Minimum Distance Estimators in Logistic Regression under Complex Designs

II Conference of European Statistics Stakeholders

E. Castilla, N. Martin and L. Pardo (*Complutense University of Madrid)*

Budapest, October 21, 2016

Outline

1. Multinomial logistic regression (MLR) with complex survey
2. Pseudo minimum phi-divergence estimators for the MLR model with complex survey
3. "Design effect matrix" and "Design effect" for MLR with sample survey design
4. Simulation study
5. References

Multinomial logistic regression (MLR) with complex survey

- Let $\mathbf{Y}$ be the response variable with $d+1$ categories
  - The population is divided in H strata $(h = 1, ..., H)$
  - In each stratum h, there are n_h clusters $(i = 1, ..., n_h)$
  - Cluster i of the stratum h has m_{hi} units
  - $\mathbf{y}_{hij} = \left(y_{hij1}, ..., y_{hij,d+1}\right)^T \equiv$ classification vectors
    - If $y_{hijr} = 1$ and $y_{hijs} = 0$ for $s \in \{1, ..., d+1\} - \{r\}$, the unit j selected from the cluster i of the stratum h falls in the category r.
  - The response variable $\mathbf{Y}$ depends on k explanatory variables

$$\mathbf{x}_{hij} = \left(x_{hij1}, ..., x_{hijk}\right)^T$$

(Explanatory variables for the unit j in the cluster i of the stratum h)
  - $w_{hi} \equiv$ Sampling weight from the cluster i of the stratum h.

Multinomial logistic regression (MLR) with complex survey

- The expectation of the element r, Y_{hijr}, of $\mathbf{Y}_{hij} = (Y_{hij1}, ..., Y_{hij,d+1})^T$, is given by

$$\pi_{hijr}(\boldsymbol{\beta}) = \begin{cases} \dfrac{\exp\{\mathbf{x}_{hij}^T \boldsymbol{\beta}_r\}}{1 + \sum_{s=1}^{d} \exp\{\mathbf{x}_{hij}^T \boldsymbol{\beta}_s\}}, & r = 1, ..., d \\ \dfrac{1}{1 + \sum_{s=1}^{d} \exp\{\mathbf{x}_{hij}^T \boldsymbol{\beta}_s\}}, & r = d+1 \end{cases},$$

  - $\boldsymbol{\beta}_r = (\beta_{1r}, ..., \beta_{kr})^T \in \mathbb{R}^k$, $r = 1, ..., d$ and $\boldsymbol{\beta}_{d+1} = (0, ..., 0)^T$.
  - $\Theta = \{\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, ..., \boldsymbol{\beta}_d^T)^T, \boldsymbol{\beta}_j = (\beta_{j1}, ..., \beta_{jk})^T \in \mathbb{R}^k,$ $j = 1, ..., d\} = \mathbb{R}^{dk}$.

- We shall assume

$$\pi_{hijr}(\boldsymbol{\beta}) = \pi_{hir}(\boldsymbol{\beta}), \quad j = 1, ..., m_{hi},$$

(All the individuals in the cluster i of the stratum h have the the same explanatory variables $\mathbf{x}_{hi} = (x_{hi1}, ..., x_{hik})^T$)

Pseudo Maximum likelihood estimator

- $\widehat{\mathbf{Y}}_{hi} = \sum_{j=1}^{m_{hi}} \mathbf{Y}_{hij} = \left(\sum_{j=1}^{m_{hi}} Y_{hij1}, \ldots, \sum_{j=1}^{m_{hi}} Y_{hij,d+1} \right)^T = (\widehat{Y}_{hi1}, \ldots, \widehat{Y}_{hi,d+1})^T$
  (The number of units in the cluster i of the stratum h)
- We define the following theoretical probability vector, $\boldsymbol{\pi}(\boldsymbol{\beta})$, by

  $$\left(\tfrac{w_{11}m_{11}}{\tau}\boldsymbol{\pi}_{11}^T(\boldsymbol{\beta}), \ldots, \tfrac{w_{1n_1}m_{1n_1}}{\tau}\boldsymbol{\pi}_{1n_1}^T(\boldsymbol{\beta}), \ldots, \tfrac{w_{H1}m_{H1}}{\tau}\boldsymbol{\pi}_{H1}^T(\boldsymbol{\beta}), \ldots, \tfrac{w_{Hn_H}m_{Hn_H}}{\tau}\boldsymbol{\pi}_{Hn_H}^T(\boldsymbol{\beta})\right)^T$$

  with

  $$\tau = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi} m_{hi}$$

- We shall also consider the non-parametric probability vector

  $$\begin{aligned} \widehat{\mathbf{p}} &= \frac{1}{\tau}(\widehat{\mathbf{Y}}_1^T, \ldots, \widehat{\mathbf{Y}}_H^T)^T \\ &= \frac{1}{\tau}(w_{11}\widehat{\mathbf{Y}}_{11}^T, \ldots, w_{1n_1}\widehat{\mathbf{Y}}_{1n_1}^T, \ldots, w_{H1}\widehat{\mathbf{Y}}_{H1}^T, \ldots, w_{Hn_H}\widehat{\mathbf{Y}}_{Hn_H}^T)^T. \end{aligned}$$

Pseudo Maximum likelihood estimator

- The Kullback-Leibler divergence between the probability vectors $\widehat{\mathbf{p}}$ and $\boldsymbol{\pi}(\boldsymbol{\beta})$ is given by

$$d_{K-L}\left(\widehat{\mathbf{p}}, \boldsymbol{\pi}(\boldsymbol{\beta})\right) = \frac{1}{\tau}\sum_{h=1}^{H}\sum_{i=1}^{n_h} w_{hi} \sum_{s=1}^{d+1} \widehat{y}_{his} \log \frac{\widehat{y}_{his}}{m_{hi}\pi_{his}(\boldsymbol{\beta})}$$
$$= K - \mathcal{L}(\boldsymbol{\beta})$$

  - Being

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{h=1}^{H}\sum_{i=1}^{n_h} w_{hi} \log \boldsymbol{\pi}_{hi}^{T}(\boldsymbol{\beta})\,\widehat{\mathbf{y}}_{hi}$$

  the **Pseudo Loglikelihood**

- The **Pseudo Maximum Likelihood Estimator** of parameter $\boldsymbol{\beta}$ can be defined by

$$\widehat{\boldsymbol{\beta}}_P = \arg\min_{\boldsymbol{\beta}\in\Theta} d_{K-L}\left(\widehat{\mathbf{p}}, \boldsymbol{\pi}(\boldsymbol{\beta})\right).$$

Pseudo Minimum Phi-divergence Estimator

- Phi-divergence measures,

$$d_\phi\left(\widehat{\mathbf{p}}, \boldsymbol{\pi}\left(\boldsymbol{\beta}\right)\right) = \frac{1}{\tau}\sum_{h=1}^{H}\sum_{i=1}^{n_h} w_{hi} m_{hi} \sum_{s=1}^{d+1} \pi_{his}\left(\boldsymbol{\beta}\right)\phi\left(\frac{\widehat{y}_{his}}{m_{hi}\pi_{his}\left(\boldsymbol{\beta}\right)}\right),$$

where $\phi \in \Phi^*$ is the class of all convex functions $\phi\left(x\right)$, defined for $x > 0$, such that at $x = 1$, $\phi\left(1\right) = 0$, $\phi''\left(1\right) > 0$, and at $x = 0$, $0\phi\left(0/0\right) = 0$ and $0\phi\left(p/0\right) = \lim_{u\to\infty}\phi\left(u\right)/u$.

Definition

We consider the MLR model with complex survey.The **Pseudo Minimum Phi-divergence Estimator** of parameter $\boldsymbol{\beta}$ is defined as

$$\widehat{\boldsymbol{\beta}}_{\phi,P} = \arg\min_{\boldsymbol{\beta}\in\Theta} d_\phi\left(\widehat{\mathbf{p}}, \boldsymbol{\pi}\left(\boldsymbol{\beta}\right)\right).$$

Pseudo Minimum Phi-divergence Estimator

Theorem

Let $\widehat{\boldsymbol{\beta}}_{\phi,P}$ *the* **pseudo minimum phi-divergence estimator** *of parameter* $\boldsymbol{\beta}$, $n = \sum_{h=1}^{H} n_h$ *the total of clusters in all the strata of the sample and* η_h^* *an unknown proportion obtained as* $\lim_{n\to\infty} \frac{n_h}{n} = \eta_h^*$, $h = 1, ..., H$. *Then we have*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\phi,P} - \boldsymbol{\beta}_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_{dk}, \mathbf{H}^{-1}\left(\boldsymbol{\beta}_0\right)\mathbf{G}\left(\boldsymbol{\beta}_0\right)\mathbf{H}^{-1}\left(\boldsymbol{\beta}_0\right)\right),$$

($\mathbf{H}(\boldsymbol{\beta}) \equiv$ *Fisher information matrix)*

Design effect matrix and design effect in the MLR model

- Let $\widehat{\boldsymbol{\beta}}_{\phi}$ denote the minimum phi-divergence estimator of $\boldsymbol{\beta}$ for multinomial sampling. It can be seen that

$$\lim_{n\to\infty} \mathbf{V}[\sqrt{n}\widehat{\boldsymbol{\beta}}_{\phi}] = \mathbf{H}^{-1}\left(\boldsymbol{\beta}_0\right).$$

- The **"design effect matrix"** for the MLR with sample survey design is defined as

$$\lim_{n\to\infty} \mathbf{V}[\sqrt{n}\widehat{\boldsymbol{\beta}}_{\phi,P}]\mathbf{V}^{-1}[\sqrt{n}\widehat{\boldsymbol{\beta}}_{\phi}] = \mathbf{H}^{-1}\left(\boldsymbol{\beta}_0\right)\mathbf{G}\left(\boldsymbol{\beta}_0\right)$$

- The **"design effect"**, for the MLR model with sample survey design is defined as

$$\nu(\boldsymbol{\beta}_0) = \frac{1}{dk}\mathrm{trace}\left(\mathbf{H}^{-1}(\boldsymbol{\beta}_0)\mathbf{G}(\boldsymbol{\beta}_0)\right).$$

Design effect matrix and design effect in the MLR model

- The design effect is specially interesting for models such that

$$\mathbf{E}[\widehat{\mathbf{Y}}_{hi}] = m_h \boldsymbol{\pi}_{hi}\left(\boldsymbol{\beta}_0\right) \quad \text{and} \quad \mathbf{V}[\widehat{\mathbf{Y}}_{hi}] = \nu_{m_h} m_h \Delta(\boldsymbol{\pi}_{hi}\left(\boldsymbol{\beta}_0\right)), \qquad (1)$$
$$\nu_{m_h} = 1 + \rho_h^2 (m_h - 1),$$

  - $\nu_{m_h} \equiv$ Parameter of overdispersion
  - $\rho_h^2 \equiv$ Intra-cluster correlation coefficient
  - Clusters have equal size in the strata, $m_{hi} = m_h$, $h = 1, ..., H$, $i = 1, ..., n_h$.
- Examples of distributions of $\widehat{\mathbf{Y}}_{hi}$ verifying (1) are the so-called "overdispersed multinomial distributions" (Dirictlet-multinomial, Random-clumped, m-inflated distribution)
- After obtaining the pseudo minimum phi-divergence estimator of parameter $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}_{\phi,P}$, the interest will be in estimating the **intra-cluster correlation coefficient** as well as the **parameter of overdispersion**.

Design effect matrix and design effect

Theorem

Assume $w_{hi} = w_h$, $i = 1, ..., n_h$. *An estimator of the* **parameter of overdispersion** *based on the* "**linearization method of Binder**" *is*

$$\widehat{\nu}_{m_h}(\widehat{\boldsymbol{\beta}}_{\phi,P}) = \frac{1}{dk}\mathrm{trace}\left(\left(\sum_{i=1}^{n_h} m_h \Delta(\boldsymbol{\pi}^*_{hi}(\widehat{\boldsymbol{\beta}}_{\phi,P})) \otimes \mathbf{x}_{hi}\mathbf{x}_{hi}^T\right)^{-1} \times \sum_{i=1}^{n_h} \left(\mathbf{v}_{hi}(\widehat{\boldsymbol{\beta}}_{\phi,P}) - \overline{\mathbf{v}}_h(\widehat{\boldsymbol{\beta}}_{\phi,P})\right)\left(\mathbf{v}_{hi}(\widehat{\boldsymbol{\beta}}_{\phi,P}) - \overline{\mathbf{v}}_h(\widehat{\boldsymbol{\beta}}_{\phi,P})\right)^T\right)$$

with $\mathbf{v}_{hi}(\widehat{\boldsymbol{\beta}}_{\phi,P}) = \mathbf{r}^*_{hi}(\boldsymbol{\beta}) \otimes \mathbf{x}_{hi}$ *and* $\overline{\mathbf{v}}_h(\widehat{\boldsymbol{\beta}}_{\phi,P}) = \frac{1}{n_h}\sum_{k=1}^{n_h} \mathbf{v}_{hk}(\widehat{\boldsymbol{\beta}}_{\phi,P})$, *and an estimator of the* **intra-cluster correlation coefficient** *is*

$$\widehat{\rho}_h^2(\widehat{\boldsymbol{\beta}}_{\phi,P}) = \frac{\widehat{\nu}_{m_h}(\widehat{\boldsymbol{\beta}}_{\phi,P}) - 1}{m_h - 1}$$

Design effect matrix and design effect

Theorem

Let $\widehat{\boldsymbol{\beta}}_{\phi,P}$ *the pseudo minimum phi-divergence estimate of parameter* $\boldsymbol{\beta}$ *for a multinomial logistic regression model with "overdispersed multinomial distribution". An estimator of the* **parameter of overdispersion** *based on the* **"method of moments"** *is given by*

$$\widetilde{\nu}_{m_h}(\widehat{\boldsymbol{\beta}}_{\phi,P}) = \frac{1}{n_h d} \sum_{i=1}^{n_h} \sum_{s=1}^{d+1} \frac{\left(\widehat{y}_{his} - m_h \pi_{his}(\widehat{\boldsymbol{\beta}}_{\phi,P})\right)^2}{m_h \pi_{his}(\widehat{\boldsymbol{\beta}}_{\phi,P})}$$

and a estimator of the **intra-cluster correlation coefficien**t *based on the* **"method of moments"** , *is*

$$\widetilde{\rho}_h^2(\widehat{\boldsymbol{\beta}}_{\phi,P}) = \frac{\widetilde{\nu}_{m_h}(\widehat{\boldsymbol{\beta}}_{\phi,P}) - 1}{m_h - 1}.$$

Simulation study: Minimum phi-divergence estimators

- The **pseudo minimum phi-divergence estimator**

$$\widehat{\boldsymbol{\beta}}_{\phi,P} = \arg\min_{\boldsymbol{\beta}\in\Theta} d_\phi\left(\widehat{\mathbf{p}}, \boldsymbol{\pi}\left(\boldsymbol{\beta}\right)\right)$$

$$d_\phi\left(\widehat{\mathbf{p}}, \boldsymbol{\pi}\left(\boldsymbol{\beta}\right)\right) = \frac{1}{\tau}\sum_{h=1}^{H}\sum_{i=1}^{n_h} w_{hi} m_{hi} \sum_{s=1}^{d+1} \pi_{his}\left(\boldsymbol{\beta}\right) \phi\left(\frac{\widehat{y}_{his}}{m_{hi}\pi_{his}\left(\boldsymbol{\beta}\right)}\right),$$

- Divergence measure of Cressie-Read

$$\phi_\lambda(x) = \begin{cases} \frac{1}{\lambda(1+\lambda)}\left[x^{\lambda+1} - x - \lambda(x-1)\right], & \lambda \in \mathbb{R} - \{-1, 0\} \\ x\log x - x + 1, & \lambda = 0 \\ -\log x + x - 1 & \lambda = -1 \end{cases}.$$

- $\lambda \in \{0, \frac{2}{3}, 1, 1.5, 2, 2.5\}$

Experiment of simulation

- $\widehat{\mathbf{Y}}_i \equiv$ Described by Dirichlet-multinomial (DM), Random-clumped (RC), m-inflated (m-I))
- $H = 1$ (One stratum). Different number of clusters in the stratum and different size in each cluster
- $d = 3$ (Four classes) and $K = 4$.
- The true probability associated with the cluster i is $\boldsymbol{\pi}_i(\boldsymbol{\beta}_0) = (\pi_{i1}(\boldsymbol{\beta}_0), \pi_{i2}(\boldsymbol{\beta}_0), \pi_{i3}(\boldsymbol{\beta}_0), \pi_{i4}(\boldsymbol{\beta}_0))^T$, where

$$\boldsymbol{\pi}_i(\boldsymbol{\beta}_0) = \frac{\exp\{\mathbf{x}_i^T \boldsymbol{\beta}_r^0\}}{\sum_{s=1}^{d+1} \exp\{\mathbf{x}_i^T \boldsymbol{\beta}_s^0\}},$$

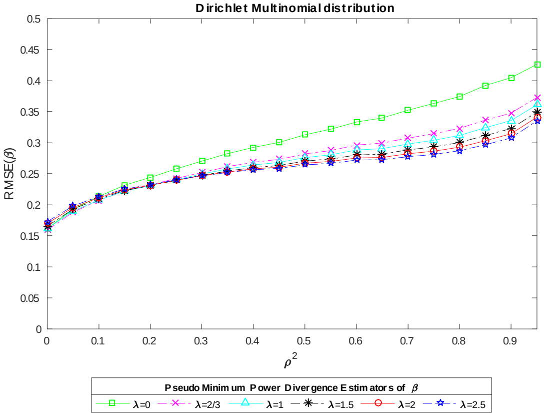
$\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \boldsymbol{\beta}_4^T)^T$, with $\boldsymbol{\beta}_1^T = (-0.3, -0.1, 0.1, 0.2)$, $\boldsymbol{\beta}_2^T = (0.2, -0.2, -0.2, 0.1)$, $\boldsymbol{\beta}_3^T = (-0.1, 0.3, -0.3, 0.1)$, $\boldsymbol{\beta}_4^T = (0, 0, 0, 0)$

- $\mathbf{x}_i \overset{ind}{\sim} \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\mu} = (1, -2, 1, 5)^T$, $\Sigma = \mathrm{diag}\{0, 25, 25, 25\}$, $i = 1, \ldots, n$,

Scenarios

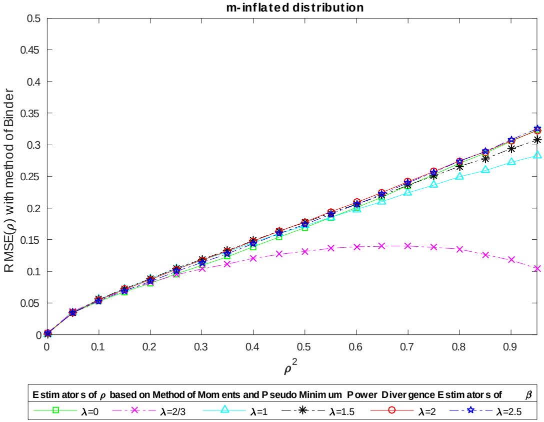
- We consider 5 different scenarios
  - Scenario 1: $n = 60$, $m = 21$, $\rho^2 \in \{0.05i\}_{i=0}^{19}$, DM, RC and m-I distributions
  - Scenario 2: $n \in \{10i\}_{i=1}^{15}$, $m = 21$, $\rho^2 = 0.25$, RC distribution
  - Scenario 3: $n = 60$, $m \in \{10i\}_{i=1}^{10}$, $\rho^2 = 0.25$, RC distribution
  - Scenario 4: $n = 60$, $m \in \{10i\}_{i=1}^{10}$, $\rho^2 = 0.75$, RC distribution
  - Scenario 5: $n = 20$, $m \in \{10i\}_{i=1}^{10}$, $\rho^2 = 0.25$, RC distribution

Conclusions for parameteres (Mean square error)



Conclusions for intra-cluster correlation coeficient (Mean square error)

- The best estimator of ρ^2 with Binder's method is obtained with $\lambda = 2/3$



References

Binder, D. A. (1983). On the variance of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, *279–292.*

Castilla, E., Martin, N. and Pardo, L. (2016). Pseudo minimum phi-divergence estimator for multinomial logistic regression with complex sample design. https://arxiv.org/abs/1606.01009.

Morel, G. (1989). Logistic regression under Complex Survey Designs. *Survey Methodology*, **15**, *203–223.*

Pardo, L. (2005). *Statistical Inference Based on Divergence Measures. Statistics: Texbooks and Monographs.* Chapman & Hall/CRC, New York.

SAS Institute Inc. (2013). `SAS/STAT`® `13.1` User's Guide. Cary, NC.