Measure Algebras and Applications

Shrinkage estimators' properties in regression models using L_1 penalized norm

Seyed Kamran Ghoreishi[1]

1. Department of Statistics, University of Qom, Qom, Iran. Email: atty_ghoreishi@yahoo.com

Article Info

Article type:
Research Article

Article history:
Received: 15 May 2023
Received in revised form:
7 August 2023
Accepted: 13 August 2023
Published Online:
30 September 2023

Keywords:
Hierarchical models,
Shrinkage estimators,
Regression models,
High-dimensional datasets

2020 Mathematics Subject Classification:
62F15, 62J07

ABSTRACT

In this paper, we first introduce a two-level hierarchical model with a linear structure. We use the moment, maximum likelihood, and SURE estimators to obtain the regression coefficients shrinkage estimators. Since, regression models have vast applications in high-dimensional datasets, using sparsity assumption, we discuss the asymptotic properties of the regression estimators under L_2 error norm and L_1 penalty norm.

Cite this article: Ghoreishi, S.K. (2023). Shrinkage estimators' properties in regression models using L_1 penalized norm. *Measure Algebras and Applications*, 1(1), 97–106. http://doi.org/10.22091/MAA.2023.9446.1003

Publisher: University of Qom
DOI: 10.22091/MAA.2023.9446.1003

Extended Abstract

Shrinkage estimators have found vast applications in many disciplines. These estimators were first introduced by James and Stein (1961) and Stein (1962). Many statisticians have shown different characteristics of these estimators. The first version of two-level hierarchical models is called the homoscedastic hierarchical model with the following structure

$$\begin{aligned} &Y_i \sim N(\theta_i, A) \qquad i = 1, 2, \ldots, n \\ &\theta_i \sim N(\mu, \lambda). \end{aligned}$$

The second version of two-level hierarchical models, heteroscedastic hierarchical models, is as follows

$$\begin{aligned} &Y_i \sim N(\theta_i, A_i) \qquad i = 1, 2, \ldots, n \\ &\theta_i \sim N(\mu, \lambda). \end{aligned}$$

Several authors have considered discussing the asymptotic properties of the heteroscedastic hierarchical models too. Among others, we can refer to Xie et al (2012-2016), Ghoreishi and Meshkani (2014), Baranchik (1970), Cai and Zijian (2016), and Shantia and Ghoreishi (2020).

In this paper, we address the following heteroscedastic hierarchical model

$$\begin{aligned} &Y_i \sim N(\theta_i, A_i) \qquad i = 1, 2, \ldots, n \\ &\theta_i \sim N(\mu, \lambda), \\ &\theta_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} \end{aligned}$$

with a linear structure for θ_is. We will use the moment, maximum likelihood, and SURE methods to obtain the shrinkage estimate

$$\hat{\theta}_i = \frac{\hat{\lambda}}{\hat{\lambda} + A_i} Y_i + \frac{A_i}{\hat{\lambda} + A_i} \hat{\mu}$$

and also to estimate the regression coefficients $\beta_1, \cdots, \beta_p$.

A common assumption for high-dimensional regression models, $p = O(n)$, is the sparsity assumption for the vector of the regression coefficient. That is,

$$card(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_0 = s_0 \ll p,$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$.

Assuming the sparsity, the asymptotic properties of the resulting estimators will be investigated under the error function L_2 and the penalized function L_1 given by

$$\arg \min_{\boldsymbol{\beta} \in \mathbb{R}^P} \left\{ \frac{1}{2n} \|W^{1/2}(\hat{\boldsymbol{\theta}} - \mathbf{X}\boldsymbol{\beta})\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{0.1}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n)^T$, $\mathbf{X}$ is the regression matrix, and $\mathbf{W} = diag\left(\frac{\hat{\lambda}+A_1}{\hat{\lambda}A_1}, \cdots, \frac{\hat{\lambda}+A_n}{\hat{\lambda}A_n} \right)$.

Assuming that $\hat{\boldsymbol{\beta}}$ is the solution of (0.1) and $\boldsymbol{\beta}^*$ is the real vector, we will construct the L_2-upper bound

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \le \frac{4}{\gamma} \sqrt{\max_j \{ \frac{\lambda + A_j}{\lambda A_j} \} \frac{s_0 \ln p}{n}},$$

for $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$.

جبرهای اندازه و کاربردها
سال ۱۴۰۲، دوره ۱، شماره ۱، ص ۱۰۶-۹۷

خواص مجانبی برآوردهای انقباضی در مدل‌های رگرسیونی با استفاده از تابع تاوانیده با نرم L_1

سید کامران قریشی [۱]

۱. گروه آمار، دانشکده علوم پایه، دانشگاه قم، قم، ایران. رایانامه: atty_ghoreishi@yahoo.com

اطلاعات مقاله

نوع مقاله:
مقالهٔ پژوهشی

تاریخ دریافت: ۱۴۰۲/۲/۲۵
تاریخ بازنگری: ۱۴۰۲/۵/۱۶
تاریخ پذیرش: ۱۴۰۲/۵/۲۲
تاریخ انتشار: ۱۴۰۲/۷/۸

کلمات کلیدی:
مدل‌های سلسله‌مراتبی،
برآوردهای انقباضی،
مدل‌های رگرسیونی،
داده‌های با ابعاد بزرگ

رده‌بندی ریاضی:
62F15, 62J07

چکیده

در این مقاله، ابتدا یک مدل دوسطحی با ساختار خطی معرفی خواهیم کرد. از برآوردهای گشتاوری، ماکسیمم درست‌نمایی و $SURE$، برای برآورد انقباضی پارامترهای مدل و برآورد پارامترهای مدل رگرسیونی استفاده خواهیم کرد. به دلیل اینکه مدل‌های رگرسیونی کاربرد فراوانی برای تحلیل داده‌های با ابعاد بزرگ دارند، لذا با فرض تُنک بودن بردار ضرایب مدل رگرسیونی، خواص مجانبی برآوردکننده‌های حاصل تحت تابع خطای L_2 و تابع تاوانیده با نرم L_1 بررسی خواهند شد.

استناد: قریشی، سید کامران. (۱۴۰۲). خواص مجانبی برآوردهای انقباضی در مدل‌های رگرسیونی با استفاده از تابع تاوانیده با نرم L_1. جبرهای اندازه و کاربردها، ۱(۱)، ۱۰۶-۹۷.

http://doi.org/10.22091/MAA.2023.9446.1003

ناشر: دانشگاه قم.

۱ مقدمه

امروزه برآوردهای انقباضی کاربرد وسیعی در حوزه‌های مختلف تحلیل داده‌ها دارند. این برآوردها ابتدا توسط جیمز و اشتاین (۱۹۶۱) معرفی شدند. اشتاین (۱۹۶۲) از این برآوردها برای تخمین هم‌زمان میانگین چند جامعهٔ نرمال استفاده نمود. به‌این‌ترتیب برآوردهای انقباضی مبنایی برای گسترش مدل‌های سلسله‌مراتبی با توزیع نرمال شدند. پس از معرفی برآوردهای چروکیده، مطالعهٔ خواص آنها همواره مورد توجه دانشمندان بوده است. تاکنون آمارشناسان مختلفی سعی نموده‌اند تا خواص بیزی، خواص بیز تجربی، و خواص تابع مخاطره برآوردهای انقباضی را در مدل‌های سلسله‌مراتبی همگن نرمال که به‌صورت زیر تعریف می‌شوند، بررسی کنند

$$\begin{aligned} &Y_i \sim N(\theta_i, A) \qquad i = ۱, ۲, \ldots, n \\ &\theta_i \sim N(\mu, \lambda). \end{aligned} \tag{۱.۱}$$

شی و همکاران (۲۰۱۲ و ۲۰۱۶) مدل‌های همگن فوق را به مدل‌های سلسله‌مراتبی ناهمگن دوسطحی

$$\begin{aligned} &Y_i \sim N(\theta_i, A_i) \qquad i = ۱, ۲, \ldots, n \\ &\theta_i \sim N(\mu, \lambda), \end{aligned} \tag{۲.۱}$$

تعمیم و خواص مجانبی مخاطره برآوردکننده‌ها را مورد بررسی قرار دادند، که در آن واریانس‌های A_i کمیت‌های معلوم ولی نابرابر هستند. مدل‌های سلسله‌مراتبی دوسطحی (۲.۱) توسط قریشی و مشکانی (۲۰۱۴) به مدل‌های سلسله‌مراتبی ناهمگن با ساختار تابعی برای واریانس‌های هر دو سطح مدل تعمیم و خواص مجانبی آنها را تحقیق نمودند. به‌منظور استفادهٔ مناسب‌تر از برآوردهای انقباضی در تحلیل داده‌های آماری، طی سال‌های اخیر، آمارشناسان صورت‌های مختلف از مدل‌های سلسله‌مراتبی (۲.۱) را توسعه و خواص برآوردهای متناظر را بررسی نموده‌اند که از آن جمله می‌توان به برانچیک (۱۹۷۰)، کاتی و زنچینگ (۲۰۱۶)، و شنتیا و قریشی (۲۰۲۰) اشاره نمود.

در این مقاله، به برآوردهای مدل رگرسیونی چندگانه $\theta_i = \beta_۱ X_{۱i} + \beta_۲ X_{۲i} + \cdots + \beta_p X_{pi}$ تحت مدل دوسطحی با ساختار خطی

$$\begin{aligned} &Y_i \sim N(\theta_i, A_i) \qquad i = ۱, ۲, \ldots, n \\ &\theta_i \sim N(\mu, \lambda), \\ &\theta_i = \beta_۱ X_{۱i} + \beta_۲ X_{۲i} + \cdots + \beta_p X_{pi} \end{aligned} \tag{۳.۱}$$

خواهیم پرداخت. از برآورد انقباضی θ_i برای برآورد پارامترهای رگرسیونی β_j استفاده خواهیم کرد. خواص این برآوردکننده‌ها با شرط $p = O(n)$ و فرض مدل رگرسیونی تُنک به دست خواهند آمد.

ساختار این مقاله به شرح زیر است:

در بخش ۲ روش‌های برآوردهای بیز تجربی ابرپارامترهای μ و λ را مرور خواهیم کرد. بخش ۳ به بررسی خواص برآوردهای انقباضی مدل رگرسیونی تُنک

$$\theta_i = \beta_۱ X_{۱i} + \beta_۲ X_{۲i} + \cdots + \beta_p X_{pi}$$

با شرط $p = O(n)$ اختصاص دارد. برای یافتن برآورد پارامترهای مدل رگرسیون تُنک بالا، از تابع تاوان محدب

$$p_\lambda(\beta_۱, \cdots, \beta_p) = \sum_{j=۱}^{p} |\beta_j| = \|\boldsymbol{\beta}\|_۱$$

استفاده خواهیم کرد که در آن $\boldsymbol{\beta} = (\beta_۱, \cdots, \beta_p)^T$.

۲ برآوردهای بیز تجربی ابرپارامترهای μ و λ

طبق مدل مفروض (۳.۱) و با استفاده از قضیهٔ بیز، چگالی حاشیه‌ای Y_i و چگالی پسین θ_i به‌صورت زیر به دست می‌آیند

$$Y_i \text{ توزیع حاشیه‌ای } : Y_i \sim N(\mu, \lambda + A_i)$$

$$\theta_i \text{ توزیع شرطی } : \theta_i | Y_i \sim N\left(\frac{\lambda}{\lambda + A_i} Y_i + \frac{A_i}{\lambda + A_i}\mu,\ \frac{\lambda A_i}{\lambda + A_i}\right)$$

یکی از برآوردهای مناسب بیزی برای پارامتر θ_i میانگین چگالی پسین

$$\hat{\theta}_i = \frac{\lambda}{\lambda + A_i} Y_i + \frac{A_i}{\lambda + A_i} \mu$$

است. در تحلیل بیز تجربی روش‌های مختلفی برای برآورد ابرپارامترهای μ و λ وجود دارند که در ادامه به طور مختصر به ۳ روش پرکاربرد اشاره می‌کنیم. برای آشنایی با روش‌های برآورد بیز تجربی به افرون و موریس (۱۹۷۳)، برگر (۱۹۷۶-۱۹۹۶)، موریس (۱۹۸۳) مراجعه نمایید.

۱. روش گشتاوری: برآوردهای بیز تجربی ابرپارامترهای μ و λ در روش گشتاوری به‌صورت زیر داده می‌شوند

$$\hat{\mu}^M = \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \quad , \hat{\lambda}^M = \left(\frac{\sum_{i=1}^{n} \left((Y_i - \bar{Y})^2 - A_i \right)}{n} \right)_+$$

که در آن a_+ برابر a است اگر $a > 0$ و برابر صفر است اگر $a < 0$. در این حالت برآورد انقباضی θ_i برابر خواهد بود با

$$\hat{\theta}_i^M = \frac{\hat{\lambda}^M}{\hat{\lambda}^M + A_i} Y_i + \frac{A_i}{\hat{\lambda}^M + A_i} \bar{Y}. \tag{1.2}$$

۲. روش ماکسیمم درست‌نمایی: برآوردهای ماکسیمم درست‌نمایی ابرپارامترهای μ و λ از طریق حل دستگاه معادلات زیر به دست می‌آیند

$$\begin{cases} \sum_{i=1}^{n} \frac{1}{\lambda + A_i} (Y_i - \mu) = 0, \\ \sum_{i=1}^{n} \left(\frac{(Y_i - \mu)^2}{(\lambda + A_i)^2} - \frac{1}{\lambda + A_i} \right) = 0. \end{cases}$$

درصورتی‌که دستگاه معادلات فوق دارای جواب باشد برآورد انقباضی θ_i برابر خواهد بود با

$$\hat{\theta}_i^L = \frac{\hat{\lambda}^L}{\hat{\lambda}^L + A_i} Y_i + \frac{A_i}{\hat{\lambda}^L + A_i} \hat{\mu}^L. \tag{2.2}$$

۳. روش $SURE$: برآوردهای $SURE$ ابرپارامترهای μ و λ بر اساس تابع ضرر

$$l_n(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^{n} (\theta_i - \hat{\theta}_i)^2$$

به دست می‌آیند. برای این منظور ابتدا لازم است برآورد نااریب تابع مخاطرهٔ متناظر $l_n(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ یعنی

$$R_n(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^{n} \frac{A_i}{(\lambda + A_i)^2} \left(A_i (\theta_i - \mu)^2 + \lambda^2 \right),$$

را که به‌صورت

$$SURE(\mu, \lambda) = \frac{1}{p} \sum_{i=1}^{p} \frac{A_i}{(\lambda + A_i)^2} \left(A_i (Y_i - \mu)^2 + \lambda^2 - A_i^2 \right)$$

است، به دست آوریم. برآوردهای $SURE$ ابرپارامترهای μ و λ با مینیمم کردن تابع $SURE(\mu, \lambda)$ به دست می‌آیند. به عبارت دیگر داریم

$$(\hat{\mu}^S, \hat{\lambda}^S) = \underset{\mu,\ \lambda > 0}{\arg\min}\ SURE(\mu, \lambda),$$

و در نتیجه برآورد $SURE$ پارامتر θ_i برابر خواهد بود با

$$\hat{\theta}_i^S = \frac{\hat{\lambda}^S}{\hat{\lambda}^S + A_i} Y_i + \frac{A_i}{\hat{\lambda}^S + A_i} \hat{\mu}^S. \tag{3.2}$$

براساس یافته‌های شی و همکاران (۲۰۱۲) برآوردهای $SURE$ دارای خاصیت مینیمم مخاطره هستند؛ لذا در عمل اقبال بیشتری برای استفاده از این نوع برآوردها وجود دارد، هر چند نباید عملاً از خواص مطلوب برآوردکننده‌های گشتاوری و ماکسیمم درست‌نمایی نیز غافل بود. در این مقاله، از هر سه نوع برآوردکننده برای بررسی خواص پارامترهای مدل رگرسیونی $\theta_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}$ بهره خواهیم جست.

۳ خواص برآوردهای انقباضی پارامترهای مدل رگرسیونی

برای بررسی خواص برآوردکننده‌های رگرسیونی چندگانهٔ انقباضی، مجدداً مدل دوسطحی با ساختار خطی (۳.۱) را در نظر بگیرید. نمایش ماتریسی این مدل به‌صورت

$$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} \tag{۱.۳}$$

است، که در آن $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n)^T$، $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$ و $\mathbf{X}$ ماتریس متغیرهای رگرسیونی است.

برای برآورد پارامترهای $\boldsymbol{\beta}$ در مدل رگرسیونی (۱.۳) کافی است یکی از برآوردهای (۱.۲)-(۳.۲) را در معادله (۱.۳) جای‌گذاری و سپس بردار پارامتر $\boldsymbol{\beta}$ را برآورد نمود. در صورت ثابت بودن p، برآورد موزون بردار پارامتر $\boldsymbol{\beta}$ به‌صورت

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\hat{\boldsymbol{\theta}}, \tag{۲.۳}$$

به دست می‌آید که در آن ماتریس قطری $\mathbf{W}$ شامل وزن‌ها و به‌صورت $\mathbf{W} = diag\left(\frac{\hat{\lambda}+A_1}{\hat{\lambda}A_1}, \cdots, \frac{\hat{\lambda}+A_n}{\hat{\lambda}A_n}\right)$ داده می‌شود. به دلیل این که موضوع بحث این مقاله به بررسی خواص این نوع برآوردکننده‌ها اختصاص ندارد؛ لذا در اینجا از پرداختن به آن خودداری می‌کنیم.

فرض تُنک بودن در مدل‌های رگرسیونی با ابعاد بزرگ به این معنی است که تعداد زیادی از متغیرهای مستقل رگرسیونی هیچ تأثیری بر متغیر پاسخ نداشته و به عبارتی ضرایب آنها برابر صفر است؛ بنابراین مدل رگرسیونی $\hat{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\beta}$ تُنک است هرگاه داشته باشیم

$$card(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_0 = s_0 \ll p.$$

فرض کنید می‌خواهیم بردار تُنک $\hat{\boldsymbol{\beta}}$ را در مدل رگرسیون موزون (۱.۳) چنان بیابیم که

$$\begin{cases} \hat{\boldsymbol{\beta}} \in \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \\ \\ \mathbf{W}^{1/2}\hat{\boldsymbol{\theta}} = \mathbf{W}^{1/2}\mathbf{X}\hat{\beta} \end{cases} \tag{۳.۳}$$

که در آن رابطهٔ رگرسیونی $\mathbf{W}^{1/2}\hat{\boldsymbol{\theta}} = \mathbf{W}^{1/2}\mathbf{X}\hat{\beta}$ پس از اعمال وزن‌ها $\mathbf{W}^{1/2}$ از رابطهٔ $\hat{\boldsymbol{\theta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ حاصل می‌شود.

همان‌طور که از رابطه (۳.۳) به دست می‌آید، یافتن $\hat{\boldsymbol{\beta}}$ معادل حل یک مسئلهٔ خطی است که با الگوریتم‌های شناخته‌شده قابل انجام است.

به دلیل آنکه در مدل‌های رگرسیونی با ابعاد بزرگ، ماتریس رگرسیونی $\mathbf{X}_{n\times p}$ پررتبه ستونی نیست؛ لذا شرط وجود جواب آن است که $\mathbf{W}^{1/2}\mathbf{X}$ در خاصیت فضای صفر مقید (RN) صدق کند. این فضا به‌صورت زیر تعریف می‌شود.

تعریف ۱.۳. برای $S \subset \{1, 2, \ldots, p\}$، ماتریس $\mathbf{Z} \in \mathbb{R}^{n\times p}$ نسبت به S در خاصیت فضای صفر مقید صدق می‌کند هرگاه داشته باشیم

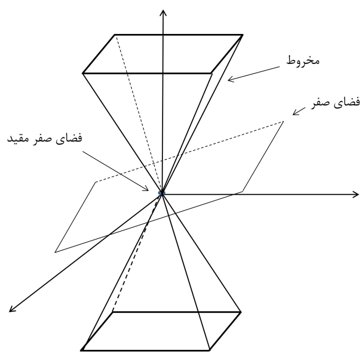
$$RN(S) = \left\{\boldsymbol{\Delta} \in \mathbb{R}^p \middle| \mathbf{Z}\boldsymbol{\Delta} = \mathbf{0}\right\} \cap \left\{\boldsymbol{\Delta} \in \mathbb{R}^{\mathbf{p}} \middle| \|\boldsymbol{\Delta}_{\mathbf{S}^c}\|_{\mathbf{1}} \leq \|\boldsymbol{\Delta}_{\mathbf{S}}\|_{\mathbf{1}}\right\} = \left\{\mathbf{0}\right\}.$$

مجموعهٔ $\left\{\boldsymbol{\Delta} \in \mathbb{R}^p \middle| \mathbf{Z}\boldsymbol{\Delta} = \mathbf{0}\right\}$ را با $N(\mathbf{Z})$ و مجموعهٔ $\left\{\boldsymbol{\Delta} \in \mathbb{R}^p \middle| \|\boldsymbol{\Delta}_{S^c}\|_1 \leq \|\boldsymbol{\Delta}_S\|_1\right\}$ را با $C(S)$ نمایش می‌دهیم و به آن مجموعهٔ مخروطی گوییم. یادآوری می‌کنیم که S^c مجموعهٔ متمم مجموعهٔ S است. لازم به توضیح است که از لحاظ هندسی یک مجموعهٔ مخروطی در فضای سه‌بعدی دارای نمایش هندسی

$$C(S) = \left\{(\Delta_1, \Delta_2, \Delta_3)' \in \mathbb{R}^3 \middle| |\Delta_1| + |\Delta_2| \leq |\Delta_3|\right\}$$

است که مطابق نمودار (۱)، $RN(S)$ در واقع اشتراک مجموعهٔ مخروطی با صفحه $N(\mathbf{Z})$ است که از مبدأ مختصات می‌گذرد. با توجه به این تعریف، برای (۳.۳) گزارهٔ زیر برقرار است:

گزاره ۲.۳. برآورد $\hat{\beta}$ که با رابطه (۳.۳) داده می‌شود جواب یکتا در بین تمام بردارهای تُنک با بُعد $|S| = s_0$ خواهد بود هرگاه داشته باشیم $\mathbf{W}^{1/2}\mathbf{X} \in RN(S)$.



شکل ۱: نمودار مخروط و فضای صفر در فضای سه‌بعدی

اثبات. برای مدل رگرسیون موزون (۳.۳) فرض کنید β^* مقدار واقعی و $\hat{\boldsymbol{\beta}}$ برداری است که از رابطه (۳.۳) به دست می‌آید. حال تعریف می‌کنیم

$$\hat{\boldsymbol{\Delta}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \tag{۴.۳}$$

در این صورت داریم

$$\mathbf{W}^{1/2}\mathbf{X}\hat{\boldsymbol{\Delta}} = \mathbf{W}^{1/2}\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = \mathbf{W}^{1/2}\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta}^* = \mathbf{0},$$

که نشان می‌دهد $\hat{\boldsymbol{\Delta}} \in N(\mathbf{W}^{1/2}\mathbf{X})$. برای اینکه نشان دهیم $\hat{\boldsymbol{\Delta}} \in C(S)$، از شرط تُنک بودن بردار $\boldsymbol{\beta}^*$ و رابطه (۳.۳) داریم

$$\|\hat{\boldsymbol{\beta}}\|_1 \leq \|\boldsymbol{\beta}^*\|_1 = \|\boldsymbol{\beta}^*_S\|_1. \tag{۵.۳}$$

در اینجا دقت داریم که بردار تُنک $\boldsymbol{\beta}^*$ دارای s_0 عضو غیرصفر و $p - s_0$ عضو صفر است. همچنین $\boldsymbol{\beta}^*_S$ شامل همهٔ مؤلفه‌های غیرصفر بردار $\boldsymbol{\beta}^*$ است. در نتیجه از رابطه (۵.۳) داریم

$$\|\hat{\boldsymbol{\beta}}\|_1 = \|\boldsymbol{\beta}^* + \hat{\boldsymbol{\Delta}}\|_1 = \|\boldsymbol{\beta}^*_S + \hat{\boldsymbol{\Delta}}_S\|_1 + \|\hat{\boldsymbol{\Delta}}_{S^c}\|_1. \tag{۶.۳}$$

علاوه‌بر این از خاصیت مثلثی نرم‌ها داریم

$$\|\boldsymbol{\beta}^*_S + \hat{\boldsymbol{\Delta}}_S\|_1 + \|\hat{\boldsymbol{\Delta}}_{S^c}\|_1 \geq \|\boldsymbol{\beta}^*_S\|_1 - \|\hat{\boldsymbol{\Delta}}_S\|_1 + \|\hat{\boldsymbol{\Delta}}_{S^c}\|_1, \tag{۷.۳}$$

و در نتیجه از روابط (۵.۳)-(۷.۳) خواهیم داشت $\|\hat{\boldsymbol{\Delta}}_{S^c}\|_1 \leq \|\hat{\boldsymbol{\Delta}}_S\|_1$ که نتیجه می‌دهد $\hat{\boldsymbol{\Delta}} \in C(S)$.

بنابر نتایج فوق داریم

$$\hat{\boldsymbol{\Delta}} \in N(\mathbf{W}^{1/2}\mathbf{X}) \cap C(S) \Longrightarrow \hat{\boldsymbol{\Delta}} = 0 \Longrightarrow \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^*.$$

رابطهٔ بالا نشان می‌دهد که الگوریتم تکراری (۳.۳) همواره منجر به جواب دقیق رگرسیون بدون خطای $\mathbf{W}^{1/2}\hat{\boldsymbol{\theta}} = \mathbf{W}^{1/2}\mathbf{X}\hat{\boldsymbol{\beta}}$ تحت تابع جریمه با نرم L_1 می‌شود، بولمن (۲۰۱۳) و بولمن و وان دی‌گیر (۲۰۱۱).

□

همان‌طور که تاکنون گفته شده شرط وجود یکتایی جواب برای معادلهٔ رگرسیون بدون خطای (۳.۳) آن است که مجموعهٔ $RN(S)$ شامل تک عضو صفر باشد. در حالت کلی برای بررسی خاصیت RN در مدل رگرسیونی تُنک (۱.۳) روش مستقیم و بسیار کارآمد وجود دارد که بر گزارهٔ زیر استوار است:

گزاره ۳.۳. برای ماتریس گوسی یا زیرگوسی $\mathbf{W}^{1/2}\mathbf{X} \in \mathbb{R}^{n\times p}$ با سطرهای $i.i.d.$ و با ماتریس کوواریانس $\mathbf{\Sigma}$ که در آن $\kappa^2 = \max_j \mathbf{\Sigma}_{jj}$ و برای هر بردار غیرصفر $\boldsymbol{\beta} \in \mathbb{R}^P$ داریم

$$\frac{\|\mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta}\|_2^2}{n} \geq c_1\|\mathbf{\Sigma}^{1/2}\boldsymbol{\beta}\|_2^2 - c_2\kappa^2\frac{\log\left(ep(\frac{\|\boldsymbol{\beta}\|_2^2}{\|\boldsymbol{\beta}\|_1^2})^2\right)}{n}\|\boldsymbol{\beta}\|_1^2 \tag{۸.۳}$$

که در آن c_1، c_2 و c_3 ثابت‌های حقیقی مقدار هستند و نامساوی فوق با احتمال حداقل $1 - 2e^{-c_3 n}$ برقرار است.

اثبات. براساس سکوتی و یوو (۲۰۱۰) و رودلسون و ژائوو (۲۰۱۲) اثبات واضح است. □

برای تحقیق در خصوص چگونگی استفاده از نامساوی (۸.۳) برای بررسی خاصیت RN، با فرض تُنک بودن بردار $\mathbf{\Delta}$ داریم

$$\mathbf{\Delta} \in C(S) = \left\{\mathbf{\Delta} \in \mathbb{R}^p;\ \|\mathbf{\Delta}_{S^c}\|_1 \leq \|\mathbf{\Delta}_S\|_1\right\}.$$

برای یافتن شرطی که بتواند خاصیت $\hat{\mathbf{\Delta}} \in N(\mathbf{W}^{1/2}\mathbf{X})$ را تضمین کند از تساوی $\|\mathbf{\Delta}\|_1 = \|\mathbf{\Delta}_{S^c}\|_1 + \|\mathbf{\Delta}_S\|_1$، خاصیت تُنک بودن بردار پارامترهای مدل رگرسیونی و نامساوی کشی-شوارتز داریم

$$\|\mathbf{\Delta}\|_1 \leq 2\|\mathbf{\Delta}_S\|_1 \leq 2\sqrt{s_0}\|\mathbf{\Delta}\|_2. \tag{۹.۳}$$

همچنین داریم

$$\|\mathbf{\Sigma}^{1/2}\boldsymbol{\beta}\|_2^2 = \boldsymbol{\beta}^T\mathbf{\Sigma}\boldsymbol{\beta} \geq \lambda_{\min}(\mathbf{\Sigma})\|\boldsymbol{\beta}\|_2^2,$$

که در آن $\lambda_{\min}(\mathbf{\Sigma})$ کوچک‌ترین مقدارویژهٔ ماتریس $\mathbf{\Sigma}$ است. اکنون با استفاده از نامساوی‌های (۸.۳) و (۹.۳) و شرط $c_1 = c_2 = c$ می‌توان نتیجه گرفت

$$\frac{\|\mathbf{W}^{1/2}\mathbf{X}\mathbf{\Delta}\|_2^2}{n} \geq c\left(\lambda_{\min}(\mathbf{\Sigma})\|\mathbf{\Delta}\|_2^2 - 4\kappa^2\frac{s_0\log p}{n}\|\mathbf{\Delta}\|_2^2\right) \geq c\|\mathbf{\Delta}\|_2^2\left(\lambda_{\min}(\mathbf{\Sigma}) - 4\kappa^2\frac{s_0\log p}{n}\right).$$

با تعریف $\gamma = c\big(\lambda_{\min}(\Sigma) - 4\kappa^2\frac{s_0\log p}{n}\big)$، نامساوی بالا به‌صورت زیر خواهد بود

$$\frac{\|\mathbf{W}^{1/2}\mathbf{X}\mathbf{\Delta}\|_2^2}{n} \geq \gamma\|\mathbf{\Delta}\|_2^2 \tag{۱۰.۳}$$

براساس رابطه (۱۰.۳)، برای ثابت مثبت c، مقدارویژهٔ مقید (RE) با شرط $n > cs_0\log p$ به دست می‌آید.
اکنون با استفاده از رابطه (۱۰.۳) می‌توان دقت برآوردکنندهٔ $\hat{\boldsymbol{\beta}}$ برای مقدار واقعی $\boldsymbol{\beta}^*$، که با رابطهٔ

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^P}\left\{\frac{1}{2n}\|\mathbf{W}^{1/2}(\hat{\boldsymbol{\theta}} - \mathbf{X}\boldsymbol{\beta})\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\right\} \tag{۱۱.۳}$$

داده می‌شود را ارزیابی نمود.

قضیه ۴.۳. برای نمونهٔ به اندازهٔ کافی بزرگ، برآوردکنندهٔ $\hat{\boldsymbol{\beta}}$ که با رابطه (۱۱.۳) به دست می‌آید تحت بعضی شرایط برای عناصر ماتریس $\mathbf{X}$ در L_2 به مقدار واقعی $\boldsymbol{\beta}^*$ همگراست و داریم

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq \frac{4}{\gamma}\sqrt{\max_j\left\{\frac{\lambda + A_j}{\lambda A_j}\right\}\frac{s_0\ln p}{n}}.$$

اثبات. طبق رابطه (۱۱.۳) داریم

$$\frac{1}{2n}\|\mathbf{W}^{1/2}(\hat{\boldsymbol{\theta}} - \mathbf{X}\hat{\boldsymbol{\beta}})\|_2^2 \leq \frac{1}{2n}\|\mathbf{W}^{1/2}(\hat{\boldsymbol{\theta}} - \mathbf{X}\boldsymbol{\beta}^*)\|_2^2.$$

با تعریف $\hat{\boldsymbol{\Delta}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$ و ساده کردن عبارت بالا، نامساوی زیر به دست می‌آید

$$\frac{1}{n}\|\mathbf{W}^{1/2}\mathbf{X}\hat{\boldsymbol{\Delta}}\|_2^2 \leq \frac{2}{n}\langle\hat{\boldsymbol{\Delta}}, (\mathbf{W}\mathbf{X})^{\mathbf{T}}\mathbf{U}\rangle$$

که در آن $\mathbf{U} = \hat{\boldsymbol{\theta}} - \mathbf{X}\hat{\boldsymbol{\beta}}$.

اکنون از رابطه (۳.۱۰) و استفاده از نامساوی هلدر داریم

$$\gamma\|\hat{\boldsymbol{\Delta}}\|_2^2 \leq \frac{1}{n}\|\mathbf{W}^{1/2}\mathbf{X}\hat{\boldsymbol{\Delta}}\|_2^2 \leq \frac{2}{n}\langle\hat{\boldsymbol{\Delta}}, (\mathbf{W}\mathbf{X})^{\mathbf{T}}\mathbf{U}\rangle \leq 2\|\hat{\boldsymbol{\Delta}}\|_1\left\|\frac{(\mathbf{W}\mathbf{X})^{\mathbf{T}}\mathbf{U}}{n}\right\|_\infty.$$

حال اگر در نامساوی بالا از $\|\hat{\boldsymbol{\Delta}}\|_1 \leq 2\sqrt{s_0}\|\hat{\boldsymbol{\Delta}}\|_2$، که از نامساوی کشی به دست می‌آید، استفاده کنیم خواهیم داشت

$$\gamma\|\hat{\boldsymbol{\Delta}}\|_2^2 \leq 4\sqrt{s_o}\|\hat{\boldsymbol{\Delta}}\|_2\left\|\frac{(\mathbf{W}\mathbf{X})^{\mathbf{T}}\mathbf{U}}{n}\right\|_\infty \Longrightarrow \gamma\|\hat{\boldsymbol{\Delta}}\|_2 \leq 4\sqrt{s_o}\left\|\frac{(\mathbf{W}\mathbf{X})^{\mathbf{T}}\mathbf{U}}{n}\right\|_\infty.$$

در نتیجه

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq \frac{4}{\gamma}\sqrt{s_0}\left\|\frac{(\mathbf{W}\mathbf{X})^{\mathbf{T}}\mathbf{U}}{n}\right\|_\infty \leq \frac{4}{\gamma}\sqrt{s_0}\max_j\left\{\frac{\lambda + A_j}{\lambda A_j}\right\}\left\|\frac{\mathbf{X}^{\mathbf{T}}\mathbf{U}}{n}\right\|_\infty.$$

بنابراین، با قبول بعضی فرضیات برای جملات ماتریس $\mathbf{X}$، همگرایی برآوردکنندهٔ $\hat{\boldsymbol{\beta}}$ را در L_2 می‌توان تضمین نمود. به‌عنوان مثال اگر عناصر ماتریس $\mathbf{X}$ دارای توزیع نرمال استاندارد باشند و داشته باشیم $\|\mathbf{U}\|_2 \leq \sqrt{n\max_j\{\frac{\lambda A_j}{\lambda+A_j}\}}$، در این صورت داریم

$$\left\|\frac{\mathbf{X}^{\mathbf{T}}\mathbf{U}}{n}\right\|_\infty \leq \sqrt{\max_j\{\frac{\lambda A_j}{\lambda + A_j}\}\frac{\ln p}{n}},$$

و در نتیجه داریم

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq \frac{4}{\gamma}\sqrt{\max_j\{\frac{\lambda + A_j}{\lambda A_j}\}\frac{s_0\ln p}{n}}.$$

□

References

[1] Baranchik, A.J. (1970). A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution. *Ann. Math. Statist*, 41, 642–645. DOI: https://doi.org/10.1214/aoms/1177697104.

[2] Berger, J. (1976). Admissible Minimax Estimation of a Multivariate Normal Mean With Arbitrary Quadratic Loss. *The Annals of Statistics*, 4, 223–226. DOI: https://doi.org/10.1214/aos/1176343356.

[3] Berger, J., & Strawderman, W.E. (1996). Choice of Hierarchical Priors:Admissibility in Estimation of Normal Means. *The Annals of Statistics*, 24, 931–951. DOI: https://doi.org/10.1214/aos/1032526950.

[4] Brown, L.D. (1971). Admissible Estimators, Recurrent Diffusions, and Insoluble Boundary Value Problems. *Ann. Math. Statist*, 42, 855–903. DOI: https://doi.org/10.1214/aoms/1177693318.

[5] Buhlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 19, 1212–1242. DOI: https://doi.org/10.3150/12-BEJSP11.

[6] Buhlmann, P., & van de Geer, S. (2011). Statistics for high-dimensional data. *Springer-Verlag*. DOI: https://doi.org/10.1007/978-3-642-20192-9.

[7] Cai, T.T., & Zijian, Guo. (2016). Accuracy assessment for high-dimensional linear regression. *arXiv preprint arXiv: 1603.03474*. DOI: https://doi.org/10.48550/arXiv.1603.03474.

[8] Efron, B., & Morris, C. (1973). Stein's Estimation Rule and Its Competitors: An Empirical Bayes Approach. *J. Amer. Statist. Assoc*, 68, 117–130. DOI: https://doi.org/10.2307/2284155.

[9] Ghoreishi, S.K., & Meshkani, M.R. (2014). On SURE estimators in hierarchical models assuming heteroscedasticity for both levels of a two-level normal hierarchical model. *J. of Multivariate Analysis*, 132, 129–137. DOI: https://doi.org/10.1016/j.jmva.2014.08.001.

[10] James, W., & Stein, C.M. (1961). Estimation With Quadratic Loss. *Proceedings of the 4th Berkeley Symposium on Probability and Statistics*, 1, 367–379. DOI: https://doi.org/10.1007/978-1-4612-0919-5_30.

[11] Morris, C. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *J. Amer. Statist. Assoc*, 78, 47–55. DOI: https://doi.org/10.2307/2287098.

[12] Shanita, V., & Ghoreishi, S.K. (2020). Empirical Estimation for Sparse Double-Heteroscedastic Hierarchical Normal Models. *Journal of Statistical Theory and Applications*, 19, 148–161. DOI: https://doi.org/10.2991/jsta.d.200422.001.

[13] Stein, C.M. (1962). Confidence Sets for the Mean of a Multivariate Normal Distribution (with discussion). *J. Roy. Statist. Soc. Ser. B*, 24, 265–296. DOI: https://doi.org/10.1111/j.2517-6161.1962.tb00458.x.

[14] Xie, X., Kou, S.C., & Brown, L.D. (2012). SURE Estimates for a Heteroscedastic Hierarchical Model. *J. Amer. Statist. Assoc*, 107, 1465–1479. DOI: https://doi.org/10.1080/01621459.2012.728154.

[15] Xie, X., Kou, S.C., & Brown, L.D. (2016). Optimal shrinkage estimation of mean parameters in family of distributions with quadratic variance. *The Annals of Statistics*, 44, 564–597. DOI: https://doi.org/10.1214/15-AOS1377.