مجله مدل‌سازی پیشرفته ریاضی
سال ۱۴۰۰، دوره ۱۱، شماره ۲، ص ۳۶۵–۳۸۱

شاپا چاپی: ۲۲۵۱–۸۰۸۸
DOI: 10.22055/jamm.2021.35112.1859

مدل‌سازی رگرسیونی به روش تی لاسو بیزی

زهرا خادم بشیری ، علی شادرخ [×]، مسعود یارمحمدی

گروه آمار، دانشگاه پیام نور، صندوق پستی ۱۹۳۹۵–۴۶۹۷ ، تهران، ایران

تاریخ دریافت: ۱۳۹۹/۶/۳۰ تاریخ پذیرش: ۱۳۹۹/۱۱/۲
دبیر مسئول: غلامرضا محتشمی برزادران

چکیده: انتخاب مدل بهینه یکی از بحث‌های مهم در مدل‌های رگرسیونی است. هدف روش‌های انتخاب مدل بهینه در مدل‌های رگرسیونی این است که متغیرهای توضیحی مهم و متغیرهای قابل اغماض را تعیین نموده و رابطه‌ی بین متغیر پاسخ و متغیرهای توضیحی را به‌طور ساده‌تر بیان کند. با توجه به محدودیت‌های فرآیندهای کلاسیک انتخاب متغیر از قبیل انتخاب گام به گام، می‌توان از روش‌های رگرسیون تاوانیده استفاده کرد. یکی از مدل‌های رگرسیون تاوانیده، رگرسیون لاسو است که در آن فرض می‌شود خطاها از توزیع نرمال پیروی می‌کنند. برای تحلیل آماری مجموعه داده‌ها در حضور مشاهدات دور افتاده، می‌توان به جای توزیع نرمال از توزیع تی–استیودنت برای خطا استفاده کرد. در این مقاله، روش انتخاب متغیری تحت عنوان مدل رگرسیون تی لاسو بیزی برای تحلیل داده‌ها در حضور مشاهدات دور افتاده، پیشنهاد می‌دهیم. مدل رگرسیون تی لاسو بیزی با دو نمایش متفاوت از تابع چگالی پیشین لاپلاس برای ضرایب مدل رگرسیونی مورد بررسی قرار می‌گیرد، به این صورت که ابتدا تابع چگالی لاپلاس به صورت توزیع آمیخته–مقیاس نرمال و سپس به صورت توزیع آمیخته–مقیاس یکنواخت نمایش داده می‌شود. سپس با استفاده از روش‌های شبیه‌سازی و تحلیل داده‌های واقعی، ارجحیت روش رگرسیون تی لاسو بیزی با نمایش تابع چگالی لاپلاس به صورت آمیخته–مقیاس یکنواخت نسبت به نمایش آمیخته–مقیاس نرمال نشان داده می‌شود.

واژه‌های کلیدی: رگرسیون لاسو بیزی، رگرسیون تاوانیده، نمایش آمیخته– مقیاس یکنواخت، الگوریتم گیبز، انتخاب متغیر.

رده‌بندی ریاضی: 62F15; 62J07;82B80

۱ مقدمه

تحلیل رگرسیون از پرکاربردترین روش‌ها برای برازش مدل‌ها به داده‌ها است. از ساده‌ترین و در عین حال کاراترین روش‌های برآورد پارامترهای مدل رگرسیونی، روش کمترین توان‌های دوم خطا است. یکی از ایرادهای وارد به روش کمترین توان‌های دوم خطا، توانایی تفسیر آن است. 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، زیر مجموعه کوچکی از متغیرها که دارای بیش‌ترین تاثیر بوده را انتخاب و به برآورد آن‌ها پرداخته

[×] نویسنده مسئول مقاله: رایانامه: Shadrokh.ali@gmail.com

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 مواجه 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 گرفته اند. 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 طوری‌که می‌توان امیدوار بود در نهایت میانگین توان دوم خطا کاهش یابد. [۳] از جمله برآوردگرهای انقباضی رایج در برآورد پارامترهای مدل رگرسیونی می‌توان به برآوردگر ریج اشاره نمود که توسط هورل و کنارد معرفی شد [۹]. هورل و کنارد برآوردگر رگرسیونی «ریج» را معرفی کردند که دروازه ورود به دنیای «برآوردگرهای تاوانیده» بر اساس روش منظم‌سازی [۲۷] بود. رگرسیون ریج، مقدمه‌ای برای راهیابی به دنیای مسئله برآورد و انتخاب متغیر است. این رگرسیون، با مشکل هم‌خطی در مدل‌های خطی مبارزه می‌کند و بر این اساس برآورد تاوانیده متولد شد. با توجه به حضور همه متغیرها در برآوردگر ریج، تفسیرپذیری آن به سادگی امکان پذیر نیست. عضو دیگری از این کلاس، برآوردگر لاسو [۱۰] است. تیبشیرانی به جای استفاده از تاوان $L_۲$ توان‌های دوم خطا را نسبت به تابع تاوان $L_۱$ کمینه نمود. این برآوردگر منجر به پیدایش برآوردگرهای جدیدی مانند برآوردگر مشتق پذیر کوتاه شده هموار [۲۶] ، الاستیک‌نت [۲۸] لاسوی سازوار [۲۹] ، لاسوی آستانه سخت [۲۵] گردید. استفاده از تاوان $L_۱$ هر ضریب را به سمت صفر منقبض کرده و متغیرهای اضافی را دقیقا صفر می‌کند . در واقع، برآوردگر لاسو، همزمان هم انتخاب متغیر انجام می‌دهد و هم ضرایب را منقبض می‌کند. یک کاربرد جالب از برآوردگرهای لاسو در مدل‌های تنک (مدل‌هایی که تعداد پارامترهای صفر آن زیاد باشد) است [۱]. کاربرد دیگر برآوردگر لاسو، زمانی است که بعد فضای پارامتر بیشتر از بعد فضای نمونه باشد. به دلیل وجود تعداد زیادی از متغیرها در مدل‌های با بعد بالا، تفسیر این مدل‌ها بسیار مشکل است. لذا مسئله انتخاب متغیر نقش بسیار مهمی را در مدل‌سازی آماری با بعد بالا ایفا می‌کند. اخیرا در این زمینه مطالعات زیادی صورت گرفته که برای مثال می‌توان به منابع [۴] و [۲] اشاره کرد. علی‌رغم مزایای برآوردگر لاسو، عملکرد روش لاسو به عنوان روشی برای انتخاب مدل بهینه در حالتی که مشاهدات شامل داده دور افتاده هستند و توزیع متغیر خطا نرمال در نظر گرفته می‌شود، ضعیف است [۵] . با توجه به اینکه مشاهدات دور افتاده تاثیر زیادی بر روی مدل برازش شده و استنباط‌های مربوط به آن دارند، استفاده از برآوردگرهای استوار نسبت به حضور داده‌های دور افتاده، از اهمیت زیادی برخوردار است. بنابراین جمعی از محققان به برآورد انقباضی که نسبت به حضور داده‌های دور افتاده استوار هستند، پرداخته و توزیع تی-استیودنت را به عنوان جایگزینی مناسب برای توزیع نرمال در نظر گرفتند [۶] و [۷] . در این مقاله، روش رگرسیون تی لاسو بیزی، در حالتی‌که مشاهدات شامل داده‌های دور افتاده هستند یا توزیع خطا رفتاری غیر نرمال دارد، پیشنهاد می‌شود. در این راستا، در بخش دوم مقاله مروری بر مدل رگرسیون لاسو بیزی نرمال خواهیم داشت. در بخش سوم، به جزییات مدل رگرسیون تی لاسو بیزی با دو نمایش متفاوت از تابع چگالی لاپلاس به صورت آمیخته-مقیاس نرمال و آمیخته-مقیاس یکنواخت پرداخته شده است و مدل بیز سلسله مراتبی به دست می‌آید. در ادامه، الگوریتم گیبز برای برآورد پارامترهای مدل رگرسیون تی لاسو بیزی مورد بررسی قرار می‌گیرد. در بخش چهارم، با استفاده از روش‌های شبیه‌سازی و تحلیل داده‌های واقعی، دو روش مورد بررسی قرار گرفته و نشان خواهیم داد که مدل رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس یکنواخت، عملکرد رضایت‌بخشی در مقایسه با مدل با نمایش آمیخته-مقیاس نرمال دارد. مقایسه‌ی مدل‌ها براساس معیار اطلاع انحراف و میانگین توان دوم خطا و بررسی همگرایی الگوریتم گیبز به روش هیدلبرگر و ولچ انجام شده است. بخش پنجم نیز به نتایج مقاله اختصاص داده شده است.

۲ مدل رگرسیون لاسو بیزی

مدل رگرسیون خطی چندگانه به صورت $\mathbf{y} = \boldsymbol{\beta} X + \varepsilon$ و $\varepsilon \sim N(\circ, \sigma^۲ I)$ را در نظر بگیرید که $\mathbf{y} = (y_۱, ..., y_n)^T$ برداری از مقادیر پاسخ، $X = (x_۱, ..., x_n)^T$ ماتریسی $n \times p$ از متغیرهای توضیحی $x_i \in R^p$، $\boldsymbol{\beta} = (\beta_۱, ..., \beta_p)^T$ برداری از ضرایب رگرسیونی و ε برداری از خطاهای تصادفی است. در مدل رگرسیون خطی چندگانه 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. در حضور همبستگی بین متغیرهای توضیحی، مقادیر واریانس‌ها و کواریانس‌ها برای برآوردگرهای کمترین توان‌های دوم خطای ضرایب رگرسیونی بزرگ خواهد شد. در واقع اعضای روی قطر اصلی ماتریس $(X'X)^{-۱}$ عبارتند از $۱/(۱ - R_j^۲), j = ۱, ..., p$ که در آن $R_j^۲$ ضریب تعیین چندگانه از رگرسیون متغیر توضیحی jام نسبت به $p - ۱$ متغیر توضیحی باقیمانده است که در صورت وجود همخطی شدید میان متغیرهای توضیحی، $R_j^۲$ نزدیک به واحد خواهد بود و چون واریانس ضریب رگرسیونی jام برابر با $V(\hat{\beta}_j) = (۱ - R^۲)^{-۱}\sigma^۲$ است، همخطی شدید موجب می‌شود که واریانس برآورد کمترین توان‌های دوم ضریب رگرسیونی β_j بسیار زیاد شود [۸]. برای بهبود دقت پیش‌بینی برآوردگر کمترین توان‌های دوم خطا، روش‌های رگرسیون کمترین توان‌های دوم خطا با تاوان‌های مختلف ایجاد شده است. یکی از روش‌های رگرسیون تاوانیده، رگرسیون ریج [۹] است. رگرسیون ریج مجموع توان‌های دوم خطا را با محدودیت $\sum_{j=۱}^{p} |\beta_j|^۲ \leq t$ مینیمم می‌کند. مدل رگرسیون ریج با انقباض ضرایب برآوردگر کمترین توان‌های دوم خطا، اغلب دقت پیش‌بینی بهتری فراهم می‌کند اما توانایی انتخاب متغیر را نداشته و همه‌ی متغیرهای توضیحی را در مدل نگه می‌دارد. در میان روش‌های رگرسیون تاوانیده، پرکاربردترین روش آماری، روش لاسو است [۱۰]. در روش لاسو

برآورد پارامترها در رابطه

$$\beta_{LASSO} = arg_{\beta} min\{(\mathbf{y} - X\beta)^T(\mathbf{y} - X\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j|\}$$

صدق می‌کند، که $\lambda \geq 0$ پارامتر تعدیل‌کننده‌ی غیر منفی است و برای کنترل سطح انقباض به کار می‌رود. به این معنی که اگر مقدارش برابر صفر باشد، مدل به رگرسیون عادی تبدیل شده و همه‌ی متغیرهای توضیحی در مدل حضور خواهند داشت و اگر مقدار آن افزایش یابد، تعداد متغیرهای توضیحی در مدل کاهش می‌یابد. تعیین مقدار برای این پارامتر معمولا توسط روش اعتبارسنجی متقابل انجام می‌شود. یک مسئله مهم در روش رگرسیون لاسو، تخمین خطاهای استاندارد برای برآورد پارامترهای ضرایب رگرسیونی است، زیرا الگوریتم‌هایی از قبیل الگوریتم لارس [۳۰]، فقط برآوردهای نقطه‌ای پارامترهای رگرسیونی را فراهم کرده و استفاده از بوت استرپ برای محاسبه برآورد خطاهای استاندارد و در نهایت محاسبه بازه اطمینان برای پارامترهای مدل رگرسیونی از لحاظ محاسباتی مشکل است [۳۱]. روش بوت استرپ را نایت و فو [۲۴] برای برآورد خطای استاندارد مورد استفاده قرار داده و تأکید کردند، وقتی برآورد برخی از ضرایب رگرسیونی صفر می‌شوند، برآوردهای بوت‌استرپ خطای استاندارد به طور 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 [۱۰]، 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. طبق پیشنهاد تیبشیرانی [۱۰]، درحالتی‌که توزیع‌های پیشین لاپلاس مستقل برای ضرایب رگرسیونی به صورت $\pi(\boldsymbol{\beta}|\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_j|/\sqrt{\sigma^2}}$ در نظر گرفته شوند، برآورد لاسو به عنوان برآورد نمای توزیع پسین β به صورت $\hat{\boldsymbol{\beta}_L} = arg_{\boldsymbol{\beta}} max\ \pi(\boldsymbol{\beta}|y)$ است. تیبشیرانی اشاره داشت که توزیع پیشین لاپلاس را می‌توان به عنوان آمیخته‌ای مقیاسی از توزیع نرمال و گاما و با معرفی بردار پارامترهای $\boldsymbol{\tau} = (\tau_1, ..., \tau_p)$ به صورت

$$\boldsymbol{\beta}|\sigma^2,\ \tau \sim N_p(0_p, \sigma^2 D_{\boldsymbol{\tau}}), \qquad \tau_j \sim Exp(\frac{\lambda}{2})$$

در نظر گرفت، که $D_{\boldsymbol{\tau}} = Diag(\tau_1, ..., \tau_p)$ است [۱۱]. به دلیل مثبت بودن پارامتر σ^2 چگالی پیشین گاما به صورت $\pi(\sigma^2) \propto 1/\sigma^2$ و مستقل از $\boldsymbol{\tau}$ برای پارامتر σ^2 پیشنهاد شده و برای برآورد پارامترهای مدل رگرسیونی از الگوریتم نمونه‌گیرگیبز استفاده می‌شود. با توجه به توصیه تیبشیرانی، این چارچوب مدل، پیشین‌ها و نمونه‌گیر گیبز تحت عنوان مدل رگرسیون لاسو بیزی توسط پارک و کسلا [۱۲] مورد بررسی قرار گرفته است.

۳ مدل رگرسیون تی لاسو بیزی

اغلب در مدل‌های رگرسیون خطی، توزیع خطاها به صورت $\varepsilon_i \sim N(0, \sigma^2)$ در نظر گرفته می‌شود. به طور کلی، علاوه بر توزیع نرمال، خطاها می‌توانند از هر توزیع دیگری نیز باشند. برای مثال توزیع t–استیودنت که یک توزیع دم سنگین است و تعمیمی مفید از توزیع نرمال می‌باشد که می‌تواند برای مدل‌سازی آماری در حضور داده‌های دور افتاده استفاده شود. البته سازگاری پیش‌بینی رگرسیون لاسو به فرضیه‌ی نرمال بودن خطاها نیاز ندارد وکافی است خطاها دارای میانگین صفر و واریانس متناهی باشند [۲۲]. در این پژوهش، با در نظرگرفتن توزیع خطاها به صورت $\varepsilon_i \overset{i.i.d}{\sim} t_v(0, \sigma^2)$ ، به تعمیم مدل رگرسیون لاسو بیزی برای مشاهدات در حضور داده دور افتاده و تحت عنوان مدل رگرسیون تی لاسو بیزی پرداخته می‌شود. قبل از بررسی جزئیات مدل رگرسیون تی لاسو بیزی، به بیان سه لم کاربردی و مهم در این زمینه پرداخته می‌شود.

قضیه ۱.۳. (نمایش توزیع تی– استیودنت به صورت توزیع آمیخته–مقیاس نرمال)
فرض کنید $\mathbf{y}|l$ دارای توزیع نرمال با میانگین ۰ و واریانس l باشد. تابع توزیع l را نیز به صورت گامای معکوس با پارامترهای $v/2$ و $v/2$ در نظر بگیرید. بنابراین توزیع حاشیه‌ای $\mathbf{y}$ ، توزیع t– استیودنت با درجه آزادی v است [۱۳].
اثبات: به پیوست (آ) رجوع شود.

قضیه ۲.۳. (نمایش توزیع لاپلاس به صورت توزیع آمیخته–مقیاس نرمال)
فرض کنید $E(1)$ نشان‌دهنده‌ی توزیع نمایی استاندارد با میانگین یک، $Laplace(0, 1)$، بیانگر توزیع لاپلاس استاندارد با میانگین صفر و واریانس یک و $N(\mu, \sigma^2)$ دارای توزیع نرمال با میانگین μ و واریانس σ^2 است. بنابراین داریم:

$$V \sim 2E(1), \quad L|V \sim N(0, V) \quad \Rightarrow \quad L \sim Laplace(0, 1)$$

به عبارت دیگر، اگر متغیر تصادفی E با توزیع نمایی استاندارد، مستقل از متغیر تصادفی Z با توزیع نرمال استاندارد باشد، آنگاه

$$L = \sqrt{2E}Z \sim Laplace(0, 1).$$

با در نظر گرفتن پارامتر مقیاس λ، داریم:

$$V \sim ۲/\lambda^۲ E(۱), \quad L|V \sim N(\circ, V) \Rightarrow L \sim \lambda Laplace(\circ, ۱).$$

اثبات: به پیوست (آ) رجوع شود.

قضیه ۳.۳. (نمایش چگالی لاپلاس به صورت توزیع آمیخته–مقیاس یکنواخت)
چگالی لاپلاس می‌تواند به صورت آمیخته‌ای مقیاسی از توزیع یکنواخت نوشته شود به طوری که چگالی آمیخته، یک توزیع گاما در نظر گرفته شود، یعنی:

$$\frac{\lambda}{۲}\exp^{-\lambda|x|} = \int_{u>|x|}^{\circ} \frac{۱}{۲\mu}\frac{\lambda^۲}{\Gamma(۲)} u^{۲-۱} e^{-\lambda\mu} d\mu \quad , \quad \lambda > \circ.$$

اثبات: به پیوست (آ) رجوع شود.

۱.۳ مدل بیز سلسله مراتبی برای رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس نرمال

مدل رگرسیونی $y_i = x_i^T\boldsymbol{\beta} + \varepsilon_i,\ \varepsilon_i \overset{i.i.d}{\sim} t_v(\circ, \sigma^۲),\ i = ۱, ..., n$ را در نظر بگیرید. در این مدل، x_i^Tها بردار p بعدی متغیرهای کمکی و $\boldsymbol{\beta} = (\beta_۱, ..., \beta_p)^T$، بردار p بعدی از پارامترهای نامعلوم، $\sigma^۲$ پارامتر مقیاس و درجه آزادی v میزان سنگینی دم‌های توزیع را تعیین می‌کند. با در نظر گرفتن تابع چگالی پیشینی لاپلاس با نمایش آمیخته–مقیاس نرمال برای بردار ضرایب رگرسیونی $\boldsymbol{\beta}$ به صورت

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

و چگالی گامای معکوس با پارامتر r و γ به صورت

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

برای پارامتر $\sigma^۲$، مدل بیز سلسله مراتبی به صورت

$$\begin{aligned}
&\mathbf{y}|X, \boldsymbol{\beta}, v, l \sim N(X\boldsymbol{\beta}, l\sigma^۲, I_n),\\
&l|v \sim IG(\frac{v}{۲}, \frac{v}{۲}),\\
&\beta_i|\tau_۱^۲, ..., \tau_p^۲,\ \sigma^۲ \sim N(\circ_p, \sigma^۲ D_{\boldsymbol{\tau}}),\ D_{\boldsymbol{\tau}} = Diag(\tau_۱^۲, ..., \tau_p^۲),\ j = ۱, ..., p,\\
&\tau_j^۲ \sim \frac{\lambda^۲}{۲}e^{-\lambda^۲\tau_j^۲/۲},\\
&\sigma^۲ \sim G(r, \gamma)
\end{aligned}$$

را خواهیم داشت. با استفاده از مدل بیز سلسله مراتبی، توزیع‌های شرطی کامل پسین برای پارامترهای مدل رگرسیون تی لاسو بیزی برای اجرای الگوریتم گیبز به صورت

$$\begin{aligned}
&\boldsymbol{\beta}|\mathbf{y}, X, \sigma^۲, \boldsymbol{\tau}, l \sim N_p((X'L^{-۱}X + D\tau^{-۱})^{-۱}y, \sigma^۲(X'L^{-۱}X + D\tau^{-۱})^{-۱}),\\
&l|y, X, \beta, \sigma^۲ \sim \prod_{i=۱}^{n} I\Gamma(\frac{۱}{۲} + \frac{v}{۲}, \frac{(y_i - X\beta)}{۲\sigma^۲} + \frac{v}{۲}),\\
&\sigma^۲|\mathbf{y}, X, \boldsymbol{\beta}, \tau, \gamma, r \sim I\Gamma(\frac{n}{۲} + \frac{p}{۲} + r, \frac{(\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta})\boldsymbol{\beta}^T D^{-۱}\tau\boldsymbol{\beta}}{۲} + \gamma),\\
&(۱/\tau_j^۲)|\boldsymbol{\beta}, \sigma^۲, \lambda \sim \prod_{j=۱}^{p} IG(\sqrt{\frac{\lambda^۲\sigma^۴}{\beta_j^۲}}, \lambda^۲)
\end{aligned}$$

به دست می‌آیند. برای بررسی جزئیات محاسبه‌ی این توابع به پیوست (ب) مراجعه شود.

۲.۳ مدل بیز سلسله مراتبی برای رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس یکنواخت

با در نظر گرفتن تابع چگالی پیشینی لاپلاس با نمایش آمیخته-مقیاس یکنواخت برای بردار ضرایب رگرسیونی $\boldsymbol{\beta}$ و چگالی گامای معکوس برای پارامتر $\sigma^۲$ به صورت $\pi(\sigma^۲) = ۱/\sigma^۲$، مدل بیز سلسله مراتبی به صورت

$$\begin{aligned}
&\mathbf{y}|X, \boldsymbol{\beta}, v, l \sim N(X\boldsymbol{\beta}, l\sigma^۲, I_n),\\
&l|v \sim IG(\frac{v}{۲}, \frac{v}{۲}),\\
&\boldsymbol{\beta}|u, \sigma^۲ \sim \prod_{j=۱}^{p} Uniform(-\sqrt{\sigma^۲}u_j, \sqrt{\sigma^۲}u_j),\\
&u|\lambda \sim \prod_{j=۱}^{p} Gamma(۲, \lambda),\\
&\sigma^۲ \sim \pi(\sigma^۲)
\end{aligned}$$

را خواهیم داشت. براساس مدل بیز سلسله مراتبی، توزیع پسین توام پارامترها برابر

$$\begin{aligned}
&\pi(\boldsymbol{\beta}, \boldsymbol{u}, \lambda, \sigma^۲, v, l|\mathbf{y}, x) \propto \pi(\mathbf{y}|x, \boldsymbol{\beta}, \sigma^۲, v, l)\pi(\boldsymbol{\beta}|\boldsymbol{u}, \sigma^۲)\pi(\boldsymbol{u}|\lambda)\pi(l|v)\pi(\sigma^۲)\\
&\propto \prod_{i=۱}^{n}[\frac{۱}{\sqrt{۲\pi l_i\sigma^۲}} exp(-\frac{۱}{۲l_i\sigma^۲}(y_i - X\beta)^۲) \times \frac{(\frac{v}{۲})^{\frac{v}{۲}}}{\Gamma(\frac{v}{۲})} l_i^{-\frac{v}{۲}-۱} exp(\frac{-\frac{v}{۲}}{l_i})]\\
&\times \prod_{j=۱}^{p} \frac{۱}{\sqrt{\sigma^۲}} I\{|\beta_j| < \sqrt{\sigma^۲}u_j\} e^{-\lambda u_j} \times \frac{۱}{\sigma^۲}.
\end{aligned}$$

است. با استفاده از مدل بیز سلسله مراتبی می‌توان توزیع‌های شرطی کامل پسین برای پارامترهای مدل را برای اجرای الگوریتم گیبز به دست آورد. با معرفی $\boldsymbol{u} = (u_۱, ..., u_p)$، توابع چگالی شرطی کامل پسین به صورت

$$\begin{aligned}
&\boldsymbol{\beta}|\boldsymbol{y}, X, \boldsymbol{u}, \lambda, \sigma^۲, v, l \sim N_p((X'L^{-۱}X)^{-۱}(X'L^{-۱}X), \sigma^۲(X'L^{-۱}X)^{-۱}) \times \prod_{j=۱}^{p} I\{|\beta_j| < \sqrt{\sigma^۲}u_j\},\\
&\boldsymbol{u}|\mathbf{y}, X, \boldsymbol{\beta}, \lambda, v, \sigma^۲, l \sim \prod_{j=۱}^{p} e^{-\lambda u_j} I\{u_j > \frac{|\beta_j|}{\sqrt{\sigma^۲}}\} \sim \prod_{j=۱}^{p} Exponential(\lambda) I\{u_j > \frac{|\beta_j|}{\sqrt{\sigma^۲}}\},\\
&\sigma^۲|\mathbf{y}, X, \boldsymbol{\beta}, \boldsymbol{u}, \lambda, l \sim I\Gamma(\frac{n-۱+p}{۲}, (\frac{۱}{۲}(\mathbf{y} - X\boldsymbol{\beta})^T L^{-۱}(\mathbf{y} - X\boldsymbol{\beta})))I(\sigma^۲ > max_j(\frac{\beta_j^۲}{u_j^۲})),\\
&l|\mathbf{y}, X, \boldsymbol{\beta}, \boldsymbol{u}, \lambda, \sigma^۲ \sim \prod_{i=۱}^{n} I\Gamma(\frac{۱}{۲} + \frac{v}{۲}, \frac{(y_i - X\boldsymbol{\beta})}{۲\sigma^۲} + \frac{v}{۲}),
\end{aligned}$$

هستند که $I(.)$ تابع نشانگر است. برای بررسی جزئیات محاسبه‌ی این توابع به پیوست (ب) مراجعه شود.
با استفاده از نمایش سلسله مراتبی مدل، تابع چگالی پسین برای λ به شرط $\boldsymbol{\beta}$ به صورت

$$\pi(\lambda|\boldsymbol{\beta}) \propto \lambda^{۲p} exp\{-\lambda \sum_{j=۱}^{p} |\beta_j|\}\pi(\lambda)$$

است. با در نظر گرفتن چگالی پیشین گاما با پارامترهای a و b برای λ، توزیع پسین شرطی نیز توزیع گاما به صورت

$$\lambda|\mathbf{y}, X, \boldsymbol{\beta}, \sigma^۲ \propto \lambda^{a+۲p-۱} exp\{-\lambda(b + \sum_{j=}^{p} |\beta_j|\},$$

به دست می‌آید. بنابراین پارامتر تنظیم λ همراه با دیگر پارامترهای مدل با استفاده از الگوریتم گیبز و تولید نمونه‌هایی از توزیع گاما با پارامترهای $a + ۲p$ و $b + \sum_{j=}^{p} |\beta_j|$ به‌روز می‌شود. در این مقاله، پارامتر تنظیم λ به عنوان میانگین توزیع پسین و با در نظر گرفتن مقادیر $a = ۱$ و $b = ۰$ برای پارامترهای چگالی پیشین، برآورد می‌شود.

۴ تحلیل داده‌ها

۱.۴ مطالعه شبیه سازی

هدف اصلی این بخش، ارزیابی عملکرد دو روش رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس یکنواخت و روش رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس نرمال می‌باشد. با استفاده از محاسبه‌ی معیارهای DIC [۱۴] و MSE برای خطاهای پیش‌بینی مدل در دو مثال شبیه‌سازی، ارزیابی انجام می‌شود. معیار DIC تعمیمی از معیار AIC برای مسائل انتخاب مدل بیزی است. معیار DIC به صورت $DIC = \overline{D(\boldsymbol{\theta})} + P_d$ تعریف می‌شود که $\overline{D(\boldsymbol{\theta})} = -۲LogL(\boldsymbol{\theta}|y)$ را انحراف نامیده و تابعی از $\boldsymbol{\theta}$، بردار پارامترهای مدل رگرسیونی، است. عبارت اول، به صورت امید انحراف تحت تابع چگالی پسینی پارامتر به صورت

$$\overline{D(\boldsymbol{\theta})} = E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] = E_{\theta|y}[-۲LogL(\boldsymbol{\theta}|\mathbf{y})]$$

تعریف می‌شود. مؤلفه دوم، تعداد پارامترهای مؤثر یا P_d را اندازه می‌گیرد که به صورت:

$$P_d = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}) = E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] - D(E_{\boldsymbol{\theta}|\mathbf{y}}[\boldsymbol{\theta}]) = E_{\theta|\mathbf{y}}[-۲lnL(\boldsymbol{\theta}|\mathbf{y})] + lnL(\boldsymbol{\theta}|\mathbf{y}).$$

تعریف می‌شود. با دوباره‌آرایی عبارت P_d داریم: $\bar{D} = D(\bar{\boldsymbol{\theta}}) + P_d$. پس DIC را می‌توان به صورت

$$DIC = D(\bar{\boldsymbol{\theta}}) + ۲P_d$$

تعریف کرد. محاسبه‌ی DIC با استفاده از الگوریتم های زنجیر مارکف مونت کارلو از قبیل گیبز اغلب بدیهی و ساده است. برآورد $\bar{D}$ با میانگین گرفتن از مقادیر شبیه‌سازی شده از $D(\boldsymbol{\theta})$ محاسبه می‌شود. مقدار پارامترهای مؤثر P_d را نیز می‌توان با محاسبه‌ی $D(\boldsymbol{\theta})$ در میانگین نمونه‌ای مقادیر شبیه‌سازی شده‌ی پارامترها و کم کردن آن از برآورد $D(\bar{\boldsymbol{\theta}})$ به دست آورد.
برای اجرای الگوریتم گیبز ابتدا متغیر پاسخ مرکزی شده و متغیرهای توضیحی استاندارد می‌شوند، به طوری که میانگین صفر و واریانس واحد داشته باشند. سپس مراحل زیر انجام می‌شود:

۱- داده‌های شبیه‌سازی تحت مدل

$$y_i = x_i^T\beta + \varepsilon_i, \quad \varepsilon_i \overset{i.i.d}{\sim} t_v(۰, \sigma^۲), \quad i = ۱, ..., n$$

و با در نظر گرفتن $x_i \sim N(۰, \Sigma)$ تولید می‌شود به طوری که ماتریس کواریانس Σ به صورت $\Sigma_{jk} = \rho^{|j-k|}$ و $\rho = ۰/۷$ است.

۲- خطاهای ε_i شامل دو توزیع تی-استیودنت متفاوت با نمایش آمیخته-مقیاس نرمال وآمیخته-مقیاس یکنواخت با درجات آزادی برابر با $v = ۲, ۵, ۱۰, ۱۰۰۰$ است. همچنین دو مقدار متفاوت برای $\sigma = (۹, ۲۵)$ در نظر گرفته می‌شود.

۳- هر نمونه شبیه‌سازی شده به دو قسمت مجموعه داده آموزشی و مجموعه داده آزمون تقسیم می‌شود.

در هر دو روش، مجموعه آموزشی و مجموعه آزمون، هر دو شامل ۵۰ مشاهده، تولید شده و پارامتر ضرایب $\boldsymbol{\beta}$ به صورت $\boldsymbol{\beta} = (۳, -۳, ۵, ۴, -۲/۸, -۳/۲, ۰, ۰, ۰, ۰)$ با ۶ پارامتر غیرصفر و ۴ پارامتر صفر در نظر گرفته شده است.

۵- مدل‌ها روی مجموعه داده آموزشی برازش شده و مقادیر DIC و MSE برای مجموعه داده آزمون محاسبه می‌شوند.

برای هر انتخاب توزیع خطا، شبیه‌سازی با ۱۰۰۰ اجرا انجام شده است. نتایج DIC و MSE برای مجموعه داده آزمون در جدول ۱ برای دو روش مختلف انتخاب مدل ذکر شده است. با توجه به مقادیر DIC و MSE در جدول ۱، در هر دو حالت $\sigma^۲ = ۸۱$ و $\sigma^۲ = ۲۲۵$ روش رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس یکنواخت برای هر ۴ مقدار درجه آزادی بهتر از روش رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس نرمال عمل می‌کند. البته با افزایش درجه آزادی و نزدیک شدن به توزیع نرمال یا به عبارت دیگر برای $v = ۱۰۰۰$ اختلاف مقادیر DIC و MSE در هر دو روش تقریبا کم می‌شود. به عبارتی می‌توان گفت کارایی روش رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس یکنواخت برای توزیع تی استیودنت با درجه آزادی $v = ۵$ بهتر از شرایط مدل با درجات آزادی $v = ۲, ۱۰, ۱۰۰$ است. با هدف انتخاب متغیر، ناحیه‌های اعتبار مرتفع‌ترین چگالی پسین (HPD) برای پارامترهای مدل محاسبه شده است. اگر ناحیه اعتبار پارامتر مقدار صفر را شامل شود، می‌توان متغیر موردنظر را در مدل‌سازی در نظر نگرفت. با توجه به نتایج جدول ۲ ، در روش رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس یکنواخت می‌توان مقدار پارامترهای $\beta_۷, \beta_۸, \beta_۹, \beta_{۱۰}$ را صفر قرار داد. اما در روش رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس نرمال، طبق ناحیه‌های اعتبار به‌دست آمده، ضرایب $\beta_۲, \beta_۵, \beta_۷, \beta_۸, \beta_۹, \beta_{۱۰}$ برای حذف از مدل انتخاب می‌شوند. همچنین اگر از مدل رگرسیون لاسوی بیزی معمولی با توزیع خطای نرمال، برای مدل‌سازی داده‌های بوستون استفاده شود، تعداد ضرایب

بیشتری شامل $\beta_{۱}, \beta_{۲}, \beta_{۵}, \beta_{۷}, \beta_{۸}, \beta_{۹}, \beta_{۱۰}$ برای حذف از مدل پیشنهاد می‌شود. بنابراین می‌توان گفت در زمینه انتخاب متغیر نیز روش رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس یکنواخت بهتر عمل می‌کند.

جدول ۱: نتایج شبیه سازی بر اساس ۱۰۰۰ تکرار و تعداد مشاهدات $n = ۱۰۰$

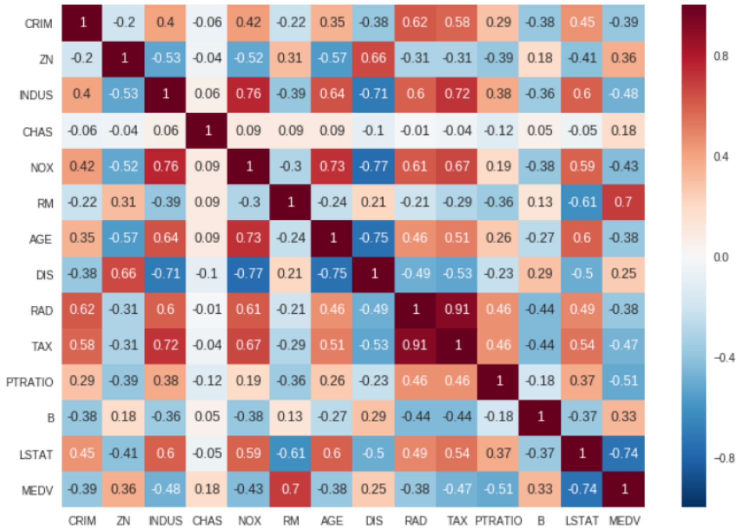
مدل آماری	درجه آزادی				
		$v = ۲$	$v = ۵$	$v = ۱۰$	$v = ۱۰۰۰$
مدل رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس نرمال (DIC,MSE)	$\sigma^{۲} = ۸۱$	$(۴۵۰, ۱۰۵/۷)$	$(۴۴۹/۲, ۳۸/۲)$	$(۳۷۶/۳, ۵۰/۸)$	$(۲۷۴/۱, ۵/۵)$
	$\sigma^{۲} = ۲۲۵$	$(۲۲۸/۱, ۳۰/۷)$	$(۵۷۴/۸, ۷۸/۵)$	$(۳۵۴/۹, ۳۱/۷)$	$(۲۷۳/۵, ۶/۲)$
مدل رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس یکنواخت (DIC,MSE)	$\sigma^{۲} = ۸۱$	$(۳۵۸/۸, ۱۷/۱)$	$(۲۴۵/۰۱, ۳/۳)$	$(۲۴۱/۷, ۳/۸)$	$(۲۲۷/۳, ۱/۸)$
	$\sigma^{۲} = ۲۲۵$	$(۳۶۰/۷, ۱۵/۸)$	$(۲۳۹/۲, ۳/۶)$	$(۲۴۱/۴, ۳/۷)$	$(۲۵۳/۸, ۱/۵)$

جدول ۲: ناحیه‌های اعتبار برای پارامترهای مدل رگرسیونی تی لاسو بیزی با نمایش آمیخته–مقیاس یکنواخت

$\beta_{۱}$	$\beta_{۲}$	$\beta_{۳}$	$\beta_{۴}$	$\beta_{۵}$
$(۱/۴۳, ۳/۳۰)$	$(۰/۱۰, ۳/۶۰)$	$(۰/۱۴, ۵/۱۳)$	$(۱/۳۴, ۵/۸۱)$	$(-۲/۸۵, -۰/۱۰)$
$\beta_{۶}$	$\beta_{۷}$	$\beta_{۸}$	$\beta_{۹}$	$\beta_{۱۰}$
$(۱/۸۹, ۶/۸۱)$	$(-۱/۵۸, ۲/۲۳)$	$(-۰/۶۹, ۲/۵۸)$	$(-۱/۸۱, ۱/۶۰)$	$(-۱/۷۸, ۱/۴۲)$

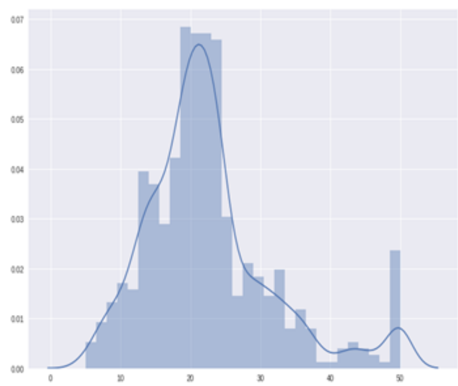
۲.۴ داده‌های واقعی

در این بخش، عملکرد رگرسیون تی لاسو بیزی با دو نمایش متفاوت آمیخته–مقیاس نرمال وآمیخته–مقیاس یکنواخت با استفاده از مجموعه داده‌های بوستون بررسی می‌شود. گزارش داده‌های بوستون توسط هریسون و رابینفلد [۱۵] به منظور بررسی تأثیر چندین ویژگی روی قیمت‌های مسکن و پیدا کردن مناسب‌ترین متغیرهای توضیحی در مجله‌ی اقتصاد و مدیریت منتشر شد. مجموعه داده‌های بوستون از طریق https://stat.ethz.ch/R-manual/R-devel/library/MASS/html/Boston.html قابل دسترسی است. مجموعه داده‌های بوستون شامل ۵۰۶ مشاهده $(n = ۵۰۶)$ و ۱۳ متغیر توضیحی $(p = ۱۳)$ است. متغیر پاسخ، مقدار میانه‌ی خانه‌های تحت مالکیت در حومه‌ی شهر بوستون (MEDV) است. برای برآورد رابطه بین متغیر پاسخ و ۱۳ متغیر توضیحی از دو روش رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس نرمال و رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس یکنواخت با درجات آزادی ۲ ، ۵ ، ۱۰ و ۱۰۰۰ استفاده می‌کنیم. مجموعه مقادیر درجه آزادی شامل طیف بزرگی از توزیع‌های دم‌سنگین $(v = ۲)$ تا توزیع نرمال $(v = ۱۰۰۰)$ است. با توجه به نمودار ماتریس همبستگی در شکل ۱، اثراتی از وجود همبستگی بین متغیرهای توضیحی مشاهده می‌شود.

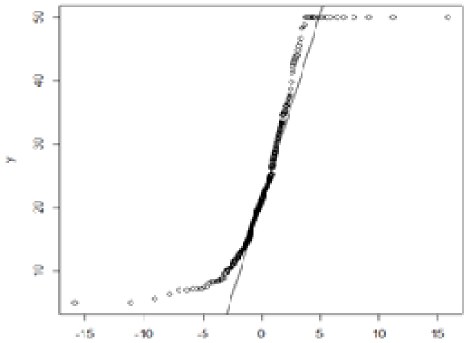


شکل ۱: نمودار ماتریس همبستگی

با توجه به نمودار توزیع داده‌ها در شکل ۲ نیز چنین استدلال می‌شود که داده ها با کمی چولگی همراه است. همچنین با توجه به مقدار احتمال $P - Value = ۰٫۰۰۳$ به دست آمده از آزمون شاپیرو و ویلک [۱۶] در سطح خطای ۵ درصد، فرض صفر که نرمال بودن داده‌ها است، رد شده و توزیع داده‌های بوستون از توزیع نرمال متفاوت است. نمودار چندک–چندک در شکل ۳ نیز مناسب بودن توزیع تی–استیودنت برای مجموعه داده‌ها را نشان می‌دهد.



شکل ۲: نمودار توزیع داده‌های بوستون



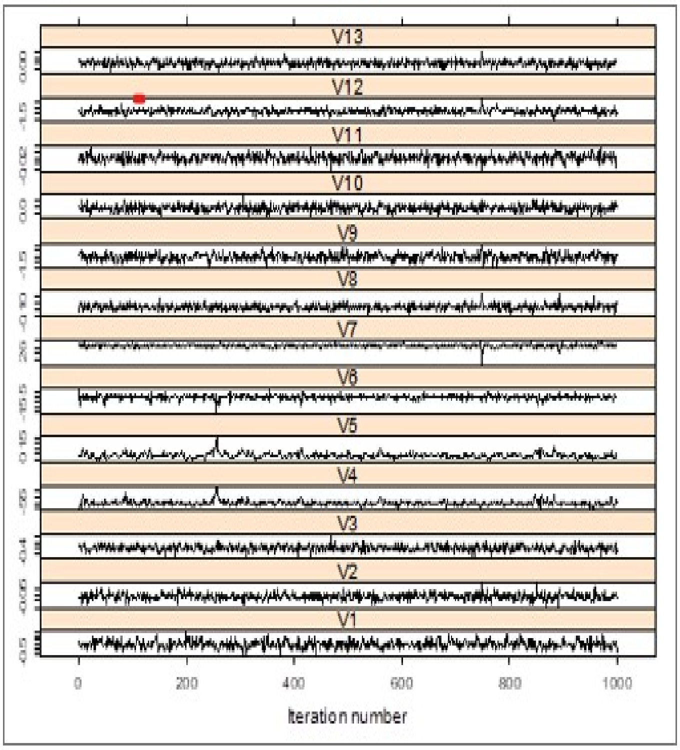
شکل ۳: نمودار چندک – چندک توزیع تی – استیودنت

ابتدا متغیر پاسخ متمرکز شده و متغیرهای توضیحی استاندارد می‌شوند به طوری که میانگین صفر و واریانس یک داشته باشند. دو روش رگرسیون تی لاسو بیزی به مجموعه داده‌ی آموزشی برازش داده شده و به منظور ارزیابی عملکرد دو روش و انتخاب مدل بهینه، معیارهای MSE و DIC برای مجموعه داده آزمون محاسبه می‌شوند، که نتایج را در جدول ۳ مشاهده می‌کنید. نتایج با در نظر گرفتن توزیع تی–استیودنت با درجات آزادی متفاوت $\upsilon = ۲, ۵, ۱۰, ۱۰۰۰$ برای توزیع خطاهای مدل در روش رگرسیون تی لاسو بیزی به دست آمده است. با توجه به جدول ۳، عملکرد روش تی لاسو بیزی با نمایش آمیخته–مقیاس یکنواخت بهتر از روش دیگر است و میزان معیارهای DIC و MSE برای مقادیر درجه آزادی مختلف، اختلاف قابل توجهی دارند.

جدول ۳: تحلیل داده های بوستون– مقادیر DIC و MSE برای دو روش رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس نرمال و آمیخته–مقیاس یکنواخت

مدل آماری	درجه آزادی			
	$\upsilon = ۲$	$\upsilon = ۵$	$\upsilon = ۱۰$	$\upsilon = ۱۰۰۰$
مدل رگرسیون تی لاسو بیزی با نمایش آمیخته نرمال (DIC,MSE)	(۱۹۰۱/۲, ۸۵/۲)	((۱۸۶۱/۷, ۸۳/۴)	(۱۸۳۹/۹, ۷۸/۳)	(۱۶۶۱/۲, ۳۴/۴)
مدل رگرسیون تی لاسو بیزی با نمایش آمیخته یکنواخت (DIC,MSE)	(۱۸۰۶/۱, ۶۶/۸)	(۱۵۶۷/۱, ۴۲/۱)	(۱۶۳۲/۳, ۳۷/۲)	(۱۶۲۴/۳, ۳۲/۱)

از نتایج ناحیه‌های اعتبار مرتفع‌ترین چگالی پسین برای پارامترهای مدل رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس یکنواخت نتیجه می‌شود که متغیرهای AGE و INDUS انتخاب مناسبی برای حذف از مدل می‌باشند. با محاسبه‌ی میانگین توان دوم خطای پیش‌گویی یا MSPE برای مدل با حضور تمام متغیرهای توضیحی و مدل با حذف متغیرهای AGE و INDUS ، مشاهده می‌شود که مقدار میانگین توان دوم خطای پیش‌گویی برای مدل کاهش‌یافته کمتر است. بنابراین می‌توان گفت مدل رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس یکنواخت در انتخاب متغیرهای توضیحی موثر عملکرد مناسبی داشته است.
آمیختگی زنجیر مارکف نمونه‌های به دست آمده برای پارامترهای مدل با استفاده از الگوریتم گیبز، بیانگر میزان همگرایی زنجیر به دست آمده از الگوریتم گیبز است و نمودار اثر معیار شهودی خوبی برای ارزیابی ویژگی آمیختگی زنجیر است [۱۷]. در شکل ۴ برای داده‌های بوستون، نمودار اثر در حالتی‌که $\upsilon = ۵$ است، مشاهده می‌شود. با توجه به نمودار، نمونه‌های به دست آمده از الگوریتم گیبز برای پارامترهای مدل، فضای توزیع پسین را به سرعت پیموده و همگرایی خوبی دارند. همچنین طبق آزمون تشخیص همگرایی هیدلبرگر و ولچ [۱۸]، زنجیر مارکف از یک توزیع ایستا پیروی می‌کند. به طور کلی با توجه به نتایج شهودی، روش رگرسیون تی لاسو بیزی بر اساس نمایش آمیخته–مقیاس یکنواخت از لحاظ ویژگی انتخاب مدل و دقت پیش‌بینی بهتر از روش رگرسیون تی لاسو بیزی بر اساس نمایش آمیخته–مقیاس نرمال بوده و با افزایش درجه آزادی توزیع تی– استیودنت و نزدیک شدن توزیع خطاهای مدل به توزیع نرمال، تفاوت بین دو روش کمتر می‌شود.



شکل ۴: نمودار اثر نمونه‌های به دست آمده از الگوریتم گیبز برای ضرایب رگرسیونی در مدل رگرسیونی تی لاسو بیزی با نمایش آمیخته یکنواخت

۵ نتیجه‌گیری

وقتی خطا در مدل‌های رگرسیون خطی از توزیع نرمال پیروی کند، برآوردگرهای کمترین توان‌های دوم خطا به مقادیر داده‌های دور افتاده حساس بوده و این موضوع ما را به انتخاب توزیع خطایی استوار برای مدل رگرسیونی ترغیب می‌کند. جهت رفع این مشکل، توزیع تی-استیودنت به عنوان یک توزیع جایگزین پیشنهاد می‌شود. در این مقاله، مدل رگرسیون تی لاسو بیزی به منظور تعمیم روش رگرسیون لاسو بیزی نرمال برای ارائه‌ی برآوردگرهایی استوار در حالت غیر نرمال بودن خطای مدل رگرسیونی پیشنهاد شده است. مدل رگرسیون تی لاسو بیزی با دو نمایش متفاوت آمیخته-مقیاس نرمال و آمیخته-مقیاس یکنواخت برای چگالی لاپلاس در نظر گرفته شده و مدل‌های بیز سلسله مراتبی برای اجرای الگوریتم نمونه گیبز ارائه شده است. طبق مطالعات شبیه‌سازی و تحلیل داده‌های واقعی، با توجه به معیارهای MSE و DIC مدل رگرسیون تی لاسو بیزی با نمایش آمیخته-مقیاس یکنواخت برای چگالی پیشین لاپلاس عملکرد رضایت‌بخشی در مقایسه با نمایش آمیخته-مقیاس نرمال دارد. در آینده پژوهش در زمینه‌ی انتخاب مدل برای مدل رگرسیون تی لاسو بیزی در ابعاد بالا مورد توجه است.

فهرست منابع

[۱] نوروزی راد، م. و آرشی، م. (۱۳۹۶). مطالعه رفتار حدی برآوردگرهای انقباضی در مدل رگرسیون تاوانیده با نرم مستطیلی. علوم آماری، ۱۱(۱)، ۱۷۴-۱۴۹.

[۲] کاظمی، م. شاهسونی، د. و آرشی، م. (۱۳۹۷). انتخاب متغیر و تشخیص ساختار در بعد بالا برای مدل های جمعی خطی جزئی. علوم آماری، ۱۲(۲)، ۵۱۲-۴۸۵.

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

[۴] تعاونی، م. و آرشی، م. (۱۳۹۹). انتخاب متغیر در مدل های خطی جزئی با اثرات آمیخته برای داده های طولی. علوم آماری، ۱۴(۲)، ۳۸۸-۳۶۷.

[5] Lange, K.L., etal. (1989). Robust statistical modeling using the t distribution. Journal of the American statistical association, **84**, 881-896.

[6] Liu, C. and Rubin, D. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. StatisticaSinica, **5**, 19-39.

[7] Lin, J.G, etal. (2009). Heteroscedasticity diagnostics for t linear regression models. Metrika, **70**, 59-77.

[8] Neter, J., Kutner, H., Wasserman W. and Nachtcheim. (1996). Applied Linear Regression Models. McGraw-Hill College.

[9] Hoerl, A.E., and Kennard, R.W. (1970). Ridge Regression: Biased Estimation for Non orthogonal Problems. Technometrics,**12**, 55-67.

[10] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society, **58**, 267-288.

[11] Andrews, D.F., and Mallows, C.L. (1974). Scale Mixtures of Normal Distributions. Journal of the Royal statistical series, **36**, 99-102.

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

[13] Wang, J.J.J., etal. (2011). Stochastic volatility models with leverage and heavy-tailed distributions, A Bayesian approach using scale mixtures. Computational Statistics and Data Analysis, **55**, 852-862.

[14] Spiegelhalter, D.J., etal. (2002). Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, **64**, 583-639.

[15] Harrison, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management, **5**, 81-102.

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

[17] Gelman, A., etal. (2003). Bayesian data analysis. Chapman Hall, London.

[18] Heidelbeger P. and Welch PD. (1981). A spectral method for confidence interval generation and run length control in simulations. Comm.ACM, **24**, 233-245

[19] Ding, P. and blitzstein J.K.(2018). On the Gaussian Mixture Representation of the Laplace Distribution.The American Statistician, **72**, 172-174.

[20] Mallick, H. and Yi, N. (2014). A New Bayesian Lasso. Statistics and Its Interface, **7**, 571-582.

[21] Saleh, A.K. Md. Ehsanes, Arashi, M., and tabatabaey, S.M.M. (n.d.).(2014). Statistical Infrence for Models with Multivariate t-Distributed Errors. In Statistical Infrence for Models with Multivariate t-Distributed Errors. John Wiley, New Jersy.

[22] Hlavackova-Schindler, K. (2016). Prediction Cosistency of Lasso Regression Does Not Need Normal Errors. British Journal of Mathematics Computer Science, **19(4)**, 2231-0851.

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

[24] Knight, K. and Fu, W. (2000). asymptotics for LASSO-type Estimators. Annals of Statistics, **28**, 1356-1378.

[25] Belloni, A. and Chernozhukov, V. (2013). Least Squares after Model Selection in High-Dimensional Sparse Models. Bernoulli, **19**, 521-547.

[26] Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. Journal of the American Statistical Association, **96**, 1348-1360.

[27] Tikhonov, A.N. (1963). solution of Incorrectly Formulated Problems and the Regularization Mehod. Doklady Akademii Nauk SSSR, **151**, 501-504.

[28] Zou, H. and Hastie, T. (2005). Regularization and variable Selection via the Elastic Net. Journal of the Royal Statistical Society, Series B, **67**, 301-320.

[29] Zou, H. (2006). The Adaptive Lasso and its Oracle Properties. Journal of the American Statistical Association, **101**, 1418-1429.

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

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

پیوست آ

تعریف ۱: متغیر تصادفی به صورت $\mathbf{y} \sim t_v(\mu, \sigma^۲)$ را گوییم دارای توزیع t با درجه آزادی v است اگر چگالی احتمال آن به صورت

$$f_v\left(y_i|\mu, \sigma^۲\right) = \left(\frac{۱}{\pi v \sigma^۲}\right)^{\frac{۱}{۲}} \left(\frac{\Gamma([v+۱]/۲)}{\Gamma(v/۲)}\right) \left(۱ + \frac{(y_i-\mu)^۲}{v\sigma^۲}\right)^{-\left(\frac{v+۱}{۲}\right)}$$

باشد که μ پارامتر مکان و $\sigma^۲$ پارامتر مقیاس است. درجه آزادی v میزان سنگینی دم‌های توزیع را تعیین می‌کند. برای $v > ۱$، میانگین توزیع برابر μ و برای $v > ۲$، واریانس توزیع برابر $\frac{\sigma^۲ v}{v-۲}$ می باشد. حالت خاص $v = ۱$، توزیع کوشی و $v = \infty$ توزیع نرمال معمولی است. (برای جزئیات بیشتر در خصوص کاربرد توزیع تی – استیودنت در مدل‌های رگرسیونی به [۲۱] مراجعه نمایید).

لم ۱: برای یافتن توزیع حاشیه‌ای متغیر $\mathbf{y}$ داریم:

$$\begin{aligned}
f_{\mathbf{y}}(y) &= \int_۰^\infty f_{\mathbf{y}}(l|y)dl \propto \int_۰^\infty \frac{۱}{\sqrt{l}} exp\left(-\frac{y^۲}{۲l}\right) l^{-\frac{v}{۲}-۱} exp\left(-\frac{v}{۲l}\right) dl \\
&= \int_۰^\infty l^{-\frac{v+۱}{۲}-۱} exp\left(-\frac{y^۲+v}{۲l}\right) dl \\
&= \Gamma\left(\frac{v+۱}{۲}\right) \left(\frac{y^۲+v}{۲l}\right)^{-\frac{v+۱}{۲}}
\end{aligned}$$

خط آخر متناسب با توزیع تی–استیودنت با درجه آزادی v است. می‌توان توزیع تی–استیودنت $t_v(\mu, \sigma^۲)$ را به صورت سلسله مراتبی

$$\begin{aligned}
&\mathbf{y}|\mu, \sigma^۲ v, l \sim N(\mu, l\sigma^۲ I) \\
&l|v \sim IG\left(\frac{v}{۲}, \frac{v}{۲}\right)
\end{aligned}$$

نیز نمایش داد.

لم۲: برای اثبات لم ۲ نیازمند محاسبه‌ی چگالی حاشیه‌ای متغیر تصادفی L به صورت

$$f(l)=\int_{۰}^{\infty}\frac{۱}{\sqrt{۲\pi v}}exp\left(-\frac{l^{۲}}{۲v}\right)\times\frac{\lambda^{۲}}{۲}exp\left(-\frac{\lambda^{۲}v}{۲}\right)dv=-\frac{\lambda}{۲}exp\left(-\lambda|l|\right).$$

می‌باشیم. با مربع کامل کردن عبارت نمایی در رابطه (انتگرال) داریم:

$$f(l)=\int_{۰}^{\infty}\frac{\lambda^{۲}e^{-\lambda|l|}}{۲\sqrt{۲\pi}}\frac{۱}{\sqrt{v}}exp\{-\frac{۱}{۲}\left(\frac{|l|}{\sqrt{v}}-\lambda\sqrt{v}\right)^{۲}\}dv$$

با تغییر متغیر $u=\sqrt{v}$ و $dv=۲udu$ عبارت

$$f(l)=\frac{\lambda^{۲}e^{-\lambda|l|}}{\sqrt{۲\pi}}\int_{۰}^{\infty}exp\{-\frac{۱}{۲}\left(\frac{|l|}{\sqrt{u}}-\lambda\sqrt{u}\right)^{۲}\}du$$

به دست می آید. همچنین با تغییر متغیر

$$\eta=\frac{|l|}{u}-\lambda u,\qquad \frac{du}{d\eta}=\frac{-۱+\dfrac{\eta}{\sqrt{\eta^{۲}+۴\lambda|l|}}}{۲\lambda},$$

داریم:

$$f(l)=\frac{\lambda^{۲}e^{-\lambda|l|}}{۲\lambda\sqrt{۲\pi}}\int_{-\infty}^{\infty}exp\left\{۱-\frac{\eta}{\sqrt{\eta^{۲}+۴\lambda|l|}}\right\}d\eta,$$
$$\frac{\lambda^{۲}e^{-\lambda|l|}}{۲}\int_{-\infty}^{\infty}\frac{۱}{\sqrt{۲\pi}}exp\left(-\frac{\eta^{۲}}{۲}\right)d\eta-\frac{\lambda^{۲}e^{-\lambda|l|}}{۲\sqrt{\pi}}\int_{-\infty}^{\infty}exp\left(-\frac{\eta^{۲}}{۲}\right)\frac{\eta}{\sqrt{\eta^{۲}+۴\lambda|l|}}d\eta=\frac{\lambda e^{-\lambda|l|}}{۲},$$

که عبارت آخر از توزیع نرمال استاندارد پیروی کرده و انتگرال از تابعی فرد است [۱۹].

لم ۳: می‌دانیم

$$\int_{z>\frac{|x|}{\sqrt{\sigma^{۲}}}}\lambda e^{-\lambda z}dz=e^{-\lambda\frac{|x|}{\sqrt{\sigma^{۲}}}},$$

بنابراین، تابع چگالی احتمال از توزیع لاپلاس با میانگین صفر و واریانس $\sqrt{\sigma^{۲}}/\lambda$ می تواند به صورت

$$\frac{\lambda}{۲\sqrt{\sigma^{۲}}}=e^{-\lambda z}dz=e^{-\lambda\frac{|x|}{\sqrt{\sigma^{۲}}}}=\frac{\lambda}{۲\sqrt{\sigma^{۲}}}\int_{u>\frac{|x|}{\sqrt{\sigma^{۲}}}}\lambda e^{-\lambda u}du$$
$$=\int_{-u\sqrt{\sigma^{۲}}<x<u\sqrt{\sigma^{۲}}}\frac{۱}{۲u\sqrt{\sigma^{۲}}}\frac{\lambda^{۲}}{\Gamma(۲)}u^{۲-۱}e^{-\lambda u}du,$$

نوشته شود و در نتیجه برابری اثبات می شود [۲۰].

پیوست ب

نمونه‌گیر گیبز برای مدل رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس نرمال

با توجه به چگالی های پیشینی که در مدل رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس نرمال برای پارامترها در نظر گرفته شد، توزیع پسین به شرط y برای پارامترها به صورت

$$
\begin{aligned}
&\pi\left(\boldsymbol{\beta},\sigma^{2},v,l|\mathbf{y},X\right)\propto\pi\left(y|X,\boldsymbol{\beta},\sigma^{2},v,l\right)\pi(l|v)\pi(\boldsymbol{\beta}|\sigma^{2})\pi(\sigma^{2})\\
&\propto\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi l_i\sigma^{2}}}exp\left(\frac{1}{2l_i\sigma^{2}}(y_i-X\boldsymbol{\beta})^{2}\right)\frac{(v/2)^{v/2}}{\Gamma(v/2)}l_i^{\frac{v}{2}-1}exp\left(\frac{-v/2}{l_i}\right)\\
&\left[\prod_{j=1}^{p}\frac{1}{\sqrt{2\pi\tau_j^{2}\sigma^{2}}}e^{-\boldsymbol{\beta}_j^{2}/2\tau_j^{2}\sigma^{2}}\frac{\lambda^{2}}{2}e^{-\frac{\lambda^{2}}{2}}\right]\left[\left(\frac{1}{\sigma^{2}}\right)^{r+1}exp\left(-\frac{v}{\sigma^{2}}\right)\right]\\
&\propto exp\left(-\frac{1}{2\sigma^{2}}(\mathbf{y}-X\boldsymbol{\beta})^{T}L^{-1}(\mathbf{y}-X\boldsymbol{\beta})\right)\prod_{i=1}^{n}\frac{(v/2)^{v/2}}{\Gamma(v/2)}l_i^{v/2-1}exp\left(\frac{-v/2}{l_i}\right)\\
&\times\left[\left(\frac{1}{\sqrt{2\pi\tau_j^{2}\sigma^{2}}}\right)^{\frac{p}{2}}exp\left(-\frac{1}{2\sigma^{2}}\boldsymbol{\beta}^{T}D_{\tau}^{-1}\boldsymbol{\beta}\right)\prod_{j=1}^{p}\frac{\lambda^{2}}{2}e^{-\lambda^{2}/2\tau_j^{2}}\right]\left[\left(\frac{1}{\sigma^{2}}\right)^{r+1}exp\left(\frac{-\gamma}{\sigma^{2}}\right)\right]
\end{aligned}
$$

است. توزیع توام پسین به دست آمده را به منظور استنباط در مورد پارامترهای مدل رگرسیونی بیزی با استفاده از الگوریتم گیبز استفاده می‌کنیم. ابتدا برای تولید نمونه‌هایی از تابع چگالی پسینی با استفاده از الگوریتم گیبز، عباراتی که شامل $\boldsymbol{\beta}$، σ^2، τ_j^2 و l_i هستند را جدا کرده و توابع چگالی شرطی مرتبط با هر پارامتر را به دست می‌آوریم.

برای یافتن چگالی پسینی حاشیه‌ای پارامتر $\boldsymbol{\beta}$، ابتدا عباراتی از چگالی پسینی توأم که شامل پارامتر $\boldsymbol{\beta}$ هستند را در نظر می گیریم:

$$
\begin{aligned}
&exp\left(-\frac{1}{2\sigma^{2}}(\mathbf{y}-X\boldsymbol{\beta})^{T}L^{-1}(\mathbf{y}-X\boldsymbol{\beta})\right)exp\left(-\frac{1}{2\sigma^{2}}\boldsymbol{\beta}^{T}D_{\tau}^{-1}\boldsymbol{\beta}\right)\\
&=exp\left[-\frac{1}{2\sigma^{2}}\left((\mathbf{y}-X\boldsymbol{\beta})^{T}L^{-1}(\mathbf{y}-X\boldsymbol{\beta})+\boldsymbol{\beta}^{T}D_{\tau}^{-1}\boldsymbol{\beta}\right)\right]\\
&=exp\left[-\frac{1}{2\sigma^{2}}\left(\mathbf{y}'L^{-1}\mathbf{y}-\boldsymbol{\beta}'X'L^{-1}\mathbf{y}-\mathbf{y}'L^{-1}X\boldsymbol{\beta}+\boldsymbol{\beta}'X'L^{-1}X\boldsymbol{\beta}+\boldsymbol{\beta}^{T}D_{\tau}^{-1}\boldsymbol{\beta}\right)\right]\\
&=exp\left[-\frac{1}{2\sigma^{2}}\left(\mathbf{y}'L^{-1}\mathbf{y}-2\boldsymbol{\beta}'X'L^{-1}\mathbf{y}+\boldsymbol{\beta}'X'L^{-1}X\boldsymbol{\beta}+\boldsymbol{\beta}^{T}D_{\tau}^{-1}\boldsymbol{\beta}\right)\right]\\
&=exp\left[-\frac{1}{2\sigma^{2}}\left(\mathbf{y}'L^{-1}\mathbf{y}-2\boldsymbol{\beta}'X'L^{-1}\mathbf{y}+\boldsymbol{\beta}'(X'L^{-1}X+D_{\tau}^{-1})\boldsymbol{\beta}\right)\right]
\end{aligned}
$$

با در نظر گرفتن $A=X'L^{-1}X+D_{\tau}^{-1}$ داریم:

$$
\begin{aligned}
&exp\left[-\frac{1}{2\sigma^{2}}\left(\boldsymbol{\beta}'A\boldsymbol{\beta}-2\mathbf{y}'L^{-1}X\boldsymbol{\beta}+\mathbf{y}'L^{-1}\mathbf{y}\right)\right]\\
&=exp\left[-\frac{1}{2\sigma^{2}}\left[\left(\boldsymbol{\beta}-A^{-1}X^{T}L^{-1}\mathbf{y}\right)^{T}\left(\boldsymbol{\beta}-A^{-1}X^{T}L^{-1}\mathbf{y}\right)\right]+\mathbf{y}^{T}\left(L^{-1}-XA^{-1}X^{T}\right)\mathbf{y}\right]\\
&\propto exp\left[-\frac{1}{2\sigma^{2}}\left(\boldsymbol{\beta}-A^{-1}X^{T}L^{-1}\mathbf{y}\right)^{T}A\left(\boldsymbol{\beta}-A^{-1}X^{T}L^{-1}\mathbf{y}\right)\right]
\end{aligned}
$$

بنابراین می‌توان نتیجه گرفت که β دارای توزیع نرمال چند متغیره با میانگین $A^{-1}X^TL^{-1}\mathbf{y}$ و σ^2A^{-1} واریانس می‌باشد. در ادامه برای یافتن تابع چگالی پسینی حاشیه‌ای σ^2، عبارات شامل σ^2 را از تابع چگالی پسینی توأم جدا می‌کنیم:

$$\left[\prod_{i=1}^{p}\frac{1}{\sqrt{2\pi l_i\sigma^2}}\right]\left[exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-X\boldsymbol{\beta})^TL^{-1}(\mathbf{y}-X\boldsymbol{\beta})\right)\right]$$
$$\times\left[\prod_{j=1}^{p}\frac{1}{\sqrt{2\pi\tau_j^2\sigma^2}}exp\left(\frac{-\beta_j^2}{2\tau_j^2\sigma^2}\right)\right]\left[\left(\frac{1}{\sigma^2}\right)^{r+1}exp\left(\frac{-\gamma}{\sigma^2}\right)\right]$$

سپس توزیع σ^2 گامای معکوس با پارامترهای $\left(\frac{n}{2}+\frac{p}{2}+r\right)$ و $\gamma+\dfrac{(\mathbf{y}-X\boldsymbol{\beta})^T(\mathbf{y}-X\boldsymbol{\beta})+\boldsymbol{\beta}^TD_\tau^{-1}\boldsymbol{\beta}}{2}$ می‌باشد. در ادامه عباراتی که شامل τ_j^2 هستند را در نظر می‌گیریم:

$$\prod_{j=1}^{p}\frac{1}{\sqrt{2\pi\tau_j^2\sigma^2}}exp\left(\frac{-1}{2\tau_j^2\sigma^2}\beta_j^2\right)\frac{\lambda^2}{2}exp\left(\frac{-\lambda^2}{2}\tau_j^2\right)$$
$$\propto(\tau_j)^{\frac{-1}{2}}exp\left[\frac{-1}{2}\left(\frac{\beta_j^2/\sigma^2}{\tau_j^2}\right)+\lambda^2\tau_j^2\right]$$

که متناسب با چگالی معکوس متغیر تصادفی گوسین است. چگالی $\eta_j^2=\dfrac{1}{\tau_j^2}$ نیز متناسب است با

$$(\eta_j^2)^{\frac{-3}{2}}exp\left[\frac{-1}{2}\left(\frac{\beta_j^2}{\sigma^2}\eta_j^2+\frac{\lambda}{\eta_j}\right)\right]$$
$$\propto(\eta_j^2)^{\frac{-3}{2}}exp\left\{-\frac{\beta_j^2\left(\eta_j^2-\sqrt{\frac{\lambda^2\sigma^2}{\beta_j^2}}\right)^2}{2\sigma^2\eta_j^2}\right\}$$

که اگر با فرم معمولی تابع گوسین معکوس به صورت

$$f(x)=\sqrt{\frac{\lambda'}{2\pi}}x^{-\frac{3}{2}}exp\left(-\frac{\lambda'(x-\mu')^2}{2(\mu')^2x}\right)$$

مقایسه شود، نتیجه می‌شود که توزیع $\frac{1}{\tau_j^2}$ گوسین معکوس با پارامتر میانگین $\mu'=\sqrt{\dfrac{\lambda^2\sigma^2}{\beta_j^2}}$ و پارامتر مقیاس $\lambda'=\lambda^2$ است. μ' پارامتر میانگین و λ' پارامتر مقیاس است.

حال عباراتی که شامل l_i هستند را در نظر می‌گیریم:

$$\pi(l|\mathbf{y},X,\boldsymbol{\beta},\mu,\lambda,v)\propto\pi(l|\mu)\pi(\mathbf{y}|X,\boldsymbol{\beta},\sigma^2,v,l)$$
$$\propto\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi l_i}}exp\left(-\frac{1}{2l_i\sigma^2}(y_i-X\boldsymbol{\beta})^2\right)l_i^{\frac{-v}{2}-1}exp\left(\frac{-v}{2l_i}\right)$$
$$=\prod_{i=1}^{n}l_i^{-\frac{1}{2}-\frac{v}{2}-1}exp\left[-\frac{1}{l_i}\left(\left(\frac{y_i-X\boldsymbol{\beta}}{\sqrt{2}\sigma}\right)^2+\frac{v}{2}\right)\right]$$
$$\sim I\Gamma\left(\frac{1}{2}+\frac{v}{2},\frac{(y_i-X\boldsymbol{\beta})^2}{2\sigma^2}+\frac{v}{2}\right)$$

بنابراین پارامتر l_i دارای توزیع گامای معکوس با پارامترهای $\frac{v}{2}+\frac{1}{2}$ و $\frac{v}{2}+\dfrac{(y_i-X\boldsymbol{\beta})^2}{2\sigma^2}$ است.

نمونه‌گیر گیبز برای مدل رگرسیون تی لاسو بیزی با نمایش آمیخته–مقیاس یکنواخت:

توزیع شرطی کامل پارامتر $\boldsymbol{\beta}$ به شرط $\mathbf{y}, X, u, \lambda, l, \sigma^2$ به صورت

$$\pi(\boldsymbol{\beta}|\mathbf{y}, X, \lambda, u, l, \sigma^2) \propto exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}-X\boldsymbol{\beta})^T L^{-1}(\mathbf{y}-X\boldsymbol{\beta})\right]\prod_{j=1}^{p} I\left\{|\beta_j| < \sqrt{\sigma^2 u_j}\right\}$$
$$\propto exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}-(X'L^{-1}X)^{-1}(X'L^{-1}\mathbf{y})\right)^T X'L^{-1}X\left(\boldsymbol{\beta}-(X'L^{-1}X)^{-1}(X'L^{-1}\mathbf{y})\right)\right]$$
$$\times\prod_{j=1}^{p} I\left\{|\beta_j| < \sqrt{\sigma^2 u_j}\right\} \sim N_p\left((X'L^{-1}X)^{-1}(X'L^{-1}\mathbf{y}), \sigma^2(X'L^{-1}X)^{-1}\right)$$
$$\times\prod_{j=1}^{p} I\left\{|\beta_j| < \sqrt{\sigma^2 u_j}\right\}$$

است. بنابراین می توان گفت دارای توزیع شرطی کامل پارامتر β به صورت توزیع نرمال چندمتغیره‌ی بریده شده است. به طور مشابه توزیع شرطی کامل پارامتر σ^2 به صورت

$$\pi(\sigma^2|\mathbf{y}, X, \boldsymbol{\beta}, u, \lambda, l) \propto \pi(\mathbf{y}|X, \boldsymbol{\beta}, \sigma^2, v, l)\pi(\boldsymbol{\beta}|u, \sigma^2)\pi(\sigma^2)$$
$$\propto\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\left(\frac{1}{\sqrt{\sigma^2}}\right)^p\left(\frac{1}{\sigma^2}\right) exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}-X\boldsymbol{\beta})^T L^{-1}(\mathbf{y}-X\boldsymbol{\beta})\right]$$
$$\times\prod_{j=1}^{p} I\left\{\sigma^2 > max_j\left(\frac{\beta_j^2}{u_j^2}\right)\right\}$$

به دست آمده که دارای توزیع گامای معکوس بریده شده چپ است. توزیع شرطی کامل برای u نیز به صورت

$$\pi(u|\mathbf{y}, X, \boldsymbol{\beta}, \lambda, v, \sigma^2, l) \propto \pi(\boldsymbol{\beta}|u, \sigma^2)\pi(u|\lambda)$$
$$\propto\prod_{j=1}^{p} e^{\lambda u_j} I\left\{u_j > \frac{|\beta_j|}{\sqrt{\sigma^2}}\right\} \sim \prod_{j=1}^{p} Exponential(\lambda) I\left\{u_j > \frac{|\beta_j|}{\sqrt{\sigma^2}}\right\}$$

است، یعنی توزیع شرطی کامل u به صورت توزیع نمایی بریده شده چپ است. در نهایت، توزیع شرطی کامل از پارامتر l به صورت

$$\pi(l|\mathbf{y}, X, \boldsymbol{\beta}, u, \lambda, v) \propto \pi(l|u)\pi(\mathbf{y}|X, \boldsymbol{\beta}, \sigma^2, v, l)$$
$$\propto\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi l_i}} exp\left(-\frac{1}{2l_i\sigma^2}(y_i - X\boldsymbol{\beta})^2\right) l_i^{-\frac{v}{2}-1} exp\left(\frac{-v}{2l_i}\right)$$
$$=\prod_{i=1}^{n} l_i^{-\frac{1}{2}-\frac{v}{2}-1} exp\left[-\frac{1}{l_i}\left(\left(\frac{y_i - X\boldsymbol{\beta}}{\sqrt{2}\sigma}\right)^2 + \frac{v}{2}\right)\right]$$

با توزیع گامای معکوس است.

Regression Modeling Via T-Lasso Bayesian Method

Zahra Khadem-Bashiri, Ali Shadrokh, [†], Masoud Yarmohammadi

Department of Statistics, Payame Noor University, P.O.Box 19395-4697, Tehran, Iran

Received: 2020/9/20 Accepted: 2021/1/21

Communicated by: G.R. Mohtashami Borzadaran

Abstract: Choosing the optimal model is one of the important issues in regression models. The purpose of optimal model selection methods in regression models is to determine important explanatory variables and negligible variables and to express the relationship between response variable and explanatory variables more simply. Due to the limitations of classical variable selection processes such as stepwise selection, penalized regression methods can be used. One of the penalized regression models is Lasso regression in which the errors are assumed to follow a normal distribution. For statistical analysis of the data set in the presence of outlier observations, the student's t distribution for error can be used and robust estimators can be provided. In this article, a variable selection method called Bayesian T-Lasso regression model is proposed based on Lasso Bayesian regression model in the presence of outlier observations in the data. The Bayesian T-Lasso regression model is presented with two different representations of the Laplace density function for the regression model coefficients, At the first the Laplace density function is represented as a scale mixture of normal distribution and then a scale mixture of uniform distribution. We demonstrate the utility of our Bayesian T-Lasso regression using simulation methods and real data analysis.

Keywords:Bayesian Lasso Regression, Penalized regression, Scale Mixture of uniform, Gibbs Sampling Algorithm, Variable Selection.

Mathematics Subject Classification (2010):62F15, 62J07,82B80.

[†] **Corresponding author:** Shadrokh.ali@gmail.com