Phase I outlier detection in profiles
with binary data based on penalized likelihood

by Wooyoul Na, Jaehyeong Ahn

Department of Education, Konkuk University
Department of Applied Statistics, Konkuk University

jayahn0104@gmail.com

Outlier detection in profiles

Phase 1 outlier detection in profiles with binary data based on penalized likelihood

Outlier detection in profiles

Data

$$j = 1, \cdots, n$$

$$X_j = (x_{j0}, \cdots, x_{j(p-1)})^T$$

$$i = 1, \cdots, m$$

$$y_{ij} \sim BIN(N_{ij}, \pi_{ij})$$

$$N_{ij} = 1, \cdots, n * m$$

π_{ij}: Probability that the quality characteristic of interest fails in the jth experimental setting of the ith profile

Outlier detection in profiles

Design Matrix

$$X^* = \begin{pmatrix} X_j & 0 & \cdots & 0 \\ 0 & X_j & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_j \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ \vdots & 0 & \cdots & 0 \\ X_n & 0 & \cdots & 0 \\ 0 & X_1 & \cdots & 0 \\ 0 & \vdots & \cdots & 0 \\ 0 & X_n & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & X_1 \\ 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & X_n \end{pmatrix}$$

Outlier detection in profiles

Model

$$logit(\pi_{ij}) = ln\frac{\pi_{ij}}{1-\pi_{ij}} = X_j^T\beta_i$$

$$\pi_{ij} = \frac{exp(X_j^T\beta_i)}{1+exp(X_j^T\beta_i)}$$

Log-likelihood

$$\ell(\beta) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(ln\binom{N_{ij}}{y_{ij}} + y_{ij}ln\pi_{ij} + (N_{ij}-y_{ij})ln(1-\pi_{ij})\right)$$

Maximized likelihood estimator

$$\hat{\beta}_i = argmax_{(\beta)}\ell(\beta)$$

Outlier detection in profiles

Idea

$$\hat{\beta}_0 = median\left\{\hat{\beta}_1, \cdots, \hat{\beta}_m\right\}$$

$$\beta_i = \hat{\beta}_0 + \delta_i, \quad \delta_i = \beta_i - \hat{\beta}_0$$

$$\delta = (\delta_1^T \cdots \delta_m^T)^T$$

Ω: set of outliers $\subset \{1, \cdots, m\}$

$$\delta_i = \begin{cases} 0, & \text{if } i \notin \Omega \\ \delta_\Omega, & \text{if } i \in \Omega \end{cases}$$

Methodology

Original Estimator(Initial)

$$\ell(\delta) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(ln \binom{N_{ij}}{y_{ij}} + y_{ij} ln \pi_{ij} + (N_{ij} - y_{ij}) ln(1 - \pi_{ij}) \right)$$

$$\pi_{ij} = \frac{exp(X_j^T(\hat{\beta}_0 + \delta_i))}{1 + exp(X_j^T(\hat{\beta}_0 + \delta_i))}, \quad \hat{\beta}_0 + \delta_i = \beta_i$$

$$\ell(\delta) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(ln \binom{N_{ij}}{y_{ij}} + y_{ij} X_j^T(\hat{\beta}_0 + \delta_i) - N_{ij} ln(1 + exp(X_j^T(\hat{\beta}_0 + \delta_i))) \right)$$

$$\hat{\delta}^{(0)} = argmax_{\delta} \ell(\delta)$$

Methodology

1. Group-type Penalized Outlier Detection(GPOD)

(adapt Group-LASSO penalty)

Penalized loss function

$$PL(\delta) = -\ell(\delta) + P_\lambda$$

$$PL(\delta) = -\ell(\delta) + \lambda \sum_{i=1}^{m} ||\delta_i||_{K_i},$$

where $||\delta_i||_{K_i} = (\delta_i^T K_i \delta_i)^{\frac{1}{2}}$

$$\hat{\delta}_\lambda = argmin_\delta PL(\delta)$$

Optimal Lambda

$$BIC_\lambda = -\ell(\hat{\delta}_\lambda) + \tfrac{1}{2} d_\lambda ln(n),$$

$$d_\lambda = \sum_i \mathrm{I}(||\hat{\delta}_{\lambda i}|| > 0) + \sum_i \frac{||\hat{\delta}_{\lambda i}||}{||\hat{\delta}_i^{(0)}||}(p-1)$$

$$\lambda^{(*)} = argmin_{(\lambda)} BIC_\lambda$$

$$\Rightarrow \hat{\delta}_{\lambda(*)}, \quad \hat{\Omega}^* = \{i : \hat{\delta}_{\lambda(*)i} \neq 0\}$$

Methodology

Calculation

$$\frac{\partial PL(\delta)}{\partial \delta} = -{X^*}^T(y-u) + \Lambda\delta = 0$$

$$X^* = diag\{X, \cdots, X\} \text{ with m identical blocks}$$

$$X = (x_1, \cdots, x_n)^T$$

$$y = (y_1^T, \cdots, y_m^T)^T, \quad y_i = (y_{i1, \cdots, y_{in}})^T$$

$$u = (u_1^T, \cdots, u_m^T)^T, \quad u_i = (N_{i1}\pi_{i1}, \cdots, N_{in}\pi_{in})^T$$

$$\Lambda = \lambda\sqrt{p} * diag\left\{\frac{1}{||\delta_1||}\mathrm{I}_p, \cdots, \frac{1}{||\delta_m||}\mathrm{I}_p\right\}$$

$$\left({X^*}^T \hat{W} X^* + \Lambda\right)\delta = {X^*}^T \hat{W} q$$

$$\hat{\delta} = \left(X^* \hat{W} X^* + \Lambda\right)^{-1} {X^*}^T \hat{W} q$$

$$\hat{W} = diag\left\{\hat{W}_1, \cdots, \hat{W}_m\right\}$$

$$\hat{W}_i = diag\left\{N_{i1}\hat{\pi}_{i1}(1-\hat{\pi}_{i1}), \cdots, N_{in}\hat{\pi}_{in}(1-\hat{\pi}_{in})\right\}$$

$$q = \text{logit}(\pi) - X^*\hat{\beta}_0^* + \hat{W}^{-1}(y-u)$$

$$q = (q_1^T, \cdots, q_m^T)^T, \; q_i = (q_{i1}, \cdots, q_{in})^T$$

$$\pi = (\pi_1^T, \cdots, \pi_m^T), \; \pi_i = (\pi_{i1}, \cdots, \pi_{in})^T$$

$$\hat{\beta}_0^* = \left(\hat{\beta}_0^T, \cdots, \hat{\beta}_0^T\right)^T$$

Methodology

Algorithm

Step1) Obtain the maximized likelihood estimator for each profile

$$\hat{\beta}_i, (i = 1, \cdots, m), \quad \hat{\beta}_0 = median\{\hat{\beta}_i, \cdots, \hat{\beta}_m\}$$

Step2) Obtain the original estimator $\hat{\delta}_0^{(0)}$ by maximize $\ell(\delta)$

$$\hat{\delta}^{(0)} = argmax_{\delta} \ell(\delta)$$

Step3) Set the range of λ and obtain $\hat{\delta}_\lambda$ by $argmin_\delta PL(\delta)$

(a) Let $\hat{\delta}^{(0)}$ be the starting point $\hat{\delta}^0 = \hat{\delta}^{(0)}$

(b) at the kth iteration (k ≥ 0), *calculate*
$q^k, \hat{W}^k, \Lambda^k$ based on $\hat{\delta}^k$, and then update the estimation of δ

$$\hat{\delta}^{(k+1)} = \left(X^* \hat{W}^k X^* + \Lambda^k\right)^{-1} X^{*T} \hat{W}^{(k)} q^{(k)}$$

(c) Repeat step (b) until converge $\quad \frac{||\hat{\delta}^{(k)} - \hat{\delta}^{(k-1)}||_1}{||\hat{\delta}^{(k)}||_1} \leq \epsilon$, $\epsilon = 10^{-4}$

Step4) Get BIC to determine $\lambda^{(*)}$

Step5) obtain $\hat{\delta}_{\lambda^{(*)}}$ and $\hat{\Omega}^* = \{i : \hat{\delta}_{\lambda(*)i} \neq 0\}$

Methodology

2. Directional Penalized Outlier Detection(DPOD)

(adapt Adaptive-LASSO penalty)

Assumption: The shift only occurs in β_{i0}

$$\delta_{0\sim} = (\delta_{01\sim}, \cdots, \delta_{0m\sim})^T$$

$$\pi_{ij} = \frac{exp(\delta_{0i\sim} x_{j0} + X_j^T \hat{\beta}_0)}{1 + exp(\delta_{0i\sim} x_{j0} + X_j^T \hat{\beta}_0)}$$

$$\ell(\delta_{0\sim}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(ln \binom{N_{ij}}{y_{ij}} + y_{ij}(\delta_{0i\sim} x_{j0} + X_j^T \hat{\beta}_0) - \right.$$

$$\mathrm{N}_{ij}\, ln(1 + exp(\delta_{0i\sim} x_{j0} + X_j^T \hat{\beta}_0))$$

$$\hat{\delta}_{0\sim}^{(0)} = argmax_{(\delta_{0\sim})} \ell(\delta_{0\sim})$$

Methodology

Desgin Matrix

$$X^*\beta = \begin{pmatrix} 1 & 0 & \cdots & X_1 \\ \vdots & 0 & \cdots & \vdots \\ 1 & 0 & \cdots & X_n \\ 0 & 1 & \cdots & X_1 \\ 0 & \vdots & \cdots & \vdots \\ 0 & 1 & \cdots & X_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & X_1 \\ 0 & 0 & \vdots & \vdots \\ 0 & 0 & 1 & X_n \end{pmatrix} \begin{pmatrix} \delta_{01} \\ \delta_{02} \\ \vdots \\ \delta_{0m} \\ \hat{\beta}_0 \end{pmatrix}$$

Methodology

2. Directional Penalized Outlier Detection(DPOD)

(adapt Adaptive-LASSO penalty)

Penalized loss function

$$PL(\delta_{0\sim}) = -\ell(\delta_{0\sim}) + \sum_{i=1}^{m} P_{\lambda i}(|\delta_{0i\sim}|)$$

$$PL(\delta_{0\sim}) = -\ell(\delta_{0\sim}) + \lambda \sum_{i=1}^{m} \hat{w}_i|\delta_{0i\sim}|,$$

where $\hat{w} = \frac{1}{|\hat{\delta}_0^{(0)}\sim|}$

$$\hat{\delta}_{0\lambda\sim} = argmin_{\delta} PL(\delta_{0\sim})$$

Optimal Lambda

$$BIC_{\lambda} = -\ell(\hat{\delta}_{0\lambda\sim}) + \frac{1}{2} d_{\lambda\sim} ln(n),$$

$$d_{\lambda\sim} = \sum_i \mathrm{I}(||\hat{\delta}_{0\lambda i\sim}|| > 0) + \sum_i \frac{||\hat{\delta}_{0\lambda i}\sim||}{||\hat{\delta}_{0i}^{(0)}\sim||}(p-1)$$

$$\lambda^{(*)} = argmin_{(\lambda)} BIC_{\lambda}$$

$$\Rightarrow \hat{\delta}_{0\lambda^{(*)}}\sim, \quad \hat{\Omega}^*\sim = \{i : \hat{\delta}_{0\lambda^{(*)}i}\sim \neq 0\}$$

Simulation

Simulation in R!