Complutense University of Madrid

Multinomial Generalized Linear Models

María del Carmen Pardo Llorente

Department of Statistics and O.R. (I)
Faculty of Mathematics,
Complutense University of Madrid

November 2009

Contents

Outline

- **Multinomial models**
- **Statistical Inference for multinomial models**
  Estimators and Phi-divergence Test
- **Diagnostics Tools for the Multinomial Generalized Linear Models**
- **Computational study for some ordinal models**

Models for Nominal responses

- $Y \in \{1,..., J\}$ Response variable
- $\mathbf{x}^T = (x_1, \ldots, x_m) \in \mathbb{R}^m$ Explicative variables
- $$U_r = \theta_{r0} + \mathbf{x}^T \boldsymbol{\theta}_r + \epsilon_r, \; r = 1, \ldots, J$$

  where $\boldsymbol{\theta}_r = (\theta_{r1}, \ldots, \theta_{rm})$ and $\epsilon_1, \ldots, \epsilon_J$ are v.a.i.i.d. (F)

$$\mathbf{Y} = r \Leftrightarrow U_r = \max_{j=1,\ldots,J} U_j$$

$$\Pr(\mathbf{Y} = r \mid \mathbf{x}^T) = \int_{-\infty}^{\infty} \prod_{s \neq r} F(\gamma_{s0} + \mathbf{x}^T \boldsymbol{\gamma}_s + \epsilon) f(\epsilon) d\epsilon$$

where $\gamma_{s0} = \theta_{r0} - \theta_{s0}$ y $\boldsymbol{\gamma}_s = \boldsymbol{\theta}_r - \boldsymbol{\theta}_s$

- For

$$F(x) = \exp(-\exp(-x))$$

we obtain the **multinomial logit model** given by

$$\Pr(\mathbf{Y} = r \mid \mathbf{x}^T) = \frac{\exp(\beta_{r0} + \mathbf{x}^T \boldsymbol{\beta}_r)}{1 + \sum_{s=1}^{J-1} \exp(\beta_{s0} + \mathbf{x}^T \boldsymbol{\beta}_s)}$$

where $\beta_{s0} = \theta_{s0} - \theta_{J0}, \; \boldsymbol{\beta}_s = \boldsymbol{\theta}_s - \boldsymbol{\theta}_J$

Models for Ordinal Responses (I)

Cumulative models

$$\mathbf{Y} = r \Leftrightarrow \alpha_{r-1} < U \leq \alpha_r,\ r = 1, \ldots, J,$$

where $-\infty = \alpha_0 < \alpha_1 < \ldots < \alpha_J = \infty$.

$$U = -\mathbf{x}^T \boldsymbol{\beta} + \epsilon,$$

where ϵ is r.v. (F).

$$\Pr(\mathbf{Y} \leq r/\mathbf{x}^T) = \Pr(U \leq \alpha_r) = F(\alpha_r + \mathbf{x}^T \boldsymbol{\beta})$$

- For $F(x) = (1 + \exp(-x))^{-1}$ $x \in \mathbb{R}$, we obtain the **cumulative logistic model**

$$\Pr(\mathbf{Y} \leq r/\mathbf{x}^T) = \frac{\exp(\alpha_r + \mathbf{x}^T \boldsymbol{\beta})}{1 + \exp(\alpha_r + \mathbf{x}^T \boldsymbol{\beta})},\ r = 1, \ldots, J-1.$$

- For $F(x) = 1 - \exp(-\exp(x))$ $x \in \mathbb{R}$, we obtain the **grouped Cox model**

$$\Pr(\mathbf{Y} \leq r/\mathbf{x}^T) = 1 - \exp(-\exp(\alpha_r + \mathbf{x}^T \boldsymbol{\beta})),\ r = 1, \ldots, J-1$$

- For $F(x) = \exp(-\exp(-x))$ $x \in \mathbb{R}$, we obtain the **extreme maximal-value distribution model**

$$\Pr(\mathbf{Y} \leq r/\mathbf{x}^T) = \exp(-\exp(-(\alpha_r + \mathbf{x}^T \boldsymbol{\beta}))),\ r = 1, \ldots, J-1$$

Models for Ordinal Responses (II)

Sequential Models

$$U_r = -\mathbf{x}^T\boldsymbol{\beta} + \epsilon_r, \quad r = 1, \ldots, J-1$$

where ϵ_r is r.v. (F).

The response mechanism starts in category 1 and the first step is determined by

$$\mathbf{Y} = 1 \Leftrightarrow U_1 \leq \alpha_1,$$

where α_1 is a threshold parameter. If $U_1 \leq \alpha_1$, the process stops. If not, the process is continuing in the form

$$\mathbf{Y} = 2 \text{ given } \mathbf{Y} \geq 2 \Leftrightarrow U_2 \leq \alpha_2,$$

and so on. Therefore, the complete mechanism is specified by

$$\mathbf{Y} > r \text{ given } \mathbf{Y} \geq r \Leftrightarrow U_r > \alpha_r, \; r = 1, \ldots, J-1$$

where $\alpha_J = \infty$.

The probabilities of the model are given by

$$\Pr\left(\mathbf{Y} = r / \mathbf{x}^T\right) = F\left(\alpha_r + \mathbf{x}^T\boldsymbol{\beta}\right)\prod_{s=1}^{r-1}\left(1 - F\left(\alpha_s + \mathbf{x}^T\boldsymbol{\beta}\right)\right), \; r = 1, \ldots, J$$

where $\prod_{s=1}^{0}(.) = 1$

Models for Ordinal responses (III)

- For $F(x) = (1+\exp(-x))^{-1}$, we obtain the **sequential logistic model**

$$\Pr\left(\mathbf{Y}=r/\mathbf{Y}\geq r,\mathbf{x}^{T}\right)=\frac{\exp\left(\alpha_r+\mathbf{x}^{T}\boldsymbol{\beta}\right)}{1+\exp\left(\alpha_r+\mathbf{x}^{T}\boldsymbol{\beta}\right)}$$

- For $F(x)=1-\exp(-\exp(x))$, we obtain the **extreme minimal-value distribution sequential model**

$$\Pr(\mathbf{Y}=r/\mathbf{Y}\geq r,\mathbf{x}^{T}) = 1-\exp(-\exp(\alpha_r+\mathbf{x}^{T}\boldsymbol{\beta})),\ r=1,\ldots,J-1$$

This model is a special parametric form of the grouped Cox model.

- For $F(x)=1-\exp(-x)$, we obtain the **exponential sequential model**

$$\Pr\left(\mathbf{Y}=r/\mathbf{Y}\geq r,\mathbf{x}^{T}\right)=1-\exp\left(-\left(\alpha_r+\mathbf{x}^{T}\boldsymbol{\beta}\right)\right)$$

Models for Ordinal responses (IV)

Adjacent categories logits model

$$\Pr\left(\mathbf{Y} = r / \mathbf{Y} \in \{r, r+1\}, \mathbf{x}^T\right) = F\left(\alpha_{r0} + \mathbf{x}^T \boldsymbol{\beta}\right)$$

with $F(x) = (1 + \exp(-x))^{-1}$

- It can be rewritten as a multinomial model with an ajusted design matrix

$$\log\left(\frac{\Pr(\mathbf{Y} = r/\mathbf{x}^T)}{\Pr(\mathbf{Y} = J/\mathbf{x}^T)}\right) = \sum_{k=r}^{J-1} (\alpha_{k0} + \mathbf{x}^T \boldsymbol{\beta}) = \beta_{r0} + \mathbf{u}_r^T \boldsymbol{\beta}$$

with $\beta_{r0} = \sum_{k=r}^{J-1} \alpha_{k0}$ and $\mathbf{u}_r = (J - r)\mathbf{x}$

- It can be generalized substituting the logistic distribution function F by every distribution function strict monotone increasing

Multinomial Generalized Linear Models (I)

- $\mathbf{Y} \in \{1, \ldots, J\}$

- $\mathbf{x}_i^T = (x_{i1},..., x_{im}), \ \ i = 1,..., N$

- Multinomial GLM assumed that $\boldsymbol{\mu}_i = E[\mathbf{Y}/\mathbf{x}_i^T]$ is determinated by a linear predictor

$$\boldsymbol{\eta}_i = \mathbf{Z}_i^T \boldsymbol{\beta}$$

in the form

$$\boldsymbol{\mu}_i = \boldsymbol{h}(\boldsymbol{\eta}_i) = \boldsymbol{h}(\mathbf{Z}_i^T \boldsymbol{\beta})$$

where $\boldsymbol{h}$ is a vectorial response function, $\mathbf{Z}_i$ is a $(p \times J - 1)$-design matrix obtained from $\mathbf{x}_i$ and $\boldsymbol{\beta}$ is a p -dimensional vector of unknown parameters.

- $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ m.a.s. with $\mathbf{Y}_i \equiv M\left(n(\mathbf{x}_i); \pi_1\left(\mathbf{Z}_i^T \boldsymbol{\beta}\right),..., \pi_{J-1}\left(\mathbf{Z}_i^T \boldsymbol{\beta}\right)\right)$

  being $\pi_r\left(\mathbf{Z}_i^T \boldsymbol{\beta}\right) = P(\mathbf{Y_i} = r \mid \mathbf{x}_i^T), \ r = 1,..., J-1$ and $\boldsymbol{\pi}\left(\mathbf{Z}_i^T \boldsymbol{\beta}\right)^T = \left(\pi_1\left(\mathbf{Z}_i^T \boldsymbol{\beta}\right),..., \pi_{J-1}\left(\mathbf{Z}_i^T \boldsymbol{\beta}\right)\right)$

- All the above models can be written as a multinomial GLM.

Multinomial logit and adjacent categoríes logits models

- Response function $\boldsymbol{h} = (h_1,\ldots,h_{J-1})$ with

$$h_r(\eta_{i1},\ldots,\eta_{iJ-1}) = \frac{\exp(\eta_{ir})}{1+\sum_{s=1}^{J-1}\exp(\eta_{is})},\quad r = 1,\ldots,J-1$$

or link function $\boldsymbol{g} = (g_1,\ldots,g_{J-1})$ with

$$g_r(\pi_1(\eta_i),\ldots,\pi_{J-1}(\eta_i)) = \log\frac{\pi_r(\eta_i)}{1-(\pi_1(\eta_i)+\ldots+\pi_{J-1}(\eta_i))},\quad r = 1,\ldots,J-1$$

- **Multinomial logit model**

$$\boldsymbol{\beta}^T = (\beta_{10},\boldsymbol{\beta}_1^T,\ldots,\beta_{J-10},\boldsymbol{\beta}_{J-1}^T)$$

$$\mathbf{Z}_i^T = \begin{pmatrix} 1 & \mathbf{x}_i^T & & & & & & \\ & & 1 & \mathbf{x}_i^T & & & & \\ & & & & \cdot & & & \\ & & & & & \cdot & & \\ & & & & & & \cdot & \\ & & & & & & & 1 \quad \mathbf{x}_i^T \end{pmatrix}$$

- **Adjacent categoríes logits model**

$$\boldsymbol{\beta}^T = (\beta_{10},\ldots,\beta_{J-10},\boldsymbol{\beta}^T)$$

$$\mathbf{Z}_i^T = \begin{pmatrix} 1 & & & & & (J-1)\mathbf{x}_i^T \\ & 1 & & & & (J-2)\mathbf{x}_i^T \\ & & \cdot & & & \cdot \\ & & & \cdot & & \cdot \\ & & & & \cdot & \cdot \\ & & & & 1 & \mathbf{x}_i^T \end{pmatrix}$$

Multinomial Generalized Linear Models (III)

Cumulative models

- Link function $\boldsymbol{g} = (g_1, \ldots, g_{J-1})$ with

$$g_r(\pi_1(\eta_i), \ldots, \pi_{J-1}(\eta_i)) = F^{-1}(\pi_1(\eta_i) + \ldots + \pi_r(\eta_i)), \; r = 1, \ldots, J-1,$$

$$\mathbf{Z}_i^T = \begin{pmatrix} 1 & & & & & \mathbf{x}_i^T \\ & 1 & & & & \mathbf{x}_i^T \\ & & \cdot & & & \cdot \\ & & & \cdot & & \cdot \\ & & & & \cdot & \cdot \\ & & & & 1 & \mathbf{x}_i^T \end{pmatrix}$$

and $\boldsymbol{\beta}^T = (\alpha_1, \ldots, \alpha_{J-1}, \boldsymbol{\beta}^T)$

Sequential models

- Link function $\boldsymbol{g} = (g_1, \ldots, g_{J-1})$ is given by

$$g_r(\pi_1(\eta_i), \ldots, \pi_{J-1}(\eta_i)) = F^{-1}(\pi_r(\eta_i)/(1 - \pi_1(\eta_i) - \ldots - \pi_{r-1}(\eta_i))), \; r = 1, \ldots, J-1$$

and the parameter vector as well as the design matrix match with the cumulative model

Outline

- Multinomial models
- **Statistical Inference for multinomial models**
  Estimators and Phi-divergence Test
- Diagnostics Tools for the Multinomial Generalized Linear Models
- Computational study for some ordinal models

Minimum ϕ–divergence estimator. Asymptotic properties (I)

- MLE of $\boldsymbol{\beta}$ is obtained maximizing the log-likelihood function of the sample $\boldsymbol{y}_1,\dots,\boldsymbol{y}_N$

$$L_N(\boldsymbol{\beta};\boldsymbol{y}_1,\dots,\boldsymbol{y}_N) = Cte + \sum_{i=1}^{N} l_{N,i}(\boldsymbol{\beta};\boldsymbol{y}_i)$$

with

$$l_{N,i}(\boldsymbol{\beta};\boldsymbol{y}_i) = \log\left(\frac{\boldsymbol{\pi}(\mathbf{Z}_i^T\boldsymbol{\beta})^T}{\pi_J(\mathbf{Z}_i^T\boldsymbol{\beta})}\right)\boldsymbol{y}_i + n(\boldsymbol{x}_i)\log(\pi_J(\mathbf{Z}_i^T\boldsymbol{\beta}))$$

where by $\log(\boldsymbol{x}^T)$ we call $(\log(x_1),\dots,\log(x_{J-1}))$ with $\boldsymbol{x} = (x_1,\dots,x_{J-1})^T$, $\boldsymbol{y}_i = (y_{1i},\dots,y_{J-1i})^T$, $i = 1,\dots,N$ and $\pi_J(\mathbf{Z}_i^T\boldsymbol{\beta}) = 1 - \pi_1(\mathbf{Z}_i^T\boldsymbol{\beta}) - \dots - \pi_{J-1}(\mathbf{Z}_i^T\boldsymbol{\beta})$

- $$D_{Kullback}(\hat{\mathbf{p}},\mathbf{p}(\boldsymbol{\beta})) = \sum_{l=1}^{J}\sum_{i=1}^{N}\frac{y_{li}}{n}\log\frac{\frac{y_{li}}{n}}{\pi_l(\mathbf{x}_i)\frac{n(\mathbf{x}_i)}{n}}$$

being

$$\hat{\mathbf{p}} = \left(\frac{y_{11}}{n},\dots,\frac{y_{J1}}{n},\frac{y_{12}}{n},\dots,\frac{y_{J2}}{n},\dots,\frac{y_{1N}}{n},\dots,\frac{y_{JN}}{n}\right)^T,$$

with $y_{Ji} = n(\mathbf{x}_i) - \sum_{s=1}^{J-1} y_{si}$, $i = 1,\dots,N$, $n = n(\mathbf{x}_1) + \dots + n(\mathbf{x}_N)$ and

$$\boldsymbol{p}(\boldsymbol{\beta}) = \left(\frac{n(\boldsymbol{x}_1)}{n}\tilde{\boldsymbol{\pi}}(\mathbf{Z}_1^T\boldsymbol{\beta})^T,\dots,\frac{n(\boldsymbol{x}_N)}{n}\tilde{\boldsymbol{\pi}}(\mathbf{Z}_N^T\boldsymbol{\beta})^T\right)^T$$

being $\tilde{\boldsymbol{\pi}}(\mathbf{Z}_i^T\boldsymbol{\beta})^T = (\pi_1(\mathbf{Z}_i^T\boldsymbol{\beta}),\dots,\pi_J(\mathbf{Z}_i^T\boldsymbol{\beta}))$

Minimum ϕ–divergence estimator. Asymptotic properties (II)

- $$D_{Kullback}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\boldsymbol{\beta})) = Cte - \frac{1}{n}\log\prod_{i=1}^{N}\prod_{l=1}^{J}\pi_l(\boldsymbol{Z}_i^T\boldsymbol{\beta})^{y_{li}} = Cte - \frac{1}{n}L_N(\boldsymbol{\beta};\boldsymbol{y}_1,\ldots,\boldsymbol{y}_N)$$

$$Maximize\ L_N(\boldsymbol{\beta};\boldsymbol{y}_1,\ldots,\boldsymbol{y}_N) \equiv Minimize\ D_{Kullback}\left(\widehat{\boldsymbol{p}},\boldsymbol{p}(\boldsymbol{\beta})\right)$$

- $$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} D_{Kullback}(\widehat{\mathbf{p}},\mathbf{p}(\boldsymbol{\beta}))$$

where

$$\Theta = \left\{\boldsymbol{\beta}^T = \left(\beta_1,\ldots,\beta_p\right) : \ \beta_s \in \mathbb{R}, s = 1,..,p\right\} \qquad (1)$$

Minimum ϕ–divergence estimator. Asymptotic properties (III)

- $$D_\phi(\widehat{\mathbf{p}},\mathbf{p}(\boldsymbol{\beta})) = \sum_{l=1}^{J}\sum_{i=1}^{N}\pi_l(\mathbf{x}_i)\frac{n(\mathbf{x}_i)}{n}\phi\left(\frac{\frac{y_{li}}{n}}{\pi_l(\mathbf{x}_i)\frac{n(\mathbf{x}_i)}{n}}\right)$$

  with $\phi\in\Phi$ y Φ is the class of the convex functions $\phi(x),\ x>0,$ such as for $x=1,\ \phi(1)=0,\ \phi'(1)=0,\ \phi''(1)>0,$ for $x=0,\ 0\,\phi(0/0)=0$ and $0\phi(p/0)=p\lim_{u\to\infty}\phi(u)/u.$

Definition 1

Let $\mathbf{Y}_i,\ i=1,\ldots,N$, be independent random variables with multinomial distributions of parameters $\left(n(\mathbf{x}_i);\pi_1(\mathbf{Z}_i^T\boldsymbol{\beta}),\ldots,\pi_{J-1}(\mathbf{Z}_i^T\boldsymbol{\beta})\right)$. The *minimum* ϕ *- divergence estimator* of $\boldsymbol{\beta}$ for the model $\boldsymbol{\mu}_i = \boldsymbol{h}(\boldsymbol{\eta}_i) = \boldsymbol{h}(\mathbf{Z}_i^T\boldsymbol{\beta})$ is given by

$$\widehat{\boldsymbol{\beta}}_\phi = \arg\min_{\boldsymbol{\beta}\in\Theta} D_\phi(\widehat{\mathbf{p}},\mathbf{p}(\boldsymbol{\beta})).$$

Note that for $\phi(x) = x\log x - x + 1$ we obtain the MLE.

- Let

$$\Delta_{JN} = \left\{q=(q_1,\ldots,q_{JN})^T\ :\ \sum_{i=1}^{JN}q_i=1,\ q_i>0,\ i=1,\ldots,JN\right\}$$

Minimum ϕ–divergence estimator. Asymptotic properties (IV)

Theorem 1

Let $\mathbf{Y}_i, i = 1,\ldots, N$, be independent random variables with multinomial distribution of parameters $\left(n(\mathbf{x}_i); \pi_1\left(\mathbf{Z}_i^T\boldsymbol{\beta}\right),\ldots,\pi_{J-1}\left(\mathbf{Z}_i^T\boldsymbol{\beta}\right)\right)$. Assuming that $\widetilde{\boldsymbol{\pi}} : \Theta \to \Delta_{JN}$ has continuous second partial derivatives in a neighborhood of $\boldsymbol{\beta}^0$ and $\phi(t) \in \Phi$ twice continuously differentiable for $t > 0$ with $I_{F,n}$ positive definite in $\boldsymbol{\beta}^0$. Under these conditions, the minimum ϕ-divergence estimator of $\boldsymbol{\beta}$ is unique in a neighbohood of $\boldsymbol{\beta}^0$ for the model $\boldsymbol{\mu}_i = \boldsymbol{h}(\boldsymbol{\eta}_i) = \boldsymbol{h}(\mathbf{Z}_i^T\boldsymbol{\beta})$ and satisfying that

$$\widehat{\boldsymbol{\beta}}_\phi = \boldsymbol{\beta}^0 + \boldsymbol{I}_{F,n}(\boldsymbol{\beta}^0)^{-1}\boldsymbol{Z}Diag\left((\boldsymbol{C}_{n,i}^0)_{i=1,..,N}\right)Diag\left(\boldsymbol{p}(\boldsymbol{\beta}^0)^{-1/2}\right)(\widehat{\boldsymbol{p}}-\boldsymbol{p}(\boldsymbol{\beta}^0)) + \|\widehat{\boldsymbol{p}}-\boldsymbol{p}(\boldsymbol{\beta}^0)\|\boldsymbol{\alpha}_1(\widehat{\boldsymbol{p}};\widehat{\boldsymbol{p}}-\boldsymbol{p}(\boldsymbol{\beta}^0))$$

where $\mathbf{I}_{F,n}(\boldsymbol{\beta}) = \mathbf{Z}\mathbf{V}_n(\boldsymbol{\beta})\mathbf{Z}^T$

being $\mathbf{Z} = (\mathbf{Z}_1,\ldots,\mathbf{Z}_N)$ and $\mathbf{V}_n(\boldsymbol{\beta}) = Diag\left(\mathbf{V}_{n,1}(\boldsymbol{\beta}),\ldots,\mathbf{V}_{n,N}(\boldsymbol{\beta})\right)$

$$\mathbf{V}_{n,i}(\boldsymbol{\beta}) = \frac{n(\mathbf{x}_i)}{n}\frac{\partial\boldsymbol{\pi}(\boldsymbol{\eta}_i)}{\partial\boldsymbol{\eta}_i}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta})\frac{\partial\boldsymbol{\pi}(\boldsymbol{\eta}_i)}{\partial\boldsymbol{\eta}_i^T} \qquad (2)$$

$$\boldsymbol{C}_{n,i}^0 = (\boldsymbol{C}_{n,i})_{\boldsymbol{\beta}=\boldsymbol{\beta}^0} = \left[\left(\frac{n(\boldsymbol{x}_i)}{n}\right)^{1/2}\frac{\partial\widetilde{\boldsymbol{\pi}}(\boldsymbol{\eta}_i)}{\partial\boldsymbol{\eta}_i^T}Diag\left(\widetilde{\boldsymbol{\pi}}(\boldsymbol{\eta}_i)^{-1/2}\right)\right]_{\boldsymbol{\beta}=\boldsymbol{\beta}^0}, \; i = 1,\ldots,N$$

and

$$\boldsymbol{\alpha}_1 : \mathbb{R}^{JN\times JN} \to \mathbb{R}^p \text{ verifies that } \boldsymbol{\alpha}_1\left(\boldsymbol{p}; \boldsymbol{p}-\boldsymbol{p}(\boldsymbol{\beta}^0)\right) \to \mathbf{0} \; as \; \boldsymbol{p} \to \boldsymbol{p}(\boldsymbol{\beta}^0).$$

Minimum ϕ–divergence estimator. Asymptotic properties (V)

Theorem 2

Under the assumptions of Theorem 1 and assuming that $n(\mathbf{x}_i) \to \infty,\ i=1,\ldots N$ such that $n(\mathbf{x}_i)/n \to \lambda_i > 0,\ i = 1,\ldots N,$ it holds

a)

$$\sqrt{n}\left(\widehat{\boldsymbol\beta}_\phi - \boldsymbol\beta^0\right) \underset{n\to\infty}{\overset{L}{\to}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_{F,\lambda}(\boldsymbol\beta^0)^{-1}\right)$$

where $\mathbf{I}_{F,\lambda}(\boldsymbol\beta) = \lim_{n\to\infty} \mathbf{I}_{F,n}(\boldsymbol\beta).$

b)

$$\sqrt{n}\left(\mathbf{p}\left(\widehat{\boldsymbol\beta}_\phi\right) - \mathbf{p}(\boldsymbol\beta^0)\right) \underset{n\to\infty}{\overset{L}{\to}} \mathcal{N}(\mathbf{0}, \Sigma_2(\boldsymbol\beta^0))$$

where $\Sigma_2(\boldsymbol\beta^0) = \mathbf{S}_\lambda(\boldsymbol\beta^0)\mathbf{Z}^T\mathbf{I}_{F,\lambda}(\boldsymbol\beta^0)^{-1}\mathbf{Z}\mathbf{S}_\lambda^T(\boldsymbol\beta^0)$

with $\mathbf{S}_\lambda\left(\boldsymbol\beta^0\right) = \lim_{n\to\infty} \mathbf{S}_n\left(\boldsymbol\beta^0\right)$ and $\mathbf{S}_n(\boldsymbol\beta) = Diag\left(\frac{n(\mathbf{x}_1)}{n}\frac{\partial\widetilde{\boldsymbol\pi}(\boldsymbol\eta_1)}{\partial\boldsymbol\eta_1}, \ldots, \frac{n(\mathbf{x}_N)}{n}\frac{\partial\widetilde{\boldsymbol\pi}(\boldsymbol\eta_N)}{\partial\boldsymbol\eta_N}\right)$

Restricted minimum ϕ –divergence estimator. Asymptotic distribution

- $$\Theta_0 = \left\{ \boldsymbol{\beta} \in \Theta \,/\, \mathbf{K}^T \boldsymbol{\beta} = \boldsymbol{\xi} \right\}$$

  with Θ defined in (1) and $\mathbf{K}^T$ is any matrix of r rows and p columns with $\operatorname{rank}(\mathbf{K}^T) = r < p$ and $\boldsymbol{\xi}$ is a vector, of order r of specified constants.

Definition 2

Let $\mathbf{Y}_i$, $i = 1,\ldots,N$, be independent random variables with multinomial distribution of parameters $\left(n(\mathbf{x}_i);\pi_1\left(\mathbf{Z}_i^T\boldsymbol{\beta}\right),\ldots,\pi_{J-1}\left(\mathbf{Z}_i^T\boldsymbol{\beta}\right)\right)$. The restricted *minimum ϕ- divergence estimator* of $\boldsymbol{\beta}$ *in* Θ_0 for the model $\boldsymbol{\mu}_i = \boldsymbol{h}(\boldsymbol{\eta}_i) = \boldsymbol{h}(\mathbf{Z}_i^T\boldsymbol{\beta})$ is

$$\widehat{\boldsymbol{\beta}}_\phi^{\Theta_0} = \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}_0} D_\phi(\widehat{\mathbf{p}},\mathbf{p}(\boldsymbol{\beta})).$$

Theorem 3

Under the assumptions of Theorem 2. The restricted minimum ϕ- divergence estimator of $\boldsymbol{\beta}$ to Θ_0 for the model $\boldsymbol{\mu}_i = \boldsymbol{h}(\boldsymbol{\eta}_i) = \boldsymbol{h}(\mathbf{Z}_i^T\boldsymbol{\beta})$ is unique in a neighborhood of $\boldsymbol{\beta}^0$ and satisfing that

a) $$\widehat{\boldsymbol{\beta}}_\phi^{\Theta_0} = \boldsymbol{\beta}^0 + \boldsymbol{H}_n(\boldsymbol{\beta}^0)\boldsymbol{I}_{F,n}(\boldsymbol{\beta}^0)^{-1}\boldsymbol{Z}Diag\left((\boldsymbol{C}_{n,i}^0)_{i=1,\ldots,N}\right)Diag\left(\boldsymbol{p}(\boldsymbol{\beta}^0)^{-1/2}\right)(\widehat{\boldsymbol{p}}-\boldsymbol{p}(\boldsymbol{\beta}^0)) + \|\widehat{\boldsymbol{p}}-\boldsymbol{p}(\boldsymbol{\beta}^0)\|\boldsymbol{\alpha}_3(\widehat{\boldsymbol{p}};\widehat{\boldsymbol{p}}-\boldsymbol{p}(\boldsymbol{\beta}^0))$$

where $\mathbf{H}_n\left(\boldsymbol{\beta}^0\right) = \mathbf{I} - \mathbf{I}_{F,n}\left(\boldsymbol{\beta}^0\right)^{-1}\mathbf{K}\left(\mathbf{K}^T\mathbf{I}_{F,n}\left(\boldsymbol{\beta}^0\right)^{-1}\mathbf{K}\right)^{-1}\mathbf{K}^T$ and

$\boldsymbol{\alpha}_3 : \mathbb{R}^{JN\times JN} \to \mathbb{R}^p$ verifies that $\boldsymbol{\alpha}_3\left(\boldsymbol{p};\boldsymbol{p}-\boldsymbol{p}\left(\boldsymbol{\beta}^0\right)\right) \to \mathbf{0}$ *as* $\boldsymbol{p} \to \boldsymbol{p}\left(\boldsymbol{\beta}^0\right)$.

b) $$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_\phi^{\Theta_0} - \boldsymbol{\beta}^0\right) \underset{n\to\infty}{\overset{L}{\to}} \mathcal{N}\left(\mathbf{0}, \mathbf{H}_\lambda(\boldsymbol{\beta}^0)\mathbf{I}_{F,\lambda}(\boldsymbol{\beta}^0)^{-1}\mathbf{H}_\lambda(\boldsymbol{\beta}^0)^T\right)$$

Hypothesis Test (I)

- $$H_0 : \mathbf{K}^T\boldsymbol{\beta} = \boldsymbol{\xi} \quad against \quad H_1 : \mathbf{K}^T\boldsymbol{\beta} \neq \boldsymbol{\xi} \qquad (3)$$

where $\mathbf{K}^T$ is any matrix of r rows and p columns with rank $(\mathbf{K}^T) = r < p$ and $\boldsymbol{\xi}$ is a vector, of orden r of specifed constants.

- Particular case

$$H_0 : \boldsymbol{\beta}_r = \mathbf{0} \quad \text{against} \quad H_1 : \boldsymbol{\beta}_r \neq \mathbf{0},$$

with $\boldsymbol{\beta}_r$ a subvector of $\boldsymbol{\beta}$ of dimension r and $\mathbf{K}$ is given by

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & 1 \end{pmatrix}_{r \times p}.$$

- $$LRT = 2nD_{Kullback}\left(\mathbf{p}\left(\widehat{\boldsymbol{\beta}}\right), \mathbf{p}\left(\widehat{\boldsymbol{\beta}}^{\Theta_0}\right)\right) + o_p(1)$$

o equivalentently

$$LRT = 2n\left(D_{Kullback}\left(\widehat{\mathbf{p}}, \mathbf{p}\left(\widehat{\boldsymbol{\beta}}^{\Theta_0}\right)\right) - D_{Kullback}\left(\widehat{\mathbf{p}}, \mathbf{p}\left(\widehat{\boldsymbol{\beta}}\right)\right)\right).$$

Hypothesis Test (II)

- $$T_n^{\phi_1,\phi_2} = \frac{2n}{\phi_1''(1)} D_{\phi_1}\left(\mathbf{p}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right), \mathbf{p}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{\Theta_0}\right)\right) \quad (4)$$

  and

  $$S_n^{\phi_2} = \frac{2n}{\phi_2''(1)} \left(D_{\phi_2}\left(\widehat{\mathbf{p}}, \mathbf{p}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{\Theta_0}\right)\right) - D_{\phi_2}\left(\widehat{\mathbf{p}}, \mathbf{p}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right)\right). \quad (5)$$

  For $\phi_2(x) = x\log x - x + 1$ in (5) we obtain LRT

Theorem 4

Under the assumptions of Theorem 2 and $\phi_1 \in \Phi$. The asymptotic distribution of the family of statistics given in (4) and (5) is a chi-square with r degrees of freedom under H_0.

- To reject H_0 at level α if $T_n^{\phi_1,\phi_2}\left(\text{ó } S_n^{\phi_2}\right) > c$ where c is such that

  $$P\left(T_n^{\phi_1,\phi_2}\left(\text{ó } S_n^{\phi_2}\right) \geq c / H_0\right) = \alpha;\ \alpha \in (0,1)$$

  By Theorem 4, c can be chosen as the $(1-\alpha)-$ quantil of a χ_r^2

Hypothesis Test (III)

- $$H_{1,n} : \mathbf{K}^T \boldsymbol{\beta}_n - \boldsymbol{\xi} = n^{-1/2} \boldsymbol{\delta}$$

  where $\mathbf{K}^T$ is any matrix of r rows and p columns with $\operatorname{rank}(\mathbf{K}^T) = r < p$, $\boldsymbol{\xi}$ is a vector of order r of specific constants and $\delta \in \mathbb{R}^P$

Theorem 5

Under the assumptions of Theorem 2 and $\phi_1 \in \Phi$. The asymptotic distribution of the family given in (4) and (5) is a noncentral chi-square under $H_{1,n}$ with r degrees of freedom and noncentrality parameter

$$\mu = \boldsymbol{\delta}^T \left(\mathbf{K}^T \mathbf{I}_{F,\lambda} (\boldsymbol{\beta}^0)^{-1} \mathbf{K} \right)^{-1} \boldsymbol{\delta}.$$

- An approximation of the power function of the test statistic $T_n^{\phi_1,\phi_2}\left(\text{ó } S_n^{\phi_2}\right)$ under the alternatives $H_{1,n}$ is obtained by

  $$\Pr\left(\chi_r^2(\mu) > \chi_{r,1-\alpha}^2\right)$$

  with $\chi_r^2(\mu)$ a noncentral chi-square with r degrees of freedom and noncentrality parameter μ.

Outline

- Multinomial models
- Statistical Inference for multinomial models
  Estimators and Phi-divergence Test
- **Diagnostics Tools for the Multinomial Generalized Linear Models**
  - *Overall test statistics*
  - *Variable Selection Methods*
  - *Diagnostic Tools*
- Computational study for some ordinal models

Overall test statistics (I)

- $$X^2 = \sum_{i=1}^{N} X_i^2,$$

  where

$$X_i^2 = \sum_{l=1}^{J} \frac{\left(y_{li} - n(\mathbf{x}_i)\pi_l\left(\mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}\right)\right)^2}{n(\mathbf{x}_i)\pi_l\left(\mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}\right)} \qquad (6)$$

- $$D = 2\sum_{i=1}^{N} d_i,$$

  where

$$d_i = \sum_{l=1}^{J} y_{li} \log \frac{y_{li}}{n(\mathbf{x}_i)\pi_l\left(\mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}\right)}.$$

Overall test statistics (II)

- $$H_0 : \mathbf{p} = \mathbf{p}(\boldsymbol{\beta}) \qquad (7)$$

$$B_n^{\phi_1,\phi_2} = \frac{2n}{\phi_1''(1)} D_{\phi_1}\left(\widehat{\mathbf{p}}, \mathbf{p}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right).$$

For $\phi_2(x) = x\log x - x + 1$ y $\phi_1(x) = \frac{1}{2}(x-1)^2$ or $\phi_1(x) = x\log x - x + 1$

we obtain X^2 or D, respectively.

Theorem 6

Let $\mathbf{Y}_i,\ i = 1,\dots,N$, be independent random variables with multinomial distribution of parameters $\left(n(\mathbf{x}_i); \pi_1\left(\mathbf{Z}_i^T\boldsymbol{\beta}\right),\dots,\pi_{J-1}\left(\mathbf{Z}_i^T\boldsymbol{\beta}\right)\right)$. Under the assumptions of Theorem 2 and $\phi_1 \in \Phi$.
The asymptotic distribution of $B_n^{\phi_1,\phi_2}$, under the null hypothesis given en (7), is a chi-square with $(J-1)N - p$ degrees of freedom.

Variable Selection Methods

Backward Stepwise Procedure

1) A maximal model *M* is tested against a set *U of admissible* submodels of *M* obtained by removing parameters associated to an explicative variable

$$H_0 \ : \ \mathbf{K}^T\beta = \beta_r = \mathbf{0} \quad \text{against} \quad H_1 \ : \ \mathbf{K}^T\beta = \beta_{\mathbf{r}} \neq \mathbf{0},$$

with $\boldsymbol{\beta}_r$ a subvector of $\boldsymbol{\beta}$ of dimension r

2) The p-values α_L of the submodels $L \in U$ are calculated

3) Then the submodel L_0 with

$$\alpha_{L_0} = \max_{L \in U} \alpha_L \text{ and } \alpha_{L_0} > \alpha_{out},$$

α_{out} a prechosen exclusion level, is selected.

4) L_0 is redefined as M, and the backward procedure is applied iteratively to admisible submodels with one less explicative variable.

5) It terminates if there is no p-valor $\alpha_L > \alpha_{out}$.

Outlying Points (I)

- $\mathbf{R}^T = (\mathbf{r}_1^T, \ldots, \mathbf{r}_N^T)$ with

$$\mathbf{r}_i = n(\mathbf{x}_i)^{-1/2} \boldsymbol{\Sigma}_i^{-1/2}\left(\widehat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_i - n(\mathbf{x}_i)\boldsymbol{\pi}\left(\mathbf{Z}_i^T \widehat{\boldsymbol{\beta}}\right)\right)$$

being $X_i^2 = (\mathbf{r}_i)^T \mathbf{r}_i.$ given in (6).

- $$\mathbf{r}_i^{\phi_2} = n(\mathbf{x}_i)^{-1/2} \boldsymbol{\Sigma}_i^{-1/2}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right) \boldsymbol{e}_i^{\phi_2},\ i = 1, \ldots, N$$

with

$$\mathbf{e}_i^{\phi_2} = \left(\mathbf{y}_i - n(\mathbf{x}_i)\boldsymbol{\pi}\left(\mathbf{Z}_i^T \widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right),\ i = 1, \ldots, N.$$

- $$\widetilde{\mathbf{r}}_i = n(\mathbf{x}_i)^{-1/2} \widetilde{\boldsymbol{\Sigma}}_i^{-1/2}\left(\widehat{\boldsymbol{\beta}}\right)\left(\widetilde{\mathbf{y}}_i - n(\mathbf{x}_i)\widetilde{\boldsymbol{\pi}}\left(\mathbf{Z}_i^T \widehat{\boldsymbol{\beta}}\right)\right)$$

where $\widetilde{\mathbf{y}}_i = (y_{1i}, \ldots, y_{Ji})^T$ y $\widetilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta}) = (\sigma_{sr}(\boldsymbol{\beta}))_{s,r=1,\ldots,J}$ with

$$\sigma_{sr}(\boldsymbol{\beta}) = \begin{cases} \pi_r\left(Z_i^T\boldsymbol{\beta}\right)\left(1 - \pi_r\left(Z_i^T\boldsymbol{\beta}\right)\right) & r = s \\ -\pi_r\left(Z_i^T\boldsymbol{\beta}\right)\pi_s\left(Z_i^T\boldsymbol{\beta}\right) & r \neq s \end{cases}$$

that verifies $X_i^2 = (\widetilde{\mathbf{r}}_i)^T \widetilde{\mathbf{r}}_i.$ A generalized inverse of $\widetilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta})$ is given by $Diag\left(\widetilde{\boldsymbol{\pi}}(\mathbf{Z}_i^T\boldsymbol{\beta})^{-1}\right)$

Outlying Points(II)

- $$\widetilde{\mathbf{r}}_i^{\phi_2} = Diag\left(\left(n(\mathbf{x}_i)\widetilde{\boldsymbol{\pi}}\left(\mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right)^{-1/2}\right)\widetilde{\mathbf{e}}_i^{\phi_2}$$

  with

$$\widetilde{\mathbf{e}}_i^{\phi_2} = \left(\widetilde{\mathbf{y}}_i - n(\mathbf{x}_i)\widetilde{\boldsymbol{\pi}}\left(\mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right)$$

Definition 3

The generalized residual (ϕ_1, ϕ_2) -divergence is defined by the J-dimensional vector $\widetilde{\mathbf{r}}_i^{\phi_1,\phi_2}$ with s -coordinate

$$\sqrt{\frac{2n(\mathbf{x}_i)}{\phi_1''(1)}}\, sign\left(y_{si} - n(\mathbf{x}_i)\pi_s\left(\mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right)\left(\pi_s\left(\mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\phi_1\left(\frac{y_{si}}{n(\mathbf{x}_i)\pi_s\left(\mathbf{Z}_i^T\widehat{\boldsymbol{\beta}}_{\phi_2}\right)}\right)\right)^{1/2}$$

and verifies

$$B_n^{\phi_1,\phi_2} = \sum_{i=1}^{N}\left(\widetilde{\mathbf{r}}_i^{\phi_1,\phi_2}\right)^T\widetilde{\mathbf{r}}_i^{\phi_1,\phi_2}.$$

Leverage Points (I)

Definition 4

The GLM hat matrix, where the parameters are estimated by using the minimum ϕ_2 −divergence estimator, is given by

$$\boldsymbol{H}\left(\widehat{\beta}_{\phi_2}\right)=\boldsymbol{V}_n\left(\widehat{\beta}_{\phi_2}\right)^{1/2}\boldsymbol{Z}^T\boldsymbol{I}_{F,n}\left(\widehat{\beta}_{\phi_2}\right)^{-1}\boldsymbol{Z}\boldsymbol{V}_n\left(\widehat{\beta}_{\phi_2}\right)^{1/2}.$$

The square matrices $\boldsymbol{H}$ and $\boldsymbol{M}=\boldsymbol{I}-\boldsymbol{H}$ are projection blocks matrices, where each block $\boldsymbol{H}^{ij}\left(\widehat{\beta}_{\phi_2}\right)$ and $\boldsymbol{M}^{ij}\left(\widehat{\beta}_{\phi_2}\right)(i,j=1,\ldots,N)$, is $(J-1)-$dimensional.

Properties

i) For all $i=1,\ldots N,$

$$0\leq Det\left(\boldsymbol{M}^{ii}\left(\widehat{\beta}_{\phi_2}\right)\right)<1,$$

ii)

$$Trace\left(\boldsymbol{H}\left(\widehat{\beta}_{\varphi_2}\right)\right)=p$$

Leverage Points (II)

- The variability of $\widehat{\boldsymbol{\beta}}_{\phi_2}$ is given by volume of the asymptotic confidence ellipsoid for $\boldsymbol{\beta}$ which is proportional to

$$\left(Det\left(\boldsymbol{Z}\widetilde{\boldsymbol{V}}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\boldsymbol{Z}^T\right)^{-1}\right)^{1/2}$$

where $\widetilde{V}(\boldsymbol{\beta}) = n\mathbf{V}_n(\boldsymbol{\beta})$

If $\boldsymbol{x}_i$ is deleted, then the volume of the confidence ellipsoid is proportional to

$$\left(Det\left(\boldsymbol{Z}_{(i)}\widetilde{\boldsymbol{V}}^{(i)}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\boldsymbol{Z}_{(i)}^T\right)^{-1}\right)^{1/2}$$

where the subscript (i) Indicates that the $\boldsymbol{x}_i$ contribution to the corresponding matrix has been removed.

A point with a value near zero of

$$Det\left(\boldsymbol{M}^{ii}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right) = \frac{Det\left(\boldsymbol{Z}_{(i)}\widetilde{\boldsymbol{V}}^{(i)}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\boldsymbol{Z}_{(i)}^T\right)}{Det\left(\boldsymbol{Z}\widetilde{\boldsymbol{V}}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\boldsymbol{Z}^T\right)} < 1$$

Is defined as a leverage point for the GLM.

- A reasonable rule of thumb for detecting leverage points is to consider that $\boldsymbol{x}_i$ is a leverage point if

$$Trace\left(\boldsymbol{H}^{ii}\left(\widehat{\boldsymbol{\beta}}_{\varphi_2}\right)\right) > 2p/N$$

Leverage Points (III)

- $$\widehat{\boldsymbol{\beta}}_{\phi_2} \approx \mathcal{N}\left(\boldsymbol{\beta}^0, \left(\boldsymbol{Z}\widetilde{\boldsymbol{V}}(\boldsymbol{\beta}^0)\boldsymbol{Z}^T\right)^{-1}\right)$$

  and

  $$\widehat{\boldsymbol{\beta}}_{\phi_2}^{(i)} \approx \mathcal{N}\left(\boldsymbol{\beta}^0, \left(\boldsymbol{Z}_{(i)}\widetilde{\boldsymbol{V}}^{(i)}(\boldsymbol{\beta}^0)\boldsymbol{Z}_{(i)}^T\right)^{-1}\right).$$

- $$D_{\phi(a)}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}, \widehat{\boldsymbol{\beta}}_{\phi_2}^{(i)}\right) = \int_{\mathbb{R}^p} f_{\widehat{\boldsymbol{\beta}}_{\phi_2}^{(i)}}(\boldsymbol{x})\phi_{(a)}\left(\frac{f_{\widehat{\boldsymbol{\beta}}_{\phi_2}}(\boldsymbol{x})}{f_{\widehat{\boldsymbol{\beta}}_{\phi_2}^{(i)}}(\boldsymbol{x})}\right)d\boldsymbol{x}$$

  where

  $$\phi_{(a)}(x) = \left(a(a+1)\right)^{-1}\left(x^{a+1} - x - a(x-1)\right);\ a \neq 0, a \neq -1,$$
  $$\phi_{(0)}(x) = \lim_{a\to 0}\phi_{(a)}(x),\ \phi_{(-1)}(x) = \lim_{a\to -1}\phi_{(a)}(x)$$

  Therefore, the new measure to detect leverage points is given by

  $$I_a^{(i)}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right) = \frac{1}{a(a+1)}\left[\frac{Det\left(\boldsymbol{I} + a\boldsymbol{H}^{ii}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right)^{-1/2}}{Det\left(\boldsymbol{I} - \boldsymbol{H}^{ii}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)\right)^{a/2}} - 1\right] \quad a \neq 0, a \neq -1,$$

  $$I_0^{(i)}\left(\widehat{\boldsymbol{\beta}}_{\varphi_2}\right) = -\frac{1}{2}\left\{Trace\left(\boldsymbol{H}^{ii}\left(\widehat{\boldsymbol{\beta}}_{\varphi_2}\right)\right) + \log\left(Det\left(\boldsymbol{I} - \boldsymbol{H}^{ii}\left(\widehat{\boldsymbol{\beta}}_{\varphi_2}\right)\right)\right)\right\}$$

  $$\boldsymbol{I}_{-1}^{(i)}\left(\widehat{\boldsymbol{\beta}}_{\varphi_2}\right) = \frac{1}{2}Trace\left(\left(\boldsymbol{Z}\widehat{\boldsymbol{V}}\left(\widehat{\boldsymbol{\beta}}_{\varphi_2}\right)\boldsymbol{Z}^T\right)\left(\boldsymbol{Z}_{(i)}\widehat{\boldsymbol{V}}^{(i)}\left(\widehat{\boldsymbol{\beta}}_{\varphi_2}\right)\boldsymbol{Z}_{(i)}^T\right)^{-1} - \boldsymbol{I}\right) + \log\left(Det\left(\boldsymbol{M}^{ii}\left(\widehat{\boldsymbol{\beta}}_{\varphi_2}\right)\right)\right).$$

Influential points

- Cook distance generalization

$$C_{\phi_2}^{(i)} = \frac{1}{p}\left(\widehat{\boldsymbol\beta}_{\phi_2} - \widehat{\boldsymbol\beta}_{\phi_2}^{(i)}\right)^T Cov\left(\widehat{\boldsymbol\beta}_{\phi_2}\right)^{-1}\left(\widehat{\boldsymbol\beta}_{\phi_2} - \widehat{\boldsymbol\beta}_{\phi_2}^{(i)}\right)$$

with $\widehat{\boldsymbol\beta}_{\phi_2}^{(i)}$ the minumum ϕ_2 –divergence estimator of $\boldsymbol\beta$ when the i-th observation is eliminated

$$Cov\left(\widehat{\boldsymbol\beta}_{\phi_2}\right) \approx \left(\boldsymbol{Z}\widetilde{\boldsymbol{V}}(\boldsymbol\beta^0)\boldsymbol{Z}^T\right)^{-1}.$$

A particular case of $C_{\phi_2}^{(i)}$ indicate the influence of the i-th observation.

- The influence of the i-th observation in the estimated model, is given by

$$D_{\phi_1,\phi_2}^{(i)} \equiv \frac{2n}{\phi_1''(1)}\left(D_{\phi_1}\left(\boldsymbol{p}\left(\widehat{\boldsymbol\beta}_{\phi_2}\right),\boldsymbol{p}\left(\widehat{\boldsymbol\beta}_{\phi_2}^{(i)}\right)\right) + D_{\phi_1}\left(\boldsymbol{p}\left(\widehat{\boldsymbol\beta}_{\phi_2}^{(i)}\right),\boldsymbol{p}\left(\widehat{\boldsymbol\beta}_{\phi_2}\right)\right)\right).$$

For $J = 2$ and $g_1\left(\pi_1\left(\eta_i\right)\right) = \log\left(\pi_1\left(\eta_i\right)/\left(1-\pi_1\left(\eta_i\right)\right)\right) \equiv$ Binomial Logit Model

$\phi_1(x) = \phi_2(x) = x\log x - x + 1$ (Johnson(85)).

- $$D_{\phi_1,\phi_2}^{(i)} = 2pC_{\phi_2}^{(i)} + o_P(1)$$

Diagnostics Tools for the Multinomial Generalized Linear Models

Outline

- Multinomial models
- Statistical Inference for multinomial models
  Estimators and Phi-divergence Test
- Diagnostics Tools for the Multinomial Generalized Linear Models
- **Computational study for some ordinal models**

Preliminaries

- Cressie and Read (1984) divergence family

$$\phi_{(a)}(x) = (a(a+1))^{-1}(x^{a+1} - x - a(x-1));\ a \neq 0, a \neq -1,$$

$$\phi_{(0)}(x) = \lim_{a\to 0} \phi_{(a)}(x),\ \phi_{(-1)}(x) = \lim_{a\to -1} \phi_{(a)}(x),$$

- $J = 4,\ m = 2,\ N = 8,\ Y = \{1,2,3,4\},\ x_1 = \{1,0\},\ x_2 = \{1,2,3,4\}$

  $\boldsymbol{\beta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2) = \left(-2, -1.2, 0.5, 0.05, 0.1\right)^T$

- $M = 10000$

- $\mathbf{n} = (n_1, \ldots, n_N)^T \in \mathbf{N} = \{n^1, n^2, n^3, n^4, n^5, n^6, n^7, n^8, n^9\}$ with

$$\begin{cases} n^1 = (15,15,15,15,15,15,15,15), \\ n^2 = (20,20,20,20,20,20,20,20), \\ n^3 = (30,30,30,30,30,30,30,30), \\ n^4 = (50,50,50,50,50,50,50,50), \\ n^5 = (60,60,60,60,60,60,60,60), \\ n^6 = (25,25,25,25,10,10,10,10), \\ n^7 = (50,50,50,50,15,15,15,15), \\ n^8 = (60,60,30,30,20,20,15,15), \\ n^9 = (40,40,15,15,5,5,25,25). \end{cases}$$

- Models: Cumulative logit, Grouped Cox, Adjacent categories logit and sequencial logit.

Computational study of $\hat{\beta}_{\phi}$

- To compare the members of the minimum Cressie and Read divergence estimators,

$$mse_j^a = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\beta}_{j(i)}^a - \beta_j^*\right)^2, \quad j = 1, \dots 5$$

where $\beta_j^* = \alpha_j, j = 1, 2, 3, \beta_4^* = \beta_1, \beta_5^* = \beta_2$ and $\widehat{\beta}_{j(i)}^a$ is the minimum Cressie and Read divergence estimator of β_j^* in the i –th sample.

- Mean square error

$$mse_{\beta}^a = \sum_{j=1}^{5}\frac{mse_j^a}{5}.$$

a	-1/2	0	2/3	1	2	3
Cumulative Logit Model	0.1550	0.1139	0.1016	**0.0998**	0.1002	0.1038
Grouped Cox Model	0.0767	0.0518	0.0451	**0.0494**	0.0457	0.0488
Adjacent categories logits model	0.3021	0.2149	0.2045	0.1991	**0.1983**	0.2047
Sequential Logit Model	0.1312	0.0926	0.0811	0.0791	**0.0783**	0.0806

Computational Study of $T_n^{\phi_1,\phi_2}$ (I)

- $$H_0 : \boldsymbol{K}^T\beta = \xi \qquad H_{\delta,n} : \boldsymbol{K}^T\beta - \xi = n^{-1/2}\delta$$

  where $\boldsymbol{K}^T = \begin{pmatrix}0 & 0 & 0 & 0 & 1\end{pmatrix}$ and $\xi = 0$

  $\delta = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7) = (-3, -2, -1, 0, 1, 2, 3)$

  $\alpha = 0.05$

- Simulated exact size and power of $T_n^{\phi_{(a)},\phi_{(1)}}$

$$po_{\delta_s}(a) = \frac{\left(\text{Number of } T_{n,j}^{\varphi_{(a)},\varphi_{(1)}} > \chi^2_{1,1-\alpha} \mid H_{\delta_s,n}\right)}{M}, \quad s = 1, \ldots, 7$$

where $T_{n,j}^{\phi_{(a)},\phi_{(1)}}, j = 1, \ldots, M$ is the estatistic $T_n^{\phi_{(a)},\phi_{(1)}}$ for the simulated j-th sample.

- Mean Gradient relative to the size

$$g(a) = \frac{1}{po_{\delta_4}(a)} \sqrt{\frac{1}{6} \sum_{\substack{s=1 \\ s \neq 4}}^{7} \left(\frac{po_{\delta_s}(a) - po_{\delta_4}(a)}{\delta_s} \right)^2}$$

Computational Study of $T_n^{\phi_1,\phi_2}$ (II)

Cumulative logit model

g(a)	-2	-1	-1/2	0	2/3	1	2	3
n^1	44.4819	47.1860	48.6279	49.2324	50.1314	**50.4039**	**50.5173**	49.4364
n^2	23.4513	24.6677	25.4068	25.4864	**26.2709**	26.2614	**26.6013**	25.9934
n^3	27.0857	28.1178	28.2602	28.2553	**28.2828**	28.2762	**28.5545**	28.0547
n^4	37.0492	37.6535	37.7806	38.1274	**38.2978**	38.1445	**38.2426**	38.0297
n^5	39.2572	39.7275	40.1961	40.2161	**40.5701**	**40.6220**	40.5241	40.5594
n^6	21.9263	23.2113	24.1904	24.9838	**25.3208**	**25.2050**	25.2012	24.8734
n^7	29.6630	30.5987	31.0217	31.5172	31.8953	**31.9148**	**31.9419**	31.3458
n^8	27.0147	28.2998	28.4423	28.7721	**28.9977**	**28.9501**	28.6696	28.0662
n^9	19.5508	21.6672	22.3575	22.4905	**22.8260**	**22.7089**	22.4385	21.3991

Grouped Cox Model

g(a)	-2	-1	-1/2	0	2/3	1	2	3
n^1	36.0441	41.9568	44.9706	47.1426	49.8293	**50.5341**	**51.6817**	50.0850
n^2	44.0068	48.7944	51.1479	52.6469	54.5833	**55.1527**	**55.7987**	55.0878
n^3	54.3651	58.1452	60.2436	62.5046	64.5607	**65.4431**	**66.3471**	65.1180
n^4	65.8072	76.6023	82.1046	86.0702	90.9755	**92.2622**	**94.3574**	91.4423
n^5	76.8838	82.2297	85.1973	88.3949	91.3026	**92.5505**	**93.8290**	92.0907
n^6	39.8076	44.7121	47.6115	50.5469	53.2101	**53.9158**	**54.8638**	53.8001
n^7	57.0605	60.9817	62.9464	64.4001	65.7420	**66.7354**	**66.8479**	66.7232
n^8	53.8019	57.8208	59.0997	60.5480	61.5705	62.1667	**63.0607**	**62.2964**
n^9	40.8188	46.3999	48.4080	48.8193	50.0774	**50.8626**	**51.1997**	48.9823

Computational Study of $T_n^{\phi_1,\phi_2}$ (III)

Adjacents categories logits models

g(a)	-2	-1	-1/2	0	2/3	1	2	3
n^1	36.4747	43.5753	46.7695	49.1414	**51.1056**	**51.3475**	51.0639	47.2147
n^2	42.4033	46.6992	48.7976	50.7110	52.1903	**52.8812**	**52.4900**	50.4163
n^3	51.9000	56.0381	57.6921	59.6484	60.8357	**62.0119**	**62.1661**	60.4330
n^4	65.8266	69.3734	70.9338	71.2846	72.6984	**73.4429**	**73.5507**	72.0045
n^5	70.5963	74.4051	75.8464	77.2558	**78.2505**	78.1238	**78.1859**	77.4970
n^6	39.4611	45.3279	47.4061	49.5395	50.2566	**51.1361**	**51.9987**	49.3958
n^7	53.6800	59.8008	61.4509	62.9226	65.6789	**65.8613**	**66.4009**	66.2323
n^8	50.4614	54.3394	56.1293	57.6412	59.2159	**59.2718**	**59.3852**	57.8434
n^9	31.1820	34.5363	36.1894	37.6215	**38.3070**	**38.3118**	36.7530	35.2671

Sequential logít model

g(a)	-2	-1	-1/2	0	2/3	1	2	3
n^1	25.3224	27.7350	29.2740	29.5193	30.2525	**30.7483**	**31.2650**	29.8938
n^2	29.6248	31.9191	32.8167	33.7971	34.3677	**34.4347**	**34.3803**	33.7805
n^3	37.1715	38.8045	39.5891	39.5580	40.3393	**40.3453**	**40.6454**	39.8253
n^4	48.7218	50.1605	50.6723	51.0151	**51.3743**	51.3059	**51.3865**	51.0876
n^5	52.5081	54.6469	55.6535	56.0797	56.2392	**56.2529**	**56.3208**	56.0516
n^6	29.9185	31.6859	32.9666	33.4379	34.0883	**34.4622**	**34.5578**	34.0637
n^7	41.4344	43.2670	43.5818	43.9205	**43.9733**	**43.9655**	43.8901	43.7811
n^8	43.2656	44.2236	44.8142	45.2637	**46.2121**	**46.2050**	45.5023	44.9058
n^9	29.6463	32.0798	32.9404	33.2821	**34.2249**	**34.3780**	33.8714	32.3461

Computational Study of $T_n^{\phi_1,\phi_2}$ (IV)

- $$\pi_{\text{Asymptotic}} = \Pr\left(\chi_1^2(\mu) > \chi_{1,1-\alpha}^2\right),$$

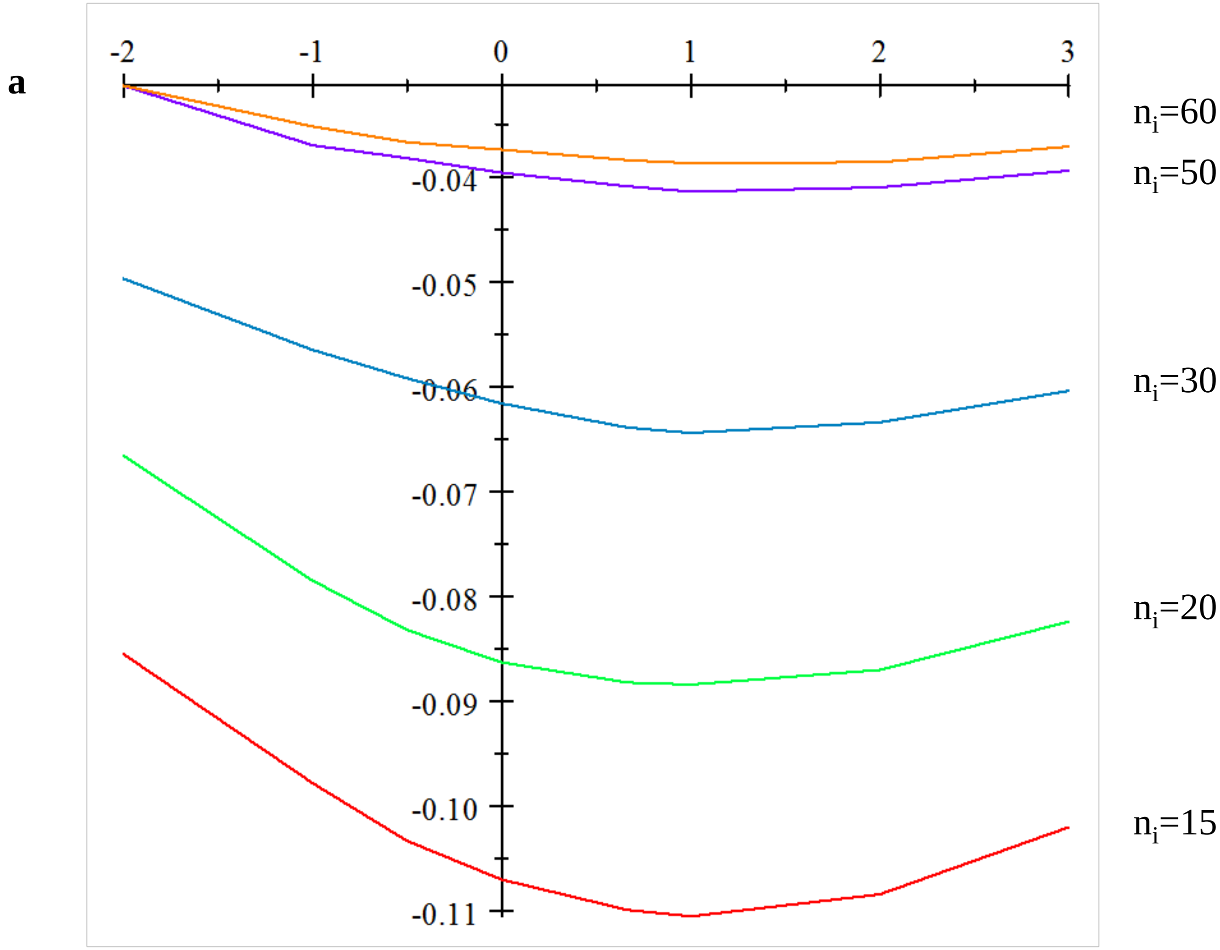
  with $\mu = \boldsymbol{\delta}^T\left(\boldsymbol{K}^T\boldsymbol{I}_{F,\lambda}(\boldsymbol{\beta}^0)^{-1}\boldsymbol{K}\right)^{-1}\boldsymbol{\delta}.$

- $\delta = \delta_1 = -3.0$

- $\mathbf{n} = (n_1,\ldots,n_N)^T \in \{n^1, n^2, n^3, n^4, n^5\}$ with
$$\begin{cases} n^1 = (15,15,15,15,15,15,15,15), \\ n^2 = (20,20,20,20,20,20,20,20), \\ n^3 = (30,30,30,30,30,30,30,30), \\ n^4 = (50,50,50,50,50,50,50,50), \\ n^5 = (60,60,60,60,60,60,60,60). \end{cases}$$

Computational Study of $T_n^{\phi_1,\phi_2}$ (V)

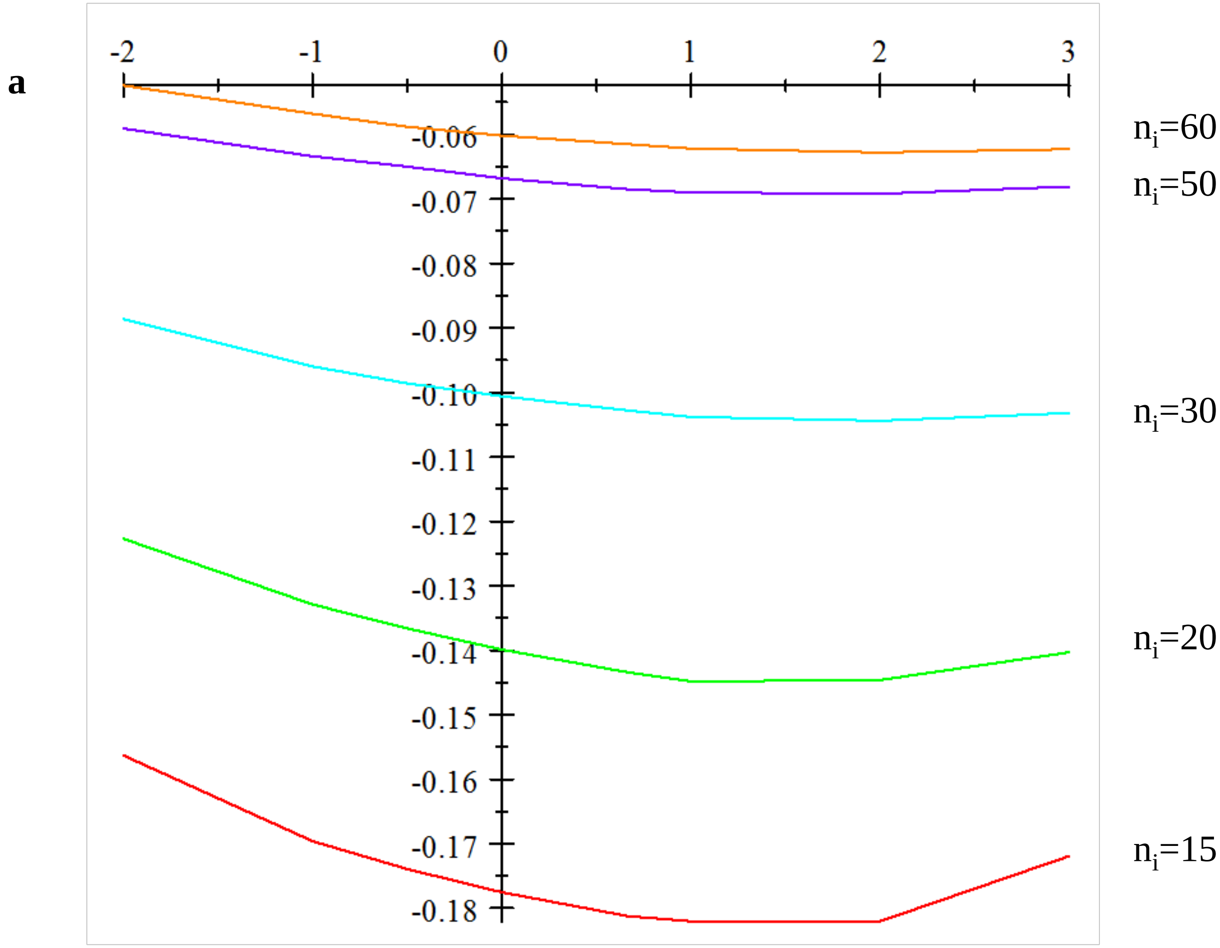
Cumulative Logit model



Exact Power – Asymptotic Power

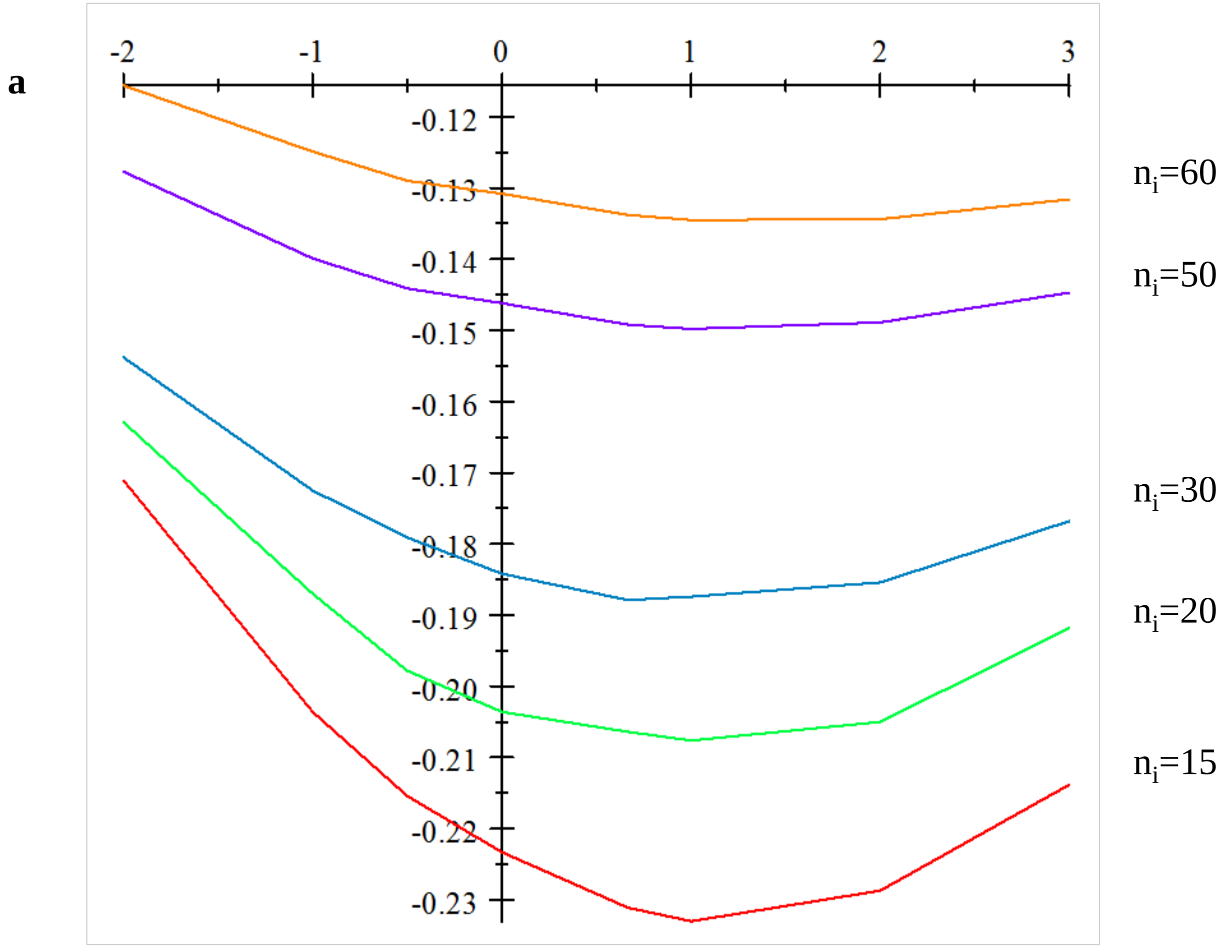
Computational Study of $T_n^{\phi_1,\phi_2}$ (VI)

Grouped Cox Model



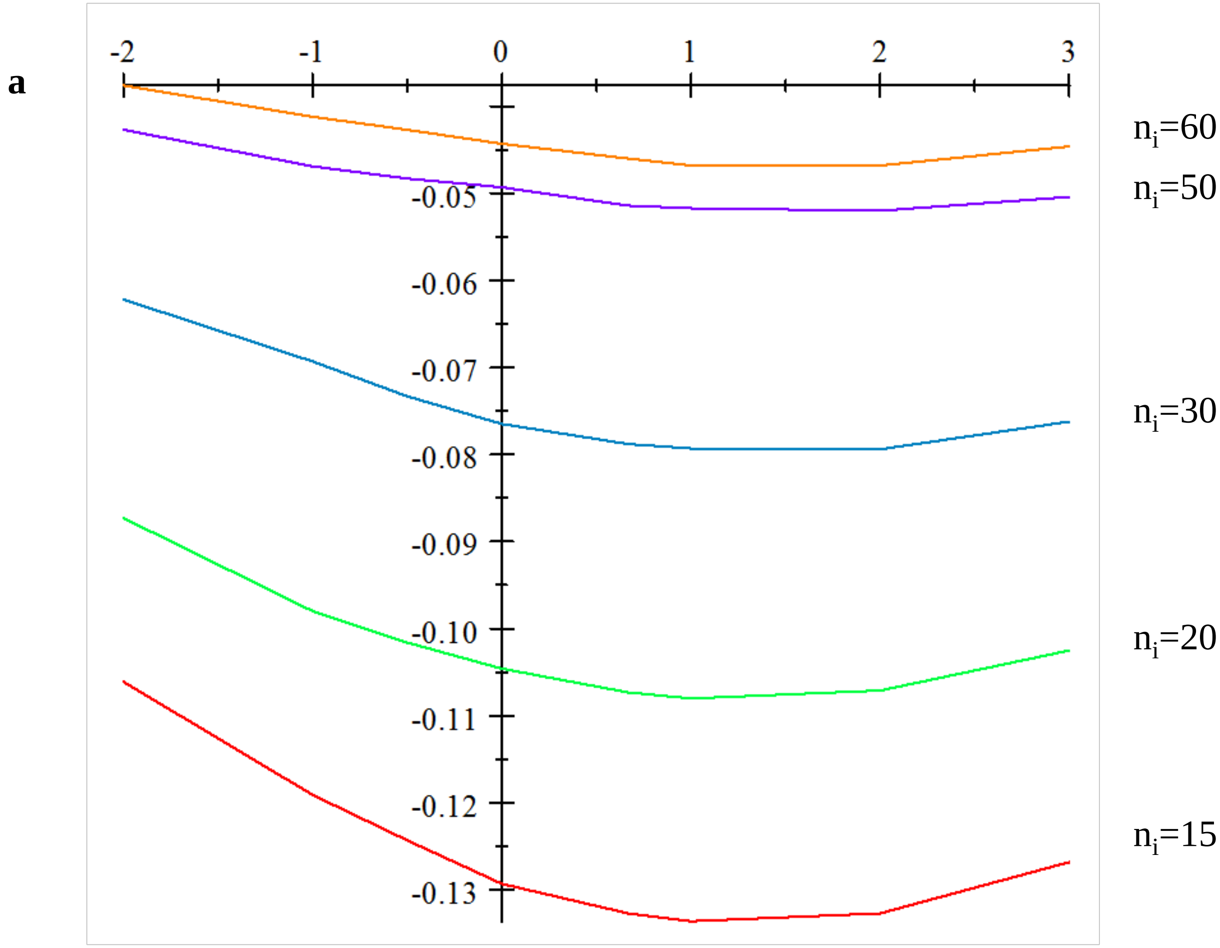
Exact Power – Asymptotic Power

Adjacent Categories Logit Model



Exact Power – Asymptotic Power

Sequential Logit Model



Exact Power – Asymptotic Power

Multinomial Generalized Linear Models

Prof. Dr. María del Carmen Pardo Llorente
mcapardo@mat.ucm.es
www.mat.ucm.es/~mcapardo

Department of Statistics and O.R. I
Faculty of Mathematics,
Complutense University of Madrid
Plaza de Ciencias, 3
Ciudad Universitaria. 28040-Madrid. SPAIN

November 2009