مجله علوم آماری، بهار و تابستان ۱۴۰۰
جلد ۱۵، شماره ۱، ص ۸۱ – ۹۶
DOI: 10.29252/jss.15.1.149
مقاله پژوهشی

مدل رگرسیون لاسو بیزی با خطای نامتقارن در بعد بالا

زهرا خادم بشیری، علی شادرخ، مسعود یارمحمدی

گروه آمار، دانشگاه پیام نور، تهران

تاریخ دریافت: ۱۳۹۸/۱۰/۱۷ تاریخ پذیرش و انتشار: ۱۳۹۹/۰۶/۰۲

چکیده: یکی از بحث‌های چالشی در مدل‌های رگرسیونی انتخاب مدل بهینه است، بدین شکل که چگونه می‌توان متغیرهای توضیحی مهم و متغیرهای قابل اغماض را مشخص کرده و رابطه بین متغیر پاسخ و متغیرهای توضیحی را به‌طور ساده‌تر بیان نمود. با توجه به محدودیت‌های مربوط به انتخاب متغیر به روش کلاسیک نظیر انتخاب گام به گام، می‌توان از روش‌های رگرسیون تاوانیده استفاده کرد. یکی از مدل‌های رگرسیون تاوانیده، مدل رگرسیونی لاسو است که در آن فرض می‌شود خطاها از توزیع نرمال پیروی می‌کنند. در این مقاله، مدل رگرسیون لاسو بیزی با خطایی با توزیع نامتقارن و وجود متغیرهای توضیحی از بعد بالا معرفی می‌شود. سپس با شبیه‌سازی و تحلیل داده‌های واقعی، عملکرد مدل پیشنهادی مورد بحث و بررسی قرار می‌گیرد.

واژه‌های کلیدی: رگرسیون تاوانیده، رگرسیون لاسوی بیزی، توزیع لاپلاس نامتقارن، توزیع آلفا چوله‌نرمال.

۱ مقدمه

مدل رگرسیون خطی چندگانه

$$\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \varepsilon, \qquad (۱)$$

را در نظر بگیرید، که در آن X یک ماتریس $n \times (p+۱)$، $\boldsymbol{\beta}$ بردار پارامترهای $p+۱$ بعدی و ε متغیر تصادفی خطا با $p+۱$ بعد بوده که معمولا از توزیع نرمال با میانگین صفر و واریانس $\sigma^۲$ پیروی می‌کند.

آدرس الکترونیکی نویسنده مسئول مقاله: علی شادرخ، shadrokh.ali@gmail.com
کد موضوع‌بندی ریاضی (۲۰۱۰): 62F15 62G35, 65C10.

در صورتی‌که تعداد متغیرهای توضیحی در مدل رگرسیونی بیشتر از تعداد مشاهدات باشد، احتمال وجود همبستگی میان متغیرهای توضیحی زیاد شده و در نتیجه ممکن است ماتریس $X'X$ وارون‌پذیر نبوده و برآورد یکتایی برای پارامتر $\boldsymbol{\beta}$ به دست نیاید. جستجو برای یک مدل رگرسیونی بهینه را انتخاب متغیر یا انتخاب زیر مجموعه می‌نامند. روش‌های کلاسیک زیادی برای انتخاب متغیر وجود دارند که برخی از آن‌ها مانند رگرسیون گام‌به‌گام براساس دنباله‌ای از آزمون‌های فرض یا برآوردهایی از نوع میانگین توان دوم خطا یا دیگر ملاکهای زیان هستند. این روش‌ها بیشتر بر موضوع انتخاب متغیر متمرکز شده و به برآورد ضرایب رگرسیونی نمی‌پردازند. همچنین در این روش‌ها، تغییرات کوچک در داده‌ها موجب می‌شود یک متغیر به جای دیگری انتخاب شده و نتایج انتخاب‌ها متفاوت شوند، به عبارت دیگر این روش‌ها نسبت به تغییرات کوچک داده‌ها پایدار نیستند. به عنوان یک روش جایگزین برای انتخاب مدل بهینه، می توان از روش‌های انقباضی نظیر رگرسیون ستیغی[1] (هورل و کنارد، ۱۹۷۰) و رگرسیون لاسو[2] (LASSO) (تیبشیرانی، ۱۹۹۶) استفاده کرد. در این روش‌ها، برآوردهای ضرایب رگرسیونی محدود شده و به سمت صفر کاهش می‌یابند. این کاهش می‌تواند به طور معنی‌داری واریانس ضرایب رگرسیونی را کاهش دهد.

در روش رگرسیون ستیغی برآورد ضرایب رگرسیونی به صفر میل کرده و از آنجایی که مقادیر آن دقیقا صفر نشده، از مدل رگرسیونی (۱) به طور کامل حذف نمی‌گردند. بنابراین روش رگرسیون ستیغی را نمی‌توان برای هدف انتخاب متغیر استفاده کرد. روش رگرسیون لاسو به عنوان روشی برای انتخاب متغیر و برآورد پارامتر به طور همزمان، مورد توجه قرار گرفته است. از مزایای اصلی و مهم رگرسیون لاسو می‌توان افزایش دقت پیش‌بینی و بهبود تفسیر مدل رگرسیونی را نام برد. علی‌رغم مزایای این روش، برآوردگر لاسو محدودیت‌هایی نیز داشته و در مواردی که گروهی از متغیرهای توضیحی همبستگی بالایی دارند، این برآوردگر قادر به انتخاب یکی از متغیرهای همبسته بوده و از انتخاب دیگر متغیرهای توضیحی صرف‌نظر می‌کند. برای حل این مسئله، زو و هستی (۲۰۰۵) روش رگرسیون شبکه ارتجاعی[3] را ارائه دادند که در آن ترکیبی از تابع تاوان رگرسیون ستیغی و رگرسیون لاسو استفاده شده است. مسئله دیگر در روش رگرسیون لاسو، تخمین خطاهای استاندارد برای برآورد پارامترهای ضرایب رگرسیونی است، زیرا الگوریتم‌هایی از قبیل الگوریتم لارس (افرون و همکاران، ۲۰۰۴) فقط برآوردهای نقطه‌ای پارامترهای رگرسیونی را فراهم کرده و استفاده از بوت استرپ برای محاسبهء برآورد خطاهای استاندارد و در نهایت محاسبهء بازه اطمینان برای پارامترهای مدل رگرسیونی از لحاظ محاسباتی مشکل است (مالیک و نینجان، ۲۰۱۳). روش بوت

[1] Ridge Regression
[2] Least Absolute Shrinkage and Selection Operator
[3] Elastic net regression

استرپ را نایت و فو (۲۰۰۰) برای برآورد خطای استاندارد مورد استفاده قرار داده و تأکید کردند، وقتی برآورد برخی از ضرایب رگرسیونی صفر می‌شوند، برآوردهای بوت استرپ خطای استاندارد به‌طور مجانبی اریب هستند. در این میان روش‌های بیزی علاوه بر اینکه برآورد خطاهای استاندارد برای ضرایب رگرسیونی را فراهم می‌کنند، در حالاتی که تعداد مشاهدات از پارامترها کمتر هستند، نیز عملکرد خوبی دارند. رگرسیون لاسوی بیزی (تیبشیرانی، ۱۹۹۶)، یک روش انتخاب مدل است که در آن خطاها دارای توزیع نرمال فرض شده و چگالی پیشینی برای ضرایب مدل رگرسیونی نیز توزیع لاپلاس متقارن در نظر گرفته شده است. با توجه به توصیهٔ تیبشیرانی، پارک و کسلا (۲۰۰۸) استفاده از تابع چگالی لاپلاس متقارن را پیشنهاد دادند. در عمل فرض نرمال بودن مانده‌ها در مدل‌های خطی می‌تواند محدود شود. مسئلهٔ انحراف از فرض نرمال بودن وقتی رخ می‌دهد که نمونه شامل داده‌های دور افتاده باشد یا توزیع خطاها نامتقارن باشند. یک روش مواجه با داده‌های غیر نرمال، بررسی روش‌های تبدیل‌کننده مناسب و نرمال کردن آن‌ها است (مارچنکو و گنتون، ۲۰۱۰).

روش دیگر در برخورد با داده‌های غیرنرمال، جستجوی روش‌های انعطاف‌پذیرتر برای مدل‌سازی داده‌های غیرنرمال است. به عبارت دیگر، توزیع‌هایی با پارامترهای اضافی برای تنظیم این عدم تقارن استفاده شوند. برای مثال می‌توان توزیع خطاها با دم‌های سنگین‌تر را در نظر گرفت. توزیع لاپلاس، دم‌های سنگین‌تری نسبت به نرمال داشته و می‌تواند انتخاب مناسبی برای توزیع خطاها باشد. مزیت مهم دیگر توزیع لاپلاس نسبت به دیگر توزیع‌های دم‌سنگین نظیر توزیع تی - استیودنت آن است که همه گشتاورهای مرکزی آن متناهی است. همچنین برای اصلاح کردن عدم تقارن، می‌توان توزیع های چوله‌تی ، چوله‌نرمال را در نظر گرفت که نسخه‌های چوله‌ای از توزیع‌های تی-استیودنت و نرمال هستند (مارچنکو و گنتون، ۲۰۱۰) . به طور کلی، خانواده توزیع‌های چوله‌بیضوی ارائه شده توسط ساهو و دی (۲۰۰۳)، عدم تقارن در کلاسی از توزیع‌های متقارن بیضوی را لحاظ می‌کنند. ساده‌ترین نمایش از خانواده چوله بیضوی همان‌طور که توسط آزالینی (۱۹۸۵) معرفی شده، توزیع چوله‌نرمال است که در مقایسه با توزیع نرمال، علاوه بر پارامترهای مقیاس و مکان، پارامتر شکل برای تنظیم عدم تقارن توزیع را نیز دارد که برای مطالعه رفتار نامتقارن مجموعه داده‌های تجربی در حوزه‌های مختلف استفاده می‌شود. الیورا (۲۰۱۰) خانواده‌ای از توزیع‌ها را معرفی کرد که هر دو حالت تک مدی بودن و دو مدی بودن شکل را حمایت می‌کنند. این خانواده جدید توزیع را آلفا چوله‌نرمال [1] نامیده و با نماد $ASN(\alpha)$ نشان می‌دهند، که در آن α نشان‌دهندهٔ پارامتر عدم تقارن و $ASN(0)$ بیانگر توزیع نرمال متقارن است.

[1] Alpha Skew Normal Distribution

هدف مقاله حاضر، بیان تحلیل بیزی متفاوتی برای داده‌های نامتقارن و با خطای آلفا چوله‌نرمال است. در این راستا در بخش ۲ به توسیع روش لاسوی بیزی در مواردی که توزیع خطاهای مدل چوله است، پرداخته و سپس مدل برای مسئله رگرسیونی با بعد بالا ارایه می‌شود. در بخش ۳ پارامتر تنظیم‌کننده مدل رگرسیونی تاوانیده را برآورد کرده و در بخش ۴ با استفاده از روش‌های شبیه‌سازی، دو مثال در شرایط مختلف از نظر همبستگی میان متغیرهای توضیحی، تعداد پارامترها و تعداد مشاهدات مدل مورد بررسی قرار می‌گیرند. در ادامه با تحلیل داده‌های واقعی به ارزیابی روش‌های رگرسیونی پرداخته می‌شود.

۲ مدل رگرسیون لاسوی بیزی نامتقارن

در مدل رگرسیون خطی چندگانه (۱) اغلب اوقات برآوردگر کمترین توان‌های دوم خطا انتخاب مناسبی است، اما در مواردی که متغیرهای توضیحی همبستگی بالا داشته باشند، این برآوردگر کارایی لازم را ندارد (آرست و همکاران، ۱۳۹۸). در حضور همبستگی بین متغیرهای توضیحی، مقادیر واریانس‌ها و کواریانس‌ها برای برآوردگرهای کمترین توان‌های دوم خطای ضرایب رگرسیونی، بزرگ خواهد شد. در واقع اعضای قطر ماتریس $(X'X)^{-1}$ عبارتند از $\frac{1}{1-R_j^2}, j = 1, \ldots, p$ ، که در آن R_j^2 ضریب تعیین چندگانه از رگرسیون متغیر توضیحی j ام نسبت به $p-1$ متغیر توضیحی باقیمانده است که در صورت وجود همخطی شدید میان متغیرهای توضیحی، R_j^2 نزدیک به واحد خواهد بود و چون واریانس ضریب رگرسیونی j ام برابر با $V(\hat{\beta}_j) = (1 - R_j^2)^{-1}\sigma^2$ است، همخطی شدید موجب می‌شود که واریانس برآورد کمترین توان‌های دوم ضریب رگرسیونی β_j بسیار زیاد شود (نتر و همکاران ، ۱۹۹۶).

روش‌های رگرسیون تاوانیده برای بهبود برآوردهای کمترین توان‌های دوم خطا ارایه شده‌اند. در این روش‌ها با افزایش کمی اریبی، واریانس مدل کاهش می‌یابد. یکی از روش‌های تاوانیده که اخیرا توجه زیادی را به خود جلب کرده، مدل رگرسیون لاسو است. دلیل محبوبیت این روش آن است که می‌تواند به طور همزمان برآورد پارامترهای رگرسیونی و انتخاب متغیرهای مهم را انجام داده و منجر به مدل‌های رگرسیونی دقیق با تفسیرپذیری بالا شود. به عبارت دیگر در این روش، شانس آمدن متغیرهایی که توان پیش‌بینی پایینی دارند، در حضور دیگر متغیرهای توضیحی در مدل، کم است. بنابراین این روش را می‌توان برای هدف انتخاب متغیر استفاده کرد. دو الگوریتم بسیار مهم برای محاسبهٔ برآورد لاسو، الگوریتم‌های لارس و مختصات مسیر است (افرون و همکاران، ۲۰۰۴) . با وجود مزایای روش رگرسیون لاسو، در حالت $p \gg n$، این روش با مشکل مواجه است. زیرا روش لاسو نمی‌تواند بیش از اندازه نمونه n، متغیرهای توضیحی را انتخاب کند و حداکثر قادر به انتخاب n متغیر است (زو و هستی، ۲۰۰۵) . برآورد

لاسو برای پارامترهای مدل رگرسیون خطی چندگانه را می‌توان به صورت برآورد بیزی تحت چگالی پیشینی لاپلاس متقارن برای پارامترهای مدل رگرسیونی نیز تفسیر کرد. پارک و کسلا (۲۰۰۸) در روش لاسو بیزی، تابع چگالی لاپلاس را به صورت آمیخته‌ای از چگالی‌های نرمال بیان کرده و نمونه‌گیر گیبس را برای تولید نمونه‌هایی از توزیع پسینی اتخاذ نمودند.

۲.۱ مدل رگرسیون لاسوی بیزی

در مدل رگرسیون خطی چندگانه (۱)، توزیع خطاها نرمال با میانگین صفر و واریانس σ^2 است و مجموع توان دوم مانده‌ها نیز به صورت

$$RSS = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2,$$

خواهد بود. هدف روش کمترین توان‌های دوم این است که مقادیری از $\beta_0, \ldots, \beta_p$ را بیابد که کمیت RSS را مینیمم کند. در مدل رگرسیون لاسو با در نظر گرفتن محدودیت $\lambda \sum_{j=1}^{p} |\beta_j|$ ، برآوردهای ضرایب رگرسیونی محاسبه می‌شوند. در واقع عبارت $RSS + \lambda \sum_{j=1}^{p} |\beta_j|$ نسبت به β مینیمم می‌شود. λ پارامتر تعدیل کننده غیر منفی است و برای کنترل سطح انقباض به کار می‌رود. یعنی اگر مقدارش برابر صفر باشد، همهٔ متغیرها در مدل حضور خواهند داشت و اگر مقدار آن افزایش یابد، تعداد متغیرهای مستقل در مدل کاهش می‌یابد. مقدار این پارامتر معمولا توسط روش اعتبارسنجی متقابل تعیین می‌شود. در مدل رگرسیون لاسو بیزی برای پارامترهای β_j توزیع پیشینی لاپلاس به صورت $\pi(\boldsymbol{\beta}) = \prod_{j=1}^{p} \frac{\lambda}{2} e^{-\lambda|\beta_j|}$، در نظر گرفته می‌شود. بنابراین توزیع پسینی به شرط $\tilde{y}$ برای پارامترها به صورت

$$\pi(\boldsymbol{\beta}, \sigma^2|\tilde{y}) \propto \pi(\sigma^2)(\sigma^2)^{-(n-1)/2} \exp\{\frac{-1}{2\sigma^2}(\tilde{y} - \mathbf{X}\boldsymbol{\beta})^T(\tilde{y} - \mathbf{X}\boldsymbol{\beta}) - \lambda \sum_{j=1}^{p} |\beta_j|\},$$

است، که در آن $\tilde{y} = (y - \bar{y})$. برای هر مقدار ثابت σ^2، در صورت وجود نما برای توزیع پسین، برآورد بیز همان نمای پسین خواهد بود که به λ و انتخاب چگالی پیشینی برای σ^2 بستگی دارد (پارک و کسلا، ۲۰۰۸).

۲.۲ مدل رگرسیون لاسوی بیزی با توزیع خطای آلفا چوله‌نرمال

در روش رگرسیون لاسو بیزی، فرض می‌شود خطاها دارای توزیع نرمال بوده و توزیع مشاهدات به صورت $\mathrm{y}|\boldsymbol{\beta}\ ,\sigma^2\ \sim\ \mathrm{N}(\mathrm{X}\boldsymbol{\beta},\sigma^2\mathrm{I_n})$ در نظر گرفته می‌شود. همچنین متناسب با توزیع مشاهدات، چگالی پیشینی پارامترهای $\boldsymbol{\beta}$ لاپلاس متقارن است. اما در عمل توزیع خطاها همیشه نرمال نیست. بنابراین در مواردی که توزیع خطاها غیرنرمال است، رگرسیون لاسو بیزی کارایی لازم را برای انتخاب بهترین مدل ندارد. در این بخش فرض می‌شود مشاهدات $\boldsymbol{Y}$ دارای توزیع آلفا چوله‌نرمال الیورا (۲۰۱۰) به صورت

$$f_{\mathrm{y}}(\boldsymbol{Y}|\beta,\sigma^{۲},k)=\prod_{i=۱}^{n}(\frac{(۱-ky_i)^{۲}+۱}{۲+k^{۲}})\phi(\boldsymbol{Y}),$$
$$\phi(\boldsymbol{Y})=\frac{۱}{\sqrt{۲\pi\sigma^{۲}}}\exp\{\frac{-۱}{۲\sigma^{۲}}(\boldsymbol{Y}-X\boldsymbol{\beta})^{T}(\boldsymbol{Y}-X\boldsymbol{\beta})\},$$

است، به طوری‌که $k\in R$ همان پارامتر α در توزیع آلفا چوله‌نرمال بوده و میزان چولگی توزیع را تعیین می‌کند. با در نظر گرفتن تابع چگالی پیشینی لاپلاس نامتقارن

$$\pi(\boldsymbol{\beta}|\sigma^{۲})\propto\prod_{j=۱}^{p}\frac{\lambda}{\sigma^{۲}}\frac{k}{۱+k^{۲}}\begin{cases}\exp(-\frac{k\lambda}{\sigma^{۲}}\beta_j) & \beta_j<۰\\ \exp(\frac{\lambda}{k\sigma^{۲}}\beta_j) & \beta_j\geq ۰,\end{cases}$$

برای بردار ضرایب رگرسیونی $\boldsymbol{\beta}$ چگالی گامای معکوس با پارامتر a,γ برای پارامتر $\sigma^{۲}$ به صورت

$$\pi(\sigma^{۲})=\frac{\gamma^{a}}{\Gamma(a)}(\frac{۱}{\sigma^{۲}})^{a+۱}\exp(-\frac{\gamma}{\sigma^{۲}}),$$

و چگالی پسینی توأم $\boldsymbol{\beta},\sigma^{۲}$ عبارت است از

$$\pi(\boldsymbol{\beta},\sigma^{۲}|\boldsymbol{Y})\propto\pi(\sigma^{۲})(\sigma^{۲})^{-\frac{n-۱}{۲}}\exp\{\frac{-۱}{۲\sigma^{۲}}(\boldsymbol{Y}-X\boldsymbol{\beta})^{T}(\boldsymbol{Y}-X\boldsymbol{\beta})\}\pi(\boldsymbol{\beta}|\sigma^{۲}).\quad(۲)$$

قضیه ۱. (کوتز و همکاران، ۲۰۰۱) اگر $Z\sim N(۰,۱)$، $X\sim Asymmetric\ Laplace(k,۱)$، $W\sim Exp(۱)$، همچنین Z و W مستقل باشند، آنگاه $X\overset{d}{=}kW+\sqrt{W}Z$.

برهان: متغیر تصادفی W با تابع چگالی e^{-w} را در نظر بگیرید. تابع مشخصه عبارت $kW+W^{\frac{۱}{۲}}Z$

را به شرط W می‌توان به صورت

$$E\Big[e^{it\left(kW+\sqrt{W}Z\right)}\Big]=\int_{۰}^{+\infty}e^{itkw}E\big(e^{it\sqrt{w}Z}\big)e^{-w}dw,$$

به دست آورد. توجه داشته باشید که $E(e^{it\sqrt{w}Z})=\phi_z(t\sqrt{w})=e^{-\frac{۱}{۲}t^۲w}$ و $\phi_z(s)=e^{-\frac{۱}{۲}s^۲}$ تابع مشخصه نرمال استاندارد است. بنابراین

$$E\big[e^{it\left(kW+\sqrt{W}Z\right)}\big]=\int_{۰}^{+\infty}e^{itkw}e^{-\frac{۱}{۲}t^۲w}e^{-w}dw=\int_{۰}^{+\infty}e^{-w(۱+\frac{۱}{۲}t^۲-itk)}dw\ ,$$

که با حل این انتگرال به سادگی تابع مشخصه تابع چگالی لاپلاس نامتقارن به صورت

$$\psi(t)=\frac{۱}{۱+\frac{۱}{۲}t^۲-itk},$$

به دست می‌آید. چون بین توزیع‌های احتمال و توابع مشخصه رابطه یک به یک وجود دارد یا به عبارت دیگر هر دو متغیر تصادفی $X_۱,X_۲$ دارای توزیع احتمال یکسان هستند اگر و فقط اگر تابع مشخصه‌های این دو یکسان باشند، می‌توان نتیجه گرفت $X\overset{d}{=}kW+W^{\frac{۱}{۲}}Z$. با توجه به قضیه ۱ رابطه

$$f_X(x)=\int_{۰}^{+\infty}f_Z(\frac{x-kw}{\sqrt{w}})\frac{۱}{\sqrt{w}}f_W\big(w\big)dw=\int_{۰}^{+\infty}\frac{۱}{\sqrt{۲\pi w}}e^{-\frac{۱}{۲}(\frac{x-kw}{\sqrt{w}})^۲}e^{-w}\ dw,$$

برای تابع چگالی لاپلاس نامتقارن برقرار است. بنابراین با فرض $\tau_j^۲=w$ در قضیه ۱، چگالی پیشینی لاپلاس نامتقارن مرتبط با پارامتر ضرایب مدل رگرسیونی (۱) را به صورت

$$\pi(\boldsymbol{\beta},\tau^{\mathbf{2}}|\lambda,\sigma^{\mathbf{2}})\propto\prod_{\mathbf{j=1}}^{\mathbf{p}}\pi(\boldsymbol{\beta}|\tau_{\mathbf{j}}^{\mathbf{2}},\sigma^{\mathbf{2}})\pi(\tau_{\mathbf{j}}^{\mathbf{2}}|\lambda),$$

می‌توان نوشت. فرض می‌شود $\tau_j^۲$ دارای توزیع نمایی با پارامتر $\frac{\lambda^۲}{۲}$ است. بنابراین

$$\boldsymbol{\beta}\big|\tau_۱^۲,\ldots,\ \tau_p^۲,\ \sigma^۲\sim N_p(k\tau,\sigma^{\mathbf{2}}\mathbf{D}_\tau),\tau=(\tau_{\mathbf{1}}^{\mathbf{2}},\ldots,\tau_{\mathbf{p}}^{\mathbf{2}}),\mathbf{D}_\tau=\mathbf{diag}(\tau).\qquad(۳)$$

با در نظر گرفتن چگالی (۳) در چگالی پسینی توأم (۲)، توزیع توأم به‌دست آمده، به منظور استنباط در مورد پارامترهای مدل رگرسیونی بیزی استفاده می‌شود. ابتدا برای تولید نمونه‌هایی از تابع چگالی پسینی توأم با استفاده از الگوریتم گیبس، باید توابع چگالی پسینی حاشیه‌ای برای پارامترهای $\sigma^۲, \tau, \boldsymbol{\beta}$ محاسبه شوند. برای یافتن چگالی پسینی حاشیه‌ای پارامتر $\boldsymbol{\beta}$، ابتدا عباراتی از چگالی پسینی توأم (۲) که شامل $\boldsymbol{\beta}$ هستند در نظر گرفته می‌شود:

$$
\begin{aligned}
&\exp\big(\frac{-۱}{۲\sigma^۲}(\boldsymbol{Y}-X\boldsymbol{\beta})^T(\boldsymbol{Y}-X\boldsymbol{\beta})\pi(\boldsymbol{\beta}|\sigma^۲)\big)\\
&=\prod_{j=۱}^{p}\frac{۱}{\sqrt{۲\pi\sigma^۲\tau_j^۲}}\exp\big(-\frac{۱}{۲\sigma^۲\tau_j^۲}(\beta_j-k\tau_j^۲)^۲\big)\\
&\sim\exp\big(\frac{-۱}{۲\sigma^۲}(\boldsymbol{Y}-X\boldsymbol{\beta})^T(\boldsymbol{Y}-X\boldsymbol{\beta})\big)\exp\big(\frac{-۱}{۲\sigma^۲}(\boldsymbol{\beta}-k\tau)^{\mathbf{T}}\mathbf{D}_\tau^{-\mathbf{۱}}(\boldsymbol{\beta}-\mathbf{k}\tau)\big)\\
&=\exp\Big[-\frac{۱}{۲\sigma^۲}\big[(\boldsymbol{\beta}'D_\tau^{-۱}\boldsymbol{\beta}+(k\tau)'\mathbf{D}_\tau^{-\mathbf{۱}}(\mathbf{k}\tau)-(\mathbf{k}\tau)'\mathbf{D}_\tau^{-\mathbf{۱}}\boldsymbol{\beta}-\boldsymbol{\beta}'\mathbf{D}_\tau^{-\mathbf{۱}}\mathbf{k}\tau)\big]\Big]\\
&=\exp\Big[-\frac{۱}{۲\sigma^۲}\big[\mathrm{y}'\mathrm{y}+\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}-\mathbf{۲}\boldsymbol{\beta}'(\mathbf{X}'\boldsymbol{Y}+\mathbf{D}_\tau^{-\mathbf{۱}}\mathbf{k}\tau)\big]\Big]\\
&\sim\exp\Big[\frac{-۱}{۲\sigma^۲}[(\boldsymbol{\beta}-A^{-۱}(X'\boldsymbol{Y}+D_\tau^{-۱}(k\tau)))]^{\mathbf{T}}\mathbf{A}(\boldsymbol{\beta}-\mathbf{A}^{-\mathbf{۱}}(\mathbf{X}'\boldsymbol{Y}+\mathbf{D}_\tau^{-\mathbf{۱}}(\mathbf{k}\tau)))\\
&+\mathrm{y}'(\mathbf{I}-(\mathbf{X}'+(\mathbf{k}\tau)'\mathbf{D}_\tau^{-\mathbf{۱}})\mathbf{A}^{-\mathbf{۱}}(\mathbf{X}+\mathbf{D}_\tau^{-\mathbf{۱}}\mathbf{k}\tau))\boldsymbol{Y}\Big],
\end{aligned}
$$

بنابراین تابع چگالی پسینی حاشیه‌ای برای بردار پارامتر ضرایب $\boldsymbol{\beta}$ ، چگالی نرمال با پارامتر میانگین $A^{-۱}(X'\boldsymbol{Y}+D_\tau^{-۱}(k\tau))$ و واریانس $\sigma^۲A^{-۱}$ است. برای یافتن تابع چگالی پسینی حاشیه‌ای $\sigma^۲$ عبارات شامل $\sigma^۲$ از تابع چگالی پسینی توأم (۲) جدا می‌شوند، بنابراین

$$
\begin{aligned}
&\big(\frac{۱}{\sqrt{۲\pi\sigma^۲}}\big)^n\exp\big\{\frac{-۱}{۲\sigma^۲}(\boldsymbol{Y}-X\boldsymbol{\beta})^T(\boldsymbol{Y}-X\boldsymbol{\beta})\big\}\\
&\times\prod_{j=۱}^{p}\exp\big(\frac{۱}{\sqrt{۲\pi\sigma^۲\tau_j^۲}}\big)\exp\big(\frac{-۱}{۲\sigma^۲}(\boldsymbol{\beta}-k\tau)^{\mathbf{T}}\mathbf{D}_\tau^{-\mathbf{۱}}(\boldsymbol{\beta}-\mathbf{k}\tau)\big)\big(\frac{\mathbf{1}}{\boldsymbol{\sigma^۲}}\big)^{\mathbf{a+1}}\exp\big(\frac{-\gamma}{\boldsymbol{\sigma^۲}}\big)\\
&\propto(\sigma^۲)^{-\frac{n}{۲}-\frac{p}{۲}-a-۱}\exp\Big[-\frac{۱}{\sigma^۲}\big(\frac{(\boldsymbol{Y}-X\boldsymbol{\beta})^T(\boldsymbol{Y}-X\boldsymbol{\beta})}{۲}\\
&+\frac{۲\sigma^۲(\boldsymbol{\beta}-k\tau)^{\mathbf{T}}\mathbf{D}_\tau^{-\mathbf{۱}}(\boldsymbol{\beta}-\mathbf{k}\tau)}{۲}+\gamma\big)\Big].
\end{aligned}
$$

در نتیجه

$$\sigma^{۲}|\boldsymbol{\beta},\tau \sim \mathbf{I\Gamma}\big(\frac{\mathbf{n}}{\mathbf{2}}+\frac{\mathbf{p}}{\mathbf{2}}+\mathbf{a},\frac{(\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathbf{T}}(\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta})+(\boldsymbol{\beta}-\mathbf{k}\tau)^{\mathbf{T}}\mathbf{D}_{\tau}^{-\mathbf{1}}(\boldsymbol{\beta}-\mathbf{k}\tau)}{\mathbf{2}}+\gamma\big).$$

برای یافتن تابع چگالی پسینی حاشیه‌ای $\tau_j^{۲}$، عبارات شامل $\tau_j^{۲}$ از تابع چگالی پسینی (۲) جدا شده و داریم:

$$\prod_{j=۱}^{p}(۲\pi\sigma^{۲}\tau_j^{۲})^{-\frac{۱}{۲}}\exp\big(-\frac{۱}{۲\tau_j^{۲}}(\beta_j-k\tau_j^{۲})^{۲}\big)\frac{\lambda^{۲}}{۲}\exp(\frac{\lambda^{۲}}{۲}\tau_j^{۲})$$
$$\sim\prod_{j=۱}^{p}(\sigma^{۲}\tau_j^{۲})^{-\frac{۱}{۲}}\exp\big(-\frac{۱}{۲\sigma^{۲}\tau_j^{۲}}(\beta_j-k\tau_j^{۲})^{۲}\big)\exp(\frac{\lambda^{۲}}{۲}\tau_j^{۲}).$$

برای تولید نمونه از $\tau_j^{۲}$ از الگوریتم متروپلیس هستینگ استفاده می‌شود.

۳ برآورد پارامتر تنظیم λ

برای برآورد پارامتر تنظیم مدل رگرسیونی تاوانیده می‌توان از روش تعیین ابر پیشین برای پارامتر λ استفاده کرد. در این روش می‌توان به جای λ ، $\lambda^{۲}$ در نظر گرفته و کلاسی از چگالی‌های پیشینی گاما به صورت

$$\pi(\lambda^{۲})\propto\frac{\delta^{۲}}{\Gamma(r)}(\lambda^{۲})^{r-۱}e^{-\delta\lambda^{۲}},r>۰,\delta>۰,$$

برای $\lambda^{۲}$ تعیین نمود. این چگالی پیشینی را نیز وارد الگوریتم گیبس کرده و طبق روال قبل عباراتی که شامل $\lambda^{۲}$ هستند را جدا کرده و به صورت

$$(\lambda^{۲})^{p+r-۱}\exp(-\lambda^{۲}(\frac{۱}{۲}\sum_{j=۱}^{p}\tau_j^{۲}+\delta)).$$

نوشت. با توجه به فرم این تابع، می‌توان گفت تابع چگالی توزیع حاشیه‌ای پسینی برای $\lambda^{۲}$ توزیع گاما با پارامتر شکل $p+r$ و پارامتر مقیاس $\sum_{j=۱}^{p}\frac{\tau_j^{۲}}{۲}+\delta \quad ,\delta>۰$ است. لازم به ذکر است که تابع چگالی پسینی حاشیه‌ای پارامترهای $\sigma^{۲},\tau,\boldsymbol{\beta}$ بدون تغییر باقی می‌مانند.

۴ مطالعه شبیه‌سازی

در یک مطالعه شبیه‌سازی چهار مدل رگرسیونی بیزی، شامل رگرسیون ستیغی بیزی هیانگ (۱۹۷۵)، رگرسیون شبکه ارتجاعی بیزی لی و لین (۲۰۱۰)، رگرسیون لاسوی بیزی پارک و کسلا (۲۰۰۸) و رگرسیون لاسوی بیزی نامتقارن برای انتخاب مدل بهینه رگرسیونی، با هم مقایسه می‌شوند. طبق فرمولبندی رگرسیون ستیغی بیزی، مدل بیز سلسله مراتبی به صورت

$$\begin{aligned} \mathrm{y}|\mathrm{X},\boldsymbol{\beta},\sigma^2 &\sim N_n(X\boldsymbol{\beta},\sigma^{۲}I_n),\\ \boldsymbol{\beta}|\tau^{\mathbf{2}} &\sim \prod_{j=۱}^{p} N(۰,\tau_j^{۲}),\\ \tau_j^{۲} &\sim Inv-\chi^{۲}(v,s^{۲}), \end{aligned}$$

و در مدل رگرسیون شبکه ارتجاعی بیزی، مدل سلسله مراتبی به صورت

$$\begin{aligned} \mathrm{y}|\beta\,,\sigma^2 &\sim N\,(\,X\beta\,,\,\sigma^{۲}\,I_n\,),\\ \beta|\tau,\sigma^{۲} &\sim \prod_{j=۱}^{\rho} N\left(۰\;,\;(\frac{\lambda_{۲}}{\sigma^{۲}}\frac{\tau_j}{\tau_j-۱})^{-۱}\right),\\ \tau|\sigma^{۲} &\sim \prod_{j=۱}^{\rho} TG\,(\frac{۱}{۲}\;,\;\frac{۸\lambda_{۲}\sigma^{۲}}{\lambda_{۱}^{۲}}\;,\;(۱,\,\infty)\,),\\ \pi(\sigma^{۲}) &\sim \frac{۱}{\sigma^{۲}}, \end{aligned}$$

تعریف می‌شود که TG بیانگر توزیع گامای بریده شده[1] است. در ادامه شبیه‌سازی با دو مثال متفاوت در شرایط مختلف از نظر همبستگی میان متغیرهای توضیحی، تعداد پارامترها و تعداد مشاهدات مدل بررسی شده است. برای حل مسئلهٔ انتخاب مدل بیزی بهینه، ملاکهایی از قبیل DIC بردلی و همکاران (۲۰۰۲) ، BIC اصلاح شده (mBIC) بوگدان و همکاران (۲۰۰۴) و BIC تعمیم یافته (EBIC) چن و چن (۲۰۰۸) پیشنهاد شده‌اند.

فرض کنید برای مشاهدات $\boldsymbol{Y}$، $f(t|\theta)$ تابع چگالی و $\theta=(\theta_{\mathbf{1}},\ldots,\theta_{\mathbf{p}})$ بردار پارامترهای مدل

[1] Truncated Gamma distribution

است. ملاک EBIC به صورت

$$\text{EBIC} = -۲\log L(\theta|\boldsymbol{Y}) + p\log n + ۲\gamma\log p,$$

تعریف شود، که در آن $L(\theta|\boldsymbol{Y})$ تابع درستنمایی و $\gamma \in [۰, ۱]$ پارامتر تنظیم‌کننده است. ملاک BIC حالت خاصی از EBIC با $\gamma = ۰$ و ملاک mBIC نیز معادل با EBIC با $\gamma = ۱$ است. ملاک DIC نیز تعمیمی از ملاک AIC برای مسائل انتخاب مدل بیزی است و به صورت

$$\text{DIC} = \overline{D(\theta)} + P_D,$$

تعریف می‌شود، که در آن $\overline{D(\theta)} = -۲\log L(\theta|\boldsymbol{Y})$ را انحراف نامیده و تابعی از θ است. عبارت اول، به صورت امید انحراف تحت تابع چگالی پسینی پارامتر به صورت

$$\overline{D(\theta)} = E_{\theta|\boldsymbol{Y}}\big[D(\theta)\big] = E_{\theta|\boldsymbol{Y}}\big[-۲\log\ L(\theta|\boldsymbol{Y})\big],$$

تعریف می‌شود. مؤلفه دوم، تعداد پارامترهای مؤثر یا P_D را اندازه می‌گیرد که به صورت اختلاف بین میانگین پسین انحراف و انحراف محاسبه شده در میانگین پسین پارامترها $(\overline{\theta})$ تعریف می‌شود. یعنی

$$P_D = \overline{D(\theta)} - D(\overline{\theta}) = E_{\theta|\boldsymbol{Y}}[D(\theta)] - \mathbf{D}(\mathbf{E}_{\theta|\boldsymbol{Y}}[\theta]) = \mathbf{E}_{\theta|\boldsymbol{Y}}[-\mathbf{2}\ln\mathbf{L}(\theta|\boldsymbol{Y})] + \mathbf{2}\ln\mathbf{L}(\overline{\theta}|\boldsymbol{Y}).$$

با دوباره‌آرایی عبارت P_D داریم: $\overline{D} = D(\overline{\theta}) + P_D$، بنابراین $\text{D}IC = D(\overline{\theta}) + ۲P_D$ و محاسبهٔ آن با استفاده از الگوریتم‌های زنجیره مارکف مونت کارلو از قبیل گیبس اغلب بدیهی و ساده است. برآورد $\overline{D}$ با میانگین گرفتن از مقادیر شبیه‌سازی شده از $D(\theta)$ محاسبه می‌شود. مقدار پارامترهای مؤثر P_D را نیز می‌توان با محاسبهٔ $D(\theta)$ در میانگین نمونه‌ای مقادیر شبیه‌سازی شده پارامترها و کم کردن آن از برآورد $D(\overline{\theta})$ به دست آورد.

مثال ۱. مقادیر مشاهدات $X, \boldsymbol{Y}$ با در نظر گرفتن فرضیات $n = ۱۰۰, p = ۱۰$ به این صورت تولید می‌گردد که مشاهدات X را از توزیع نرمال چند متغیره استاندارد نمونه‌گیری کرده و بردار ضرایب نیز به صورت $\boldsymbol{\beta} = (۳, -۳, ۵, ۴, -۲/۸, ۳/۲, ۰, \ldots, ۰)$ در نظر گرفته می‌شود. بردار n بعدی $\mathbf{e}$ از مقادیر خطا با نمونه‌گیری از تابع چگالی آلفا چوله‌نرمال با میزان چولگی $k = ۰/۸$ و سپس طبق مدل (۱) مشاهدات

$\boldsymbol{Y}$ به دست می‌آیند. سپس مشاهدات $\boldsymbol{Y}$ را مرکزی کرده به گونه‌ای که میانگین صفر داشته باشند و همچنین مقادیر متغیر کمکی X نیز استاندارد می‌شوند. با استفاده از شبیه‌سازی و انجام الگوریتم گیبس مقادیر ملاکهای DIC و EBIC با $\gamma = ۱$ برای چهار مدل مختلف رگرسیون ریج بیزی، رگرسیون لاسوی بیزی، رگرسیون شبکه ارتجاعی بیزی و رگرسیون لاسوی بیزی نامتقارن در جدول ۱ آمده است. طبق هر دو ملاک، مینیمم مقادیر DIC و EBIC ، مدل با بهترین پیش‌بینی را فراهم می‌کند. معمولا اگر اختلاف مقادیر هر کدام از این ملاکها در دو مدل متفاوت، بیشتر از ۱۰ باشد، مدل با مقادیر DIC و EBIC کمتر به عنوان مدل بهینه انتخاب می‌شود. بنابراین طبق نتایج جدول ۱ می‌توان گفت مدل لاسوی بیزی نامتقارن نسبت به روش‌های دیگر در این داده‌های شبیه‌سازی، عملکرد بهتری در انتخاب متغیر دارد.

جدول ۱. مقادیر DIC و EBIC به ازای اندازه نمونه ۱۰۰ و تعداد پارامترها ۱۰

مدل رگرسیون	DIC	EBIC
ستیغی بیزی	۹۴/۲	۲۳۵/۳۴
لاسوی بیزی	۸۴/۲۳	۱۱۸/۶
شبکه ارتجاعی بیزی	۳۶۸/۲	۴۳۲/۸۷
لاسوی بیزی نامتقارن	۶۲/۲	۱۰۴/۲

مثال ۲. در این مثال مانند مثال ۱ عمل می‌شود، با این تفاوت که فرض می‌شود $n = ۱۰۰$، $p = ۴۰۰$، یعنی $p \gg n$ است. همچنین میان متغیرهای توضیحی، همبستگی به صورت $\mathrm{Cov}(X)_{ij} = ۰/۷^{|i-j|}$ لحاظ شده است. مقادیر DIC و EBIC با $\gamma = ۱$ حاصل از شبیه‌سازی برای چهار مدل مختلف را در جدول ۲ مشاهده می‌کنید. واضح است که ملاک DIC و EBIC برای مدل رگرسیون لاسوی بیزی نامتقارن به طور قابل ملاحظه‌ای کمتر از سایر مدل‌ها است و در نتیجه بهتر از سه مدل دیگر عمل می‌کند.

جدول ۲. مقادیر DIC و EBIC به ازای اندازه نمونه ۱۰۰و تعداد پارامترها ۴۰۰

مدل رگرسیون	DIC	EBIC
ستیغی بیزی	−۱۹۹/۳۱	۲۷۹۰/۶
لاسوی بیزی	۲۱۹/۳۱	۱۹۷۴/۵
شبکه ارتجاعی بیزی	−۳۹۶	۴۸۹۶/۸
لاسوی بیزی نامتقارن	۷۵/۹۸	۱۸۴۵/۲

۵ تحلیل داده واقعی

از آن‌جائی‌که وجود ریبوفلاوین یا ویتامین $B۲$ در بدن انسان، باعث آزادسازی انرژی از کربوهیدرات‌ها، پروتئین و چربی می‌شود و این امر در رشد و ترمیم پوست، مو، ناخن و مفاصل ضروری و باعث حفاظت سیستم ایمنی بدن در مقابل بیماری‌ها است. اندازه‌گیری میزان این ویتامین امری ضروری است. در این اندازه‌گیری متغیر پاسخ، لگاریتم میزان ریبوفلاوین است.

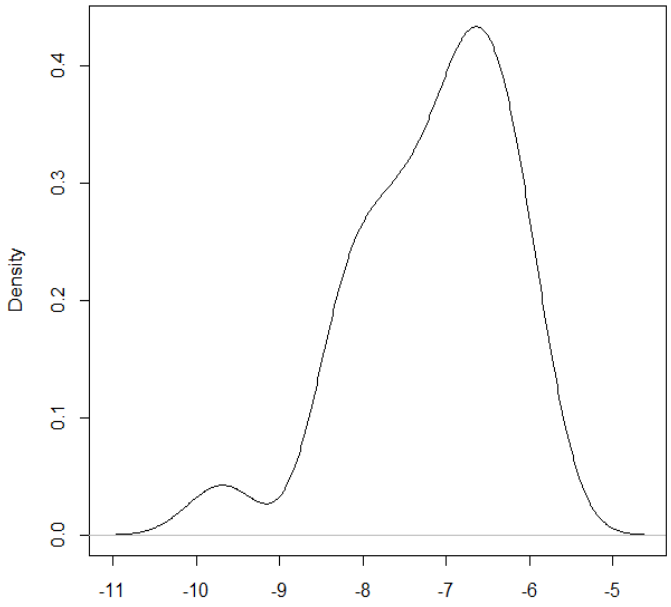
داده‌ها شامل $p = ۴۰۸۸$ ژن متفاوت برای $n = ۷۱$ نمونه است که با توجه به اندازه متغیر و اندازه نمونه، مجموعه داده‌ای با ابعاد بالا محسوب می‌شود (بولمن و همکاران، ۲۰۱۴). مجموعه داده‌های ریبوفلاوین از طریق https://rdrr.io/cran/hdi/man/riboflavin.html قابل دسترس است و نمودار توزیع آن‌ها در شکل ۱ ارائه شده است. با توجه به نمودار، مقداری چولگی در داده‌ها مشاهده می‌شود. همچنین با توجه به مقدار احتمال (P-value= 0.005) به دست آمده از آزمون شاپیرو و ویلک (۱۹۶۵) در سطح خطای ۵ درصد، فرض صفر که نرمال بودن داده‌ها است، رد شده و توزیع داده‌های ریبوفلاوین از توزیع نرمال متفاوت است. پس از اجرای روش‌های رگرسیون لاسو بیزی، رگرسیون لاسو بیزی نامتقارن، رگرسیون ستیغی بیزی و رگرسیون شبکه ارتجاعی بیزی برای داده‌های ریبوفلاوین، نتایج در جدول ۳ ارایه شده است. با توجه به مقادیر DIC، EBIC، AIC، BIC و CV در چهار روش رگرسیونی بیزی، نتیجه می‌شود که روش رگرسیون لاسو بیزی نامتقارن بهتر از بقیه روش‌ها به انتخاب مدل بهینه می‌پردازد.

جدول ۳. مقادیر ملاک‌های DIC، EBIC، AIC، BIC و CV برای داده‌های ریبوفلاوین

مدل رگرسیون	DIC	EBIC	CV	AIC	BIC
ستیغی بیزی	۶۲۵/۱۹	۱۷۸۹۲/۹	۵۶۶/۲	۶۵۰/۱	۶۵۹/۲
لاسوی بیزی	۳۵۴/۷	۱۷۸۵۰/۶	۵۶۳/۴	۳۶۲/۴	۳۶۵/۱
شبکه ارتجاعی بیزی	۵۶۵/۱	۱۷۸۷۲/۲	۵۶۵/۱	۵۷۵/۲	۵۸۰/۳
لاسوی بیزی نامتقارن	۲۶۲/۲	۱۷۶۴۵/۴	۵۶۲/۳	۲۸۹/۱	۲۹۶/۴

بحث و نتیجه‌گیری

تعمیمی از مدل رگرسیون لاسو بیزی با توزیع خطاهای آلفا چوله‌نرمال وقتی $p \gg n$ ارایه شد. در مدل پیشنهادی از تابع چگالی پیشینی لاپلاس نامتقارن برای پارامترهای ضرایب رگرسیونی استفاده شده و



شکل ۱. نمودار توزیع داده‌های ریبوفلاوین

پارامترهای مدل نیز با الگوریتم‌های MCMC برآورد شده است. با توجه به نتایج شبیه‌سازی و تحلیل داده‌های واقعی، می‌توان نتیجه گرفت که بر اساس ملاک‌های DIC، EBIC، AIC، BIC و CV مدل پیشنهادی رگرسیون لاسوی بیزی نامتقارن در مقایسه با مدل‌های رگرسیون لاسوی بیزی، رگرسیون ستیغی بیزی و رگرسیون شبکه ارتجاعی بیزی برازش بهتری برای داده‌ها فراهم می‌کند.

تقدیر و تشکر

نویسندگان مقاله از سر دبیر، داوران و ویراستار محترم مجله در ارزیابی این مقاله قدردانی و تشکر می کنند.

مراجع

آرست، م.، آرشی، م. و ربیعی، م. (۱۳۹۸)، مطالعه رفتار برآوردگر انقباضی تحت یک قید خطی در مدل رگرسیون تاوانیده، مجله علوم آماری، **۱۳**، ۱-۱۴.

Azzalini, A. (1985), A Class of Distributions Which Includes the Normal Ones, *Scandinavian Journal of Staistics*, **12**, 171-178.

Bogdan, M., Doerge, R. W. and Ghosh, J. (2004), Modifying the Schwarz Bayesian Information Creterion to Locate Multiple Interacting Quantitive Trait Loci, *Genetics*, **167**, 989-999.

Buhlmann,P., Kalisch, M. and Meier, L. (2014), High-Dimentional Statistics with a View Towards Applications in Biology,*Annual Review of Statistics and Its Applications*, **1**, 255-278.

Chen, J., and Chen, Z. (2008), Extended Bayesian Information Criteria for Model Selection with Large Model Space, *Biometrika*, **95**, 759–771.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004), Least Angle Regression, *Annals of Statistics*, **32**, 407–499.

Elal-Olivero, D. (2010), Alpha–Skew–Normal Distribution, *Proyecciones Journal of Mathematics*, **29**, 224 - 240.

Hoerl, A. E., and Kennard, R. W. (1970), Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, **12**, 55–67.

Hsiang, T. C. (1975), A Bayesian View on Ridge Regression, *The Statistician*, **24**, 267–268.

Kotz, S., Koubowski, T. and Podgorski, K. (2001), *The Laplace Distribution and Generalizations*, Birkhauser Basel, New York.

Knight, K., and Fu, W. (2000), Asymptotics for LASSO-type Estimators, *Annals of Statistics*, **28**, 1356–1378.

Li, Q., and Lin, N. (2010), The Bayesian Elastic Net, *Bayesian Annals of Statistics*, **5**, 151-170.

Mallick, H., and Nengjun,Y. (2013), Bayesian Methods for High Dimensional Linear Models, *Journal of Biometrics & Biostatistics*, **1**, 1-13.

Marchenko, V., and Genton, M. (2010), A Suite of Commands for Fitting the Skew-normal and Skew-t Models, *Stata Journal*, **10**, 507-539.

Neter, J., Kutner, H., Wasserman W. and Nachtscheim, J. (1996), *Applied Linear Regression Models*, McGraw-Hill College.

Park, T., and Casella, G. (2008), The Bayesian LASSO, *Journal of the American Statistical Association*, **103**, 681-687.

Sahu, S. K., and Dey, M. D. (2003), A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models, *Canadian Journal of Statistics*, **31**, 129-150.

Spiegelhalter, D., Bradley, P., and Van der Linde, A. (2002), Bayesian Measures of Model Complexity and Fit (with Discussion), *Journal of the Royal Statistical Society*, Series B, **64**, 583–639.

Shapiro, S. S. and Wilk, M. B. (1965), An Analysis of Variance Test, *Biometrika*, **52**, 591-611.

Tibshirani, R. (1996), Regression Shrinkage and Selection via the LASSO, *Journal of the Royal Statistical Society*, Series B, **58**, 267–288.

Zou, H., and Hastie, T. (2005), Regularization and Variable Selection via the Elastic Net, *Journal of the Royal Statistical Society* , **67**, 301–320.

Journal of Statistical Sciences, Spring and Summer, 2021
Vol. 15, No. 1, pp 81-96
DOI: 10.29252/jss.15.1.149

Bayesian LASSO Regression with Asymmetric Error in High Dimensional

Khadem Bashiri, Z., Shadrokh, A., Yarmohammadi, M.
Department of Statistics, Payame Noor University, Tehran, Iran.

Abstract: One of the most critical discussions in regression models is the selection of the optimal model, by identifying critical explanatory variables and negligible variables and more easily express the relationship between the response variable and explanatory variables. Given the limitations of selecting variables in classical methods, such as stepwise selection, it is possible to use penalized regression methods. One of the penalized regression models is the Lasso regression model, in which it is assumed that errors follow a normal distribution. In this paper, we introduce the Bayesian Lasso regression model with an asymmetric distribution error and the high dimensional setting. Then, using the simulation studies and real data analysis, the performance of the proposed model's performance is discussed.

Keywords: Penalized regression, Bayesian LASSO regression, Asymmetric Laplace distribution, Alpha-skew normal distribution.

Mathematics Subject Classification (2010): 62F15, 62G35, 65C10.