} &= \lambda_{i\ 2}^{T_2|I} \\
\lambda_{i\ 1}^{T_3|I} &= \lambda_{i\ 2}^{T_3|I} \\
\lambda_{i\ 1}^{T_4|I} &= \lambda_{i\ 2}^{T_4|I}
\end{aligned}$$

Outline

Illustration of problem

Intersection marginal and conditional independence:

$$A \perp\!\!\!\perp B \quad \cap \quad A \perp\!\!\!\perp B | C$$

If C is binary, then equivalent to union

$$A \perp\!\!\!\perp C \quad \cup \quad B \perp\!\!\!\perp C$$

so intersection nonsmooth at $A \perp\!\!\!\perp B \perp\!\!\!\perp C$.

How can we know?

Solution in this case

Again same loglinear 'effect' (of AB) restricted in two different marginal tables:

$$A \perp\!\!\!\perp B \Leftrightarrow \lambda_{i\,j}^{AB} = 0$$

$$A \perp\!\!\!\perp B|C \Leftrightarrow \left(\lambda_{i\,j\,k}^{ABC} = 0 \text{ and } \lambda_{i\,j\,*}^{ABC} = 0\right)$$

No simplification possible, so problems to be expected.

Another example:

$$\left.\begin{array}{l} A \perp\!\!\!\perp BC \mid DE \\ F \perp\!\!\!\perp BD \mid C \\ AF \perp\!\!\!\perp BE \mid DC \end{array}\right\} ABC \text{ (and } FBDC\text{) effects restricted twice!}$$

But intersection smooth, how do we know? Next theorem needed.

General conditional independence model

$$\mathcal{Q} = \cap_i \{P \in \mathcal{P} \ : \ \mathcal{A}_i \perp\!\!\!\perp \mathcal{B}_i \mid \mathcal{C}_i \, (P)\}$$

where $\mathcal{A}_i, \mathcal{B}_i, \mathcal{C}_i \subseteq \mathcal{V}$ for a set of variables $\mathcal{V}$, $\mathcal{P}$ the family of positive probability distributions for $\mathcal{V}$.

General conditional independence model

$$\mathcal{Q} = \cap_i \{P \in \mathcal{P} \,:\, \mathcal{A}_i \perp\!\!\!\perp \mathcal{B}_i \mid \mathcal{C}_i\,(P)\}$$

where $\mathcal{A}_i, \mathcal{B}_i, \mathcal{C}_i \subseteq \mathcal{V}$ for a set of variables $\mathcal{V}$, $\mathcal{P}$ the family of positive probability distributions for $\mathcal{V}$.

Any finite set of axioms incompletely describes relations among models (Studeny, 2005).

General conditional independence model

$$\mathcal{Q} = \cap_i \{P \in \mathcal{P} \,:\, \mathcal{A}_i \perp\!\!\!\perp \mathcal{B}_i \mid \mathcal{C}_i\,(P)\}$$

where $\mathcal{A}_i, \mathcal{B}_i, \mathcal{C}_i \subseteq \mathcal{V}$ for a set of variables $\mathcal{V}$, $\mathcal{P}$ the family of positive probability distributions for $\mathcal{V}$.

Any finite set of axioms incompletely describes relations among models (Studeny, 2005).

But: we can identify 'well-behaved' subsets of models (next theorem).

Identification of smooth models

Conditional independence model:

$$\mathcal{Q} = \cap_i \{P \in \mathcal{P} \ : \ \mathcal{A}_i \perp\!\!\!\perp \mathcal{B}_i \mid \mathcal{C}_i \,(P)\}$$

With $\mathbb{P}(.)$ denoting the power set, let

$$\mathbb{D}_i = \mathbb{D}_i(\mathcal{A}_i, \mathcal{B}_i, \mathcal{C}_i) = \mathbb{P}(\mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i) \setminus (\mathbb{P}(\mathcal{A}_i \cup \mathcal{C}_i) \cup \mathbb{P}(\mathcal{B}_i \cup \mathcal{C}_i))$$

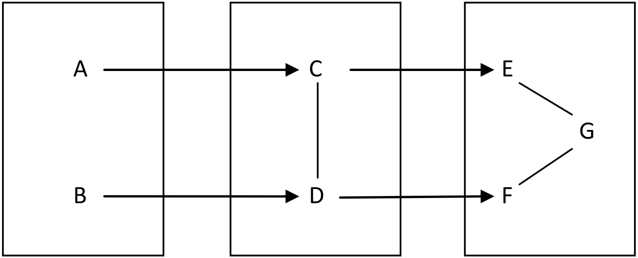
($\mathbb{D}_i$ contains loglinear ‘effects’ set to zero under *i*th CI)

Let $\mathcal{M}_1, \ldots, \mathcal{M}_m = \mathcal{V}$ be nondecreasing ordering of marginals.
For $\mathcal{E} \subseteq \mathcal{V}$, $\mathcal{M}(\mathcal{E})$ is first of the $\mathcal{M}_i$ containing $\mathcal{E}$.

Theorem

Suppose $\mathcal{C}_i \subseteq \mathcal{M}(\mathcal{E}) \subseteq \mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i$ for all i and $\mathcal{E} \in \mathbb{D}_i$. Then
** $\mathcal{Q}$ is hierarchical marginal log-linear and is hence smooth.*
** Simple formula can be given for correct df*

Chain graph whose Andersson–Madigan–Perlman interpretation is a smooth model by Theorem (family contains nonsmooth models)



Smoothness not easily verified without theorem

Further work

By incompleteness of axioms, conditional independence theory as complex as number theory. Much to be discovered!

Other results by Milan Studeny, Frantisek Matus.

Some references

Bergsma, W. P. and T. Rudas (2002). Marginal models for categorical data. Annals of Statistics, Vol. 30, 140-159

Bergsma, W. P., M. A. Croon and J. A. Hagenaars (2009). Marginal models for dependent, clustered and longitudinal categorical data. Springer NY.

Rudas, T. and W. P. Bergsma and R. Nemeth (2010). Marginal log-linear parameterization of conditional independence models. Biometrika, vol 97, issue 4, pp.1006-1012