Biometrika (2014), **xx**, x, *pp*. 1–2

Supplementary material for Variable selection in regression with compositional covariates

BY WEI LIN, PIXU SHI, RUI FENG AND HONGZHE LI

Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia 19104, U.S.A.

weilin1@mail.med.upenn.edu pixushi@mail.med.upenn.edu
ruifeng@mail.med.upenn.edu hongzhe@upenn.edu

1. ADDITIONAL SIMULATIONS

We conducted additional simulations to compare the performance of the two-step procedures formed by adding a refitting step to the proposed method and lasso (ii), respectively. The means and standard errors of performance measures are summarized in Table S1. We observe that, in the more challenging settings where the proposed method tends to miss fewer important variables than lasso (ii), it also has an advantage over lasso (ii) in terms of prediction and estimation, which is consistent with our comparisons of one-step procedures.

2. PROOFS

Proof of Proposition 1. To apply Theorem 4 of Rockafellar (1976) for the convergence of the method of multipliers, we need only verify for problem (3) that (a) Slater's condition is satisfied, and (b) there exists a constant c such that the c-sublevel set of feasible points $B_c = \{\beta\colon Q(\beta) \leq c \text{ and } \sum_{j=1}^{p} \beta_j = 0\}$ is nonempty and bounded, where $Q(\cdot)$ is the objective function in problem

Table S1. *Means and standard errors (in parentheses) of various performance measures for two two-step procedures based on* 100 *simulations*

(n, p)	Method	PE	ℓ_1 loss	ℓ_2 loss	ℓ_∞ loss	FP	FN
				$\rho = 0{\cdot}2$			
(50, 30)	Lasso (ii)	0·37 (0·01)	0·87 (0·04)	0·11 (0·01)	0·18 (0·01)	4·15 (0·28)	0·00 (0·00)
	Proposed	0·36 (0·01)	0·83 (0·04)	0·10 (0·01)	0·17 (0·01)	3·57 (0·23)	0·00 (0·00)
(100, 200)	Lasso (ii)	0·30 (0·01)	0·54 (0·03)	0·05 (0·00)	0·13 (0·00)	2·96 (0·23)	0·00 (0·00)
	Proposed	0·30 (0·00)	0·55 (0·02)	0·05 (0·00)	0·12 (0·00)	3·03 (0·24)	0·00 (0·00)
(100, 1000)	Lasso (ii)	0·34 (0·01)	0·63 (0·03)	0·10 (0·01)	0·17 (0·01)	2·84 (0·21)	0·13 (0·04)
	Proposed	0·32 (0·01)	0·60 (0·03)	0·07 (0·01)	0·14 (0·01)	3·10 (0·22)	0·04 (0·02)
				$\rho = 0{\cdot}5$			
(50, 30)	Lasso (ii)	0·38 (0·01)	1·08 (0·05)	0·16 (0·01)	0·21 (0·01)	5·00 (0·30)	0·02 (0·01)
	Proposed	0·39 (0·01)	1·08 (0·05)	0·16 (0·01)	0·21 (0·01)	4·81 (0·27)	0·02 (0·01)
(100, 200)	Lasso (ii)	0·33 (0·01)	0·76 (0·04)	0·11 (0·02)	0·17 (0·01)	4·61 (0·27)	0·09 (0·05)
	Proposed	0·31 (0·01)	0·69 (0·03)	0·07 (0·01)	0·14 (0·01)	4·60 (0·29)	0·01 (0·01)
(100, 1000)	Lasso (ii)	0·61 (0·05)	1·59 (0·10)	0·69 (0·08)	0·47 (0·03)	2·44 (0·20)	1·29 (0·13)
	Proposed	0·59 (0·07)	1·53 (0·11)	0·63 (0·10)	0·43 (0·03)	3·73 (0·29)	0·99 (0·13)

PE, prediction error; FP, number of false positives; FN, number of false negatives.

(3). Claim (a) holds since in this case Slater's condition reduces to feasibility and 0 is a feasible point. To show (b), take any $c \geq Q(0)$; then B_c is nonempty since $0 \in B_c$, and is bounded since $\|\beta\|_1 \leq c/\lambda$ for $\beta \in B_c$. Proposition 1 follows from the aforementioned result. □

Proof of Proposition 2. For $J \subset \{1, \ldots, p-1\}$, let Z^r_J denote the submatrix formed by the jth columns of Z^r with $j \in J$. Define

$$P^r = \begin{pmatrix} I_{r-1} & -1 & 0 \\ 0 & \vdots & I_{p-1-r} \\ 0 & -1 & 0 \end{pmatrix} \in \mathbb{R}^{(p-1)\times(p-1)},$$

and let $E^r \in \mathbb{R}^{(p-1)\times(p-1)}$ denote the matrix with 1s in the rth column and 0s elsewhere. Then we have $\mathrm{sgn}(\beta^*_{S_{\backslash r}}) - \mathrm{sgn}(\beta^*_r)1_{s-1} = P^r_{S_{\backslash r}S_{\backslash p}}\{\mathrm{sgn}(\beta^*_{S_{\backslash p}}) - \mathrm{sgn}(\beta^*_p)1_{s-1}\}$, $Z^r_{S_{\backslash r}} = Z^p_{S_{\backslash p}}(P^r_{S_{\backslash r}S_{\backslash p}})^{\mathrm{T}}$, and $Z^r_{S^{\mathrm{c}}} = Z^p_{S^{\mathrm{c}}} - Z^p_{S_{\backslash p}}(E^r_{S^{\mathrm{c}}S_{\backslash p}})^{\mathrm{T}}$. Furthermore,

$$\begin{aligned} C^r_{S_{\backslash r}S_{\backslash r}} &= n^{-1}(Z^r_{S_{\backslash r}})^{\mathrm{T}} Z^r_{S_{\backslash r}} = n^{-1} P^r_{S_{\backslash r}S_{\backslash p}}(Z^p_{S_{\backslash p}})^{\mathrm{T}} Z^p_{S_{\backslash p}}(P^r_{S_{\backslash r}S_{\backslash p}})^{\mathrm{T}} \\ &= P^r_{S_{\backslash r}S_{\backslash p}} C^p_{S_{\backslash p}S_{\backslash p}}(P^r_{S_{\backslash r}S_{\backslash p}})^{\mathrm{T}}, \\ C^r_{S^{\mathrm{c}}S_{\backslash r}} &= n^{-1}(Z^r_{S^{\mathrm{c}}})^{\mathrm{T}} Z^r_{S_{\backslash r}} = n^{-1}\{Z^p_{S^{\mathrm{c}}} - Z^p_{S_{\backslash p}}(E^r_{S^{\mathrm{c}}S_{\backslash p}})^{\mathrm{T}}\}^{\mathrm{T}} Z^p_{S_{\backslash p}}(P^r_{S_{\backslash r}S_{\backslash p}})^{\mathrm{T}} \\ &= n^{-1}\{(Z^p_{S^{\mathrm{c}}})^{\mathrm{T}} Z^p_{S_{\backslash p}} - E^r_{S^{\mathrm{c}}S_{\backslash p}}(Z^p_{S_{\backslash p}})^{\mathrm{T}} Z^p_{S_{\backslash p}}\}(P^r_{S_{\backslash r}S_{\backslash p}})^{\mathrm{T}} \\ &= (C^p_{S^{\mathrm{c}}S_{\backslash p}} - E^r_{S^{\mathrm{c}}S_{\backslash p}} C^p_{S_{\backslash p}S_{\backslash p}})(P^r_{S_{\backslash r}S_{\backslash p}})^{\mathrm{T}}. \end{aligned}$$

Substituting these identities into the left-hand side of (9) yields

$$\begin{aligned} & C^r_{S^{\mathrm{c}}S_{\backslash r}}(C^r_{S_{\backslash r}S_{\backslash r}})^{-1}\{\mathrm{sgn}(\beta^*_{S_{\backslash r}}) - \mathrm{sgn}(\beta^*_r)1_{s-1}\} + \mathrm{sgn}(\beta^*_r)1_{p-s} \\ &= (C^p_{S^{\mathrm{c}}S_{\backslash p}} - E^r_{S^{\mathrm{c}}S_{\backslash p}} C^p_{S_{\backslash p}S_{\backslash p}})(C^p_{S_{\backslash p}S_{\backslash p}})^{-1}\{\mathrm{sgn}(\beta^*_{S_{\backslash p}}) - \mathrm{sgn}(\beta^*_p)1_{s-1}\} + \mathrm{sgn}(\beta^*_r)1_{p-s} \\ &= C^p_{S^{\mathrm{c}}S_{\backslash p}}(C^p_{S_{\backslash p}S_{\backslash p}})^{-1}\{\mathrm{sgn}(\beta^*_{S_{\backslash p}}) - \mathrm{sgn}(\beta^*_p)1_{s-1}\} \\ &\quad - E^r_{S^{\mathrm{c}}S_{\backslash p}}\{\mathrm{sgn}(\beta^*_{S_{\backslash p}}) - \mathrm{sgn}(\beta^*_p)1_{s-1}\} + \mathrm{sgn}(\beta^*_r)1_{p-s} \\ &= C^p_{S^{\mathrm{c}}S_{\backslash p}}(C^p_{S_{\backslash p}S_{\backslash p}})^{-1}\{\mathrm{sgn}(\beta^*_{S_{\backslash p}}) - \mathrm{sgn}(\beta^*_p)1_{s-1}\} \\ &\quad - \{\mathrm{sgn}(\beta^*_r) - \mathrm{sgn}(\beta^*_p)\}1_{p-s} + \mathrm{sgn}(\beta^*_r)1_{p-s} \\ &= C^p_{S^{\mathrm{c}}S_{\backslash p}}(C^p_{S_{\backslash p}S_{\backslash p}})^{-1}\{\mathrm{sgn}(\beta^*_{S_{\backslash p}}) - \mathrm{sgn}(\beta^*_p)1_{s-1}\} + \mathrm{sgn}(\beta^*_p)1_{p-s}, \end{aligned}$$

and (10) follows similarly. □

REFERENCES

ROCKAFELLAR, R. T. (1976). Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**, 97–116.