

مطالعه رفتار برآوردگر انقباضی تحت یک قید خطی در مدل رگرسیون تاوانیده

محمد آرست، محمد آرشی، محمدرضا ربیعی

گروه آمار، دانشگاه صنعتی شاهرود

تاریخ دریافت: ۱۳۹۶/۰۸/۲۷ تاریخ پذیرش: ۱۳۹۷/۰۷/۲۳

چکیده: معمولاً در مسائل با بعد بالا، وقتی تعداد متغیرها بیشتر از تعداد مشاهدات است، برآوردگرهای جریمه شده بر پایه روش‌های انقباضی از دیدگاه خطای پیشگویی پاسخ، از کارایی بهتری نسبت به برآوردگر کمترین توان‌های دوم در برآورد ضرایب رگرسیونی برخوردار هستند. در این برآوردگرها پارامتر تنظیم‌کننده یا انقباضی نقش اساسی در انتخاب متغیر و برآورد پارامترها بازی می‌کند. برآوردگر انقباضی بریج، برآوردگری است که با تغییر پارامتر تنظیم‌کننده آن می‌توان به برآوردگرهای معروف ریج و لاسو دست یافت. در این مقاله برآوردگر انقباضی بریج را، با اعمال یک قید خطی روی بردار ضرایب رگرسیونی، به دست آورده سازگاری آن اثبات می‌شود. به علاوه در قالب یک مطالعه شبیه‌سازی و مثال واقعی کارایی آن از دیدگاه میانگین توان دوم خطا مورد ارزیابی قرار می‌گیرد. **واژه‌های کلیدی:** برآوردگر انقباضی، برآوردگر مقید، سازگاری، میانگین توان دوم خطا، بریج.

۱ مقدمه

تحلیل رگرسیونی روشی آماری برای بررسی و به مدل در آوردن ارتباط بین متغیرها است. در بسیاری از مسائل از جمله پزشکی تعداد متغیرها بسیار زیاد و تعداد مشاهدات کم هستند که به مسئله p بزرگ و n کوچک معروف است. این مسئله مشکلات بسیار زیادی در برازش مدل‌های رگرسیون به وجود می‌آورد. از طرفی بسیاری از متغیرها رفتار یکسانی با متغیرهای دیگر دارند یا در واقع برخی از متغیرها ترکیب خطی از یک یا چند متغیر دیگر هستند. از این رو، در رویارویی با این گونه مسائل، زیرمجموعه کوچکی از متغیرها

آدرس الکترونیکی نویسنده مسئول مقاله: محمد آرست، arast_amar@yahoo.com

کد موضوع‌بندی ریاضی (۲۰۱۰): 62J07

۲ برآوردگر بریج مقید

که دارای بیشترین تاثیر بوده را انتخاب و به برآورد آن‌ها پرداخته می‌شود. بنابراین انتخاب متغیر و برآورد ضرایب، اساسی‌ترین بخش در مدل‌سازی رگرسیونی است. روش‌های برآوردیابی کمترین توان‌های دوم، انتخاب متغیر به صورت پیشرو و غیره، در مواجهه با داده‌هایی که از ویژگی‌های متفاوتی برخوردار باشند، عملکرد قابل اطمینانی از خود نشان نمی‌دهند. از آسیب‌های مدل در هنگام استفاده از این روش‌ها می‌توان به عدم پایداری، دقت پیش‌بینی کم و انتخاب نادرست متغیرها، اشاره نمود. به علاوه این مشکلات زمانی که همبستگی بین متغیرهای پیش‌بین زیاد باشد، تشدید نیز می‌شوند. روش‌های انقباضی به عنوان راهکاری برای کاهش این مشکلات به خصوص زمانی که همبستگی بین متغیرهای پیش‌بین زیاد باشد، مورد توجه قرار گرفته‌اند.

جیمز و استاین (۱۹۶۱) نشان دادند، برای تابع زیان توان دوم خطا، روش‌های انقباضی کارایی برآوردگر طبیعی پارامتر مکان را افزایش می‌دهند. این روش‌ها ضرایب رگرسیونی را با اعمال محدودیت روی دامنه تغییرات آن‌ها برآورد می‌کنند. اگرچه وجود چنین محدودیت‌هایی واریانس برآوردگر را کاهش می‌دهد ولی مقداری اریبی ایجاد می‌کند، به طوری که می‌توان امیدوار بود در نهایت میانگین توان دوم خطا کاهش یابد. از جمله برآوردگرهای انقباضی رایج در برآورد پارامترهای مدل رگرسیونی می‌توان به برآوردگر بریج اشاره نمود که توسط فرانک و فریدمن (۱۹۹۳) معرفی شد. در ادامه این مقاله پس از معرفی این برآوردگر در بخش ۲، برآوردگر بریج مقید تحت یک قید خطی را در بخش ۳ ارائه کرده و رفتار مجانبی آن را از منظر سازگاری مورد مطالعه قرار داده می‌شود. در بخش ۴ رفتار برآوردگر بریج مقید در قالب چند سناریوی شبیه‌سازی شده مورد ارزیابی قرار می‌گیرد و کارایی این برآوردگر از منظر خطای پیشگویی در یک مثال واقعی در بخش ۵ بررسی می‌شود. نتایج حاصل در بخش ۶ مورد بحث قرار می‌گیرد.

۲ معرفی برآوردگرها

مدل خطی

$$Y = X\beta + \epsilon \quad (1)$$

را در نظر بگیرید، که در آن $Y = (y_1, \dots, y_n)$ بردار متغیر پاسخ، $X = (x_1, \dots, x_n)^T$ یک ماتریس $n \times p$ شامل متغیرهای پیش‌بین $x_i \in \mathbb{R}^p$ است. فرض کنید ماتریس X دارای رتبه کامل ستونی، $\beta = (\beta_1, \dots, \beta_p)^T$ بردار ضرایب رگرسیونی و $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ بردار خطاهای تصادفی

محمد آرست و همکاران ۳

است، که در آن $E(\epsilon) = 0$ و $Cov(\epsilon) = \sigma^2 I_n$. برآوردگر کمترین توان‌های دوم (OLS) پارامتر β از کمینه کردن مجموع توان‌های دوم مانده‌های

$$RSS = RSS(\beta_1, \dots, \beta_p) = \epsilon^T \epsilon = \sum_{i=1}^n \epsilon_i^2$$

به صورت $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$ حاصل می‌شود. وقتی در داده‌ها هم‌خطی وجود داشته باشد، این برآوردگر کارایی لازم را ندارد. مسئله هم‌خطی و کاربرد نیودن برآوردگر کمترین توان‌های دوم، منجر به توسعه برآوردگرهای آریب مانند برآوردگرهای انقباضی استاین، مولفه‌های اصلی، کمترین توان‌های دوم جزئی، نوع لیو، تقریباً نااریب، نوع ریچ و لاسو که برآوردگرهای بهبودیافته را ارائه می‌دهند، شده است. در دو دهه اخیر روش‌های انقباضی مختلفی برای برآورد ضرایب رگرسیونی ارائه شده است، که در بین آن‌ها روش انقباضی بریچ از توجه شایانی برخوردار شده است.

۱.۲ روش انقباضی بریچ

فرانک و فریدمن (۱۹۹۳) مجموع توان‌دوم باقیمانده RSS را تحت قید $\sum_{j=1}^p |\beta_j|^\gamma \leq t$ به‌ازای $\gamma \geq 0$ کمینه کردند و آن را به‌عنوان تعمیمی از رگرسیون ریچ به نام رگرسیون بریچ، برای مقابله با هم‌خطی ارائه نمودند. برآوردگر حاصل از حل مسئله بهینه‌سازی فوق به‌صورت

$$\hat{\beta}_{Bridge} = \underset{\beta}{\operatorname{argmin}} ((Y - X\beta)^T (Y - X\beta) + k \sum_{j=1}^p |\beta_j|^\gamma) \quad (2)$$

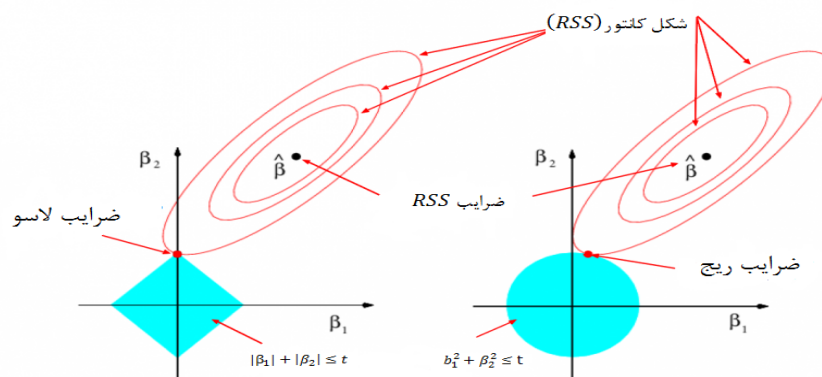
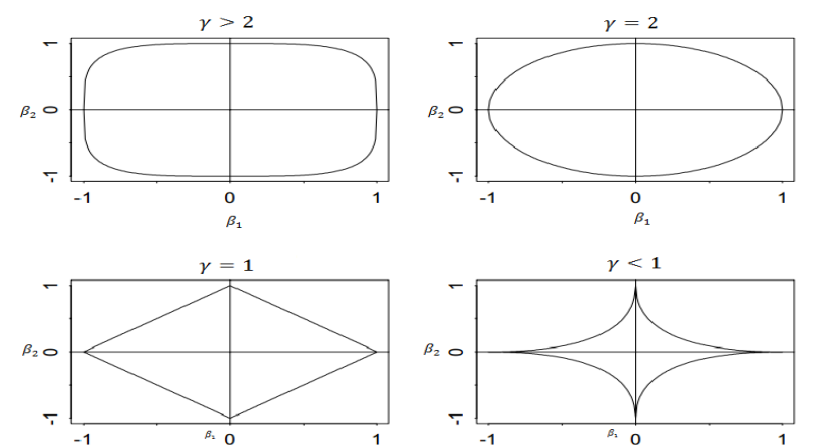
حاصل می‌شود، که در آن $k, t \geq 0$ پارامترهای کنترل بوده و میزان انقباض تحمیل شده به ضرایب را کنترل می‌کند. لازم به ذکر است که انتخاب بهینه γ منجر به افزایش کارایی برآوردگر خواهد شد. اگر $\gamma = 2$ ، این مساله به رگرسیون ریچ تبدیل می‌شود. اگر $\gamma = 1$ رگرسیون لاسو و برای $\gamma = 2$ ، برآوردگر ریچ دارای صورت بسته $\hat{\beta}_{Ridge} = (X^T X + kI)^{-1} X^T Y$ است. برخلاف روش ریچ، لاسو صورت بسته‌ای ندارد و برای هر مقدار k بردار ضرایب به‌صورت عددی به‌دست می‌آیند، اما تیشیرانی (۱۹۹۶) نشان داد

۴ برآوردگر بریج مقید

اگر ماتریس طرح متعامد ($X'X = I$) باشد، ضرایب برآوردگر لاسو به صورت

$$\hat{\beta}_{Lasso} = \text{Sgn}(\hat{\beta}_{OLS})\left(\hat{\beta}_{OLS} - \frac{k}{\gamma}I\right)^+ \quad (۳)$$

به دست می‌آید، که در آن Sgn تابع علامت و $a^+ = \max(0, a)$ است. برآوردگر بریج در حالت کلی همانند لاسو دارای صورت بسته‌ای نیست.



شکل ۱: عملکرد رگرسیون بریج در حالت $p = 2$ و $t = 1$.

شکل ۱ ناحیه تاوان ($|\beta_1|^\gamma + |\beta_2|^\gamma \leq 1$) و مجموع توان دوم مانده‌ها با مرکزیت برآورد کمترین توان‌های دوم معمولی را نشان می‌دهد. این شکل به وضوح ناتوانی روش بریج در حالت $\gamma \geq 2$ در صفر برآورد کردن ضرایب را نشان داده، که برخورد دو ناحیه نمی‌تواند در نقطه‌ای باشد که در آن یکی از ضرایب صفر است. اما زمانی که $\gamma \leq 1$ ، برآورد ضرایب به گونه‌ای انجام می‌پذیرد که بعضی از ضرایب، مربوط به متغیرهای بی‌اثر دقیقاً صفر برآورد شده و به این ترتیب متغیر مربوط به آن ضرایب از مدل خارج خواهد شد. بنابراین در این روش، برآوردیابی و انتخاب متغیر توأم صورت می‌پذیرد، که این ویژگی استفاده از روش انقباضی بریج را برای $\gamma \leq 1$ (از جمله لاسو) در بعد بالا بسیار محبوب ساخته است.

۲.۲ واریانس برآوردگر بریج

فو (۱۹۹۸) نشان داد واریانس برآوردگر بریج برای $\gamma > 1$ به صورت

$$Var(\hat{\beta}_{Bridge}) = \sigma^2 (\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y_0})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y_0})^{-1} \quad (۴)$$

است، که در آن $D(\hat{\beta}) = \text{diag}(\frac{\gamma(\gamma-1)}{\gamma} |\hat{\beta}_j|^{\gamma-2})$ و y_0 نقطه‌ای ثابت در فضای نمونه است. برای حالت خاص $\gamma = 2$ در رابطه (۴)، $D(\hat{\beta}) = \mathbf{I}$. برای $\gamma = 1$ تیبشیرانی (۱۹۹۶) تقریبی از واریانس را به صورت $Var(\hat{\beta}_{Lasso}) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X} + k\mathbf{W}^-)^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X} + k\mathbf{W}^-)^{-1}$ ارائه داد. که در آن $\hat{\sigma}^2$ برآوردی از خطای واریانس و \mathbf{W}^- معکوس تعمیم یافته $\mathbf{W} = \text{diag}(|\hat{\beta}_j|)$ است.

۳.۲ انتخاب پارامتر انقباضی γ و پارامتر تنظیم کننده k

برای انتخاب بهینه پارامترهای انقباضی γ و تنظیم کننده k از آماره اعتبارسنجی متقابل تعمیم یافته (GCV) که کراون و واهبا (۱۹۷۹) ارائه کردند، استفاده می‌شود. برای $k \geq 0$ و $\gamma \geq 1$ ، این معیار به صورت

$$GCV = \frac{1}{n} \sum_{i=1}^n \left(\frac{e_{i,k}}{1 - \frac{p(k)}{n}} \right)^2 \quad (۵)$$

تعریف می‌شود، که در آن $n_0 = \text{tr}(\mathbf{X}(\mathbf{X}^T \mathbf{X} + k\mathbf{\Omega}^-)^{-1} \mathbf{X}^T) - n_0$ تعداد $\hat{\beta}_j$ هایی که $\hat{\beta}_j = 0$ است، $\mathbf{\Omega}^-$ معکوس تعمیم یافته ماتریس $\mathbf{\Omega} = \text{diag}(\frac{\gamma}{\gamma} |\hat{\beta}_j|^{\gamma-2})$ و $e_{i,k} = y_i - \hat{y}_i^{Bridge}$. پارامترهای γ و k چنان انتخاب می‌شوند که آماره GCV کمینه مقدار خود را انتخاب کند.

۳ برآوردگر انقباضی بریج مقید

در مسائل کاربردی و مدل‌های اقتصادی، استفاده از اطلاعات اولیه در خصوص بعضی از پارامترهای مورد علاقه در مدل آماری تحت بررسی، اغلب منجر به استفاده از روش‌های پیشرفته و تعمیم‌یافته‌ای می‌شود که با توجه به قطعی یا غیرقطعی بودن این اطلاع، به کار برده می‌شوند. استفاده از اطلاعات اولیه معلوم در مدل‌های آماری، معمولاً به صورت یک قید مطرح است که در نهایت به مدل‌های مقید منتج می‌شود. برآوردگری که بر این اساس به دست می‌آید، برآوردگر مقید نامیده می‌شود. حال در مدل (۱) محدودیت خطی $R\beta = r$ را برای β در نظر بگیرید، که در آن R ماتریس معلوم محدودیت در اندازه $q \times p$ ، با $\text{rank}(R) = q < p$ و برداری r با مولفه‌های معلوم است. به مدل خطی جدید، مدل خطی مقید (RLM) گفته می‌شود. به منظور یافتن برآوردگر β ، از روشی که توسط رانو در سال ۱۹۹۵ پیشنهاد داده شد، استفاده می‌کنیم. این روش که تلفیقی از روش کمترین توان‌های دوم و لاگرانژ است، تابع

$$F(\beta, \lambda) = (\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) + 2\lambda^T(\mathbf{R}\beta - \mathbf{r})$$

را به ازای $\lambda \geq 0$ ، که تابعی خطی از مجموع توان‌های دوم خطا و محدودیت اعمال شده بر روی مدل است، کمینه می‌سازد.

با مشتق‌گیری از $F(\beta, \lambda)$ نسبت به β و λ و برابر صفر قرار دادن آن‌ها، داریم

$$\hat{\beta}_{OLS}^R = \hat{\beta}_{OLS} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \lambda.$$

از آنجا که

$$\mathbf{R}\hat{\beta}_{OLS}^R = \mathbf{R}\hat{\beta}_{OLS} - \mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \hat{\lambda} = \mathbf{r}$$

برای مدل RLM برآوردگر کمترین توان‌های دوم مقید (RLSE) به صورت

$$\hat{\beta}_{OLS}^R = \hat{\beta}_{OLS} + \mathbf{S}^{-1} \mathbf{R}^T (\mathbf{R}\mathbf{S}^{-1} \mathbf{R}^T)^{-1} (\mathbf{r} - \mathbf{R}\hat{\beta}_{OLS})$$

به دست می‌آید، که در آن $\mathbf{S} = \mathbf{X}^T \mathbf{X}$ و $\hat{\beta}_{OLS} = \mathbf{S}^{-1} \mathbf{X}^T \mathbf{Y}$ یکی از بهترین روش‌ها برای حل

محمد آرست و همکاران ۷

مسئله هم‌خطی، اعمال محدودیت اضافی روی بردار پارامترهاست. بدین منظور، تابع

$$F(\beta, k, \lambda) = (\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) + k \sum_{j=1}^p |\beta_j|^\gamma + \lambda^T(\mathbf{R}\beta - \mathbf{r}) \quad (6)$$

که مجموع توان‌های دوم خطا و دو محدودیت اعمال شده روی فضای پارامترهاست، در نظر گرفته می‌شود.

قضیه ۱: برای $\gamma > 1$ برآوردگر انقباضی بریج مقید عبارت است از

$$\hat{\beta}_{Bridge}^R = \hat{\beta}_{Bridge} + \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} (\mathbf{r} - \mathbf{R} \hat{\beta}_{Bridge}) \quad (7)$$

برهان: با مشتق گرفتن از (۶) نسبت به β و مساوی صفر قرار دادن داریم

$$\frac{\partial F}{\partial \beta} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta + k\gamma \sum_{j=1}^p |\beta_j|^{\gamma-1} + 2\mathbf{R}'\lambda = 0$$

چون تحت محدودیت اعمال شده $\mathbf{R}\hat{\beta} = \mathbf{r}$ ، می‌توان نوشت

$$\hat{\lambda} = (\mathbf{R}(\mathbf{X}'\mathbf{X} + kD(\hat{\beta})|_{y_*})^{-1} \mathbf{R}^T)^{-1} (\mathbf{R}(\mathbf{X}'\mathbf{X} + kD(\hat{\beta})|_{y_*})^{-1} \mathbf{X}'\mathbf{Y} - \mathbf{r})$$

در نهایت برآوردگر بریج مقید به صورت

$$\begin{aligned} \hat{\beta}_{Bridge}^R &= \hat{\beta}_{Bridge} - (\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y_*})^{-1} \mathbf{R}^T \\ &\times (\mathbf{R}(\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y_*})^{-1} \mathbf{R}^T)^{-1} (\hat{\beta}_{Bridge} - \mathbf{r}) \end{aligned}$$

حاصل می‌شود. با قرار دادن $\mathbf{S}_k = \mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y_*}$ ، می‌توان این برآوردگر را به صورت ساده‌تر

$$\hat{\beta}_{Bridge}^R = \hat{\beta}_{Bridge} + \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} (\mathbf{r} - \mathbf{R} \hat{\beta}_{Bridge})$$

بازنویسی کرد.

برآوردگر بریج مقید ۸

۱.۳ مطالعه رفتار برآوردگر انقباضی بریج مقید

برای هر برآوردگر β^* ، میانگین توان دوم خطا به صورت $MSE(\beta, \beta^*) = E(\beta^* - \beta)^T(\beta^* - \beta)$ به دست می‌آید.

قضیه ۲: تحت مفروضات این بخش میانگین توان دوم خطای برآوردگر بریج مقید (۷) به ازای $\gamma > 1$ به صورت

$$MSE(\beta, \hat{\beta}_{Bridge}^R) = \sigma^2 \text{tr}(M_k S M_k) + k^2 \text{tr}(M_k D^T(\hat{\beta})|_{y_0} \beta^T \beta D(\hat{\beta})|_{y_0} M_k)$$

است، که در آن

$$S_k = X^T X + kD(\hat{\beta}_{Bridge})|_{y_0} \text{ و } M_k = S_k^{-1} - S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R S_k^{-1}$$

برهان: چون

$$\begin{aligned} E(\hat{\beta}_{Bridge}^R) &= E[(X^T X + kD(\hat{\beta}_{Bridge})|_{y_0})^{-1} X^T Y] \\ &= (I - kD(\hat{\beta}_{Bridge})|_{y_0} S_k^{-1}) \beta \end{aligned}$$

با استفاده از قضیه ۱ برآوردگر بریج مقید (۷) را می‌توان به صورت

$$\begin{aligned} \hat{\beta}_{Bridge}^R &= S_k^{-1} S_k \hat{\beta}_{Bridge} + S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} (r - R S_k^{-1} S_k \hat{\beta}_{Bridge}) \\ &= M_k S_k \hat{\beta}_{Bridge} + S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} r \\ &= M_k S_k \hat{\beta}_{Bridge} + S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R \beta \\ &= M_k S_k \hat{\beta}_{Bridge} - M_k S_k \beta_0 + \beta_0 \end{aligned}$$

باز نویسی کرد. بنابراین

$$\begin{aligned} E(\hat{\beta}_{Bridge}^R) &= M_k S_k ((I - kD(\hat{\beta}_{Bridge})|_{y_0} S_k^{-1}) \beta - \beta_0) + \beta_0 \\ &= M_k S_k (\beta - \beta_0) + \beta_0 - kD(\hat{\beta}_{Bridge})|_{y_0} M_k \beta \end{aligned}$$

محمد آرست و همکاران ۹

با توجه به روابط $\mathbf{R}\beta = \mathbf{r}$ و $\mathbf{R}\beta_0 = \mathbf{r}$ و $\mathbf{M}_k \mathbf{S}_k (\beta - \beta_0) = \beta - \beta_0$ می توان نوشت

$$E(\hat{\beta}_{Bridge}^R - \beta) = -k D(\hat{\beta}_{Bridge})|_{y_0} \mathbf{M}_k \beta$$

از طرفی $Cov(\hat{\beta}_{Bridge}^R) = \sigma^2 \mathbf{M}_k \mathbf{X}^T \mathbf{X} \mathbf{M}_k = \sigma^2 \mathbf{M}_k \mathbf{S} \mathbf{M}_k$ مورد نظر به صورت

$$\begin{aligned} MSE(\hat{\beta}_{Bridge}^R) &= \text{tr}(Cov(\hat{\beta}_{Bridge}^R)) + \text{tr}(E(\hat{\beta}_{Bridge}^R - \beta)E(\hat{\beta}_{Bridge}^R - \beta)^T) \\ &= \sigma^2 \text{tr}(\mathbf{M}_k \mathbf{S} \mathbf{M}_k) + k^2 \text{tr}(\mathbf{M}_k D^T(\hat{\beta})|_{y_0} \beta^T \beta D^T(\hat{\beta})|_{y_0} \mathbf{M}_k) \end{aligned}$$

حاصل می شود.

۲.۳ رفتار مجانبی برآوردگر بریج

برای بررسی رفتار مجانبی برآوردگر بریج مقید (۷)، مدل خطی (۱) را در نظر بگیرید. برای برآورد β از تابع هدف

$$Z_n(\phi) = \frac{1}{n} RSS(\phi) + \frac{k_n}{n} \sum_{j=1}^p |\phi_j|^\gamma \quad \gamma > 0 \quad \forall k_n \quad (۸)$$

استفاده می شود، که در آن $\phi = (\phi_1, \dots, \phi_n)^T$. برای این منظور لازم است شرایط نظم

$$C_n = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \rightarrow C \quad (A1)$$

$$\frac{1}{n} \max_{1 \leq i \leq n} \mathbf{x}_i^T \mathbf{x}_i \rightarrow 0. \quad (A2)$$

در نظر گرفته شود، که در آن C ماتریس معلوم و نامنفی و C_n ماتریسی نامنفرد است و همچنین فرض کنید مقدار کمینه (۸) به ازای $\hat{\beta}_n = \phi$ حاصل شود. نایت و فو (۲۰۰۰) نشان دادند اگر $k_n = o(n)$

۱۰ برآوردگر بریج مقید

آن‌گاه برآوردگر $\hat{\beta}_n$ برای β سازگار است. به عبارت دقیق‌تر اگر $k_n \geq 0$ و $k_n/n \rightarrow 0$ آن‌گاه

$$\operatorname{argmin}(Z_n(\hat{\beta}_n)) \xrightarrow{P} \operatorname{argmin}(Z(\beta)),$$

که در آن

$$Z(\phi) = (\phi - \beta)^T C (\phi - \beta) + k_n \sum_{j=1}^p |\phi_j|^\gamma.$$

قضیه ۳: تحت مفروضات این بخش برآوردگر بریج را به صورت

$$\hat{\beta}_{Bridge}^R = \hat{\beta}_n + S_n^{-1}(k) R^T (R S_n^{-1}(k) R^T)^{-1} (r - R \hat{\beta}_n)$$

در نظر بگیرید، که در آن $k_n = o(n)$ اگر $S_n(k) = C_n + \frac{k_n}{n} D(\hat{\beta})|_{y_0}$.

$$\hat{\beta}_{Bridge}^R \xrightarrow{P} \beta + C^{-1} R^T (R C^{-1} R^T)^{-1} (r - R \beta)$$

برهان: تحت فرض $k_n = o(n)$ داریم $\hat{\beta}_n \xrightarrow{P} \beta$. همچنین تحت شرایط نظم (A۱)، $S_n(k) \rightarrow C$. نتیجه حاصل از همگرایی در احتمال تابع پیوسته مجموع حاصل می‌شود.

۴ مطالعه شبیه‌سازی

در این بخش با استفاده از یک فرایند شبیه‌سازی مونت‌کارلو به ارزیابی عملکرد و مقایسه روش کمترین توان‌های دوم و بریج (با تمرکز بر روی برآوردگرهای ریج و لاسو) در مدل رگرسیونی چندگانه در دو حالت بدون محدودیت و با محدودیت پرداخته می‌شود. از آنجا که روش‌های انقباضی برای مقابله با هم‌خطی چندگانه است، لذا مدل‌های شبیه‌سازی شده هم‌خط را بیشتر در این طرح مورد مطالعه قرار می‌گیرد. برای داشتن داده‌هایی با میزان هم‌خطی متفاوت، مک دونالد و گلارنیو (۱۹۷۵) پیشنهاد دادند که داده‌ها را به روش $x_{ij} = (1 - \nu^2)^{1/2} z_{ij} + \nu z_{ip}$ که $i = 1, \dots, n$ و $j = 1, \dots, p$ است، تولید کنیم، که در آن اعداد شبه‌تصادفی مستقل از توزیع نرمال استاندارد و ν^2 همبستگی بین هر دو متغیر توضیحی است. در ادامه در قالب چند سناریو ابتدا ویژگی‌های داده‌های شبیه‌سازی شده را بیان نموده، سپس نتایج کاربرد روش‌ها ارائه می‌شود. در تمام سناریوها مقدار k با روش GCV به دست آورده شده است.

مثال ۱: نیوهوسن و اومن (۱۹۷۱) بیان کردند در صورتی که میانگین توان دوم خطا، تابعی از β ، σ^2 و k باشد و متغیرهای توضیحی ثابت باشند، اگر β بردار ویژه نرمال شده متناظر با بزرگترین مقدار ویژه ماتریس $X'X$ تحت قید $\beta'\beta = 1$ باشد، آنگاه MSE کمینه می‌شود. بنابراین، از بردار ویژه متناظر با بزرگترین مقدار ویژه ماتریس $X'X$ برای برآورد ضرایب β_1, \dots, β_p به‌ازای $n = 5, 20, 50, 100$ ، $\nu = 0/7, 0/9, 0/99$ ، $\sigma = 0/5$ ، $p = 4, 20$ ، $\nu = 0/7, 0/9, 0/99$ استفاده می‌شود.

جدول ۱: نتایج مربوط به سناریوی ۱

$p = 20$				$p = 4$				MSE	ν
n				n					
۱۰۰	۵۰	۲۰	۵	۱۰۰	۵۰	۲۰	۵		
0/149	0/471	NA	NA	0/023	0/043	0/150	3715	OLS	
0/072	0/104	0/284	0/421	0/027	0/040	0/080	0/473	Ridge	0/7
0/159	0/483	1/282	3/617	0/021	0/440	0/144	0/850	Lasso	
0/465	1/291	NA	NA	0/062	0/126	0/407	5135	OLS	
0/063	0/0917	0/282	0/254	0/045	0/061	0/109	0/371	Ridge	0/9
0/470	1/025	1/942	5/322	0/062	0/125	0/342	1/363	Lasso	
5/691	15/31	NA	NA	0/133	0/274	0/876	4955	OLS	
0/045	0/072	0/133	0/149	0/441	0/060	0/109	0/351	Ridge	0/99
1/612	2/129	3/451	7/474	0/133	0/255	0/566	1/751	Lasso	

در ادامه روش شبیه‌سازی مشابه سناریوی ۱ در نظر می‌گرفته شده است، با این تفاوت که بردار β ساختارهای متفاوتی داشته و در یک حالت آن محدودیت خطی نیز به آن اضافه شده است.

مثال ۲: بردار β را به صورت $\beta = (\underbrace{0, \dots, 0}_{10}, \underbrace{2, \dots, 2}_{10}, \underbrace{0, \dots, 0}_{10}, \underbrace{2, \dots, 2}_{10})^T$ در نظر می‌گیریم. در این مدل گروهی از متغیرهای موثر و غیر موثر وجود دارد که همبستگی بین آن‌ها متوسط و یکسان است

مثال ۳: بردار β به صورت $\beta = (\underbrace{3, \dots, 3}_{15}, \underbrace{0, \dots, 0}_{25})^T$ در نظر گرفته می‌شود.

مثال ۴: تحت مفروضات سناریوی ۱، فرض می‌شود $r = [1/9, 3/8]^T$ ، $R = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & -1 & 2 & 0 & 0 \end{bmatrix}$ $p = 5$ و $q = 2$.

جدول ۲: نتایج مربوط به سناریوی دوم و سوم.

سناریوی ۳				سناریوی ۲				MSE	ν
n				n					
۱۰۰	۵۰	۲۰	۵	۱۰۰	۵۰	۲۰	۵		
۱۴/۹	۴۲/۰	NA	NA	۱۰/۴۱۰	۳۰/۰۲۲	NA	NA	OLS	
۷/۴۷۲	۱۶/۱۰	۸۴/۸۱	۹۷/۵۰	۴/۲۸۸	۹/۳۸۴	۴۱/۹۳	۵۱/۶۸	Ridge	۰/۷
۶/۹۷۰	۱۱/۹۴	۱۷۱/۱۵	۴۱۹/۹۷	۸/۳۳۰	۱۶/۷۷	۱۴۱/۶	۲۸۳/۳	Lasso	
۸۱/۲۷۳	۹۰/۴۴۶	NA	NA	۲۱/۲۹۴	۹۳/۲۶	NA	NA	OLS	
۲۱/۳۱	۳۰/۶۱	۸۴/۹۳	۸۸/۷۳	۱۲/۶۹	۱۸/۴۱	۴۱/۳۷	۴۳/۹۶	Ridge	۰/۹
۳۵/۵۴	۶۱/۰۶	۲۷۱/۲	۵۹۹/۸	۴۶/۱۵	۸۶/۲۳	۲۳۲/۹	۴۱۶/۶	Lasso	
۹۵/۱۴	۱۰۶/۲	NA	NA	۸۴/۱۶	۱۰۷/۶	NA	NA	OLS	
۳۷/۲۳	۴۱/۵۹	۸۶/۴۴	۱۵۹/۸۵	۳۸/۶۶	۳۸/۸۶	۴۱/۸۳	۴۲/۱۴	Ridge	۰/۹۹
۷۵/۲	۷۸/۳	۹۶/۷	۶۹۲/۶	۴۳/۸۹	۴۴/۸	۴۷/۳	۵۴/۸	Lasso	

۵ داده‌های سرطان پروستات

در این بخش کارایی برآوردگرهای پیشنهاد شده در یک مثال واقعی مورد مطالعه قرار می‌گیرد. برای این منظور از داده‌های پروستات که استیمی و همکاران (۱۹۸۹) ارائه داده‌اند، استفاده شده است. این داده‌ها وابستگی بین آنتیژن نوع خاص پروستات و اندازه کلینیکی مردانی که به پروستاتکومی مراجعه کرده‌اند را نشان می‌دهند. این نتایج نشان می‌دهد زمانی که تعداد متغیرها کمتر از مشاهدات است یا وقتی داده‌ها

جدول ۳: نتایج مربوط به داده‌های پروستات.

Lasso			Ridge			OLS			متغیر
t_0	خطای معیار	برآورد ضرایب	t_0	خطای معیار	برآورد ضرایب	t_0	خطای معیار	برآورد ضرایب	
۰/۷۲۸	۱/۴۴۲	۱/۰۴۳	-۰/۲۲۸	۰/۴۷۱	-۰/۱۰۷	۰/۵۱۶	۱/۲۹۰	۰/۶۶۹	Intercept
۵/۰۰۶	۰/۰۹۴	۰/۴۷۳	۶/۳۱۸	۰/۱۱۱	۰/۷۰۴	۶/۶۷۷	۰/۰۸۷	۰/۵۸۷	lcavol
۱/۰۷۹	۰/۱۸۰	۰/۱۹۵	۲/۶۸۷	۰/۱۷۴	۰/۴۶۹	۲/۶۷۳	۰/۱۷۰	۰/۴۵۴	lweight
۰/۰۰۰	۰/۰۱۲	۰/۰۰۰	-۱/۵۲۶	۰/۰۱۳	-۰/۰۲۱	-۱/۷۵۸	۰/۰۱۱	-۰/۰۱۹	age
۰/۰۰۰	۰/۰۶۱	۰/۰۰۰	۱/۳۹۳	۰/۰۷۰	۰/۰۹۷	۱/۸۳۲	۰/۰۵۸	۰/۱۰۷	lbph
۱/۴۵۰	۰/۲۵۹	۰/۳۷۵	۱/۵۹۹	۰/۲۷۷	۰/۴۴۴	۳/۱۳۶	۰/۲۴۴	۰/۷۶۶	svi
۰/۰۰۰	۰/۰۹۸	۰/۰۰۰	-۰/۵۹۹	۰/۱۱۸	-۰/۰۷۰	-۱/۱۵۹	۰/۰۹۱	-۰/۱۰۵	lcp
۰/۰۰۰	۰/۱۷۴	۰/۰۰۰	۱/۱۳۹	۰/۱۳۸	۰/۱۵۸	۰/۲۸۷	۰/۱۵۷	۰/۰۴۵	gleason
۰/۰۰۰	۰/۰۰۴	۰/۰۰۰	-۰/۲۱۰	۰/۰۰۵	-۰/۰۰۱	۱/۰۲۴	۰/۰۰۴	۰/۰۰۴	pgg45

دارای هم‌خطی می‌باشند، برآوردگر کمترین توان‌های دوم نمی‌تواند به خوبی ضرایب را به دست آورد و مقدار میانگین توان دوم خطای آن زیاد است. در حالی که، برآوردگر لاسو بهتر از برآوردگر کمترین توان‌های دوم ضرایب را برآورد کرده و دارای میانگین توان دوم خطای کمتری است. در مجموع، رگرسیون ریبج دارای بهترین دقت پیش‌بینی در بین سایر روش‌هاست و عیب اصلی این روش غیر صفر برآورد کردن ضرایب و عدم حذف متغیر از مدل است.

جدول ۴: نتایج مربوط به داده‌های پروستات در مدل مقید با محدودیت سناریوی پنجم.

Ridge			OLS			متغیر
t_0	خطای معیار	برآورد ضرایب	t_0	خطای معیار	برآورد ضرایب	
۳/۳۵۷	۸/۱۵۲	۶/۱۷۸	۴/۵۱۱	۰/۲۸۷	۶/۸۷۸	Intercept
-۳/۶۳۸	۰/۰۶۴	۰/۰۱۶	۲/۰۹۷	۰/۰۹۰	۰/۱۲۵	lcavol
NA	۰/۵۵۸	-۰/۲۲۷	NA	۰/۱۰۲	۱/۲۱۳	lweight
NA	۰/۰۲۶	-۰/۰۲۶	NA	۰/۰۱۲	-۰/۰۸۴	age
NA	۰/۱۴۹	۰/۱۰۰	NA	۰/۰۷۳	۰/۱۵۸	lbph
NA	۰/۱۴۹	۰/۱۰۰	NA	۰/۲۹۲	۱/۲۵۰	svi
NA	۰/۰۹۸	۰/۰۷۵	NA	۰/۱۰۳	۰/۴۱۷	lcp
NA	۰/۷۰۶	-۰/۵۸۰	NA	۰/۱۴۲	-۰/۵۹۶	gleason
NA	۰/۰۱۰	۰/۰۰۸	NA	۰/۰۰۵	۰/۰۱۱	pgg45

۶ بحث و نتیجه‌گیری

در این مقاله با توجه به اینکه وقتی در داده‌ها هم خطی وجود داشته باشد برآوردگر کمترین توان‌های دوم کارایی لازم را ندارد، برآوردگر بریج را پیشنهاد نموده و ضمن بررسی میانگین توان دوم خطای آن، با توجه به این‌که در مسائل کاربردی و مدل‌های اقتصادی اگر اطلاعات اولیه‌ای در اختیار باشد، برای آنکه این اطلاعات به مدل اضافه شود، برآوردگر بریج مقید ارائه شد. خواص این برآوردگر از منظر میانگین توان دوم خطا و همگرایی در احتمال مورد بررسی قرار گرفت. نتایج حاصل با مطالعات عددی مورد ارزیابی و تحلیل قرار گرفت، در انتها ملاحظه شد که برآوردگر لاسو از میانگین توان دوم خطای کمتر و برآوردگر بریج از دقت پیش‌بینی بیشتری برخوردار هستند.

تقدیر و تشکر

نویسندگان مقاله ضمن تشکر از اعضای محترم هیئت تحریریه مجله، از پیشنهادهای و نظرات ارزشمند داوران و ویراستار محترم مقاله که موجب ارتقاء سطح آن گردید کمال تشکر و قدردانی را دارند.

مراجع

- Horel, A. E. and Kennard, R. W. (1970), Ridge Regression Biased Estimation for Nonorthogonal Problems, *Thechnometrics*, **12**, 69-89.
- Tibshirani, R. (1966), Regression Shrinkage and Selection via the Lasso, *Journal of Royal Statistical Society*, **58**, 267-88.
- Fu, J. (1998), Penalized Regressions: The Bridge Versus the Lasso, *Journal of Computational and Graphical Statistics*, **7**, 397-416.
- James, W. and Stein, C. (1961), Estimation with Quadratic Loss, *Proceeding of the Forth Berkely Symposium*, **1**, 361-379.
- Frank, I. E. and Friedman, J. H. (1993), A Statistical View of Some Chemometrics Regression Tools (with Discussion), *Technom.*, **35**, 109-148.
- Craven, P. and Wahba, G. (1979), Smoothing Noisy Data With Spline Functions, *Numerische Mathematik*, **31**, 377-403.
- Knight, K. and Fu, W. (2000), Asymptotic for LASSO-type Estimators, *The Annals of Statistics*, **28**, 1356-1378.
- Newhouse, J. P. and Oman, S. D. (1971), An Evaluation of Ridge Estimators, *Rand Corporation*, 716-PR.
- McDonald, G. C. and Galarnaeu, D. I. and Keimel K. (1975), A Monte Carlo Evaluation of Some Ridge-Type Tstimators, *Journal of American Statistical Associety*, **20**, 407-416.