

حاشا
الرحمن الرحيم



دانشکده علوم ریاضی

رشته آمار، گرایش آمار ریاضی

پایان نامه کارشناسی ارشد

رگرسیون جریمه شده بیزی

نگارنده: سید علی اصغر رحمتی

استاد راهنما

دکتر محمد آرشی

دی ۱۳۹۷

پایان نامه خود را تقدیم می کنم به پدر و
مادر و خانواده ام که همیشه در سختی ها
کنارم بوده اند

سپاس‌گزاری...

خدا را شاکرم که به من این توانایی را داد، تا پایان‌نامه خود را به پایان برسانم. از خانواده ممنونم که وجودشان همیشه قوت قلبی برای من بوده است. همچنین از استاد گرامی آقای دکتر محمد آرشی از صمیم قلب تشکر می‌کنم که راهنمایی‌های ایشان روشنی بخش راه من در این کار بوده است. در اینجا جا دارد از دو استاد عزیز آقایان دکتر باغیثنی و دکتر شاهسونی تشکر کنم که داوری این پایان‌نامه را قبول زحمت کردند.

سید علی اصغر رحمتی

دی ۱۳۹۷

تعهد نامه

اینجانب سید علی اصغر رحمتی دانشجوی کارشناسی ارشد رشته آمار علوم ریاضی دانشگاه صنعتی شاهرود، نویسنده پایان نامه با عنوان **رگرسیون جریمه شده بیزی**، تحت راهنمایی **محمد آرشی** متعهد می شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش های دیگر پژوهش گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام “ دانشگاه صنعتی شاهرود “ یا “ Shahrood University of Technology “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آوردن نتایج اصلی پایان نامه تاثیرگذار بوده اند، در مقالات مستخرج از پایان نامه رعایت می گردد.
- در تمام مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

سید علی اصغر رحمتی

دی ۱۳۹۷

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان نامه بدون ذکر منبع مجاز نمی باشد.

چکیده

در برخی از مدل‌های رگرسیونی تعداد متغیرهای توضیحی مورد بررسی زیاد است، که در این گونه موارد پیشنهاد می‌شود از روش‌های برآورد جریمه‌شده استفاده شود. از جمله این روش‌ها می‌توان به لاسو اشاره کرد که همزمان هم انتخاب متغیر انجام داده و هم پارامترهای مدل رگرسیونی را برآورد می‌کند. حال چنانچه هدف مدل‌بندی رگرسیونی از دیدگاه بیزی باشد می‌توان از روش رگرسیون جریمه‌شده بیزی استفاده کرد. در این مدل‌بندی معمولاً توزیع پیشین طوری انتخاب می‌شود که نتیجه حاصل شبیه یکی از برآورگرهای جریمه‌شده در حالت کلاسیک است. در این پایان‌نامه رگرسیون جریمه‌شده بیزی را مورد مطالعه قرار داده و دو رهیافت لاسوی بیزی و نعل اسبی بیزی را با جزئیات بیشتری مورد بررسی قرار می‌دهیم. در ادامه با استفاده تحلیل دو نوع داده دیابت، نحوه عملکرد رگرسیون جریمه‌شده بیزی را بررسی می‌کنیم.

کلمات کلیدی: رگرسیون جریمه‌شده بیزی، لاسوی بیزی، نعل اسبی بیزی، داده‌های دیابت، توزیع آمیخته مقیاسی نرمال، توزیع نیم کوشی.

پیشگفتار

در برخی مطالعات نشان داده شده است که استفاده از روش‌های بیزی برای برآورد پارامترهای مدل رگرسیونی نتایجی بهتر از روش‌های کلاسیک ارائه می‌دهد. یک دلیل بدیهی آن این است که ورود اطلاعات پیشین در خصوص پارامترهای مدل باعث افزایش اطلاعات موجود در برآورد پارامترها می‌شود. حال شرایطی را در نظر بگیرید که تعداد متغیرهای توضیحی در مدل رگرسیون زیاد است. در این شرایط یک راهکار مناسب در مدل‌بندی استفاده از رگرسیون جریمه‌شده بیزی است. در این گونه مدل‌بندی، توزیع پیشین پارامترهای رگرسیونی طوری در نظر گرفته می‌شود که عمل انتخاب متغیر در برآورد بیزی در نظر گرفته شود. به عبارت دقیق‌تر ابرپارامترهای توزیع پیشین طوری انتخاب می‌شوند که سطح تنکی (صفر برآورد شدن پارامترها) از دیدگاه بیزی کنترل شود. از جمله رهیافت‌های معروف در این زمینه می‌توان به دو روش لاسوی بیزی و نعل اسبی بیزی اشاره کرد. روش اول توسط تیبشیرانی (۱۹۹۶) و پارک و کسلا (۲۰۰۸) در ساختار رگرسیون لاسوی بیزی معرفی شد. در حقیقت آن‌ها پیشنهاد کردند که از توزیع پیشین لاپلاس برای پارامترهای ضرایب رگرسیونی استفاده شود. در ادامه کاروالهو و همکاران (۲۰۱۰) برآوردگر نعل اسبی بیزی را ارائه کردند که در آن از توزیع پیشین پولیا-گاما برای ضرایب رگرسیونی استفاده کردند. در این پایان‌نامه رگرسیون جریمه بیزی را مورد بررسی قرار داده و حالت‌های خاص آن را با جزئیات بیشتری مطالعه می‌کنیم. مرجع اصلی این پایان‌نامه مقاله اخیر مکالیک و اشمیت (۲۰۱۶) است. همچنین به کمک نرم‌افزار R داده‌های دیابت را تحلیل کرده و یک مطالعه شبیه‌سازی انجام می‌دهیم. این مجموعه شامل چهار فصل و سه پیوست است که محتوای آن‌ها به‌طور مختصر به صورت زیر است.

- در فصل یک، به بیان مفاهیم مقدماتی مدل‌های رگرسیونی که شامل تعاریف و چند مثال کوتاه است، می‌پردازیم.
 - در فصل دو، برآوردگر لاسوی بیزی را بیان می‌کنیم.
 - در فصل سه، برآوردگر نعل اسبی بیزی مورد بررسی قرار داده می‌شود.
 - در فصل چهار، برآوردگرهای جریمه‌شده بیزی از منظر کلی گفته شده و مفاهیمی از قبیل استواری سیگنال و میزان تنکی برآورد توضیح داده می‌شود.
- پیوست‌های ارائه شده در برگزیده چند قضیه و تعاریف مورد نیاز در پایان‌نامه بوده که آخرین بخش آن شامل گزیده‌ای از برنامه‌های کامپیوتری است.

فهرست مطالب

ش	فهرست تصاویر
ث	فهرست جداول
۱	۱ مقدمات مدل‌های رگرسیونی
۲	۱.۱ رگرسیون خطی چندگانه
۳	۱.۱.۱ برآورد پارامترهای مدل
۴	۱.۲ رگرسیون خطی چندگانه جریمه‌شده
۵	۱.۲.۱ برآوردگر لاسو
۶	۱.۲.۲ برآوردگر ریج
۹	۱.۳ رگرسیون خطی تعمیم‌یافته
۱۱	۱.۳.۱ برآورد ماکزیمم درست‌نمایی
۱۴	۱.۳.۲ برآورد بیزی
۱۷	۲ لاسوی بیزی
۱۸	۲.۱ رگرسیون لاسوی بیزی
۱۸	۲.۱.۱ مدل سلسله‌مراتبی
۲۲	۲.۲ انتخاب پارامتر جریمه
۲۳	۲.۲.۱ آگوریتم EM برای انتخاب λ
۲۴	۲.۲.۲ ابرپیشین‌ها برای پارامتر جریمه لاسو
۲۴	۲.۳ تحلیل مجموعه داده دیابت
۲۷	۳ نعل اسبی بیزی
۲۷	۳.۱ برآوردگر نعل اسبی
۲۹	۳.۲ تابع چگالی نعل اسبی
۳۰	۳.۳ ارتباط با توزیع‌های پیشین مشابه
۳۲	۳.۴ استواری نسبت به سیگنال‌های بزرگ

۳۵	برآوردگرهای جریمه شده از منظر کلی	۴
۳۶	رگرسیون جریمه شده بیزی	۴.۱
۳۸	مدل بندی چندگانه	۴.۲
۴۱	رگرسیون خطی	۴.۲.۱
۴۵	رگرسیون دودوئی لجستیک	۴.۲.۲
۴۷	توزیع های پیشین	۴.۳
۵۱	شرح بسته bayesreg در R	۴.۴
۵۲	مثال	۴.۵
۵۲	داده های دیابت	۴.۵.۱
۵۴	نتیجه گیری و پیشنهادات برای آینده تحقیق	۴.۶
۵۵	مراجع	
۶۳	آ	
۶۳	روش نمونه گیری MCMC	۱.آ
۶۸	الگوریتم EM	۲.آ
۷۱	ب توزیع های آماری	
۷۱	توزیع های تک متغیره	۱.ب
۷۴	توزیع آمیخته مقیاسی نرمال چند متغیره	۲.ب
۷۴	قضایای حدی و احتمالاتی	۳.ب
۷۷	پ گزیده ای از برنامه R	

فهرست تصاویر

- ۱.۲ نمودار جعبه‌ای برای لاسوی بیزی ۲۶
- ۱.۳ چگالی‌های $p(k_i)$: نمایی دوگانه، کوشی، استرودرمن-برگر، نرمال-نمایی-گاما، نرمال-جفریز و نعل اسبی. در قطعه مستطیلی سمت چپ برای پیشین نرمال-نمایی-گاما، خط مشکی برای $c = 4$ و $d = 1$ و خط قطعه
۳۲ قطعه برای $c = \frac{1}{4}$ و $d = 1$ است.
- ۱.۴ آنالیز داده‌های دیابت برای لاسوی بیزی با پاسخ لجستیک ۵۴

فهرست جداول

۱۱	۱.۱ پارامترها و ویژگی‌های GLM های خاص که رابطه (۱۱.۱) نماد گذاری شده است. θ در اینجا پارامتر کانونی می‌باشد و همچنین $E[y] = \mu$ و $var(y) = \alpha V(\mu)$
۲۵	۱.۲ ۱۰ متغیر پایه برای ۴۴۲ بیمار دیابتی
۳۱	۱.۳ پیشین‌ها برای λ_i و k_i وابسته به تعدادی از قاعده‌های متداول انقباضی موضعی برای پیشین نرمال-نمایی-گاما، $d = 1$ فرض شده است.
۵۳	۱.۴ پیشگویی پارامترها بر اساس مدل لجستیک از دیدگاه بیزی برای داده‌های دیابت

فصل ۱

مقدمات مدل‌های رگرسیونی

مقدمه

در سال‌های اخیر، مدل خطی تعمیم‌یافته^۱ (GLM) به طور چشم‌گیری استفاده شده است و متون آماری متعددی در این زمینه به رشته تحریر درآمده‌اند. در این راستا، توجه خواننده را به چند مرجع اصلی جلب می‌کنیم. مدل‌های خطی تعمیم‌یافته، به عنوان تعمیم مدل‌های خطی، توسط نلدر و ودربرن^۲ (۱۹۷۲) معرفی شد. بررسی نظری دقیق این دسته از مدل‌ها در مک کولاک^۳ و نلدر (۱۹۸۹) آمده است. میر^۴ و همکاران (۲۰۰۲) نیز منبع خوبی برای دیدگاه شهودی این مدل‌ها است. لازم به ذکر است که بیشترین استفاده از این مدل‌ها به عنوان رقیبی قوی در برابر مدل‌های لگ خطی، در تحلیل داده‌های گسسته گسترش یافته است. برای اطلاعات بیشتر در این مورد و ارتباط با مدل‌های لگ خطی به اگری^۵ (۱۹۹۶) مراجعه کنید. در دانشگاه صنعتی شاهرود نیز پایان‌نامه‌های زیادی در جهت بررسی جنبه‌های نظری و کاربردی این مدل از دیدگاه‌های کلاسیک و بیزی به رشته تحریر درآمده‌اند. از جمله ولیپور (۱۳۹۲) مدل‌بندی داده‌های نرخ و نسبت با رگرسیون بتا و عجم (۱۳۹۳) تحلیل بیزی مدل‌های

^۱Generalized linear model

^۲Nelder and Wedderburn

^۳McCullagh

^۴Meyer

^۵Agresti

رگرسیونی بتا با پاسخ‌های وابسته را بررسی کرده‌اند. کیحانی (۱۳۹۴) مدل‌بندی پاسخ‌های نرخ و نسبت با رگرسیون مستطیلی بتا را ارزیابی کرد، بلوری (۱۳۹۵) پاسخ‌های دوسطحی نامتعادل با مدل رگرسیون مقادیر فرین تعمیم‌یافته را تحلیل کرد، عابدین‌پور (۱۳۹۶) به مطالعه یک مدل منعطف برای پیش‌گویی فضایی پاسخ‌های نرخ به کمک توزیع بتا دوجمله‌ای پرداخت، و رحمانی (۱۳۹۷) مدل خطی تعمیم‌یافته از دیدگاه کلاسیک در تحلیل داده‌های طولی، را مورد بررسی قرار داد.

برای ورود به بحث GLM ابتدا یک مدل رگرسیون خطی یا همان مدل رگرسیونی میانگین را معرفی کرده و سپس بیان می‌کنیم چطور می‌توان به جای مدل‌بندی تنها میانگین پاسخ، تابعی از میانگین پاسخ را به وسیله GLM مدل‌بندی کرد. اما از آنجایی که هدف اصلی این مطالعه برآورد جریمه‌شده از دیدگاه بیزی است در این فصل ابتدا مدل رگرسیون خطی چندگانه و برآورد جریمه‌شده را معرفی کرده و سپس در ادامه آن، به طور مختصر به بحث در GLM و برآورد در دیدگاه بیزی می‌پردازیم.

۱.۱ رگرسیون خطی چندگانه

رابطه p متغیر توضیحی و یک متغیر پاسخ در رگرسیون خطی چندگانه^۶ به صورت زیر فرمول‌بندی می‌شود

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1.1)$$

که در آن $\mathbf{y} = (y_1, \dots, y_n)^\top$ یک بردار $n \times 1$ از پاسخ‌ها، $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ یک ماتریس $n \times p$ غیر تصادفی و معلوم از n بردار p - بعدی، $\mathbf{x}_i \in \mathbb{R}^p$ ، $i = 1, \dots, n$ ، $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ بردار ضرایب رگرسیون مجهول (پارامترهای مدل) و $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$ بردار خطاهای تصادفی است. به عبارتی دیگر داریم

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

که در آن $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$.

اغلب در مسائل کاربردی، ستون اول \mathbf{X} بردار $n \times 1$ از یک‌ها، به منظور برآورد کردن مقدار ثابت در مدل رگرسیون، بوده است که در این حالت بعد ماتریس \mathbf{X} ، $n \times (p+1)$ است. در این مجموعه بدون کاسته شدن از کلیت مسئله فرض می‌کنیم عرض از مبدا از مدل حذف شده است. برای چگونگی انجام و جزئیات محاسبات به رجبلو (۱۳۹۷) مراجعه کنید.

^۶Multiple linear regression

رگرسیون خطی چندگانه تعمیمی از رگرسیون خطی ساده^۷ است. این روش پارامتری هنگامی مناسب است که فرض خطی بودن رابطه‌ی میان متغیر پاسخ و متغیرهای توضیحی دست کم به صورت تقریبی برقرار باشد. اما هنگامی که رابطه میان متغیر پاسخ و متغیرهای توضیحی نامعلوم است، استفاده از رگرسیون خطی ممکن است منجر به نتایج گمراه‌کننده‌ای شود.

۱.۱.۱ برآورد پارامترهای مدل

مدل خطی (۱.۱) را در نظر گرفته و فرض کنید ماتریس واریانس-کوواریانس مولفه خطا برابر $\sigma^2 \mathbf{I}$ باشد، به عبارتی $cov(\epsilon) = \sigma^2 \mathbf{I}$ ، که در آن $\sigma^2 > 0$ یک مقدار عددی نامعلوم و \mathbf{I} ماتریس همانی است.

برآوردگر کمترین توان دوم^۸ (LS)، برآوردگری است که مجموع توان‌های دوم مانده‌ها یعنی

$$\begin{aligned} S(\beta) &= \sum_{i=1}^n \epsilon_i^2 \\ &= \sum_{i=1}^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \\ &= (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) \\ &= \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{X}\beta + \beta^\top \mathbf{X}^\top \mathbf{X}\beta \end{aligned} \quad (2.1)$$

را مینیمم کند. با مشتق‌گیری از رابطه (۲.۱) نسبت به β داریم

$$\frac{\partial S}{\partial \beta} = 2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\beta.$$

با مساوی صفر قرار دادن $\frac{\partial S}{\partial \beta}$ ، در صورت نامنفرد بودن ماتریس $\mathbf{X}^\top \mathbf{X}$ (معکوس پذیر بودن)، برآوردگر LS بردار ضرایب رگرسیونی β به صورت زیر به دست می‌آید

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (3.1)$$

در این صورت بردار مقادیر برازش شده \hat{y} با استفاده از مقادیر مشاهده شده y چنین خواهد بود

$$\hat{y} = \mathbf{X}\hat{\beta} = \mathbf{H}\mathbf{y}$$

که در آن $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ ماتریس هت یا برازش شده نامیده می‌شود زیرا باعث می‌شود روی y یک علامت هت گذاشته شود، و از طرفی بردار مقادیر مشاهدات را به مقادیر برازش شده تصویر می‌کند. ماتریس \mathbf{H} ماتریسی $p \times p$ ، متقارن و خودتوان است. بنابراین قضیه **۱.۳.ب** رابطه زیر بین رتبه و اثر ماتریس برقرار است

$$\text{rank}(\mathbf{H}) = \text{tr}(\mathbf{H}) = p.$$

^۷ Ordinary linear regression

^۸ Least squares

از آنجایی که $cov(\mathbf{y}) = \sigma^2 \mathbf{I}$ ، ماتریس واریانس-کوواریانس $\hat{\beta}$ نیز به صورت زیر محاسبه می‌شود

$$\begin{aligned} cov(\hat{\beta}) &= cov\left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}\right) \\ &= (\mathbf{X}^T \mathbf{X}) \mathbf{X}^T cov(\mathbf{y}) \left((\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X}^T\right)^T \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T cov(\mathbf{y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \end{aligned}$$

۲.۱ رگرسیون خطی چندگانه جریمه‌شده

گاهی اوقات تعداد متغیرهای توضیحی در مسئله مورد بررسی زیاد است. بدیهی است مدل حاصل پیچیده بوده و تفسیر آن مشکل است. مثلاً در مدل (۱.۱) چنانچه p بزرگ باشد ماتریس $(\mathbf{X}^T \mathbf{X})^{-1}$ به راحتی قابل محاسبه نیست و در حالت $p > n$ وجود ندارد. در این‌گونه موارد استفاده از یک روش انتخاب متغیر می‌تواند کارساز باشد. اما از آنجایی که هدف نهایی، برآورد بردار ضرایب رگرسیونی β است، بهتر است از روش‌هایی استفاده کنیم که هم انتخاب متغیر انجام داده و هم β را برآورد کند. برای این منظور، می‌توان از رگرسیون جریمه‌شده استفاده کرد و با روشی به هدف مورد نظر رسید. دقت کنید چنانچه تعداد پارامترها زیاد باشد تابعی از β که قدرمطلق عناصر آن را جمع می‌کند نیز بزرگ می‌شود. بنابراین اگر در این گونه مسائل بزرگی این تابع را جریمه کنیم و کاهش دهیم، عملاً می‌توان به طور غیر مستقیم تعداد متغیرها را کاهش داد. به عنوان نتیجه از این بحث می‌توان به جای مینیمم محک LS، محک LS همراه با بزرگی این تابع جریمه را همزمان مینیمم کرد. این عمل باعث می‌شود به طور همزمان هم کاهش بعد انجام داده و هم بردار β را برآورد کنیم. لذا در حالت کلی در رگرسیون جریمه شده^۹، هدف مینیمم کردن تابع زیر است

$$\begin{aligned} \phi(\beta) &= S(\beta) + \lambda P(\beta) \\ &= (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda P(\beta), \end{aligned} \quad (۴.۱)$$

که در آن $P(\beta)$ تابع جریمه بر حسب β و λ ثابتی است که میزان جریمه را کنترل یا تنظیم می‌کند. با دقت در ساختار تابع (۴.۱)، استنباط می‌شود که به تابع توان دوم خطا در رگرسیون معمولی، عبارت جریمه $P(\beta)$ اضافه شده است. این عبارت جریمه، تابعی از پارامترهاست و بنا به روش رگرسیون جریمه‌شده، صورت‌های متفاوتی دارد. برای آگاهی بیشتر رجیلو (۱۳۹۷) را ببینید.

در این جا پارامتر تنظیم‌کننده^{۱۰} $\lambda > 0$ است. لازم به ذکر است λ در اینجا، موازنه‌ای بین

^۹Penalized regression

^{۱۰}Tuning parameter

خطاهای پیش‌گویی^{۱۱} و پیچیدگی مدل^{۱۲} انجام می‌دهد و انتخاب درست آن حائز اهمیت است. تعیین مقدار مناسب برای پارامتر تنظیم‌کننده در آمار نیز، یک مسئله چالش‌برانگیز است. برای جزئیات بیشتر درباره ارتباط بین مسائل بهینه‌سازی ریاضی با آمار و به‌ویژه روش‌های رگرسیونی، وانگ^{۱۳} و همکاران (۲۰۰۴)، یائو^{۱۴} (۲۰۰۴) و نوروزی راد (۱۳۹۶) را ببینید. جریمه‌های متفاوتی می‌توانند در نظر گرفته شوند که در هر حالت، برآوردگر حاصل از مینیمم‌سازی (۴.۱) نام خاصی به خود می‌گیرد. جریمه‌های $\sum_{j=1}^p |\beta_j|$ و $\sum_{j=1}^p \beta_j^2$ از معروفترین توابع جریمه بوده که به ترتیب بر اساس نرم‌های مستطیلی (L_1) و اقلیدسی (L_2) هستند. در ادامه برآوردگرهای جریمه‌شده حاصل از مینیمم‌سازی با استفاده از نرم‌های L_1 و L_2 ، به طور مختصر آورده شده است.

۱.۲.۱ برآوردگر لاسو

مدل رگرسیون خطی (۱.۱) را در نظر بگیرید. تیبشیرانی^{۱۵} (۱۹۹۶) روشی جدید برای انتخاب متغیرها پیشنهاد نمود که منجر به مدل دقیق، پایا و صرفه‌جو (تفسیرپذیرتر) می‌شود و آن را عملگر کمترین قدرمطلق انقباضی و انتخاب متغیر^{۱۶} (LASSO) نامید که ما آن را به اختصار لاسو می‌گوییم.

لاسو به حلقه طنابی گفته می‌شود که یک گاوچران به سمت هدف مشخصی پرتاب می‌کند و پس از اصابت کردن به آن، طناب کشیده شده و حلقه کوچکتر می‌شود. لذا از دیدگاه آماری لاسو، برآوردگری است که حاصل عمل انتخاب‌کنندگی، منقبض‌کنندگی و مینیمم‌کنندگی از جنس قدرمطلق است. برآوردگر لاسو، نوع جریمه شده برآوردگر LS است که بر اساس ویژگی تنکی نرم L_1 ، در سال‌های اخیر مورد توجه قرار گرفته است. برای آگاهی در خصوص ویژگی تنکی پیشنهاد می‌شود آرست (۱۳۹۵) را ببینید. برآوردگر لاسو از حل مسئله بهینه‌سازی زیر به دست می‌آید

$$\hat{\beta}_L = \min_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 \right\} \quad s.t. \quad \sum_{j=1}^p |\beta_j| \leq t, \quad (5.1)$$

که در آن t یک مقدار ثابت است. اگر $t = 0$ ، آن‌گاه متغیری در مدل وجود نخواهد داشت در حالی که اگر $t = \infty$ ، آن‌گاه مدل، کامل خواهد بود به عبارتی همه متغیرها در مدل وجود

^{۱۱} Prediction errors

^{۱۲} Model complexity

^{۱۳} Wang

^{۱۴} Yao

^{۱۵} Tibshirani

^{۱۶} Least absolute shrinkage and selection operator

دارند و هیچ متغیری انتخاب نشده است. مسئله بهینه‌سازی (۵.۱) معادل با

$$\hat{\beta}_L = \arg \min_{\beta} \left\{ (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (6.1)$$

است، که در آن λ همان پارامتر تنظیم‌کننده است و سطح تنگی^{۱۷} (تعداد پارامترهای صفر) را در برآوردگر لاسو کنترل می‌کند. وقتی $\lambda \rightarrow 0$ ، $\hat{\beta}_L \rightarrow \hat{\beta}$ ، و وقتی $\lambda \rightarrow \infty$ ، $\hat{\beta}_L = 0$. به عبارت دیگر، هر چه جریمه بزرگتری به کار برده شود، تعداد بیشتری از ضرایب به سمت صفر منقبض می‌شوند. در حقیقت λ و رفتار یا رابطه‌ای معکوس با یکدیگر دارند. به رگرسیون لاسو، به خاطر شکل تابع جریمه آن که بر اساس نرم L_1 است، رگرسیون جریمه‌شده L_1 نیز می‌گویند. انتخاب این تابع جریمه باعث می‌شود که برآوردگر لاسو، ضرایب را به سمت صفر منقبض کرده و متغیرهای اضافی را از مدل حذف کند. در واقع، برآوردگر لاسو، همزمان هم انتخاب متغیر انجام می‌دهد و هم ضرایب را منقبض کرده به عنوان برآورد β ارائه می‌دهد. تابع جریمه قدرمطلق باعث می‌شود که جواب صریحی برای برآوردگر لاسو وجود نداشته باشد. تیبشیرانی (۱۹۹۶) این برآوردگر را از طریق برنامه‌ریزی درجه دوم به‌دست آورده بود. افرون^{۱۸} و همکاران (۲۰۰۴) نوع دیگری از رگرسیون گام گام، به نام رگرسیون کمترین زاویه^{۱۹} (LAR) معرفی کردند که برآوردهای لاسویی را نتیجه می‌دهد که هزینه محاسباتی آن با برآوردهای کمترین توان دوم برابر است. برای آگاهی بیشتر نوروژی راد (۱۳۹۶) و رجبلو (۱۳۹۷) را ببینید. برای مشاهده خواص این برآوردگر از قبیل اریبی و واریانس به آرست (۱۳۹۵) مراجعه کنید.

۲.۲.۱ برآوردگر ریج

قبل از اینکه به طور مستقیم و بلافاصله محک LS را همراه با تابع جریمه نرم L_2 مینیمم کرده و برآوردگر ریج را تعریف کنیم، به طور کلی مشکل همخطی را که در ارتباط مستقیم با برآوردگر ریج است توضیح می‌دهیم.

مدل رگرسیون خطی (۱.۱) را در نظر بگیرید، اگر توان دوم تفاضل β از مقدار واقعی آن را به صورت زیر در نظر بگیریم

$$L_2^{\lambda} = (\hat{\beta} - \beta)^\top (\hat{\beta} - \beta)$$

^{۱۷}Sparsity

^{۱۸}Efron

^{۱۹}Least angle regression

$$\begin{aligned}
 E(L_1^2) &= E((\hat{\beta} - \beta)^\top (\hat{\beta} - \beta)) \\
 &= \sum_{j=1}^p E(\hat{\beta}_j - \beta_j)^2 \\
 &= \sum_{j=1}^p \text{var}(\hat{\beta}_j) \\
 &= \sigma^2 \text{tr}((\mathbf{X}^\top \mathbf{X})^{-1}) \\
 &= \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j},
 \end{aligned}$$

که در آن λ_j ها مقادیر ویژه ماتریس $\mathbf{X}^\top \mathbf{X}$ است. در صورت وجود همخطی، حداقل یکی از مقادیر ویژه بسیار کوچک خواهد بود (روزبه، ۱۳۹۲ را ببینید). در این صورت امید توان دوم فاصله خیلی بزرگ شده و برآورد LS به طور متوسط خیلی بزرگ شده و از مقدار واقعی خود بسیار متفاوت است. با توجه به اینکه

$$E(L_1^2) = E(\hat{\beta}^\top \hat{\beta}) - \beta^\top \beta$$

داریم

$$E(\hat{\beta}^\top \hat{\beta}) = \beta^\top \beta + \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j}.$$

این رابطه نشان می‌دهد که در صورت وجود همخطی، برآورد LS β بردار پارامتر به طور متوسط دارای طولی بزرگتر از مقدار واقعی می‌باشد. یک روش برای رفع مشکل، یافتن برآوردگری است که میانگین توان دوم خطای ^{۲۰} کمتری نسبت به برآوردگر ناریب $\hat{\beta}$ دارد. اگر این برآوردگر را با $\hat{\beta}^*$ نشان دهیم، علاقه‌مندیم $MSE(\hat{\beta}^*) \leq MSE(\hat{\beta})$ که در آن

$$\begin{aligned}
 MSE(\hat{\beta}^*) &= E((\hat{\beta}^* - \beta)^\top (\hat{\beta}^* - \beta)) \\
 &= \text{tr}(\text{cov}(\hat{\beta}^*)) + \text{Bias}^\top(\hat{\beta}^*) \text{Bias}(\hat{\beta}^*)
 \end{aligned}$$

و $\text{Bias}(\hat{\beta}^*) = E(\hat{\beta}^* - \beta)$ مقدار اریبی برآوردگر $\hat{\beta}^*$ است. با کوچک در نظر گرفتن میزان $\hat{\beta}^*$ ، واریانس $\hat{\beta}^*$ را می‌توان کوچک کرد، به طوری که $MSE(\hat{\beta}^*)$ کمتر از واریانس برآوردگر ناریب $MSE(\hat{\beta})$ شود. روش‌های متعددی برای غلبه با مشکل همخطی در مدل‌های خطی ارائه شده‌اند. یکی از موثرترین روش‌ها برای حل مشکل همخطی، رگرسیون ریج است که توسط هورل و کنارد ^{۲۱}

^{۲۰} Mean squares error

^{۲۱} Hoerel and Kennard

(۱۹۷۰) معرفی شد.

برآوردگر ریج از حل مسئله بهینه‌سازی زیر به دست می‌آید

$$\hat{\beta}_{Ridge} = \min_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 \right\} \quad s.t. \quad \sum_{j=1}^p \beta_j^2 \leq t \quad (7.1)$$

که در آن t یک مقدار ثابت است. برآوردگر حاصل، برآوردگر ریج^{۲۲} نام داشته، که به پارامتر ریج بستگی دارد. مسئله بهینه‌سازی (۷.۱) معادل با حل مسئله (۸.۱) بوده که برآوردگر ریج را به صورت زیر ارائه می‌دهد

$$\begin{aligned} \hat{\beta}_{Ridge} &= \arg \min_{\beta} \left\{ (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) + k \sum_{j=1}^p \beta_j^2 \right\} \quad (8.1) \\ &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}^\top \mathbf{Y} \\ &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} (\mathbf{X}^\top \mathbf{X}) \hat{\beta}_{OLS} \\ &= \mathbf{T}(k) \hat{\beta}_{OLS}, \end{aligned} \quad (9.1)$$

که در آن $\beta^\top \beta = \sum_{j=1}^p \beta_j^2$ و $\mathbf{T}(k) = (k(\mathbf{X}^\top \mathbf{X})^{-1} + \mathbf{I}_p)^{-1}$ است. واضح است که برآوردگر ریج یک تبدیل خطی از برآوردگر LS است. امید ریاضی $\hat{\beta}_{Ridge}$ عبارت است از

$$\begin{aligned} E(\hat{\beta}_{Ridge}) &= E((\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}^\top \mathbf{Y}) \\ &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}^\top \mathbf{X} \beta. \end{aligned}$$

بنابراین $\hat{\beta}_{Ridge}$ برآوردگر اریب β است. همچنین ماتریس واریانس-کواریانس آن به صورت زیر می‌باشد

$$cov(\hat{\beta}_{Ridge}) = \sigma^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1}.$$

بنابراین می‌توان MSE برآوردگر ریج را به صورت زیر نوشت

$$\begin{aligned} MSE(\hat{\beta}_{Ridge}) &= tr(cov(\hat{\beta}_{Ridge})) + Bias^\top(\hat{\beta}_{Ridge}) Bias(\hat{\beta}_{Ridge}) \\ &= \sigma^2 tr((\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1}) \\ &\quad + k^2 \beta^\top (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2} \beta. \end{aligned} \quad (10.1)$$

اولین جمله سمت راست (۱۰.۱) مجموع واریانس مولفه‌های پارامتر $\hat{\beta}_{Ridge}$ و دومین عبارت، توان دوم اریبی (ترانهاده اریبی در خودش) است. بنابراین اریبی $\hat{\beta}_{Ridge}$ با افزایش k ، افزایش و واریانس آن با افزایش k ، کاهش می‌یابد و بر عکس.

با استفاده از تعریف ب.۳.۱ ماتریس $\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p$ را می‌توان به صورت $\Gamma(\Lambda + k\mathbf{I}_p)\Gamma^\top$ تجزیه کرد که در آن، Γ یک ماتریس متعامد است و ستون‌های آن بردارهای ویژه ماتریس $\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_p$

^{۲۲}Ridge estimator

بوده و $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_p)$ λ_i ها مقادیر ویژه ماتریس $\mathbf{X}^T \mathbf{X}$ است $\lambda_1 > \lambda_2 > \dots > \lambda_p \geq 0$. لذا می توان نوشت

$$\begin{aligned} \text{MSE}(\hat{\beta}_{\text{Ridge}}) &= \sigma^2 \text{tr}((\mathbf{\Gamma}(\Lambda + k\mathbf{I}_p))^{-1} \mathbf{\Gamma}^T \mathbf{\Gamma} \Lambda \mathbf{\Gamma}^T \mathbf{\Gamma}(\Lambda + k\mathbf{I}_p) \mathbf{\Gamma}^T)) \\ &+ k^2 \beta^T \mathbf{\Gamma}(\Lambda + k\mathbf{I}_p)^{-2} \mathbf{\Gamma}^T \beta \\ &= \sigma^2 \text{tr}((\Lambda + k\mathbf{I}_p)^{-1} \Lambda (\Lambda + k\mathbf{I}_p)^{-1}) + k^2 \alpha^T (\Lambda + k\mathbf{I}_p)^{-2} \alpha \\ &= \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2}, \end{aligned}$$

که در آن $\alpha = \mathbf{\Gamma}^T \beta$. در رگرسیون ریدج، پارامتر k طوری انتخاب می شود که کاهش در واریانس بیشتر از افزایش توان دوم آریبی باشد. در این صورت MSE برآوردگر ریدج $\hat{\beta}_{\text{Ridge}}$ کمتر از واریانس برآوردگر کمترین توان های دوم $\hat{\beta}_{\text{OLS}}$ خواهد بود. هورل و کنارد (۱۹۷۰) ثابت کردند که اگر $\beta^T \beta$ کراندار باشند، مقدار غیر صفری از k وجود دارد به طوری که مخاطره $\hat{\beta}_{\text{Ridge}}$ از مخاطره برآوردگر کمترین توان های دوم $\hat{\beta}_{\text{OLS}}$ کمتر خواهد شد. همچنین برای حالت ماتریس طرح متعامد، یعنی زمانی که $\mathbf{X}^T \mathbf{X}$ ، برآوردگر ریدج یک برآوردگر انقباضی است. برای مشاهده جزئیات به رجبو (۱۳۹۷) مراجعه کنید.

۳.۱ رگرسیون خطی تعمیم یافته

تقریباً در اغلب موارد استفاده از روش های آماری، مرکز توجه روی میانگین ها و تغییرات آنهاست، که منجر به ایجاد روش های استنباط درباره میانگین ها یا درباره ماهیت آنها می شود. معمول ترین مدل رگرسیونی، مدل خطی (۱.۱) است. در یک مدل خطی، با متغیر پاسخ y و ماتریس طرح \mathbf{X} ، تابع رگرسیونی که همان میانگین شرطی $y | \mathbf{X}$ است، بر حسب ترکیب خطی از \mathbf{X} به صورت $E[y | \mathbf{X}] = \mathbf{X}\beta$ در نظر گرفته می شود. در این مدل، ضرایب β بردار ثابت های نامعلوم، یا به عبارتی اثرات ثابت نامیده می شوند.

یکی از پذیره های مدل های خطی، استقلال بین مشاهدات است. زمانی که برای پاسخ ها این پذیره برقرار نیست، تعمیمی از این مدل ها معروف به مدل های آمیخته خطی^{۲۳} (LMMs) مورد استفاده قرار می گیرد. برای در نظر گرفتن وابستگی بین پاسخ ها، استفاده از متغیرهای پنهان یا همان اثرات تصادفی یک راه رایج و مناسب است. میانگین پاسخ در یک LMM به صورت $E[y | \mathbf{X}, \mathbf{Z}, \mathbf{b}] = \mathbf{X}\beta + \mathbf{Z}\mathbf{b}$ نمایش داده می شود که در آن \mathbf{X} ماتریس طرح متناظر با اثرات ثابت β و \mathbf{Z} ماتریس طرح متناظر با بردار اثرات تصادفی \mathbf{b} است (مک کولاک و سیرل^{۲۴}، ۲۰۰۱). اثرات تصادفی معمولاً به عنوان متغیرهای تصادفی نرمال (با میانگین صفر) تلقی شده و پارامترهایی که توزیع این اثرات را توصیف می کنند، برآورد می شوند.

^{۲۳} Linear mixed models

^{۲۴} Searle

در بسیاری از مسایل، مانند پاسخ‌های دودویی و شمارشی و مدل‌هایی که در مقدمه این فصل به آن‌ها اشاره شد، استفاده از توزیع نرمال و مدل خطی مناسب نیست. زیرا مدل‌های خطی محدودیت موجود در میانگین پاسخ برای این داده‌ها را در نظر نمی‌گیرند. به عنوان مثال برای داده‌ها دودویی، میانگین که همان احتمال موفقیت می‌باشد، در فاصله (۰, ۱) قرار می‌گیرد و چنانچه از مدل خطی استفاده کنیم ممکن است برآورد میانگین پاسخ هر مقداری در مجموعه اعداد حقیقی باشد. در نتیجه برای این داده‌ها با فرض توزیع برنولی، معمولاً، از رگرسیون پروبیت یا لجستیک استفاده می‌شود. برای مرتفع ساختن این مشکل نلدر و ودربرن (۱۹۷۲) مدل خطی تعمیم‌یافته را معرفی کردند.

مدل خطی تعمیم‌یافته دارای ویژگی‌های مهمی است که در ادامه به طور مختصر تک تک آن‌ها را بررسی می‌کنیم.

- متغیر پاسخ y_i دارای توزیعی از خانواده توزیع‌های نمایی است که تابع (چگالی) احتمال آن عبارتست از

$$p(y_i | \theta_i, \alpha) = \exp\left(\frac{y_i \theta_i - b(\theta_i)}{\alpha} + c(y_i, \alpha)\right) \quad (11.1)$$

که در آن θ_i و α برای توابع $b(\cdot)$ و $c(\cdot, \cdot)$ به صورت اسکالر یا عددی بوده و به صورت ساده امید ریاضی و واریانس در زیر نشان داده می‌شوند

$$\begin{aligned} E(y_i | \theta_i, \alpha) &= \mu_i = b'(\theta_i) \\ &= \frac{db(\theta_i)}{d\theta_i} \end{aligned}$$

و

$$\begin{aligned} \text{var}(y_i | \theta_i, \alpha) &= \alpha b''(\theta_i) \\ &= \alpha \text{var}(\mu_i) \end{aligned}$$

که در آن $b'(\cdot)$ و $b''(\cdot)$ به ترتیب مشتقات اول و دوم هستند. همچنین برای $i = 1, \dots, n$ فرض می‌کنیم به ازای $i \neq j$ ، $\text{cov}(y_i, y_j | \theta_i, \theta_j, \alpha) = 0$. در ادامه جدول ۱.۱، برخی پارامترهای مدل (۱۱.۱) را در توزیع‌های خاص معروف مشخص می‌کند.

- تابع پیوند $g(\cdot)$ ارتباط بین تابع میانگین $\mu_i = E(y_i | \theta_i, \alpha)$ و پیش‌گویی خطی $\mathbf{x}_i^\top \beta$ از طریق

$$g(\mu_i) = \mathbf{x}_i^\top \beta$$

را فراهم می‌کند که در اینجا \mathbf{x}_i یک بردار $p \times 1$ از متغیرهای توضیحی و $\beta = (\beta_1, \dots, \beta_p)^\top$ یک بردار $p \times 1$ از پارامترهای رگرسیونی است. به طور خلاصه، یک GLM یک رابطه‌ی خطی را در یک مقیاس میانگین تبدیل شده و یک صورت خانواده توزیع‌های نمایی برای توزیع پاسخ مشخص می‌کند.

جدول ۱.۱: پارامترها و ویژگی‌های GLM های خاص که رابطه (۱۱.۱) نماد گذاری شده است. θ در اینجا پارامتر کانونی می‌باشد و همچنین $E[y] = \mu$ و $var(y) = \alpha V(\mu)$.

توزیع	نرمال	پواسن	برنولی	گاما
$E[Y \theta]$	θ	$exp(\theta)$	$\frac{exp(\theta)}{1+exp(\theta)}$	$-\frac{1}{\theta}$
$V(\mu)$	۱	μ	$\mu(1 - \mu)$	μ^2
$b(\theta)$	$\frac{\theta^2}{2}$	$exp(\theta)$	$log(1 + e^\theta)$	$-log(-\theta)$
$c(y, \alpha)$	$-\frac{1}{2}[\frac{y^2}{\alpha} + log(2)\pi\alpha]$	$-logy!$	۱	$\frac{log \frac{y}{\alpha}}{\alpha} - log y + log \Gamma(\alpha)$

۱.۳.۱ برآورد ماکزیمم درست‌نمایی

مدل (۱۱.۱) را در نظر بگیرید برای دستیابی به برآوردگر ماکزیمم درست‌نمایی (ML) ^{۲۵} برای پارامتر β و توزیع مجانبی آن کافی است تابع امتیاز و ماتریس اطلاع فیشر را بیابیم. برای یک نمونه تصادفی به حجم n به صورت Y_1, \dots, Y_n از خانواده نمایی (۱۱.۱) لگاریتم تابع درست‌نمایی عبارتست از

$$l(\theta) = \prod_{i=1}^n p(y_i | \theta_i, \alpha)$$

$$= \sum_{i=1}^n \left(\frac{y_i \theta_i - b(\theta_i)}{\alpha} + c(y_i, \alpha) \right),$$

که در آن $\theta = \theta(\beta) = (\theta_1(\beta), \dots, \theta_n(\beta))^T$ برداری از پارامترهای کانونی است. با استفاده از قانون زنجیره‌ای، تابع امتیاز به صورت زیر به دست می‌آید

$$F(\beta) = \frac{\partial l(\theta)}{\partial \beta}$$

$$= \sum_{i=1}^n \frac{dl_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{\partial \mu_i}{\partial \beta}$$

$$= \sum_{i=1}^n \frac{y_i - b'(\theta_i)}{\alpha} \frac{1}{V_i} \frac{\partial \mu_i}{\partial \beta}, \quad (12.1)$$

که در اینجا $var(y_i | \beta) = \alpha V_i$ و برای $i = 1, \dots, n$

$$\frac{d^2 b}{d\theta_i^2} = \frac{d\mu_i}{d\theta_i} = V_i.$$

از اینرو

$$F(\beta) = \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^T \left(\frac{y_i - E(y_i | \mu_i)}{var(y_i | \mu_i)} \right)$$

$$= \frac{\mathbf{D}^T \mathbf{V}^{-1} [\mathbf{y} - \boldsymbol{\mu}(\beta)]}{\alpha}, \quad (13.1)$$

^{۲۵}Maximum likelihood

که در آن D یک ماتریس $n \times p$ با عناصر $\frac{\partial \mu_i}{\partial \beta_j}$ برای $i = 1, \dots, n$ و $j = 1, \dots, p$ بوده و همچنین V یک ماتریس قطری $n \times n$ با عنصر قطری i ام V_i است. بدیهی است برآوردگر ML بردار ضرایب رگرسیونی β از حل $F(\beta) = 0$ حاصل می‌شود. به ازای $\theta_i = \mathbf{x}_i^\top \beta$ داریم

$$\begin{aligned} \sum_{i=1}^n \frac{\partial l_i}{\partial \beta} &= \sum_{i=1}^n \frac{dl_i}{d\theta_i} \frac{\partial \theta_i}{\partial \beta} \\ &= \frac{1}{\alpha} \sum_{i=1}^n \mathbf{x}_i^\top [y_i - \mu_i(\beta)] \end{aligned}$$

که با مساوی صفر قرار دادن نتیجه می‌شود برآوردگر $\hat{\beta}$ در رابطه زیر صدق می‌کند

$$\sum_{i=1}^n \mathbf{x}_i^\top y_i = \sum_{i=1}^n \mathbf{x}_i^\top \mu_i(\hat{\beta})$$

. قضیه زیر توزیع جانبی برآوردگر ML بردار $\theta_n = (\theta_1, \dots, \theta_n)$ را نشان می‌دهد.

قضیه ۱.۳.۱. (واکنیلد، ۲۰۱۳) فرض کنید جوابی برای مسئله زیر باشد

$$G_n(\theta) = \frac{1}{n} \sum_{i=1}^n G(\theta, y_i) = 0,$$

که در آن $G(\cdot, \cdot)$ تابعی بر حسب $\theta = (\theta_1, \dots, \theta_n)^\top$ و مشاهدات نمونه‌های تصادفی y_1, \dots, y_n بوده که دوبار مشتق‌پذیر نسبت به θ است. در نتیجه $\theta \xrightarrow{p} \hat{\theta}_n$ (سازگاری) و

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N_p[0, \mathbf{A}^{-1} \mathbf{B} (\mathbf{A}^\top)^{-1}] \quad (14.1)$$

(نرمال بودن جانبی)، که در آن

$$\mathbf{A} = \mathbf{A}(\theta) = E \left[\frac{\partial}{\partial \theta^\top} G(\theta, \mathbf{y}) \right]$$

و

$$\mathbf{B} = \mathbf{B}(\theta) = E(G(\theta, \mathbf{y})G(\theta, \mathbf{y})^\top) = \text{var}[G(\theta, \mathbf{y})].$$

برهان. با توجه به این که $\hat{\theta}_n$ جواب $G_n(\theta) = 0$ است، پس $G_n(\hat{\theta}_n) = 0$. بسط سری تیلور $G_n(\hat{\theta}_n)$ حول مقدار صحیح θ را به صورت زیر می‌نویسیم

$$0 = G_n(\hat{\theta}_n) = G_n(\theta) + (\hat{\theta}_n - \theta) \left. \frac{dG_n(\hat{\theta}_n)}{d\hat{\theta}_n} \right|_{\hat{\theta}=\theta} + \frac{1}{2} (\hat{\theta}_n - \theta)^2 \left. \frac{d^2 G_n(\hat{\theta}_n)}{d\hat{\theta}_n^2} \right|_{\hat{\theta}} \quad (15.1)$$

که در آن $\hat{\theta}$ نقطه بین θ و $\hat{\theta}_n$ است. از آنجایی که می‌خواهیم توزیع جانبی $\sqrt{n}(\hat{\theta}_n - \theta)$ را بیابیم، از رابطه (۱۵.۱) داریم

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{-\sqrt{n}G_n(\theta)}{\left. \frac{dG_n(\hat{\theta}_n)}{d\hat{\theta}_n} \right|_{\theta} + \frac{1}{2} (\hat{\theta}_n - \theta)^2 \left. \frac{d^2 G_n(\hat{\theta}_n)}{d\hat{\theta}_n^2} \right|_{\hat{\theta}}} \quad (16.1)$$

برای استفاده از قضیه حد مرکزی ب.۴.۳، توجه داشته باشید که طبق تعریف $E(G_n(\theta)) = 0$ و

$$n \times \text{var}(G_n(\theta)) = \text{var}(G(\theta, Y)) = E(G(\theta, Y)^2) = B(\theta).$$

بنابراین، با توجه به قضیه حد مرکزی و این فرض که $E(G(\theta, y)^2) < \infty$ داریم

$$\sqrt{n} G_n(\theta) \xrightarrow{d} N(0, B(\theta)) \quad (17.1)$$

اکنون خواص تابع $G(\theta)$ را از طریق رابطه (۱۶.۱) به برآوردگر $\hat{\theta}_n$ انتقال می‌دهیم. مشتق اول در رابطه (۱۶.۱) به صورت زیر است

$$\left. \frac{dG_n(\theta)}{d\theta} \right|_{\theta} = \frac{1}{n} \sum_{i=1}^n \left. \frac{d}{d\theta} G(\theta, Y_i) \right|_{\theta}$$

که طبق قانون ضعیف اعداد بزرگ ب.۲.۳ به امیدریاضی خودش همگرا است. بنابراین

$$\left. \frac{dG_n(\theta)}{d\theta} \right|_{\theta} \xrightarrow{p} E\left(\left. \frac{d}{d\theta} G(\theta, Y) \right)\right) = A(\theta)$$

با توجه به سازگاری، $\hat{\theta}_n \xrightarrow{p} \theta$ ، و این که رابطه دوم در مخرج (۱۶.۱) شامل میانگین است، به عبارتی

$$\left. \frac{d^2 G_n(\theta)}{d\theta^2} \right|_{\hat{\theta}} = \frac{1}{n} \sum_{i=1}^n \left. \frac{d^2}{d\theta^2} G(\theta, Y_i) \right|_{\hat{\theta}}$$

به امیدریاضی خودش (در صورت وجود) همگرا بوده و در این صورت

$$\left. \frac{d^2 G_n(\theta)}{d\theta^2} \right|_{\hat{\theta}} \xrightarrow{p} E\left(\left. \frac{d^2}{d\theta^2} G(\theta, Y) \right)\right).$$

از این رو، عبارت دوم در مخرج در رابطه (۱۶.۱) در احتمال همگرا به صفر بوده و با استفاده از قضیه اسلاتسکی ب.۵.۳ داریم

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, \frac{B}{A^2}\right)$$

و اثبات کامل است. □

با توجه به قضیه ۱.۳.۱، برآوردگر ML دارای توزیع جانبی به صورت

$$\mathbf{I}_n(\beta)^{\frac{1}{2}} (\hat{\beta}_n - \beta) \xrightarrow{d} N_{k+1}(0, \mathbf{I}_{k+1})$$

است که در اینجا

$$\mathbf{I}_n(\beta) = E[S(\beta)S(\beta)^T] = \frac{\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D}}{\alpha}$$

از آنجایی که در عمل \mathbf{D} و \mathbf{V} مجهول هستند از نسخه‌های برآوردشده آن‌ها استفاده می‌شود. بنابراین

$$\mathbf{I}(\hat{\beta}_n) = \frac{\hat{\mathbf{D}}^T \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}}}{\alpha}$$

که در آن \hat{V} و \hat{D} به ازای $\hat{\beta}_n$ محاسبه می‌شوند. برآوردگر واریانس برآوردگر $\hat{\beta}$ به صورت

$$\text{var}(\hat{\beta}) = \alpha(\hat{D}^T \hat{V}^{-1} \hat{D})^{-1} \quad (18.1)$$

است.

۲.۳.۱ برآورد بیزی

برای انجام استنباط آماری، دو دیدگاه استنباط بسامدی^{۲۶} (دیدگاه کلاسیک) و استنباط بیزی^{۲۷} در ادبیات آماری وجود دارند. هر کدام از آن‌ها هدف خاص خود را در استنباط آماری دنبال می‌کند. در دیدگاه بیزی، مقدار واقعی پارامتر، تحقق از یک توزیع تصادفی است. در واقع پارامترها خود متغیر تصادفی هستند. در دیدگاه بیزی اگر توزیع احتمالی پارامتر مورد بررسی به یک زیر فضا از فضای پارامتر وزن بیشتری بدهد، آن‌گاه اعتقاد بیشتری وجود دارد که پارامتر به آن زیر فضا تعلق دارد. در واقع به این اعتقاد، اعتقاد پیشین^{۲۸} و به توزیع احتمالی پارامتر توزیع پیشین می‌گویند. در این دیدگاه با ترکیب اطلاعات اولیه درباره‌ی پارامتر توسط توزیع و اطلاعات موجود در داده‌های مشاهده شده توسط درست‌نمایی، چگالی پسین^{۲۹} با روشی واحد بر اساس احتمال شرطی، به دست می‌آید. در ادامه مختصری از برآورد بیزی که برگرفته از بدخشان (۱۳۹۷) است را می‌آوریم.

فرض کنید $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ مشاهدات یک نمونه تصادفی n تایی با تابع درست‌نمایی $L(\theta | \mathbf{y})$ و $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$ بردار پارامتر دارای تابع چگالی پیشین $\pi(\theta)$ است. در این صورت، بنا بر قضیه بیز (احتمال شرطی)، توزیع پسین به صورت

$$\begin{aligned} \pi(\theta | \mathbf{y}) &= \frac{\pi(\theta).L(\theta | \mathbf{y})}{\int_{\Theta} \pi(\beta).L(\beta | \mathbf{y})d\theta} \\ &= c(\mathbf{y}).\pi(\theta).L(\theta | \mathbf{y}) \end{aligned}$$

محاسبه می‌شود، که در آن

$$c(\mathbf{y}) = \frac{1}{\int_{\Theta} \pi(\theta).L(\theta | \mathbf{y})d\theta}$$

ثابت نرمال‌ساز نامیده می‌شود که به پارامتر θ بستگی ندارد. بنابراین توزیع پسین را می‌توان به صورت تقریبی

$$\pi(\theta | \mathbf{y}) \propto L(\theta | \mathbf{y})\pi(\theta)$$

در نظر گرفت. در دیدگاه بیزی همه استنباط‌ها، مبتنی بر توزیع پسین به دست می‌آید. در ادامه استنتاج بیزی را برای GLM ها ارائه می‌دهیم.

^{۲۶}Frequentistic inference

^{۲۷}Bayesian inference

^{۲۸}Prior

^{۲۹}Posterior

تحت مفروضات بخش ۳.۱، توزیع پسین به صورت

$$\pi(\beta, \alpha | \mathbf{y}) \propto l(\beta, \alpha | \mathbf{y})\pi(\beta, \alpha)$$

در نظر گرفته می‌شود که در آن معمولاً فرض می‌شود توزیع‌های پیشین ضرایب رگرسیونی β و پارامتر مقیاس α مستقل از یکدیگرند به عبارتی فرض می‌کنیم

$$\pi(\beta, \alpha) = \pi(\beta)\pi(\alpha).$$

به یاد آورید که $\beta = (\beta_1, \dots, \beta_p)^T$ روی \mathbb{R}^p تعریف می‌شود، بنابراین یک پیشین نرمال چند متغیره برای β انتخابی واضح است. علاوه بر این، موارد محدودی وجود دارد که در آن‌ها پیشین ناسره^{۳۰} یعنی $\pi(\beta) \propto 1$ در نظر گرفته می‌شود. با این حال، باید این انتخاب با دقت انجام شود زیرا ممکن است منجر به پسین ناسره^{۳۱} شود. اگر پارامتر مقیاس، α مجهول باشد، توزیع‌های نرمال گاما یا لگ نرمال انتخاب‌های مناسبی می‌باشد. برای آگاهی بیشتر به واکنفیلد^{۳۲} (۲۰۱۳) مراجعه کنید. همچنین عجم (۱۳۹۳) که در مقدمه نیز به آن اشاره شد مرجع مناسبی برای بررسی دیدگاه بیزی در GLM است.

در ادامه یک مثال ساده از GLM را که در آن متغیر پاسخ شمارشی است با استفاده از توزیع پواسن در دیدگاه بیزی مورد بررسی قرار می‌دهیم.

مطالعه‌ای را در نظر بگیرید که افراد در ناحیه‌های مختلف با دریافت اشعه رادون در معرض بیماری خاص قرار می‌گیرند. اگر y_i تعداد افراد مبتلا به بیماری باشد آن‌گاه می‌توان فرض کرد که y_i دارای توزیع پواسن است. در این صورت در دیدگاه بیزی مدل پواسن با یک تابع پیوند خطی به صورت

$$y_i | \beta \stackrel{ind}{\sim} \text{Poisson}(E_i(\lambda - x_i)\beta_0 + x_i\beta_1)$$

در نظر گرفته می‌شود که در آن x_i و E_i به ترتیب متوسط رادون و تعداد مورد انتظار از نمونه‌ها در ناحیه i ، برای $i = 1, \dots, n$ است، و فرض کنید از یک توزیع پیشین یکنواخت برای β_0 در نظر می‌گیریم یعنی

$$\pi(\beta_0) \propto 1$$

تعریف می‌کنیم $e^\gamma = \frac{\beta_1}{\beta_0} > 0$ که به عنوان پارامتر مورد نظر در اینجا در نظر گرفته می‌شود و در این صورت

$$\mu_i = \beta_0 E_i[(\lambda - x_i) + x_i \exp(\gamma)] = \beta_0 \mu_i^*.$$

^{۳۰} Flat

^{۳۱} Improper Posterior

^{۳۲} Wakefield

پسین حاشیه‌ای برای γ عبارت است از

$$\begin{aligned} \pi(\gamma | \mathbf{y}) &= \int \pi(\beta_0, \gamma | \mathbf{y}) d\beta_0 \\ &\propto \int l(\beta_0, \gamma) d\beta_0 \times \pi(\gamma) \\ &\propto \int \exp\left(-\beta_0 \sum_{i=1}^n \mu_i^* y_i\right) \beta_0^{\sum_{i=1}^n y_i} \prod_{i=1}^n \mu_i^{*y_i} d\beta_0 \times \pi(\gamma) \\ &\propto \prod_{i=1}^n \left(\frac{E_i[(1-x_i) + x_i e^\gamma]}{\sum_{i=1}^n E_i[(1-x_i) + x_i e^\gamma]} \right)^{y_i} \times \pi(\gamma) \quad (19.1) \\ &= l(\gamma) \times \pi(\gamma), \quad (20.1) \end{aligned}$$

که در اینجا، خط سوم نشان می‌دهد که انتگرال، هسته‌ای از توزیع $\text{Gamma}(\sum_{i=1}^n y_i, \sum_{i=1}^n \mu_i^*)$ است. درست‌نمایی $l(\gamma)$ در رابطه‌ی (۲۰.۱)، از فرم چند جمله‌ای با تعداد کل موارد y_+ که میان n ناحیه با احتمالات متناسب با $E_i[(1-x_i) + x_i \exp(\gamma)]$ توزیع شده است پیروی می‌کند. برای مثال، E_i بزرگتر و x_i بزرگتر (اگر $\gamma > 0$) منجر به تخصیص بیشتر مشاهدات به ناحیه i می‌شود. در این صورت

$$l(\gamma) = \prod_{i=1}^n \left(\frac{E_i(1-x_i)}{\sum_{i=1}^n E_i(1-x_i)} \right)^{y_i} \quad (21.1)$$

وقتی که $\gamma \rightarrow -\infty$ میل کند. رابطه‌ی (۲۱.۱)، یک رابطه‌ی غیر صفر است مگر اینکه در هر ناحیه با $y_i \neq 0$ داشته باشیم $x_i = 1$. برای توضیحات بیشتر در خصوص تفسیر نتایج حالت $x_i = 1$ و $y_i \neq 0$ به واکیلد (۲۰۱۳) مراجعه کنید.

فصل ۲

لاسوی بیزی

مقدمه

برای اولین بار رگرسیون لاسوی بیزی^۱ توسط پارک و کسلا^۲ (۲۰۰۸) ارائه شد. جلوتر از آن‌ها فرناندز و استیل^۳ (۲۰۰۰) از پیشینی خاص در چهارچوب مدل‌سازی رگرسیون بیزی استفاده کردند اما پیوند و ارتباط خاصی با روش لاسو نداشت. در هر دو روش ارائه شده از توزیع آمیخته مقیاسی نرمال^۴ با توزیع نمایی دوگانه استفاده می‌شود. فرناندز و استیل (۲۰۰۰) و پارک و کسلا (۲۰۰۸) عدم اطمینان^۵ در ابر پارامترهای مدل رگرسیونی لاسوی بیزی را با در نظر گرفتن توزیع‌های پیشین برای ابرپارامترها گسترش دادند و برآوردهای نقطه‌ای ضرایب رگرسیونی^۶ را با استفاده از میانه توزیع پسین به‌دست آوردند. در این فصل مدل رگرسیونی لاسوی بیزی را معرفی کرده و تحقیقات پارک و کسلا (۲۰۰۸) را در این راستا ارائه می‌کنیم.

^۱ Bayesian lasso

^۲ Park and Casella

^۳ Fernandez and Steel

^۴ Scale mixture

^۵ Uncertainty

^۶ Regression coefficients

۱.۲ رگرسیون لاسوی بیزی

مدل رگرسیونی زیر را در نظر بگیرید

$$y = X\beta + \epsilon, \quad (1.2)$$

که در آن y یک بردار $n \times 1$ از پاسخ‌ها، $\beta = (\beta_1, \dots, \beta_p)^\top$ یک بردار $p \times 1$ از ضرایب رگرسیونی و X یک ماتریس $n \times p$ از متغیرهای توضیحی استاندارد شده و ϵ یک بردار $n \times 1$ از خطای مستقل و هم توزیع با توزیع نرمال با میانگین صفر و واریانس نامعلوم σ^2 است. تیشیرانی (۱۹۹۶) عنوان کرد که برآوردگرهای لاسو می‌توانند همانند برآوردگرهای بیز از توزیع پسین، وقتی پارامترهای رگرسیونی مستقل و هم‌توزیع با توزیع پیشین لاپلاس باشند استفاده کرد. دلیل این ارتباط، این بود که چندین نویسنده دیگر پیشنهاد استفاده از پیشینی مانند لاپلاس را داده بودند (فیگوردو^۷، ۲۰۰۳، بائه و مالیک^۸، ۲۰۰۴ و یوان و لین^۹، ۲۰۰۵). در همه تحقیق‌های انجام شده، فرض کلی این است که توزیع پیشین β (به شرط σ^2)، دارای توزیع لاپلاس به صورت زیر است

$$\pi(\beta | \sigma^2) = \prod_{j=1}^p \frac{\lambda}{\sqrt{\sigma^2}} e^{-\frac{\lambda|\beta_j|}{\sqrt{\sigma^2}}}. \quad (2.2)$$

برای $\frac{1}{\sigma^2}$ نیز یک توزیع پیشین در نظر گرفته می‌شود. دلیل این که در توزیع پسین برای β از شرطی کردن روی σ^2 استفاده می‌کنیم این است که توزیع پسین تک مدی^{۱۰} باشد زیرا در غیر این صورت نمونه‌گیری گیبز^{۱۱} کند انجام می‌شود.

۱.۱.۲ مدل سلسله‌مراتبی

اگر بتوان مدل بیزی را به صورت سلسله‌مراتبی نوشت، آن‌گاه به راحتی با استفاده از نمونه‌گیر گیبز می‌توان از توزیع پسین نمونه تولید کرد. برای این منظور در رگرسیون لاسوی بیزی باید بتوانیم توزیع لاپلاس را به صورت شرطی نمایش دهیم. با استفاده از اندروز و مالوز^{۱۲} (۱۹۷۴) می‌توان جمله‌نمایی در توزیع لاپلاس را به صورت نرمال آمیخته مقیاسی نوشت. به عبارتی از آمیزش توزیع نرمال با توزیع نمایی، توزیع لاپلاس حاصل می‌شود. به طور دقیق‌تر داریم

$$\frac{a}{\sqrt{\pi}} e^{-a|z|} = \int_0^\infty \frac{1}{\sqrt{\pi s}} e^{-\frac{z^2}{s}} \frac{a}{\sqrt{\pi}} e^{-\frac{a^2 s}{\pi}} ds \quad (3.2)$$

^۷Figueiredo

^۸Bae and Mallick

^۹Yuan and Lin

^{۱۰}Unimodal

^{۱۱}Gibbs sampler

^{۱۲}Andrews and Mallows

که در آن یک توزیع نرمال با میانگین صفر و واریانس s ($N(0, s)$) با توزیع نمایی با میانگین $\frac{\gamma}{\alpha}$ ($Exp(\frac{\alpha}{\gamma})$) آمیخته شده است. در حقیقت $S \sim Exp(\frac{\alpha}{\gamma})$ با توجه به صورت آمیخته مقیاسی، نمایش سلسله مراتبی در مدل رگرسیون لاسوی بیزی عبارت است از

$$\begin{aligned} \mathbf{y} \mid \mathbf{X}, \beta, \sigma^2 &\sim N_n(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n), \\ \beta \mid \sigma^2, \tau_1^2, \tau_2^2, \dots, \tau_p^2 &\sim N_p(0_p, \sigma^2 \mathbf{D}_\tau), \\ \mathbf{D}_\tau &= \text{diag}(\tau_1^2, \dots, \tau_p^2), \\ \sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim \pi(\sigma^2) d\sigma^2 \prod_{j=1}^p \frac{\lambda_j^2}{\gamma} e^{-\frac{\lambda_j^2 \tau_j^2}{\gamma}} d\tau_j^2, \\ \sigma^2, \tau_1^2, \dots, \tau_p^2 &> 0. \end{aligned} \quad (4.2)$$

با ادغام کردن $\tau_1^2, \dots, \tau_p^2$ ، پیشین شرطی روی β به صورت رابطه (۲.۲) است. پارک و کسلا (۲۰۰۸) از چگالی پیشین مبهم $\frac{1}{\sigma^2}$ استفاده کردند، اما در هر گام پیشین گامای معکوس برای σ^2 می‌تواند مزدوج‌سازی^{۱۳} را حفظ کند. مدل سلسله مراتبی مشابه رابطه (۳.۲) توسط سایر نویسندگان استفاده شده است. فیگرد^{۱۴} (۲۰۰۳) از یک مدل سلسله‌مراتبی در رابطه با یک الگوریتم EM برای محاسبه پسین حاشیه‌ای استفاده کرد. بایه^{۱۵} و مالیک (۲۰۰۴) نوع نمایش سلسله مراتبی و یک نمونه‌گیری گیبز متناظر برای مدل پروبیت^{۱۶} در رگرسیون دودویی^{۱۷} پیشنهاد کردند که به یک پارامتر واریانس جداگانه نیاز ندارد. در ادامه توزیع‌های پسین شرطی کامل حاشیه‌ای را برای پارامترهای مدل رگرسیونی به‌دست می‌آوریم. قضیه زیر در پارک و کسلا (۲۰۰۸) بدون جزئیات و پشت سر هم آورده شده است که در این جا روند اثبات با شرح همه مراحل آورده شده است.

قضیه ۱.۱.۲. مدل رگرسیونی خطی (۱.۲) را در نظر بگیرید که توزیع‌های بردار پاسخ و پیشین برای پارامترهای رگرسیونی به صورت مدل سلسله مراتبی (۴.۲) است. در این صورت توزیع‌های پسین شرطی کامل (که در نمونه‌گیر گیبز استفاده می‌شوند) عبارتند از

$$\begin{aligned} \beta \mid (\mathbf{X}, \mathbf{y}, \sigma^2, \tau_1^2, \dots, \tau_p^2) &\sim N_p(\mathbf{A}^{-1} \mathbf{X}^\top \mathbf{y}, \sigma^2 \mathbf{A}^{-1}) \\ \sigma^2 \mid (\mathbf{X}, \mathbf{y}, \beta, \tau_1^2, \dots, \tau_p^2) &\sim \text{IGamma}\left(\frac{n+p-1}{\gamma}, \mathbf{B}\right) \\ \frac{1}{\tau_j^2} \mid (\mathbf{X}, \mathbf{y}, \beta, \sigma^2, \dots, \tau_p^2) &\sim \text{IG}(\mu', \lambda') \end{aligned}$$

^{۱۳}Conjugacy

^{۱۴}Figueiredo

^{۱۵}Bae

^{۱۶}Probit

^{۱۷}Binary regression

که در آنها

$$\begin{aligned} \mathbf{A} &= \mathbf{X}^T \mathbf{X} + \mathbf{D}_\tau^{-1} \\ \mathbf{B} &= \frac{1}{\nu} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \frac{1}{\nu} \boldsymbol{\beta}^T \mathbf{D}_\tau^{-1} \boldsymbol{\beta} \\ \mu' &= \sqrt{\frac{\lambda^2 \sigma^2}{\beta_j^2}} \\ \lambda' &= \lambda^2 \end{aligned}$$

برهان. طبق تعریف توزیع پسین داریم

$$\begin{aligned} \pi(\boldsymbol{\beta} | \mathbf{X}, \mathbf{y}, \sigma^2, \tau_1^2, \dots, \tau_p^2) &\propto L(\boldsymbol{\beta}) \times \pi(\boldsymbol{\beta}) \\ &\propto \frac{1}{(\nu\pi)^{\frac{n}{2}}} |\sigma^2 \mathbf{I}_n|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{\nu} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\sigma^2 \mathbf{I}_n)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \times \\ &\quad \frac{1}{(\nu\pi)^{\frac{p}{2}}} |\sigma^2 \mathbf{D}_\tau|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{\nu} \boldsymbol{\beta}^T (\sigma^2 \mathbf{D}_\tau)^{-1} \boldsymbol{\beta} \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{\nu\sigma^2} \boldsymbol{\beta}^T \mathbf{D}_\tau^{-1} \boldsymbol{\beta} \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu\sigma^2} (\mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{D}_\tau^{-1} \boldsymbol{\beta}) \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu\sigma^2} \boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{X} + \mathbf{D}_\tau^{-1}) \boldsymbol{\beta} \right. \\ &\quad \left. - 2\mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X} + \mathbf{D}_\tau^{-1})^{-1} (\mathbf{X}^T \mathbf{X} + \mathbf{D}_\tau^{-1}) \boldsymbol{\beta} \right. \\ &\quad \left. + \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X} + \mathbf{D}_\tau^{-1})^{-1} \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X} + \mathbf{D}_\tau^{-1})^{-1} \mathbf{X}^T \mathbf{y} \right\} \end{aligned}$$

با تعریف $\mathbf{A} = (\mathbf{X}^T \mathbf{X} + \mathbf{D}_\tau^{-1})$ ، نتیجه می‌شود

$$\begin{aligned} \pi(\boldsymbol{\beta} | \mathbf{X}, \mathbf{y}, \sigma^2, \tau_1^2, \dots, \tau_p^2) &\propto \exp \left\{ -\frac{1}{\nu\sigma^2} ((\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{X}^T \mathbf{y})^T \mathbf{A} (\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{X}^T \mathbf{y}) - \mathbf{y}^T \mathbf{A}^{-1} \mathbf{X}^T \mathbf{y}) \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu\sigma^2} ((\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{X}^T \mathbf{y})^T \mathbf{A} (\boldsymbol{\beta} - \mathbf{A}^{-1} \mathbf{X}^T \mathbf{y})) \right\} \end{aligned}$$

پس در نتیجه $\boldsymbol{\beta}$ دارای توزیع نرمال p -متغیره با میانگین $\mathbf{A}^{-1} \mathbf{X}^T \mathbf{y}$ و ماتریس واریانس-کوواریانس $\sigma^2 \mathbf{A}^{-1}$ است. توزیع پسین شرطی کامل برای σ^2 به صورت زیر حاصل می‌شود

$$\pi(\sigma^2 | \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \tau_1^2, \dots, \tau_p^2) \propto L(\sigma^2) \times \pi(\boldsymbol{\beta} | \sigma^2, \tau_1^2, \dots, \tau_p^2) \times \pi(\sigma^2)$$

از آن جایی که

$$\begin{aligned} L(\sigma^2) &= \frac{1}{(\nu\pi)^{\frac{n}{2}}} |\sigma^2 \mathbf{I}_n|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{\nu} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\sigma^2 \mathbf{I}_n)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ &\propto (\sigma^2)^{-\frac{n-1}{2}} \times \exp \left\{ -\frac{1}{\nu\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \end{aligned}$$

$$\begin{aligned}\pi(\boldsymbol{\beta} | \sigma^2, \tau_1^2, \dots, \tau_p^2) &= \frac{1}{(\nu\pi)^{\frac{p}{2}}} |\sigma^2 \mathbf{D}_\tau|^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2} \boldsymbol{\beta}^\top (\sigma^2 \mathbf{D})^{-1} \boldsymbol{\beta} \right\} \\ &\propto (\sigma^2)^{-\frac{p}{2}} \times \exp \left\{ -\frac{1}{2\sigma^2} \boldsymbol{\beta}^\top (\mathbf{D})^{-1} \boldsymbol{\beta} \right\}\end{aligned}$$

می‌توان نوشت

$$\begin{aligned}\pi(\sigma^2 | \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \tau_1^2, \dots, \tau_p^2) &\propto (\sigma^2)^{-\frac{n-1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ &\quad \times (\sigma^2)^{-\frac{p}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \boldsymbol{\beta}^\top (\mathbf{D}_\tau)^{-1} \boldsymbol{\beta} \right\} \times \frac{1}{\sigma^2} \\ &\propto (\sigma^2)^{-(\frac{n+p-1}{2})-1} \exp \left\{ -\frac{\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \frac{1}{2} \boldsymbol{\beta}^\top (\mathbf{D}_\tau)^{-1} \boldsymbol{\beta}}{\sigma^2} \right\}\end{aligned}$$

که توزیع گامای معکوس را نتیجه می‌دهد.
در نهایت برای توزیع پسین مولفه j ام ابرپارامتر τ داریم

$$\begin{aligned}\pi(\tau_j^2 | \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) &\propto L(\tau_j^2) \times \pi(\boldsymbol{\beta} | \sigma^2, \tau_1^2, \dots, \tau_p^2) \times \pi(\tau_j^2) \\ &\propto \frac{1}{(\nu\pi)^{\frac{n}{2}}} |\sigma^2 \mathbf{I}_n|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\sigma^2 \mathbf{I}_n)^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ &\quad \times \frac{1}{(\nu\pi)^{\frac{p}{2}}} |\sigma^2 \mathbf{D}_\tau|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\beta}^\top (\sigma^2 \mathbf{D}_\tau)^{-1} \boldsymbol{\beta} \right\} \times \frac{\lambda^{\nu}}{\nu} \exp \left\{ -\frac{\lambda^{\nu}}{\nu} \tau_j^2 \right\} \\ &\propto \left(\prod_{j=1}^p (\tau_j^2)^{-\frac{1}{2}} \right) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\tau_j^2} \right\} \times \exp \left\{ -\frac{\lambda^{\nu}}{\nu} \tau_j^2 \right\} \\ &\propto (\tau_j^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \frac{\beta_j^2}{\tau_j^2} - \frac{\lambda^{\nu}}{\nu} \tau_j^2 \right\} \\ &\propto (\tau_j^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left(\frac{\beta_j^2}{\sigma^2 \tau_j^2} + \lambda^{\nu} \tau_j^2 \right) \right\}\end{aligned}$$

در این صورت

$$\pi(\tau_j^2 | \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \propto (\tau_j^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left(\frac{\beta_j^2}{\sigma^2 \tau_j^2} + \lambda^{\nu} \tau_j^2 \right) \right\}.$$

حال فرض کنید $\eta_j = \frac{1}{\tau_j^2}$. در این صورت ژاکوبی تبدیل به صورت زیر است

$$J(\tau_j \rightarrow \eta_j) = \left| \frac{d\tau_j^2}{d\eta_j^2} \right| = \left| -\frac{1}{(\eta_j^2)^2} \right|$$

پس در نتیجه داریم

$$\begin{aligned}\pi(\tau_j^\gamma | \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) &\propto (\tau_j)^{-\frac{1}{\gamma}} \exp \left\{ -\frac{1}{\gamma} \left(\frac{\beta_j^\gamma}{\sigma^2} \eta_j^\gamma + \frac{\lambda^\gamma}{\eta_j^\gamma} \right) \right\} \times J(\tau_j \rightarrow \eta_j) \\ &\propto (\eta_j^\gamma)^{-\gamma} (\eta_j^\gamma)^{-\frac{1}{\gamma}} \exp \left\{ -\frac{1}{\gamma} \frac{\beta_j^\gamma \left(\eta_j^\gamma - \sqrt{\frac{\lambda^\gamma \sigma^2}{\beta_j^\gamma}} \right)^2}{2 \sigma^2 \eta_j^\gamma} \right\} \\ &\propto (\eta_j^\gamma)^{-\frac{\gamma}{\gamma}} \exp \left\{ -\frac{1}{\gamma} \frac{\beta_j^\gamma \left(\eta_j^\gamma - \sqrt{\frac{\lambda^\gamma \sigma^2}{\beta_j^\gamma}} \right)^2}{2 \sigma^2 \eta_j^\gamma} \right\}\end{aligned}$$

که نتیجه می‌دهد $\frac{1}{\tau_j}$ دارای توزیع نرمال معکوس است. حال اگر فرض کنیم $\eta_j^\gamma = x$ ، نتیجه می‌دهد

$$f(x) = \sqrt{\frac{\lambda'}{\gamma \pi}} x^{-\frac{\gamma}{\gamma}} \exp \left\{ -\frac{\lambda' (x - \mu')^2}{2 (\mu')^2 x} \right\} \quad x > 0$$

که در آن

$$\mu' = \sqrt{\frac{\lambda^2 \sigma^2}{\beta_j^2}} \quad \lambda' = \lambda^2$$

□

و اثبات کامل است.

۲.۲ انتخاب پارامتر جریمه

پارامتر تنظیم‌کننده یا همان پارامتر جریمه در لاسوی معمولی می‌تواند به صورت اعتبارسنجی متقابل، اعتبارسنجی متقابل تعمیم‌یافته، یا بر اساس ایده پایه‌ای استاین^{۱۸} به وسیله برآورد ریسک ناریب^{۱۹} انتخاب می‌شوند (تیبشیرانی (۱۹۹۶). کسلا (۲۰۰۱) یک الگوریتم مونت کارلو^{۲۰} EM که مکمل یک نمونه‌گیری گیبز و با برآورد درست‌نمایی ماکزیمم حاشیه‌ای از ابرپیشین‌ها است پیشنهاد کرد. پارک و کسلا (۲۰۰۸) پیشنهاد کردند برای لاسوی بیزی، هر تکرار از الگوریتم EM شامل اجرای نمونه‌گیری با استفاده از مقدار λ برآورد شده از نمونه‌ای از تکرار قبلی می‌باشد. به طور مشخص، k امین تکرار پارامتر جریمه مورد استفاده در نمونه‌گیری گیبز با استفاده از $\lambda^{(k-1)}$ (برای نمونه، برآورد از $k-1$ تکرار) به صورت زیر برورسانی می‌شود

$$\lambda^{(k)} = \sqrt{\frac{2p}{\sum_{j=1}^p E_{\lambda^{(k-1)}}[\tau_j^\gamma | \mathbf{y}]}}$$

^{۱۸}Stein

^{۱۹}Unbiased risk estimate

^{۲۰}Monte Carlo algorithm

که در آن $E[\tau_j^2 | \mathbf{y}]$ امیدریاضی شرطی τ_j^2 با پارامتر $\lambda^{(k-1)}$ است. لازم به ذکر است که مقدار اولیه به صورت زیر در نظر گرفته می‌شود

$$\lambda^{(0)} = \frac{p\sqrt{\hat{\sigma}_{LS}^2}}{\sum_{j=1}^p |\hat{\beta}_j^{LS}|},$$

که در اینجا $\hat{\sigma}_{LS}^2$ و $\hat{\beta}_j^{LS}$ برآوردگرهای LS پارامترهای σ^2 و β_j هستند.

۱.۲.۲ الگوریتم EM برای انتخاب λ

روش مونت کارلویی EM برای برآورد بیز تجربی از ابر پارامترها توسط کسلا (۲۰۰۱) پیشنهاد شد که اساسا برای بررسی پارامترهایی که به عنوان « داده‌های گمشده ۲۱ » می‌باشند در نظر گرفته می‌شود و سپس از الگوریتم EM به صورت تکراری برای تقریب ابر پارامترها استفاده می‌شود. برای لاسوی بیزی، نمونه‌گیری گیبز برای برآورد مورد انتظار استفاده می‌شود. در روابط سلسله مراتبی (۴.۲) با پیشین مزدوج گامای معکوس

$$\pi(\sigma^2) = \frac{\gamma^a}{\Gamma(a)} (\sigma^2)^{-a-1} e^{-\frac{\gamma}{\sigma^2}}, \quad \sigma^2 > 0 \quad (a > 0, \gamma > 0)$$

لگاریتم درست‌نمایی « داده کامل » عبارت است از

$$\begin{aligned} & -\left(\frac{n+p-1}{2} + a + 1\right) \ln(\sigma^2) - \frac{1}{\sigma^2} \left(\frac{\|\mathbf{y} - \mathbf{X}\beta\|_2^2}{2} + \gamma \right) \\ & - \frac{1}{2} \sum_{j=1}^p \frac{\beta_j^2}{\sigma^2 \tau_j^2} + p \ln(\lambda^2) - \frac{\lambda^2}{2} \sum_{j=1}^p \tau_j^2 \end{aligned}$$

پس از حذف عبارات ثابت، که شامل λ نیست پیشین ناوردای مقیاسی $\frac{1}{\sigma^2}$ ، به طور رسمی با در نظر گرفتن $a = 0$ و $\gamma = 0$ به دست می‌آید.

براس ساختن k امین برآورد λ ، از $k-1$ برآورد قبلی آن ($\lambda^{(k-1)}$) استفاده می‌شود (یا مقدار اولیه، اگر $k=1$ باشد). در گام E شامل گرفتن مقدار مورد انتظار از لگاریتم درست‌نمایی بدست می‌آید، به طوری که شرایط روی \mathbf{y} و تحت $\lambda^{(k-1)}$ به صورت

$$Q(\lambda | \lambda^{(k-1)}) = p \ln(\lambda^2) - \frac{\lambda^2}{2} \sum_{j=1}^p E_{\lambda^{(k-1)}}[\tau_j^2 | \mathbf{y}] + (\text{شرایطی که شامل } \lambda \text{ نیست})$$

است (در نماد معمول همراه با EM). گام M برای بیان حداکثرسازی λ برای تولید برآورد بعدی، یعنی $\lambda^{(k)}$ می‌باشد. این موردی که در اینجا وجود دارد با یک تحلیل ساده حل می‌شود

$$\lambda^{(k)} = \sqrt{\frac{2p}{\sum_{j=1}^p E_{\lambda^{(k-1)}}[\tau_j^2 | \mathbf{y}]}}$$

^{۲۱}Missing data

البته، امیدریاضی شرطی فقط امیدریاضی پسین تحت ابر پارامتر $\lambda^{(k-1)}$ می‌باشد، و در نتیجه می‌توان آنها را با استفاده از میانگین نمونه از یک نمونه‌گیری گیبز برآورد کرد

۲.۲.۲ ابرپیشین‌ها برای پارامتر جریمه لاسو

جایگزین مناسب به جای روش ارائه شده در فوق استفاده از یک توزیع پیشین برای λ است. پارک و کسلا (۲۰۰۸) پیشنهاد کردند که یک توزیع پیشین گاما برای λ^2 به صورت زیر در نظر گرفته شود

$$\pi(\lambda^2) = \frac{\delta^r}{\Gamma(r)} (\lambda^2)^{r-1} e^{-\delta\lambda^2}, \quad \lambda^2 > 0, (r > 0, \delta > 0), \quad (5.2)$$

که در آن r و δ ابرپارامترهای توزیع پیشین هستند. می‌توان پیشین ناوردای^{۲۲} مقیاسی $\frac{1}{\lambda^2}$ را برای λ^2 ($r = 0, \delta = 0$) در نظر گرفت، اما منجر به پسین ناسره می‌شود. علاوه بر این، تغییرناپذیری مقیاس یک معیار خیلی قانع‌کننده‌ای نیست زیرا λ بدون واحد^{۲۳} است. وقتی از پیشین رابطه‌ی (۵.۲) در رابطه سلسله‌مراتبی (۴.۲) استفاده می‌شود، توزیع پسین شرطی کامل λ^2 ، یک توزیع گاما با پارامتر شکل $p+r$ و پارامتر مقیاس $\delta + \sum_{j=1}^p \frac{\tau_j^2}{\gamma}$ است، زیرا

$$\begin{aligned} \pi(\lambda^2 | \tau_1^2, \dots, \tau_p^2) &\propto L(\lambda^2) \times \pi(\lambda^2) \\ &\propto \prod_{j=1}^p \left(\frac{\lambda^2}{\gamma} e^{-\frac{\lambda^2}{\gamma} \tau_j^2} \right) \times \frac{\delta^r}{\Gamma(r)} (\lambda^2)^{r-1} e^{-\delta\lambda^2} \\ &\propto \left(\frac{\lambda^2}{\gamma} \right)^p e^{-\frac{\lambda^2}{\gamma} \sum_{j=1}^p \tau_j^2} \times (\lambda^2)^{r-1} e^{-\delta\lambda^2} \\ &\propto (\lambda^2)^{p+r-1} e^{-\lambda^2 (\frac{1}{\gamma} \sum_{j=1}^p \tau_j^2 + \delta)}. \end{aligned}$$

با این مشخصات، λ^2 به سادگی می‌تواند با پارامترهای دیگر در نمونه‌گیری گیبز ارتباط برقرار کند، زیرا توزیع پسین شرطی کامل از پارامترهای دیگر قابل تغییر نمی‌باشد. چگالی پسین برای λ^2 باید به صفر نزدیک شود تا به اندازه کافی سریع باشد به طوری که $\lambda^2 \rightarrow \infty$ می‌رود (برای جلوگیری از آمیختن مسائل)، اما باید به طور نسبی و با احتمال بالایی نزدیک به برآورد ماکزیمم درست‌نمایی باشد.

۳.۲ تحلیل مجموعه داده دیابت

این داده‌ها از برگرفته افرون و همکاران (۲۰۰۴) است که در آن تعداد $n = 442$ بیمار دیابتی مورد آزمایش قرار گرفته‌اند. در این مطالعه متغیر پاسخ اندازه کمی میزان پیشرفت بیماری

^{۲۲} Invariant

^{۲۳} Unitless

تحلیل مجموعه داده دیابت ۲۵

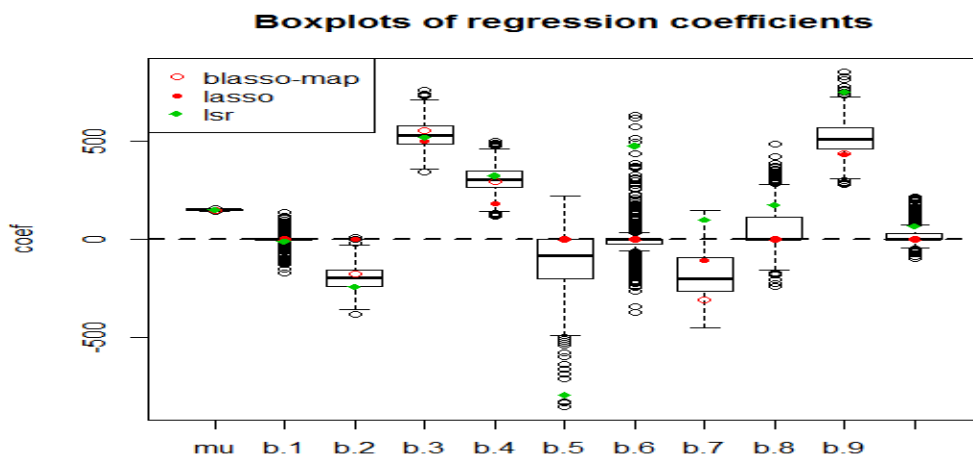
دیابت یک سال پس از زمان پایه^{۲۴} است. منظور از زمان پایه، زمانی است که متغیرهای توضیحی برای هر بیمار اندازه‌گیری شده‌اند در این مطالعه ۱۰ متغیر توضیحی برای هر بیمار اندازه‌گیری شده است. جدول ۱.۲ قسمتی از این داده‌ها را نشان می‌دهد.

جدول ۱.۲: ۱۰ متغیر پایه برای ۴۴۲ بیمار دیابتی

متغیر پاسخ	اندازه سرم‌های خونی						فشار خون	BMI	جنسیت	سن	بیمار
	x_{10}	x_9	x_8	x_7	x_6	x_5					
۱۵۱	۸۷	۴/۹	۴	۳۸	۹۳/۲	۱۵۷	۱۰۱	۳۲/۱	۲	۵۹	۱
۷۵	۶۹	۳/۹	۳	۷۰	۱۰۳/۲	۱۸۳	۸۷	۲۱/۶	۱	۴۸	۲
۱۴۱	۸۵	۴/۷	۴	۴۱	۹۳/۶	۱۵۶	۹۳	۳۰/۵	۲	۷۲	۳
۲۰۶	۸۹	۴/۹	۵	۴۰	۱۳۱/۴	۱۹۸	۸۴	۲۵/۳	۱	۲۴	۴
۱۳۵	۸۰	۴/۳	۴	۵۲	۱۲۵/۴	۱۹۲	۱۰۱	۲۳/۰	۱	۵۰	۵
۹۷	۶۸	۴/۲	۲	۶۱	۶۴/۸	۱۳۹	۸۹	۲۲/۶	۱	۲۳	۶
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
۲۲۰	۸۵	۵/۱	۵	۴۲	۱۲۵/۲	۲۰۱	۹۵	۳۰/۰	۱	۳۶	۴۴۱
۵۷	۹۲	۴/۶	۳	۹۷	۱۳۳/۲	۲۵۰	۷۱	۱۹/۶	۱	۳۶	۴۴۲

در ادامه (قالب یک نمودار) چگونگی عملکرد برآوردگر لاسوی بیزی را با دو روش لاسوی معمولی و ریج مقایسه می‌کنیم. شکل ۱.۲ نمودار جعبه‌ای میانه پسین برآوردهای لاسوی بیزی، لاسو و کمترین توان‌های دوم را نشان می‌دهد. به علاوه فاصله اعتبار ۹۵ درصد برای هر برآورد آورده شده است. در این نمودار پارامتر λ بر اساس روش ماکزیمم درست‌نمایی حاشیه ای بخش ۲.۲ تعیین شده است.

همانطور که در شکل ۱.۲ دیده می‌شود مقادیر میانه پسین بیزی (لاسوی بیزی) شبیه مقادیر برآوردهای لاسو است. به علاوه، همه برآوردهای لاسو در فاصله اعتبار ۹۵ درصد قرار دارد در حالی که برآوردهای کمترین توان دوم برای چهار متغیر خارج از فاصله اعتبار قرار دارد.



شکل ۱.۲: نمودار جعبه‌ای برای لاسوی بیزی

فصل ۳

نعل اسبی بیزی

مقدمه

همانطور که در فصل ۲ مشاهده کردیم، از دیدگاه بیزی، استفاده از توزیع پیشین لاپلاس منجر به توزیع پسینی شد که ماکزیمم کردن آن برآوردگر لاسو را نتیجه داد. رهیافتی دیگر در دیدگاه بیزی، استفاده از توزیع پیشین آمیخته مقیاسی نرمال است که در راستای انتخاب متغیر در دیدگاه بیزی می‌توان از آن استفاده کرد. در این حالت برآوردگر حاصل نعل اسبی^۱ نام دارد. در این فصل این رهیافت رامعرفی کرده و خواص آن را در مدل‌های تنک، در راستای انتخاب متغیر بررسی می‌کنیم. مطالب این فصل برگرفته از کاروالهو^۲ و همکاران (۲۰۱۰) است که اثبات قضایا و نتایج به صورت مفصل‌تری توسط نویسنده این مجموعه آورده شده است.

۱.۳ برآوردگر نعل اسبی

فرض کنید بردار مشاهدات p - بعدی با یک توزیع نرمال با میانگین $\theta = (\theta_1, \dots, \theta_p)^T$ و ماتریس واریانس-کواریانس $\sigma^2 \mathbf{I}_p$ توزیع شده که در آن σ^2 مجهول است. به عبارت دقیق‌تر $y | \theta, \sigma^2 \sim N_p(\theta, \sigma^2 \mathbf{I}_p)$ ، که در آن $y = (y_1, \dots, y_p)^T$. توزیع پیشین را به صورت زیر در نظر

^۱Horseshoe

^۲Carvalho

بگیرید

$$\pi(\theta, \sigma^2) = \pi(\theta | \sigma^2) \pi(\sigma^2)$$

که در آن σ^2 دارای توزیع پیشین جفریز به صورت

$$\pi(\sigma^2) \propto \frac{1}{\sigma^3}$$

است و $\pi(\theta | \sigma^2)$ از یک توزیع پیشین نرمال آمیخته مقیاس تبعیت می کند که صورت سلسله‌مهراتی آن عبارتست از

$$\begin{aligned} \theta_i | \lambda_i &\sim N(\circ, \lambda_i^2) \\ \lambda_i | \tau &\sim C^+(\circ, \tau) \\ \tau | \sigma &\sim C^+(\circ, \sigma) \end{aligned} \quad (1.3)$$

که در آن $C^+(\circ, a)$ توزیع نیم کوشی استاندارد روی \mathbb{R}^+ با پارامتر مقیاسی a است. از آن جایی که پیشین نیم کوشی شبیه یک نعل اسب است به پیشین نرمال آمیخته مقیاسی $\theta_i | \lambda_i \sim N(\circ, \lambda_i^2)$ با ساختار نعل اسبی هم می گویند.

با استفاده از پیشین $\theta_i | \lambda_i \sim N(\circ, \lambda_i^2)$ به ازای $\sigma^2 = \tau^2 = 1$ ، میانگین توزیع پسین عبارتست از

$$\begin{aligned} E(\theta_i | \mathbf{y}) &= \int_0^1 (1 - k_i) y_i p(k_i, \mathbf{y}) dk_i \\ &= y_i \left(\int_0^1 p(k_i, y_i) dk_i - \int_0^1 k_i p(k_i, y_i) dk_i \right) \\ &= y_i (1 - E(k_i | \mathbf{y})), \end{aligned}$$

که در اینجا $k_i = \frac{1}{(1 - \lambda_i^2)}$ بدیهی است که $E(k_i, \mathbf{y})$ می تواند به صورت یک عامل انقباضی عمل کند و باعث می شود بعضی مقادیر میانگین پسین صفر شوند.

همانطور که گفتیم پیشین نیم کوشی روی λ_i به شکل نعل اسب است. سمت چپ رابطه نعل اسبی، $k_i \approx \circ$ است که تقریباً هیچ انقباضی ندارد و سیگنال‌ها را نشان می دهد. سمت راست رابطه نعل اسبی، $k_i \approx 1$ ، تقریباً تمام انقباض را تولید می کند و نشان دهنده خطا یا نوفه است.

بر خلاف سایر روش‌های مشابه، پیشین نعل اسبی به ابر پارامترها، که معمولاً توسط کاربر انتخاب می شوند، بستگی ندارد. با این وجود، این پیشین بسیار سازگار و استوار است و عملکرد قوی در شرایط مختلف دارد. هدف در اینجا، گذشته از تعریف اینکه پیشین نعل اسبی ابزاری برای مدل سازی است، پیشنهاد یک چهارچوب نظری است که در آن مدل را می توان با سایر پیشین‌های انقباضی مشابه مقایسه کرد.

دو مزیت عمده این پیشین عبارتند از

- استواری نسبت به سیگنال‌های بزرگ
- انقباض نوفه

۲.۳ تابع چگالی نعل اسبی

تابع چگالی نعل اسبی تک متغیره فاقد صورت تحلیلی است اما برای آن‌ها کران‌هایی به دست آمده‌اند. برای سادگی در نماد گذاری، فرض می‌کنیم $\sigma^2 = \tau^2 = 1$. قضیه زیر برخی خواص تابع چگالی نعل اسبی را بیان می‌دارد.

قضیه ۱.۲.۳. فرض کنید θ دارای توزیع نعل اسبی بوده که صورت سلسله مراتبی آن در رابطه (۱.۳) آمده است و آن را با $p(\theta)$ نشان دهیم. در این صورت $p(\theta)$ (تک متغیره) دارای ویژگی‌های زیر است

۱.

$$\lim_{\theta \rightarrow 0} p(\theta) = \infty$$

۲. برای $\theta \neq 0$

$$\frac{K}{\sqrt{2}} \log \left(1 + \frac{\theta^2}{\lambda^2} \right) < p(\theta) < K \log \left(1 + \frac{\theta^2}{\lambda^2} \right) \quad (2.3)$$

که در آن $K = \frac{1}{(2\pi^3)^{\frac{1}{4}}}$

برهان. طبق تعریف صورت سلسله مراتبی (۱.۳) به ازای $\sigma^2 = \tau^2 = 1$ داریم

$$p(\theta) = \int_0^\infty \frac{1}{(2\pi\lambda^2)^{\frac{1}{4}}} \exp\left(-\frac{\theta^2}{2\lambda^2}\right) \frac{2}{\pi(1+\lambda^2)} d\lambda.$$

با اعمال تغییر متغیر $u = \frac{1}{\lambda^2}$ با ژاکوبی تبدیل $J(\lambda \rightarrow u) = -\frac{1}{2\sqrt{u^3}}$ می‌توان نوشت

$$\begin{aligned} p(\theta) &= \int_0^\infty \frac{1}{(2\pi)^{\frac{1}{4}}} u^{\frac{1}{4}} \exp\left(-\frac{\theta^2 u}{2}\right) \frac{2}{1+u} \frac{1}{2} u^{-\frac{3}{4}} du \\ &= K \int_0^\infty \frac{1}{1+u} \exp\left(-\frac{\theta^2 u}{2}\right) du. \end{aligned}$$

مجدداً با اعمال تغییر متغیر $z = 1 + u$ با ژاکوبی تبدیل $J(u \rightarrow z) = 1$ داریم

$$\begin{aligned} p(\theta) &= K e^{\frac{\theta^2}{2}} \int_1^\infty \frac{1}{z} \exp\left(-\frac{z\theta^2}{2}\right) dz \\ &= K \exp\left(\frac{\theta^2}{2}\right) E_1\left(\frac{\theta^2}{2}\right), \end{aligned}$$

که در اینجا $E_1(\cdot)$ تابع انتگرال نمایی است. به طور واضح این تابع در کران‌های بالا و پایین زیر محدود است:

$$\frac{\exp(-t)}{t} \log\left(1 + \frac{t}{t}\right) < E_1(t) < \exp(-t) \log\left(1 + \frac{1}{t}\right) \quad (3.3)$$

که برای همه $t > 0$ قسمت (۲) ثابت می‌شود. برای اطلاعات بیشتر در مورد رابطه (۳.۳) به آبرامویتز و استگان^۳ (۱۹۶۴) مراجعه کنید. بدیهی است به ازای $\theta \rightarrow 0$ نتیجه می‌شود

$$\begin{aligned} \lim_{\theta \rightarrow 0} p(\theta) &= K \int_1^\infty \frac{1}{z} dz \\ &= K \ln(z) \Big|_1^\infty \\ &= \infty \end{aligned}$$

□

با انتگرال‌گیری روی τ ، یک چگالی حاشیه‌ای برای λ_i به صورت

$$p(\lambda_i) = \frac{1}{\pi} \log \frac{\lambda_i}{\{\pi^2(\lambda_i^2 - 1)\}}$$

حاصل می‌شود.

۳.۳ ارتباط با توزیع‌های پیشین مشابه

همانطور که دیدیم پیشین نعل اسبی یک چگالی آمیخته با پارامتر مقیاس λ_i است و لذا در چندمتغیره آمیخته مقیاسی نرمال قرار می‌گیرد. هدف در این قسمت بررسی توزیع‌های پیشینی است که دارای ساختار انقباضی بوده و در آن‌ها قاعده‌های انقباضی به وسیله پارامتر مقیاسی τ مشخص می‌شوند. در این بخش، خلاصه‌ای از قاعده‌های انقباضی موضعی را معرفی می‌کنیم. پیشین آمیخته گسسته $\theta_i \sim \omega g(\theta_i) + (1 - \omega)\delta_0$ را در نظر بگیرید. آن را می‌توان به صورت یک چگالی آمیخته مقیاسی در نظر گرفت که در آن $\lambda_i \sim \omega h(\lambda_i) + (1 - \omega)\delta_0$. تابع h با توجه به چگالی g در نظر گرفته می‌شود. برای اطلاعات بیشتر به اسکات و برگر^۴ (۲۰۰۶) مراجعه کنید. پیشین t - استیودنت به صورت $\theta_i \sim t_{\xi}(\tau, \tau)$ را در نظر بگیرید. در این حالت λ_i دارای توزیع گامای معکوس به صورت $\lambda_i \sim \text{IGamma}(\frac{\xi}{\tau}, \frac{\xi}{\tau})$ است. پیشین نمایی دوگانه را در نظر بگیرید که در آن $p(\lambda_i^2 | \tau^2) = (\tau^2)^{-1} \exp\{-\frac{\lambda_i^2}{\tau^2}\}$ و $\tau^2 \sim \text{IGamma}(\frac{\xi}{\tau}, \frac{\xi}{\tau})$ است. پارک و کسلا (۲۰۰۸) و هنس (۲۰۰۹) از پیشین نمایی دوگانه در دیدگاه بیزی به عنوان جایگزینی برای لاسو استفاده کردند. پیشین نرمال جفریز که یک پیشین ناسره است و به صورت $p(\lambda_i^2) \propto \frac{1}{\lambda_i^2}$

^۳ Aramowitz and Stegun

^۴ Scott and Berger

جدول ۱.۳: پیشین‌ها برای k_i و λ_i وابسته به تعدادی از قاعده‌های متداول انقباضی موضعی برای پیشین نرمال-نمایی-گاما، $d = 1$ فرض شده است.

پیشین برای θ_i	چگالی برای λ_i	چگالی برای k_i
نمایی دوگانه	$\lambda_i \exp(\lambda_i^2/2)$	$k_i^2 \exp\{-1/(2k_i)\}$
کوشی	$\lambda_i^{-2} \exp\{1/(2\lambda_i^2)\}$	$k_i^{-1/2} (1 - k_i)^{-3/2} \exp(-k_i/\{2/(1 - k_i)\})$
استرودرمن-برگر	$\lambda_i(1 + \lambda_i^2)^{-3/2}$	$k_i^{-1/2}$
نرمال-نمایی-گاما	$\lambda_i(1 + \lambda_i^2)^{-(c+1)}$	k_i^{c-1}
نرمال-جفریز	λ_i^{-1}	$k_i^{-1}(1 - k_i)^{-1}$
نعل اسبی	$(1 + \lambda_i^2)^{-1}$	$K_i^{-1/2}(1 - k_i)^{-1/2}$

می‌باشد توسط فیگوردو (۲۰۰۳) و بایه و مالیک (۲۰۰۴) مورد مطالعه قرار گرفته است. پیشین استرودرمن-برگر (استرودرمن^۵، ۱۹۷۱ و برگر^۶، ۱۹۸۰) فاقد صورت تحلیلی است، اما در آن $\theta_i | k_i \sim N(0, k_i^{-1} - 1)$ و $k_i \sim Beta(\frac{1}{2}, 1)$ می‌باشد. جانستون و سیلورمن^۷ (۲۰۰۴) چگالی این توزیع را چگالی شبه کوشی^۸ نامیدند، و آن را به عنوان یک انتخاب امکان‌پذیر g در مدل آمیخته گسسته مطالعه کردند. پیشین‌هایی از خانواده‌های نرمال-نمایی-گاما، توسط گرفین و براون (۲۰۰۵) پیشنهاد شدند که بعداً توسط شاپیل و نیب^۹ (۲۰۰۹) مورد استفاده قرار گرفت. پیشین نرمال-نمایی-گاما بر اساس چگالی آمیخته‌نمایی است، اما مشخصات لاسو با استفاده از چگالی $Ga(c, d^2)$ برای آمیختن روی نرخ پارامتر نمایی تعمیم داده می‌شود. در این پیشین

$$p(\lambda_i^2) = \frac{c}{d^2} \left(1 + \frac{\lambda_i^2}{d^2}\right)^{-(c-1)}$$

است که در آن دو ابرپارامتر c و d می‌توانند به ترتیب وزن و مقیاس دم را کنترل کنند. جدول ۱.۳ پیشین‌های ذکر شده در بالا را به صورت توزیعی از λ_i و k_i نشان می‌دهد. همچنین شکل ۱.۳، نمودار توزیع k_i ‌ها برای برخی از پیشین‌ها را نشان می‌دهد. در جدول؟؟ ضرایب نرمال ساز حذف شده است.

بر اساس شکل ۱.۳، می‌توان به یک سری نتایج شهودی در مقایسه پیشین‌های انقباضی رسید. رفتار $p(k_i)$ در نزدیکی $k_i = 0$ ، پایدار بودن مقاومت دم را کنترل می‌کند و همچنین رفتار آن در نزدیکی $k_i = 1$ انقباض نوفه را فراهم می‌کند. پیشین نرمال-جفریز و پیشین نعل اسبی حاصل از $p(k_i)$ بدون محدودیت در نزدیکی ۱ منعکس

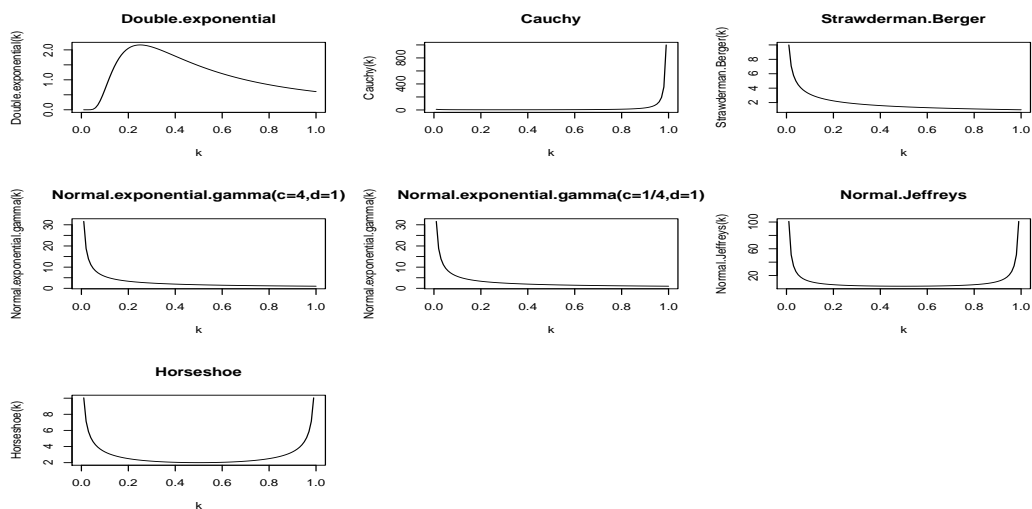
^۵Strawderman

^۶Berger

^۷Johnstone and Silverman

^۸Quasi-Cauchy

^۹Scheipl and Kneib



شکل ۱.۳: چگالی‌های $p(k_i)$: نمایی دوگانه، کوشی، استرودرمن-برگر، نرمال-نمایی-گاما، نرمال-جفریز و نعل اسبی. در قطعه مستطیلی سمت چپ برای پیشین نرمال-نمایی-گاما، خط مشکی برای $c = 4$ و $d = 1$ و خط قطعه قطعه برای $c = \frac{1}{4}$ و $d = 1$ است.

کننده قطب‌های^{۱۰} آن‌ها در نزدیکی $\theta_i = 0$ است. پیشین‌های نمایی دوگانه، استرودرمن-برگر، کوشی و نرمال-نمایی-گاما به سمت یک ثابت در $k_i = 1$ میل می‌کنند. این تفاوت‌ها برای رفتار میانگین پسین بسیار مهم و معنی‌دار است. زمانی که بردار پارامتر صحیح تنک باشد (یعنی شامل تعداد زیادی صفر یا اثرات بی‌معنی باشد) پیشین‌های دم سنگین مانند کوشی، استرودرمن-برگر، نرمال-جفریز، نعل اسبی و نرمال-نمایی-گاما با $c < 1$ حاصل از $p(k_i)$ در نزدیکی صفر می‌باشند. پیشین دم سبک‌تر، که نمایی دوگانه و نرمال-نمایی-گاما با $c \leq 1$ هستند باعث می‌شود که $p(k_i)$ در $k_i = 0$ به صفر نزدیک شود. پیشین نعل اسبی، یک تبدیل یا انتقال فشرده بین دو مقدار حداکثر را نشان می‌دهد. پیشین نرمال-جفریز، به ازای حالت مبهم $k_i \sim \text{Beta}(\epsilon, \epsilon)$ وقتی $\epsilon \rightarrow 0$ نیز حاصل می‌شود. این باعث می‌شود که دم‌ها سنگین‌تر شوند، و در $\theta = 0$ در مقایسه با پیشین نعل اسبی صورت برجسته‌تری دارد.

۴.۳ استواری نسبت به سیگنال‌های بزرگ

در این بخش در قالب نمایی برای میانگین پسین، استواری دم برآوردگر را مورد بررسی قرار می‌دهیم.

قضیه زیر رفتار استوار برآوردگر را در شرایطی که y فاصله زیادی از میانگین پسین دارد نشان می‌دهد. استواری دم برآوردگر یک خاصیت مهم در موقعیت‌های تنک است به این معنی که

^{۱۰} Poles

علاقه‌مندیم مشاهداتی که نزدیک به صفر هستند را با قدرت خیلی بیشتری نسبت به آنهایی که خیلی فاصله بیشتری نسبت به صفر دارند، منقبض کنیم.

قضیه ۱.۴.۳. (کاروالهو و همکاران، ۲۰۱۰). فرض کنید $p(|y - \theta|)$ یک تابع چگالی آمیخته مقیاسی نرمال با میانگین صفر برای θ است، یعنی $\theta | \lambda \sim N(0, \lambda^2)$ که در آن λ دارای پیشین مناسب $p(\lambda)$ است. همچنین فرض کنید که درست‌نمایی و $p(\theta)$ طوری هستند که چگالی حاشیه‌ای $m(y)$ برای همه y ‌ها متناهی است. قرار دهید

$$m^*(y) = \int_{\mathbb{R}} p(|y - \theta|) p^*(\theta) d\theta, \quad p^*(\theta) = \int_{\mathbb{R}^+} p(\theta | \lambda) p^*(\lambda) d\lambda, \quad p^*(\lambda) = \lambda^2 p(\lambda).$$

در این صورت

$$\begin{aligned} E(\theta | y) &= \frac{m^*(y)}{m(y)} \frac{d}{dy} \log m^*(y) \\ &= \frac{1}{m(y)} \frac{d}{dy} m^*(y) \end{aligned} \quad (4.3)$$

برهان.

$$\begin{aligned} \lambda^2 \frac{d}{d\theta} \{N(\theta | 0, \lambda^2)\} &= \lambda^2 \frac{d}{d\theta} \left(\frac{1}{\sqrt{2\pi}\lambda} \exp\left(-\frac{1}{2\lambda^2} \theta^2\right) \right) \\ &= \lambda^2 \frac{1}{\sqrt{2\pi}\lambda} - \frac{1}{\lambda^2} \theta \exp\left(-\frac{1}{2\lambda^2} \theta^2\right) \\ &= -\theta N(\theta | 0, \lambda^2) \end{aligned}$$

از آنجایی که $p(|y - \theta|)$ یک تابع چگالی آمیخته مقیاسی نرمال برای θ است، مشخص است که

$$E(\theta | y) = \frac{1}{m(y)} \int \theta p(y - \theta) N(\theta | 0, \lambda^2) \pi(\lambda) d\theta d\lambda.$$

با توجه به این که $\frac{d}{dy} p(y - \theta) = -\frac{d}{d\theta} p(y - \theta)$ می‌توان نوشت

$$E(\theta | y) = \frac{1}{m(y)} \int \frac{d}{dy} p(y - \theta) N(\theta | 0, \lambda^2) p^*(\lambda) d\theta d\lambda$$

از طرفی چون

$$\begin{aligned} p^*(\theta) &= \int_{\mathbb{R}^+} p(\theta | \lambda) p^*(\lambda) d\lambda \\ &= \int_{\mathbb{R}^+} N(\theta | 0, \lambda^2) p^*(\lambda) d\lambda \end{aligned}$$

پس

$$E(\theta | y) = \frac{1}{m(y)} \frac{d}{dy} \int_{\mathbb{R}} p(|y - \theta|) p^*(\theta) d\theta$$

در نتیجه داریم

$$E(\theta | y) = \frac{1}{m(y)} \frac{d}{dy} m^*(y)$$

□

و حکم ثابت است.

اگر $p(|y - \theta|)$ چگالی نرمال باشد آن گاه رابطه (۴.۳) به صورت زیر ساده می‌شود

$$E(\theta | y) = y + \frac{d}{dy} \log m(y). \quad (۵.۳)$$

روابطی شبیه رابطه (۵.۳) را می‌توان در مارسلیز (۱۹۷۵) و پولسن (۱۹۹۱) و پریکی و اسمیت (۱۹۹۲) یافت. اما این روابط برای پیشین نعل اسبی قابل استفاده نیستند، زیرا شرط کراندارگی زیر برای آن‌ها برقرار نیست.

$$\frac{K}{\tau} \log \left(1 + \frac{\tau}{\theta^2} \right) < p(\theta) < K \log \left(1 + \frac{\tau}{\theta^2} \right).$$

قضیه ۱.۴.۳ این شرایط را ساده‌تر کرده و نتیجه را به حالت‌هایی که $p(\theta)$ یک آمیخته مقیاسی نرمال با یک چگالی آمیخته مناسب و حاشیه‌ای $m(y)$ متناهی است، تعمیم می‌دهد. استواری بیزی با انتخاب پیشینی برای θ به طوری که مشتق لگاریتم چگالی پیشگو تابعی کراندار نسبت به y باشد، حاصل می‌شود. در حالت ایده آل، باید این کران به صفر همگرا شود که با توجه به رابطه (۵.۳) برای $|y|$ بزرگ نتیجه می‌شود $E(\theta | y) \approx y$. این دقیقاً همان چیزی است که تحت پیشین نعل اسبی و سایر توزیع‌های با دم به قدر کافی سنگین اتفاق می‌افتد. قضیه زیر نشان می‌دهد که توزیع پیشین نعل اسبی منجر به برآوردگری می‌شود که دارای دم استوار است.

قضیه ۲.۴.۳. (کاروالهو، پولسن و اسکات، ۲۰۱۰). فرض کنید $y \sim N(0, 1)$ و $m(y)$ نشان دهنده چگالی پیشگو تحت پیشین نعل اسبی برای پارامتر مقیاسی معلوم $\tau < \infty$ (در اینجا θ) $(\lambda \sim N(0, \tau^2 \lambda^2)$ و $\lambda \sim C^+(0, 1)$). فرض کنید $E(\theta | y)$ میانگین پسین است. در این صورت $\lim_{|y| \rightarrow \infty} \frac{d}{dy} \log m(y) = 0$ و $b_\tau \leq \infty$ که به بستگی τ بستگی دارد.

فصل ۴

برآوردگرهای جریمه شده از منظر کلی

مقدمه

رگرسیون جریمه شده بیزی مانند برآوردگرهای لاسوی بیزی و نعل اسبی بیزی تکنیکی است که به تازگی به صورت گسترده‌ای مورد توجه و پذیرش مطبوعات آماری قرار گرفته است. در حالت کلی برای انجام یک رگرسیون جریمه شده بیزی، می‌توان یک روش زنجیر مارکوف مونت کارلویی^۱ را در نرم افزار stan انجام داد که انجام آن ساده نیست. اما مکالیک و اشمیت^۲ (۲۰۱۶) بسته bayesreg را در نرم‌افزار R طراحی کردند که به راحتی می‌توان با استفاده از آن یک رگرسیون جریمه شده بیزی با توزیع پیشین انقباضی (از نوع پیوسته) را اجرا کرد. در این فصل جزئیات اجرا یک مدل بیزی را در این بسته مطرح کرده و در قالب یک مثال عددی روی داده‌های دیابت، رگرسیون جریمه شده بیزی را شرح می‌دهیم.

^۱Markov Chain Monte Carlo

^۲Schmidt

۱.۴ رگرسیون جریمه شده بیزی

رگرسیون جریمه شده بیزی تکنیکی برای تحلیل داده‌ها در بعد بالا می‌باشد. اخیراً مثال‌هایی شامل لاسوی بیزی (پارک و کسلا، ۲۰۰۸ و هنس، ۲۰۰۹)، برآوردگر نرمال-گاما (گریفین^۳ و براون، ۲۰۱۰) و برآوردگر نعل اسبی (کاروالهو، ۲۰۱۰)، برآوردگر تعمیم یافته پارتو دوگانه (آرمگن^۴ ۲۰۱۳) در این راستا بیان شده‌اند که دو مورد معروف لاسوی بیزی و برآوردگر نعل اسبی در دو فصل گذشته بررسی شدند. این برآوردگرها در مدل‌های تنک شامل توزیع پیشین برای پارامترهای رگرسیونی به طور رایج و کاربردی در تحلیل داده‌های بزرگ^۵ استفاده می‌شوند. این برآوردگرها با استفاده از توزیع‌های پیشینی که باعث تنکی در برآورد پارامترهای رگرسیونی می‌شوند (پیشین تنک) ناشی شده و به مراتب در زمینه تحلیل داده‌های بزرگ، هنگامی که اغلب متغیر توضیحی ارتباط معناداری با پاسخ ندارند، به کار می‌روند.

توزیع پیشین تنک از طریق توزیع‌های آمیخته مقیاسی نرمال در مسئله مورد نظر وارد شده و اغلب توزیع پسین بیزی متناظر مستقیماً یک تکنیک رگرسیون جریمه شده را نتیجه می‌دهد، برای مثال می‌توان به لاسو (تیبشیرانی، ۱۹۹۶ و ۲۰۰۱) و الاستیک نت (ژو و هستی، ۲۰۰۵) اشاره کرد. به منظور آگاهی بیشتر در مورد تکنیک‌های رگرسیون جریمه شده بیزی و روش‌های متناظر کلاسیک آن‌ها به پولسن و اسکات (۲۰۱۰ و ۲۰۱۲) مراجعه کنید.

تعداد زیادی از بسته‌های نرم‌افزاری برای برآوردگر استاندارد رگرسیون جریمه وجود دارند، که glmnet (فریدمن^۶، ۲۰۱۰) و ncvreg (بیهنری و هوانگ^۷، ۲۰۱۱) بیشترین محبوبیت را در اجرای عملی دارند. اجرای نرم‌افزاری بسته glmnet در لاسو و الاستیک نت برای تعمیم مدل‌های خطی است در حالی که بسته ncvreg یک الگوریتم کارآمد برای برازش MCP (ژانگ^۸، ۲۰۱۰) یا SCAD (فن و لی^۹، ۲۰۰۱) برای مدل‌های رگرسیون خطی و لجستیک می‌باشد. برای این منظور، سهم عمده رگرسیون جریمه شده بیزی، یک بسته نرم‌افزاری برای اجرای برآوردگرهای رگرسیون جریمه شده بیزی برای مدل‌های رگرسیون خطی و لجستیک است برای توضیح بسته bayesreg ابتدا مدل مورد نظر را معرفی می‌کنیم.

در حالت کلی فرض کنید $y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$ بردار پاسخ رگرسیونی و ماتریس $X = (x_1, \dots, x_n)^T \in \mathbb{R}^{n \times p}$ شامل n متغیر پیشگو p - بعدی برای $j = 1, \dots, p$ ، x_{ij} ، $i = 1, \dots, n$ باشد. در این حالت مدل رگرسیون جریمه شده بیزی که در بسته bayesreg به کار می‌رود دارای

^۳Griffin

^۴Armagan

^۵Big data

^۶Friedman

^۷Breheny and Huang

^۸Zhang

^۹Fan and Li

ساختار سلسله مراتبی زیر است

$$z_i | x_i, \beta, \beta_0, \omega_i, \sigma^2 \sim N(\mathbf{x}_i^T \beta + \beta_0, \sigma^2 \omega_i^2), \quad (1.4)$$

$$\sigma^2 \sim \pi(\sigma^2) d\sigma^2, \quad (2.4)$$

$$\omega_i^2 \sim \pi(\omega_i^2) d\omega_i^2, \quad (3.4)$$

$$\beta_0 \sim d\beta_0, \quad (4.4)$$

$$\beta_j | \lambda_j^2, \tau^2, \sigma^2 \sim N(0, \lambda_j^2 \tau^2 \sigma^2), \quad (5.4)$$

$$\lambda_j^2 \sim \pi(\lambda_j^2) d\lambda_j^2, \quad (6.4)$$

$$\tau^2 \sim \pi(\tau^2) d\tau^2, \quad (7.4)$$

که در آن برای $i = 1, \dots, n$ و $j = 1, \dots, p$ یک پارامتر جدا کننده یا عرض از مبدا، $\beta \in \mathbb{R}^p$ بردار ضرایب رگرسیونی و متغیرهای (z_1, \dots, z_n) یک مجموعه از متغیرهای وابسته به داده y بوده که در آن y ها می‌توانند پیوسته یا دودوئی باشند. در رویارویی با داده‌های بزرگ یا مه داده معمولاً مقدار p بسیار بزرگ است به طوری که از حجم نمونه بزرگتر بوده ($p > n$) و بنابراین ماتریس X پررتبه (رتبه کامل ستونی) نمی‌باشد. در سلسله روابط (۱.۴) تا (۷.۴) دو گروه داریم

- گروه اول مدل‌هایی برای توزیع نمونه‌ای از داده بوده که توسط روابط (۱.۴) تا (۳.۴) مشخص شده است.

- گروه دوم، توزیع‌های پیشین ضرایب رگرسیونی بوده که در روابط (۵.۴) تا (۷.۴) به آن اشاره شده است.

همانطور که از ساختار روابط (۱.۴) و (۵.۴) مشخص است مدل‌های آماری برای داده‌ها و توزیع‌های پیشین جزو توزیع‌های نرمال آمیخته می‌باشند (اندروز و مالوز^{۱۰}، ۱۹۷۴). تعداد زیادی از توزیع‌های آمیخته مقیاسی از نوع توزیع نرمال وجود دارند برای مثال می‌توان به چگالی لجستیک (پولسن ۲۰۱۳) و توزیع لاپلاس (اندروز و مالوز، ۱۹۷۴) و دسته‌ای یا کلاسی از توزیع z (براندورف-نیلسن^{۱۱} ۱۹۸۲) اشاره کرد. روابط (۱.۴) تا (۳.۴) به این نکته اشاره دارند که داده‌ها می‌توانند از مدل آمیخته مقیاسی تبعیت کنند و همچنین می‌توان متغیر پنهان^{۱۲} $(\omega_1, \dots, \omega_n)$ نیز داشت که به طور رایج در مدل رگرسیونی دودوئی و داده‌ای پیوسته استفاده می‌شود. از آن جمله می‌توان به رگرسیون خطی با خطای نرمال، رگرسیون خطی با خطاهای لاپلاس، رگرسیون با خطای t -استیودنت، و رگرسیون لجستیک دودوئی اشاره کرد. ابر پارامترهای $(\tau^2 > 0)$ و $(\lambda_1, \dots, \lambda_n)$ برای ضرایب رگرسیونی در مدل‌های تنک که شامل

^{۱۰} Andrews and Mallows

^{۱۱} Brandorff and Nielsen

^{۱۲} Latent

توزیع‌های پیشین هستند استفاده می‌شوند. توزیع‌های پیشین مورد توجه در اینجا، مثال‌های رایج از پیشین‌های انقباضی فراموضعی - موضعی^{۱۳} بوده که شامل رگرسیون لاسو، رگرسیون ریج و برآوردگر نعل اسبی است.

۲.۴ مدل بندی چندگانه

همانند دو فصل گذشته برای انجام رگرسیون جریمه‌شده با استفاده از روابط (۱.۴) تا (۷.۴) لازم است که توزیع پسین شرطی کامل برای تک تک پارامترها محاسبه شود که در این حالت از روش MCMC و نمونه‌گیر گیبز برای تولید نمونه‌های پسین استفاده می‌کنیم. برای پارامتر عرض از مبدا β_0 که در رابطه (۴.۴) به آن اشاره شد توزیع پیشین یکنواخت در نظر می‌گیریم، در حالی که پارامترهای رگرسیونی که دارای توزیع پیشین توام هستند با توزیع نرمال آمیخته مقیاسی در نظر گرفته شده که روابط (۵.۴) تا (۷.۴) تعیین می‌شوند. در ادامه باید توزیع‌های پسین شرطی را برای پارامتر عرض از مبدا β_0 و پارامترهای رگرسیونی β به دست آورد. فرض کنید

$$e_i = z_i - \mathbf{x}_i^T \beta - \beta_0, \quad (i = 1, \dots, n), \quad (8.4)$$

$$\Omega_n = \sigma^2 \text{diag}(\omega_1^2, \dots, \omega_n^2), \quad (9.4)$$

$$\Lambda_p = \sigma^2 \tau^2 \text{diag}(\lambda_1^2, \dots, \lambda_p^2), \quad (10.4)$$

که در آن (e_1, \dots, e_n) خطاهای مدل رگرسیون خطی، $\Omega_n \in \mathbb{R}^{n \times n}$ و $\Lambda_p \in \mathbb{R}^{p \times p}$ ماتریس‌های قطری می‌باشند. در مدل رگرسیونی خطی ساده برای داده‌های پیوسته داریم $z_i = y_i$ در حالی که در رگرسیون لجستیک که دودوئی است فرض می‌شود

$$z_i = \omega_i^2 \left(y_i - \frac{1}{\tau} \right) \quad i = 1, \dots, n. \quad (11.4)$$

در این موارد نیز، توزیع پسین شرطی برای پارامتر عرض از مبدا β_0 به صورت توزیع نرمال $N(\tilde{\mu}, \tilde{\sigma}^2)$ است که در آن

$$\tilde{\mu} = \left(\sum_{i=1}^n \frac{z_i - \mathbf{x}_i^T \beta}{\omega_i^2} \right) \left(\sum_{i=1}^n \frac{1}{\omega_i^2} \right)^{-1}, \quad \tilde{\sigma}^2 = \sigma^2 \left(\sum_{i=1}^n \frac{1}{\omega_i^2} \right)^{-1}. \quad (12.4)$$

زیرا

$$\pi(\beta_0 | \mathbf{x}_i, z_i, \beta, \omega_i^2, \sigma^2) \propto L(\beta_0) \times \pi(\beta_0)$$

از طرفی

$$L(\beta_0) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2\omega_i^2}} \times \exp \left\{ -\frac{1}{2\sigma^2\omega_i^2} (z_i - (\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0))^2 \right\}$$

$$\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (z_i - (\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0))^2 \right\}$$

در این صورت داریم

$$\begin{aligned} \pi(\beta_0 | \mathbf{x}_i, z_i, \boldsymbol{\beta}, \omega_i^2, \sigma^2) &\propto L(\beta_0) \times \pi(\beta_0) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (z_i - (\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0))^2 \right\} \times 1 \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (-2z_i \mathbf{x}_i^T \boldsymbol{\beta} - 2z_i \beta_0 + (\mathbf{x}_i^T \boldsymbol{\beta})^2 + 2\beta_0 \mathbf{x}_i^T \boldsymbol{\beta} + \beta_0^2) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (\beta_0^2 - 2z_i \beta_0 + 2\beta_0 \mathbf{x}_i^T \boldsymbol{\beta}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (\beta_0^2 - 2\beta_0 (z_i - \mathbf{x}_i^T \boldsymbol{\beta})) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (\beta_0^2 - 2\beta_0 (z_i - \mathbf{x}_i^T \boldsymbol{\beta}) + (z_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 - \right. \\ &\quad \left. (z_i - \mathbf{x}_i^T \boldsymbol{\beta})^2) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (\beta_0^2 - 2\beta_0 (z_i - \mathbf{x}_i^T \boldsymbol{\beta}) + (z_i - \mathbf{x}_i^T \boldsymbol{\beta})^2) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (\beta_0 - (z_i - \mathbf{x}_i^T \boldsymbol{\beta}))^2 \right\} \end{aligned}$$

حال اگر فرض کنیم که $\tilde{\mu} = \sum_{i=1}^n \left(\frac{z_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\omega_i^2} \right) \left(\sum_{i=1}^n \frac{1}{\omega_i^2} \right)^{-1}$ و $\tilde{\sigma}^2 = \sigma^2 \left(\sum_{i=1}^n \frac{1}{\omega_i^2} \right)^{-1}$ در این صورت $\beta_0 \sim N(\tilde{\mu}, \tilde{\sigma}^2)$.

قضیه ۱.۲.۴. (مکالیک و همکاران، ۲۰۱۶). تحت مفروضات این بخش، چگالی پسین شرطی برای پارامتر رگرسیونی β دارای توزیع نرمال $N_p(\tilde{\boldsymbol{\mu}}, \mathbf{A}_p^{-1})$ است که در آن

$$\tilde{\boldsymbol{\mu}} = \mathbf{A}_p^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} (\mathbf{z} - \beta_0 \mathbf{1}_n), \quad \mathbf{A}_p = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X} + \mathbf{A}_p^{-1}), \quad (13.4)$$

و $\mathbf{1}_n = (1, \dots, 1)$ یک بردار n -بعدی از یک‌ها و $\mathbf{A}_p \in \mathbb{R}^{p \times p}$ یک ماتریس دقت بوده که معین مثبت است.

برهان. تحت فرض مسئله داریم

$$\mathbf{z} | \mathbf{X}, \boldsymbol{\beta}, \beta_0 \mathbf{1}, \boldsymbol{\Omega}_n \sim N_n(\mathbf{X}\boldsymbol{\beta} + \beta_0 \mathbf{1}_n, \boldsymbol{\Omega}_n)$$

$$\pi(\beta|\mathbf{X}, \mathbf{z}, \beta_0 \mathbf{1}_n, \Omega_n, \Lambda_p) \propto L(\beta) \times \pi(\beta)$$

از آن جایی که

$$\begin{aligned} L(\beta) &= \frac{1}{(\sqrt{\pi})^{\frac{n}{\nu}}} |\Omega_n|^{-\frac{1}{\nu}} \exp \left\{ -\frac{1}{\nu} (\mathbf{z} - (\mathbf{X}\beta + \beta_0 \mathbf{1}_n))^\top (\Omega_n)^{-1} (\mathbf{z} - (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)) \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu} (\mathbf{z}^\top \Omega_n^{-1} \mathbf{z} - \mathbf{z}^\top \Omega_n^{-1} (\mathbf{X}\beta + \beta_0 \mathbf{1}_n) - (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)^\top \Omega_n^{-1} \mathbf{z} \right. \\ &\quad \left. + (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)^\top \Omega_n^{-1} (\mathbf{X}\beta + \beta_0 \mathbf{1}_n) \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu} (-\nu \mathbf{z} \Omega_n^{-1} (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)^\top + (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)^\top \Omega_n^{-1} (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)) \right\} \end{aligned}$$

$$\begin{aligned} \pi(\beta) &= \prod_{j=1}^p \frac{1}{\sqrt{\nu \pi \lambda_j^2 \tau^2 \sigma^2}} \exp \left\{ -\frac{1}{\nu \lambda_j^2 \tau^2 \sigma^2} \beta_j^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu} \beta^\top \Lambda_p^{-1} \beta \right\} \end{aligned}$$

می توان نوشت

$$\begin{aligned} \pi(\beta|\mathbf{X}, \mathbf{z}, \beta_0 \mathbf{1}_n, \Omega_n, \Lambda_p) &\propto L(\beta) \times \pi(\beta) \\ &\propto \exp \left\{ -\frac{1}{\nu} (-\nu \mathbf{z} \Omega_n^{-1} (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)^\top + \right. \\ &\quad \left. (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)^\top \Omega_n^{-1} (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)) \right\} \times \exp \left\{ -\frac{1}{\nu} \beta^\top \Lambda_p^{-1} \beta \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu} (-\nu \mathbf{z} \Omega_n^{-1} (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)^\top + \right. \\ &\quad \left. (\mathbf{X}\beta + \beta_0 \mathbf{1}_n)^\top \Omega_n^{-1} (\mathbf{X}\beta + \beta_0 \mathbf{1}_n) + \beta^\top \Lambda_p^{-1} \beta) \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu} (-\nu \beta^\top \mathbf{X}^\top \Omega_n^{-1} \mathbf{z} - \nu (\beta_0 \mathbf{1}_n)^\top \Omega_n^{-1} \mathbf{z} + \right. \\ &\quad \left. \beta^\top \mathbf{X}^\top \Omega_n^{-1} \mathbf{X} \beta + \beta \mathbf{X}^\top \Omega_n^{-1} (\beta_0 \mathbf{1}_n) + (\beta_0 \mathbf{1}_n)^\top \Omega_n^{-1} \mathbf{X} \beta + \right. \\ &\quad \left. (\beta_0 \mathbf{1}_n)^\top \Omega_n^{-1} (\beta_0 \mathbf{1}_n) + \beta^\top \Lambda_p^{-1} \beta) \right\} \\ &\propto \exp \left\{ -\frac{1}{\nu} (-\nu \beta^\top \mathbf{X}^\top \Omega_n^{-1} \mathbf{z} + \beta^\top \mathbf{X}^\top \Omega_n^{-1} \mathbf{X} \beta + \right. \\ &\quad \left. \nu \beta^\top \mathbf{X}^\top \Omega_n^{-1} (\beta_0 \mathbf{1}_n) + \beta^\top \Lambda_p^{-1} \beta) \right\} \end{aligned}$$

$$\begin{aligned} & \propto \exp \left\{ -\frac{1}{\varphi} (\beta^\top (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1}) \beta - 2\beta^\top \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}_n)) \right\} \\ & \propto \exp \left\{ -\frac{1}{\varphi} (\beta^\top (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1}) \beta - 2\beta^\top \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}) \times \right. \\ & \quad (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1})^{-1} (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1}) + \\ & \quad (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1})^{-1} \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}_n) - \\ & \quad \left. (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1})^{-1} \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}_n)) \right\} \\ & \propto \exp \left\{ -\frac{1}{\varphi} \left(\beta - (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1})^{-1} \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}_n) \right)^\top \right. \\ & \quad \left. (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1}) \left(\beta - (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1})^{-1} \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}_n) \right) \right\} \end{aligned}$$

حال اگر $\mathbf{A}_p = (\mathbf{X}^\top \Omega_n^{-1} \mathbf{X} + \Lambda_p^{-1})$ باشد در این صورت

$$\begin{aligned} \pi(\beta | \mathbf{X}, \mathbf{z}, \beta \circ \mathbf{1}_n, \Omega_n, \Lambda_p) & \propto \exp \left\{ -\frac{1}{\varphi} \left(\beta - \mathbf{A}_p^{-1} \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}_n) \right)^\top \mathbf{A}_p \right. \\ & \quad \left. \left(\beta - \mathbf{A}_p^{-1} \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}_n) \right) \right\} \end{aligned}$$

□ است. با در نظر گرفتن $\tilde{\mu} = \mathbf{A}_p^{-1} \mathbf{X}^\top \Omega_n^{-1} (\mathbf{z} - \beta \circ \mathbf{1}_n)$ حکم ثابت است.

از دیدگاه محاسباتی چالش عمده الگوریتم گیبز پیشنهادی، نمونه‌گیری از ضرایب رگرسیونی با توزیع نرمال $N_p(\tilde{\mu}, \mathbf{A}_p^{-1})$ وقتی تعداد متغیرهای توضیحی p زیاد است، می‌باشد. بنابراین محاسبه مستقیم معکوس ماتریس $\mathbf{A}_p(\mathbf{A}_p^{-1})$ زمانی که تعداد متغیرهای توضیحی p زیاد است، پیشنهاد نمی‌شود زیرا دقت محاسبات عددی کاهش یافته و هزینه محاسبات زیاد است. در عوض از دو الگوریتم دیگر برای نمونه‌گیری از ضرایب رگرسیونی استفاده می‌کنیم. به طور دقیق‌تر اگر نسبت $\frac{p}{n}$ کوچک‌تر از ۲ باشد از الگوریتم رو^{۱۴} (۲۰۰۱) و اگر $\frac{p}{n} > ۲$ از الگوریتم باتاچاریا^{۱۵} استفاده می‌کنیم. در بسته bayesreg زمانی که p از n خیلی بزرگتر است، نمونه‌گیری با استفاده از الگوریتم باتاچاریا (۲۰۱۶) انجام می‌شود.

وقتی که $\frac{p}{n} > ۲$ است الگوریتم باتاچاریا (۲۰۱۶) کارآمدی بیشتری نسبت به الگوریتم رو (۲۰۰۱) دارد. در ادامه، توزیع نرمال آمیخته مقیاسی در چهارچوب روابط (۱.۴) تا (۳.۴) در بسته bayesreg را برای مدل‌های رگرسیون خطی و لجستیک توضیح می‌دهیم.

۱.۲.۴ رگرسیون خطی

یک حالت ساده رگرسیون جریمه‌شده بیزی که در روابط (۱.۴) تا (۷.۴) به آن اشاره شد، مدل‌های رگرسیون خطی با خطای نرمال است. اگر در داده‌ها، داده‌های پرت^{۱۶} وجود داشته باشد، مدل‌های رگرسیونی خطی با خطای نرمال مناسب نیستند و در این صورت باید از

^{۱۴}Rue

^{۱۵}Bhattacharya

^{۱۶}Outlier

مدل‌های رگرسیونی که توزیع خطاهای آن با دم‌های سنگین است استفاده کرد. در این حالت، مدل‌های رگرسیون خطی با خطاهای لاپلاس و t -استیودنت که معادل با توزیع نرمال آمیخته مقیاسی با چگالی‌های نمایی و گاما معکوس هستند، قابل استفاده می‌باشند. در رگرسیون خطی برای داده‌های پیوسته y فرض می‌کنیم $z_i = y_i$. در همه مدل‌ها چه رگرسیون خطی و چه لجستیک، پارامتر مقیاس σ^2 است که دارای توزیع پیشین یکسان $\pi(\sigma^2) = \frac{1}{\sigma^2}$ می‌باشد. توزیع پسین شرطی برای σ^2 دارای توزیع گاما معکوس به صورت $\text{IGamma}(\hat{\alpha}, \hat{\beta})$ بوده که در آن

$$\tilde{\alpha} = \frac{n+p}{2}, \quad \tilde{\beta} = \frac{1}{2} \left(\sum_{i=1}^n \frac{e_i^2}{\omega_i^2} + \sum_{j=1}^p \frac{\beta_j^2}{\tau^2 \lambda_j^2} \right). \quad (14.4)$$

از آن جایی که

$$\pi(\sigma^2 | \mathbf{x}_i, z_i, \beta, \beta_0, \omega_i^2) \propto L(\sigma^2) \times \pi(\beta | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\sigma^2)$$

با توجه به

$$\begin{aligned} L(\sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2\omega_i^2}} \exp \left\{ -\frac{1}{2\sigma^2\omega_i^2} (z_i - (\mathbf{x}_i^T \beta + \beta_0))^2 \right\} \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{1}{\omega_i^2} (z_i - (\mathbf{x}_i^T \beta + \beta_0))^2 \right\} \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{e_i^2}{\omega_i^2} \right\} \end{aligned}$$

و

$$\begin{aligned} \pi(\beta | \lambda_j^2, \tau^2, \sigma^2) &= \prod_{j=1}^p \frac{1}{\sqrt{2\pi\lambda_j^2\tau^2\sigma^2}} \times \exp \left\{ -\frac{1}{2\sigma^2\tau^2\lambda_j^2} \beta_j^2 \right\} \\ &\propto (\sigma^2)^{-\frac{p}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\tau^2\lambda_j^2} \right\} \end{aligned}$$

نتیجه می شود

$$\begin{aligned} \pi(\sigma^2 | \mathbf{x}_i, z_i, \beta, \beta_0, \omega_i^2) &\propto L(\sigma^2) \times \pi(\beta | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{e_i^2}{\omega_i^2} \right\} \times \\ &\quad (\sigma^2)^{-\frac{p}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\tau^2 \lambda_j^2} \right\} \times \frac{1}{\sigma^2} \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{e_i^2}{\omega_i^2} \right\} \times \\ &\quad (\sigma^2)^{-\frac{p}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\tau^2 \lambda_j^2} \right\} \times (\sigma^2)^{-1} \\ &\propto (\sigma^2)^{-(\frac{n+p}{2})-1} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{j=1}^p \frac{\beta_j^2}{\tau^2 \lambda_j^2} + \sum_{i=1}^n \frac{e_i^2}{\omega_i^2} \right) \right\} \end{aligned}$$

حال اگر فرض کنیم $\tilde{\alpha} = \frac{n+p}{2}$ و $\tilde{\beta} = \left(\sum_{j=1}^p \frac{\beta_j^2}{\tau^2 \lambda_j^2} + \sum_{i=1}^n \frac{e_i^2}{\omega_i^2} \right)$ نتیجه مورد نظر در (۱۴.۴) حاصل می شود.

در ادامه نمونه گیری از متغیرهای پنهان $(\omega_1^2, \dots, \omega_n^2)$ برای توزیع هایی که خطای مدل نرمال، لاپلاس، یا t - استیودنت است، توضیح داده می شود.

- خطای نرمال: در حالت کلی که توزیع خطاها نرمال است داده ها را به جای اینکه از توزیع آمیخته مقیاسی نرمال تولید کنیم از خود نرمال تولید می کنیم و بنابراین در این حالت متغیرهای پنهان $(\omega_1^2, \dots, \omega_n^2)$ را به صورت $\omega_i^2 = 1$ به ازای $i = 1, \dots, n$ در نظر می گیریم که بدین معنی است که از متغیرهای پنهان، نمونه نمی گیریم و لذا در رابطه $\Omega_n = \sigma^2 \mathbf{I}_n$ (۹.۴).

- خطای لاپلاس: توزیع لاپلاس دم های پهن تری نسبت به توزیع نرمال دارد و معمولاً در مدل بندی داده های آمیخته و داده های با نقاط پرت به کار می رود. مهم ترین برتری توزیع لاپلاس نسبت به سایر توزیع های دم پهن مثل t - استیودنت این است که همه گشتاورهای مرکزی آن متناهی هستند. در این حالت توزیع لاپلاس به صورت یک توزیع آمیخته مقیاسی نرمال در نظر گرفته می شود که چگالی آمیختگی به صورت $\omega_i \sim \text{Exp}(1)$ (نمایی با میانگین یک) است. این انتخاب خاص توزیع آمیختگی باعث می شود باقیمانده های (e_1, \dots, e_n) از یک توزیع لاپلاس پیروی کرده که در آن پارامتر مقیاس σ^2 (رابطه (۲.۴)) معادل با واریانس باقیمانده ها، مشابه رگرسیون خطی با خطای نرمال است. توزیع پسین شرطی متغیر $\frac{1}{\omega_i^2}$ دارای توزیع نرمال معکوس $IG(\tilde{\mu}_i, \tilde{\lambda})$ با $\tilde{\mu}_i = \frac{2\sigma^2}{e_i^2}$ و $\tilde{\lambda} = 2$ به ازای $i = 1, \dots, n$ است.

- خطای t - استیودنت: در مقابل توزیع لاپلاس که معمولاً در رگرسیون خطی با دم سنگین استفاده می شود، توزیع t - استیودنت است. توزیع t - استیودنت، دم سنگین تر

از توزیع نرمال می‌باشد. وقتی درجه آزادی (در اینجا آن را با δ نشان می‌دهیم) برابر با یک ($\delta = 1$) باشد توزیع t - استیودنت به توزیع کوشی تبدیل می‌شود که از دو توزیع نرمال و لاپلاس دم‌های سنگین‌تری دارد. در کل، توزیع t - استیودنت دارای دم سنگین‌تر از توزیع لاپلاس برای همه $\delta \leq 5$ بوده اما، برخلاف توزیع لاپلاس، توزیع t - استیودنت دارای واریانس نامتناهی به ازای $\delta \leq 2$ است.

همانطور که در فصل ۳ نیز به آن اشاره شد، توزیع t - استیودنت یک توزیع آمیخته مقیاسی نرمال است. به علاوه ω_i^2 دارای توزیع گامای معکوس به صورت زیر است

$$\omega_i^2 \sim \text{IGamma}\left(\frac{\delta}{2}, \frac{\delta}{2}\right), \quad (i = 1, \dots, n). \quad (15.4)$$

برای پارامتر مقیاسی σ^2 و باقیمانده‌های (e_1, \dots, e_n) ، توزیع پسین شرطی برای متغیرهای پنهان ω_i^2 دارای توزیع گاما معکوس $\text{IGamma}(\hat{\alpha}, \hat{\beta}_i)$ است که در آن

$$\hat{\alpha} = \frac{\delta + 1}{2}, \quad \hat{\beta}_i = \frac{1}{2} \left(\frac{e_i^2}{\sigma^2} + \delta \right). \quad (16.4)$$

در این قسمت نیز مشابه اثبات سراسری که برای (۱۴.۴) آورده شده به راحتی می‌توان رابطه (۱۶.۴) را نتیجه گرفت. به عبارت دقیق‌تر از آن جایی که

$$\pi(\omega_i^2 | \mathbf{x}_i, z_i, \beta, \beta_0, \sigma^2) \propto L(\omega_i^2) \times \pi(\omega_i^2)$$

با توجه به

$$\begin{aligned} L(\omega_i^2) &\propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2\omega_i^2}} \exp\left\{-\frac{1}{2\sigma^2\omega_i^2}(z_i - (\mathbf{x}_i^\top \boldsymbol{\beta} + \beta_0))^2\right\} \\ &\propto (\omega_i^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{\omega_i^2} \left(\frac{1}{2\sigma^2}(z_i - (\mathbf{x}_i^\top \boldsymbol{\beta} + \beta_0))^2\right)\right\} \\ &\propto (\omega_i^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{\omega_i^2} \left(\frac{1}{2} \left(\frac{e_i^2}{\sigma^2}\right)\right)\right\} \end{aligned}$$

$$\begin{aligned} \pi(\omega_i^2) &\propto \frac{\left(\frac{\delta}{2}\right)^{\frac{\delta}{2}}}{\Gamma\left(\frac{\delta}{2}\right)} (\omega_i^2)^{-\frac{\delta}{2}-1} \exp\left\{-\frac{\delta}{2\omega_i^2}\right\} \\ &\propto (\omega_i^2)^{-\frac{\delta}{2}-1} \exp\left\{-\frac{1}{\omega_i^2} \left(\frac{1}{2}\delta\right)\right\} \end{aligned}$$

$$\begin{aligned}
 \pi(\omega_i^\gamma | \mathbf{x}_i, z_i, \beta, \beta_0, \sigma^\gamma) &\propto L(\omega_i^\gamma) \times \pi(\omega_i^\gamma) \\
 &\propto (\omega_i^\gamma)^{-\frac{1}{\gamma}} \exp \left\{ -\frac{1}{\omega_i^\gamma} \left(\frac{1}{\gamma} \frac{e_i^\gamma}{\sigma^\gamma} \right) \right\} \times \\
 &\quad (\omega_i^\gamma)^{-\frac{\delta}{\gamma}-1} \exp \left\{ \left(\frac{1}{\gamma} \delta \right) \right\} \\
 &\propto (\omega_i^\gamma)^{-\frac{\delta}{\gamma}-\frac{1}{\gamma}-1} \exp \left\{ -\frac{1}{\omega_i^\gamma} \left(\frac{1}{\gamma} \left(\frac{e_i^\gamma}{\sigma^\gamma} \right) \right) \right\} \\
 &\propto (\omega_i^\gamma)^{-\frac{\delta}{\gamma}-\frac{2}{\gamma}} \exp \left\{ -\frac{1}{\omega_i^\gamma} \left(\frac{1}{\gamma} \left(\frac{e_i^\gamma}{\sigma^\gamma} \right) \right) \right\} \\
 &\propto (\omega_i^\gamma)^{-\frac{\delta+1}{\gamma}-1} \exp \left\{ -\frac{1}{\omega_i^\gamma} \left(\frac{1}{\gamma} \left(\frac{e_i^\gamma}{\sigma^\gamma} \right) \right) \right\}
 \end{aligned}$$

با فرض $\bar{\alpha} = \frac{\delta+1}{\gamma}$ و $\bar{\beta} = \frac{1}{\gamma} \left(\frac{e_i^\gamma}{\sigma^\gamma} + \delta \right)$ ، نتیجه دلخواه در رابطه (۱۶.۴) حاصل می شود.

۲.۲.۴ رگرسیون دودویی لجستیک

رگرسیون لجستیک برای یافتن احتمال رویداد مبنی بر موفقیت و شکست به کار می رود که تابع پیوند این مدل رگرسیونی عضو خانواده GLM به صورت

$$\text{logit}(p) = \ln \left(\frac{p}{1-p} \right) = \mathbf{x}^\top \boldsymbol{\beta} = \sum_{j=1}^p x_{ij}^\top \beta_j, \quad i = 1, \dots, n$$

تعریف می شود که در آن

$$p = p(y_i | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\beta_1 x_{1,i} + \dots + \beta_p x_{p,i}\}}{1 + \exp\{\beta_1 x_{1,i} + \dots + \beta_p x_{p,i}\}}$$

است. از طرفی به حالتی که متغیر وابسته بیش از دو سطح داشته باشد به آن رگرسیون لجستیک چندگانه گفته می شود. در مورد، داده های دودویی $y_i \in \{0, 1\}$ ($i = 1, \dots, n$)، ارتباط بین متغیرهای توضیحی و متغیر خروجی در مدل های رگرسیون لجستیک دودویی با استفاده از رابطه زیر تعیین می گردد

$$p(y_i | x_i, \beta_0, \boldsymbol{\beta}) = \frac{1}{1 + \exp(-(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}))}. \quad (17.4)$$

نمونه گیری مستقیم از توزیع پسین پارامترهای رگرسیونی در مدل رگرسیون لجستیک دودویی سخت و مشکل است که این ناشی از روابط پیچیده ریاضی توابع لجستیک است. به تازگی، نمونه گیری غیر مستقیم روی متغیرهای کمکی^{۱۷} (یا متغیرهای پنهان) برای رگرسیون لجستیک

^{۱۷}Auxiliary

توسط هلمز و هلد^{۱۸} (۲۰۰۶)، فرورس-اشنایدر و فرورس^{۱۹} (۲۰۰۷)، گرامسی^{۲۰} و پولسن (۲۰۱۲) و پولسن (۲۰۱۳) پیشنهاد شده است، که در میان همه این‌ها، الگوریتم پولسن (۲۰۱۳) در دیدگاه محاسباتی و کارایی نمونه‌گیری، به خوبی و به سهولت اجرا می‌شود و متداول تر است. نکته مهم در الگوریتم پولسن (۲۰۱۳) این است که تابع لجستیک را به صورت یک توزیع نرمال آمیخته با یک چگالی پولیا-گاما آمیخته در نظر گرفته، به راحتی در ساختار روابط (۱.۴) تا (۷.۴) قرار می‌گیرد.

اجرای رگرسیون لجستیک بیزی از طریق نمایش سلسه‌مراتبی توزیع پولیا-گاما نیاز به نمونه‌گیری دقیقی از متغیرهای پنهان $(\omega_1, \dots, \omega_n)$ و پارامتر مقیاس σ^2 دارد. لذا برای سادگی، بر خلاف رگرسیون خطی، در این حالت σ^2 نیاز به نمونه‌گیری نداشته و مقدار ثابت تعیین شده $\sigma^2 = 1$ برای آن در نظر می‌گیریم. در ادامه مشابه پولسن و همکاران (۲۰۱۳) فرض می‌کنیم متغیرهای پنهان $(\omega_1, \dots, \omega_n)$ دارای پیشین پولیا-گاما استاندارد هستند. به عبارتی فرض می‌کنیم

$$\omega_i^{\downarrow} \sim PG(\circ, 1) \quad (18.4)$$

قضیه ۲.۲.۴. (مکالیک و همکاران، ۲۰۱۶). تحت توزیع پولیا-گاما برای متغیرهای پنهان، توزیع پسین شرطی متغیرهای پنهان $\frac{1}{\omega_i^{\downarrow}}$ ، پولیا-گاما $PG(1, \tilde{c}_i)$ بوده که در آن

$$\tilde{c}_i = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}, \quad (i = 1, \dots, n) \quad (19.4)$$

برهان. با توجه به

$$\pi(\omega_i^{\downarrow} | \mathbf{x}_i, z_i, \boldsymbol{\beta}, \beta_0, \sigma^2) \propto L(\omega_i^{\downarrow}) \times \pi(\omega_i^{\downarrow})$$

از آن جایی که

$$\begin{aligned} L(\omega_i^{\downarrow}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2\omega_i^{\downarrow}}} \exp\left\{-\frac{1}{2\sigma^2\omega_i^{\downarrow}}(z_i - (\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0))^2\right\} \\ &\propto (\omega_i^{\downarrow})^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2\omega_i^{\downarrow}}(z_i - (\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0))^2\right\} \end{aligned}$$

و

$$\pi(\omega_i^{\downarrow}) = \frac{1}{\Gamma} \sum_{i=1}^n (-1)^n \frac{1}{n(n+1)} \frac{\Gamma n}{\sqrt{2\pi(\omega_i^{\downarrow})^3}} \exp\left\{-\frac{\Gamma n^{\downarrow}}{\lambda \omega_i^{\downarrow}} - \frac{1}{\Gamma} \omega_i^{\downarrow}\right\}$$

نتیجه می‌شود

$$\begin{aligned} \pi(\omega_i^{\downarrow} | \mathbf{x}_i, z_i, \boldsymbol{\beta}, \beta_0, \sigma^2) &\propto L(\omega_i^{\downarrow}) \times \pi(\omega_i^{\downarrow}) \\ &\propto (\omega_i^{\downarrow})^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2\omega_i^{\downarrow}}(z_i - (\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0))^2\right\} \times \\ &\quad \frac{1}{\Gamma} \sum_{i=1}^n (-1)^n \frac{1}{n(n+1)} \frac{\Gamma n}{\sqrt{2\pi(\omega_i^{\downarrow})^3}} \exp\left\{-\frac{\Gamma n^{\downarrow}}{\lambda \omega_i^{\downarrow}} - \frac{1}{\Gamma} \omega_i^{\downarrow}\right\} \end{aligned}$$

^{۱۸}Holmes and Held

^{۱۹}Fruhwirth and Schnatter

^{۲۰}Gramacy

با اعمال تغییر متغیر $t = \frac{1}{\omega_i^2}$ و استفاده از ژاکوبی تبدیل $J(\omega_i^2 \rightarrow t) = -\frac{1}{t^2}$ داریم

$$\begin{aligned} p_i(t|\mathbf{x}_i, z_i, \beta, \beta_0, \sigma^2) &\propto t^{\frac{n}{2}} \times \frac{1}{\sqrt{\pi}} \sum_{i=1}^n (-1)^n \frac{1}{n(n+1)} \frac{2}{\sqrt{2\pi(\frac{1}{t})^3}} \exp\left\{\frac{4}{\lambda} n^2 t - \frac{1}{2t} - \frac{1}{2\sigma^2} t(z_i - \tilde{c}_i)\right\} \times |J(\omega_i^2 \rightarrow t)| \\ &\propto t^{\frac{n}{2}} \times \frac{1}{\sqrt{\pi}} \sum_{i=1}^n (-1)^n \frac{1}{n(n+1)} \frac{2}{\sqrt{2\pi(\frac{1}{t})^3}} \exp\left\{\frac{4}{\lambda} n^2 t - \frac{1}{2t} - \frac{1}{2\sigma^2} t(z_i - \tilde{c}_i)\right\} \times \left| -\frac{1}{t^2} \right| \\ &\propto t^{\frac{n}{2}-2} \times \frac{1}{\sqrt{\pi}} \sum_{i=1}^n (-1)^n \frac{1}{n(n+1)} \frac{2}{\sqrt{2\pi(\frac{1}{t})^3}} \exp\left\{\frac{4}{\lambda} n^2 t - \frac{1}{2t} - \frac{1}{2\sigma^2} t(z_i - \tilde{c}_i)\right\}. \end{aligned}$$

□ در نتیجه $\frac{1}{\omega_i^2} \sim \text{PG}(1, \tilde{c}_i)$ بوده و اثبات کامل است.

یک الگوریتم کارا در نمونه‌گیری از توزیع پولیا-گاما توسط ویندل و همکاران (۲۰۱۴) ارائه شده که در بسته bayesreg از آن استفاده می‌شود. لازم به ذکر است که برای نمونه‌گیری از توزیع پولیا-گاما، لازم است ابتدا بسته BayesLogit نصب شود.

۳.۴ توزیع‌های پیشین

همه توزیع‌های پیشین برای ضرایب رگرسیونی در روابط (۵.۴) تا (۷.۴) مشخص شده‌اند. باید توجه داشت که متغیرهای پنهان τ^2 و $(\lambda_1, \dots, \lambda_p)$ به عنوان پارامترهایی هستند که تنگی را کنترل می‌کنند. ابر پارامتر τ^2 نشان‌دهنده واریانس فراموضعی است. پولسن و اسکات (۲۰۱۰) اولین بار برای پارامتر τ^2 توزیع پیشین را معرفی کردند. در بسته bayesreg توزیع پیشین برای پارامترهای انقباضی فراموضعی τ به صورت زیر انتخاب می‌شود

$$\tau \sim C^+(\circ, 1), \quad (20.4)$$

که در آن $C^+(\circ, 1)$ یک توزیع نیم کوشی^{۲۱} با میانگین صفر و پارامتر مقیاسی یک است. جالب توجه این که، توزیع نیم کوشی می‌تواند به صورت توزیع گامای معکوس مقیاسی (مکالیک و اشمیت، ۲۰۱۶) نوشته شود. بنابراین (با توجه رابطه (۲۰.۴)) می‌توان نوشت

$$\tau^2 | \xi \sim \text{IGamma}\left(\frac{1}{\rho}, \frac{1}{\xi}\right), \quad \xi \sim \text{IGamma}\left(\frac{1}{\rho}, 1\right) \quad (21.4)$$

که در آن $\xi > \circ$ یک پارامتر آمیخته است و $\text{IGamma}(\hat{\alpha}, \hat{\beta})$ معرف توزیع گامای معکوس با پارامتر شکل $\hat{\alpha}$ و پارامتر مقیاسی $\hat{\beta}$ می‌باشد. از مزیت نمایش آمیخته این است که اجازه می‌دهد

^{۲۱}Half-Cauchy

که نمونه‌گیری از توزیع پسین شرطی از متغیرهای پنهان τ و $(\lambda_1, \dots, \lambda_p)$ با نمونه‌گیری گیبز انجام می‌شود. با فرض نمایش سلسه‌مراتبی متغیر پنهان در رابطه (۲۱.۴) توزیع پسین شرطی τ ، IGamma($\tilde{\alpha}$, $\tilde{\beta}$) که در آن

$$\tilde{\alpha} = \frac{p+1}{2}, \quad \tilde{\beta} = \frac{1}{\xi} + \frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2} \quad (22.4)$$

توجیه رابطه (۲۲.۴) همانند رابطه گذشته سراسر است. با توجه به

$$\begin{aligned} \pi(\tau^2 | \mathbf{x}_i, z_i, \beta, \beta_0, \omega_i^2, \sigma^2, \xi) &\propto L(\tau^2) \times \pi(\beta | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\tau^2) \\ &\propto \pi(\beta | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\tau^2) \\ &\propto \pi(\beta | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\tau^2 | \xi) \\ &\propto \prod_{j=1}^p \frac{1}{\sqrt{2\pi\lambda_j^2\tau^2\sigma^2}} \exp\left\{-\frac{1}{2\lambda_j^2\tau^2\sigma^2}\beta_j^2\right\} \times \\ &\quad \frac{(\frac{1}{\xi})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \times (\tau^2)^{-\frac{1}{2}-1} \exp\left\{-\frac{1}{\tau^2}\right\} \\ &\propto (\tau^2)^{-\frac{p}{2}} \exp\left\{-\frac{1}{2\tau^2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}\right\} \times (\tau^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{\tau^2}\right\} \\ &\propto (\tau^2)^{-\frac{p}{2}-\frac{1}{2}} \exp\left\{-\frac{1}{\tau^2}\left(\frac{1}{\xi} + \frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}\right)\right\} \\ &\propto (\tau^2)^{-\frac{p+1}{2}-1} \exp\left\{-\frac{1}{\tau^2}\left(\frac{1}{\xi} + \frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}\right)\right\} \end{aligned}$$

در صورتی که $\tilde{\alpha} = \frac{p+1}{2}$ و $\tilde{\beta} = \left(\frac{1}{\xi} + \frac{1}{2\sigma^2} \sum_{j=1}^p \frac{\beta_j^2}{\lambda_j^2}\right)$ باشد نتیجه حاصل می‌شود. به طور مشابه توزیع پسین شرطی پارامتر آمیختگی ξ ، IGamma(α^* , β^*) است که در آن

$$\alpha^* = 1, \quad \beta^* = 1 + \frac{1}{\tau^2} \quad (23.4)$$

زیرا

$$\begin{aligned} \pi(\xi | \mathbf{x}_i, z_i, \beta, \beta_0, \omega_i^2, \sigma^2, \tau^2) &\propto L(\xi) \times \pi(\beta | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\tau^2 | \xi) \times \pi(\xi) \\ &\propto \pi(\tau^2 | \xi) \times \pi(\xi) \\ &\propto \frac{(\frac{1}{\xi})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} (\tau^2)^{-\frac{1}{2}-1} \exp\left\{-\frac{1}{\tau^2}\left(\frac{1}{\xi}\right)\right\} \times \\ &\quad \frac{1}{\Gamma(\frac{1}{2})} \xi^{-\frac{1}{2}-1} \exp\left\{-\frac{1}{\xi}\right\} \\ &\propto \xi^{-\frac{1}{2}} \exp\left\{-\frac{1}{\xi}\left(\frac{1}{\tau^2}\right)\right\} \times \xi^{-\frac{1}{2}} \exp\left\{-\frac{1}{\xi}\right\} \\ &\propto \xi^{-2} \exp\left\{-\frac{1}{\xi}\left(\frac{1}{\tau^2} + 1\right)\right\} \end{aligned}$$

بدیهی است اگر فرض کنیم

$$-\alpha^* - 1 = -2, \quad \alpha^* = 1, \quad \beta^* = 1 + \frac{1}{\tau^2}$$

در این صورت $\xi \sim \text{IGamma}(\alpha^*, \beta^*)$ است.

ابریارامتر $(\lambda_1, \dots, \lambda_n)$ معادل مولفه‌های واریانس موضعی (عامل انقباض) هستند که نوع جریمه انقباضی را در ضرایب رگرسیونی مشخص می‌کنند. مشابه پولسن و اسکات (۲۰۱۲) پارامترهای واریانس پارامترهای واریانس موضعی باید توزیع پیشینی داشته باشند

- یک قطب در صفر که تضمین کننده کارآمدی پیش‌بینی کننده در به‌دست آوردن توزیع نمونه‌ای درست از پارامترهای رگرسیونی $\beta \in \mathbb{R}^p$ است.
- دارای دنباله‌های چند جمله‌ای باشد که شدیداً متغیرهای خطا را منقبض کرده و در عین حال سیگنال‌های بزرگ بدون تغییر بمانند.

جزئیات نمونه‌گیری برای برآوردگرهای ریج بیزی، لاسوی بیزی، نعل اسبی بیزی ارائه شده است. از میان این برآوردگرها، فقط برآوردگر نعل اسبی دارای دو خاصیت یا ویژگی انقباضی است.

همان‌طور که در فصل دوم دیدیم در برآوردگر لاسوی بیزی (پارک و کسلا، ۲۰۰۸ و هنس، ۲۰۰۹) ملزم به استفاده از یک توزیع پیشین لاپلاس برای ضرایب رگرسیونی هستیم به طوری که چگالی پسین در این حالت، به طور متداول با برآوردگر لاسو معادل است. توزیع لاپلاس را می‌توان به صورت یک توزیع نرمال آمیخته مقیاسی نوشت. بنابراین ابریارامترهای $(\lambda_1, \dots, \lambda_p)$ دارای توزیع نمایی با میانگین یک به صورت

$$\lambda_j^2 \sim \text{Exp}(1) \quad j = 1, \dots, p \quad (24.4)$$

هستند.

لازم به ذکر است نمایش تصادفی (۲۱.۴) و (۲۴.۴) تا حدی متفاوت از لاسوی بیزی پارک و کسلا (۲۰۰۸) است. زیرا آن‌ها ابریارامترهای بیشتری را در توزیع توام $(\lambda_1, \dots, \lambda_n)$ در نظر گرفته و بنابراین یک برتری فرمول‌بندی در اینجا نسبت به لاسوی بیزی آن‌ها این است که نیاز به سلسله‌مراتب بالاتر در پارامترگذاری نداریم.

برخلاف توزیع‌های پیشین لاسو و ریج، توزیع پیشین نعل اسبی یک قطب در صفر و دنباله‌های چندجمله‌ای دارد که این ویژگی مهم، کارایی خوبی در قلمرو داده‌های بزرگ از خود نشان می‌دهد (پولسن و اسکات، ۲۰۱۲). توزیع پیشین برای ابریارامترهای انقباضی موضعی $(\lambda_1, \dots, \lambda_p)$ دارای توزیع نیم کوشی با میانگین صفر می‌باشد

$$\lambda_j \sim C^+(0, 1), \quad (25.4)$$

که مشابه رابطه (۲۱.۴) معادل با

$$\lambda_j^2 \sim \text{IGamma}\left(\frac{1}{2}, \frac{1}{\nu_j}\right) \quad \nu_j \sim \text{IGamma}\left(\frac{1}{2}, 1\right). \quad (26.4)$$

است .

قضیه ۱.۳.۴. (مکالیک و همکاران، ۲۰۱۶). تحت مفروضات این بخش، توزیع پسین شرطی برای پارامتر انقباضی موضعی $\frac{1}{\lambda_j^2}$ ، توزیع نمایی به صورت

$$Exp\left(\frac{1}{\nu_j} + \frac{\beta_j^2}{2\tau^2\sigma^2}\right)$$

و توزیع پسین شرطی برای ابر پارامتر $\frac{1}{\nu_j}$ برابر است با

$$Exp\left(1 + \frac{1}{\lambda_j^2}\right)$$

برهان. طبق تعریف داریم

$$\begin{aligned} \pi(\lambda_j^2 | \mathbf{x}_i, z_i, \boldsymbol{\beta}, \beta_0, \omega_i^2, \sigma^2, \nu_j) &\propto L(\lambda_j^2) \times \pi(\beta_j | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\lambda_j^2) \\ &\propto \pi(\beta_j | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\lambda_j^2) \\ &\propto \frac{1}{\sqrt{2\pi\lambda_j^2\tau^2\sigma^2}} \exp\left\{-\frac{1}{2\lambda_j^2\tau^2\sigma^2}\beta_j^2\right\} \times \\ &\quad \frac{\left(\frac{1}{\nu_j}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} (\lambda_j^2)^{-\frac{1}{2}-1} \exp\left\{-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}\right)\right\} \\ &\propto (\lambda_j^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{\lambda_j^2}\left(\frac{\beta_j^2}{2\tau^2\sigma^2}\right)\right\} \times \\ &\quad (\lambda_j^2)^{-\frac{3}{2}} \exp\left\{-\frac{1}{\lambda_j^2}\left(\frac{1}{\nu_j}\right)\right\} \\ &\propto (\lambda_j^2)^{-2} \exp\left\{-\frac{1}{\lambda_j^2}\left(\frac{\beta_j^2}{2\tau^2\sigma^2} + \frac{1}{\nu_j}\right)\right\} \\ &\propto \left(\frac{1}{\lambda_j^2}\right)^2 \exp\left\{-\frac{1}{\lambda_j^2}\left(\frac{\beta_j^2}{2\tau^2\sigma^2} + \frac{1}{\nu_j}\right)\right\} \end{aligned}$$

با اعمال تغییر متغیر $t = \frac{1}{\lambda_j^2}$ با ژاکوبی تبدیل $J(\lambda_j^2 \rightarrow t) = -\frac{1}{t^2}$ نتیجه می شود

$$\begin{aligned} \pi(t | \mathbf{x}_i, z_i, \boldsymbol{\beta}, \beta_0, \omega_i^2, \sigma^2, \nu_j) &\propto t^2 \exp\left\{-t\left(\frac{\beta_j^2}{2\tau^2\sigma^2} + \frac{1}{\nu_j}\right)\right\} \times |J(\lambda_j^2 \rightarrow t)| \\ &\propto t^2 \exp\left\{-t\left(\frac{\beta_j^2}{2\tau^2\sigma^2} + \frac{1}{\nu_j}\right)\right\} \times \left|-\frac{1}{t^2}\right| \\ &\propto \exp\left\{-t\left(\frac{\beta_j^2}{2\tau^2\sigma^2} + \frac{1}{\nu_j}\right)\right\} \end{aligned}$$

که نتیجه می دهد $\frac{1}{\lambda_j^2} \sim Exp(\tilde{\beta}_j)$

به‌طور مشابه، با توجه

$$\begin{aligned}
 \pi(\nu_j | \mathbf{x}_i, z_i, \boldsymbol{\beta}, \beta_0, \omega_i^2, \sigma^2, \lambda_j^2) &\propto L(\nu_j) \times \pi(\beta_j | \lambda_j^2, \tau^2, \sigma^2) \times \pi(\lambda_j^2 | \nu_j) \times \pi(\nu_j) \\
 &\propto \pi(\lambda_j^2 | \nu_j) \times \pi(\nu_j) \\
 &\propto \frac{(\frac{1}{\nu_j})^{\frac{1}{\nu}}}{\Gamma(\frac{1}{\nu})} (\lambda_j^2)^{-\frac{1}{\nu}-1} \exp\left\{-\frac{1}{\lambda_j^2} \left(\frac{1}{\nu_j}\right)\right\} \times \\
 &\quad \frac{1}{\Gamma(\frac{1}{\nu})} \nu_j^{-\frac{1}{\nu}-1} \exp\left\{-\frac{1}{\nu_j}\right\} \\
 &\propto (\nu_j)^{-\frac{1}{\nu}} \exp\left\{-\frac{1}{\nu_j} \left(\frac{1}{\lambda_j^2}\right)\right\} \times (\nu_j)^{-\frac{1}{\nu}} \exp\left\{-\frac{1}{\nu_j}\right\} \\
 &\propto (\nu_j)^{-2} \exp\left\{-\frac{1}{\nu_j} \left(1 + \frac{1}{\lambda_j^2}\right)\right\} \\
 &\propto \left(\frac{1}{\nu_j}\right)^2 \exp\left\{-\frac{1}{\nu_j} \left(1 + \frac{1}{\lambda_j^2}\right)\right\}
 \end{aligned}$$

و اعمال تغییر متغیر $t = \frac{1}{\nu_j}$ با ژاکوبی تبدیل $J(\nu_j \rightarrow t) = -\frac{1}{t^2}$ داریم

$$\begin{aligned}
 \pi(t | \mathbf{x}_i, z_i, \boldsymbol{\beta}, \beta_0, \omega_i^2, \sigma^2, \lambda_j^2) &\propto t^2 \exp\left\{-t \left(1 + \frac{1}{\lambda_j^2}\right)\right\} \times |J(\nu_j \rightarrow t)| = \frac{1}{t^2} \\
 &\propto t^2 \exp\left\{-t \left(1 + \frac{1}{\lambda_j^2}\right)\right\} \times \left|-\frac{1}{t^2}\right| \\
 &\propto \exp\left\{-t \left(1 + \frac{1}{\lambda_j^2}\right)\right\}
 \end{aligned}$$

□ که نتیجه می‌دهد $\frac{1}{\nu_j} \sim \text{Exp}(\tilde{\beta}_j)$ و اثبات کامل است.

توزیع‌های شرطی برای ابرپارامترهای λ_j و ν_j به منظور افزایش کارایی در حالت p بزرگ در نظر گرفته شده‌اند.

۴.۴ شرح بسته bayesreg در R

روشی ارائه‌شده در کاروالهو و همکاران (۲۰۱۰) می‌تواند از طریق تابع bayesreg اجرا شود. دستور کلی به صورت زیر است

```
bayesreg <- function(formula, data, model="normal" , prior="ridge", ...)
```

که در آن

- formula و data، همان مکانیزم استاندارد داده و فرمول در مدل بندی در R است.
- model مدل خطا است که ممکن است gaussian، laplace، t و binomial باشد

• **prior** چگالی پیشین برای ضرایب رگرسیونی می باشد که می تواند به صورت مجموعه‌ای از **lasso**، **ridge** یا **hs** باشد که **hs** مخفف **horseshoe** یعنی نعل اسبی است.

نسخه **bayesreg**، جدول آماره‌های خلاصه‌های آماری را مانند سایر بسته‌های دیگر در رگرسیون ارائه می کند. در نرم افزار R، بعضی از آماره‌های خلاصه را می توان با استفاده از دستور **summary** نشان داد که در بسته **bayesreg** فراهم شده است. آماره‌های خلاصه شامل میانگین پسین، انحراف استاندارد و بازه اطمینان ۹۵٪ برای هر ضریب رگرسیونی است. همچنین برآورد آماره **t** رتبه و عامل اندازه نمونه برای p متغیر توضیحی را بیان می کند. آماره رتبه‌ای توسط الگوریتم رتبه‌بندی بیزی محاسبه و برآورد می شود و به قدرت ارتباط بین متغیر و پاسخ y مربوط می شود. برآورد رتبه r یک متغیر با مقادیر صحیح می باشد ($1 \leq r \leq p$) و در جاهایی که بیشتر متغیرهای مهم رتبه کمتری دارند مشخص می شود، و به این معنی که، یک متغیر با رتبه ($r = 1$) قویترین ارتباط را با y داشته در حالی که متغیری دارای رتبه ($r = p$) ارتباط کمتری با y برآورد دارد. عامل موثر بر اندازه نمونه کسلا و روبرت (۲۰۰۴) و گایر (۱۹۹۲) مطرح کردند این است که به عنوان یک درصد و کارایی نمونه‌گیری از زنجیره برای هر پیش‌بینی کننده بیان شده است. تابع **bayesreg** عامل موثر اندازه نمونه از خودهمبستگی زنجیره‌های نمونه‌گیری شده را محاسبه می کند. در ادامه چند مثال که در این بسته آمده است را شرح می دهیم.

۵.۴ مثال

۱.۵.۴ داده‌های دیابت

برای نشان دادن چگونگی عملکرد بسته **bayesreg** در قالب یک مثال واقعی از رگرسیون لجستیک بیزی برای تحلیل داده‌های دیابت استفاده می کنیم. این مجموعه داده متفاوت از مثال فصل ۲ است. مجموعه داده دیابت در این جا داده‌های **Pima Indians** است که توسط موسسه بین المللی دیابت و بیماری گوارشی در اختیار پایگاه یادگیری ماشین **UCI** قرار داده شده است. برای آگاهی بیشتر لیچمن (۲۰۱۳) را ببینید. این داده‌ها مشتمل بر $n = 768$ مشاهدات از بیماران زن ۲۱ ساله هندی با $p = 8$ متغیر توضیحی جمع آوری شده است، که در زیر هر کدام معرفی شده است.

• **PREG** – تعداد دفعات بارداری

• **PLAS** – غلظت گلوکز پلاسما در ۲ ساعت در آزمایش تحمل گلوکز خوراکی

• **BP** – فشار خون دیاستولیک (mm Hg)

• **SKIN** – ماهیچه سه سر ضخامت پوست (mm)

• **INS** – سرم انسولین در ۲ ساعت (mu U / ml)

● BMI – شاخص توده بدنی ($\text{kg}/(\text{m})^2$)

● PED – عملکرد ریشه دیابت

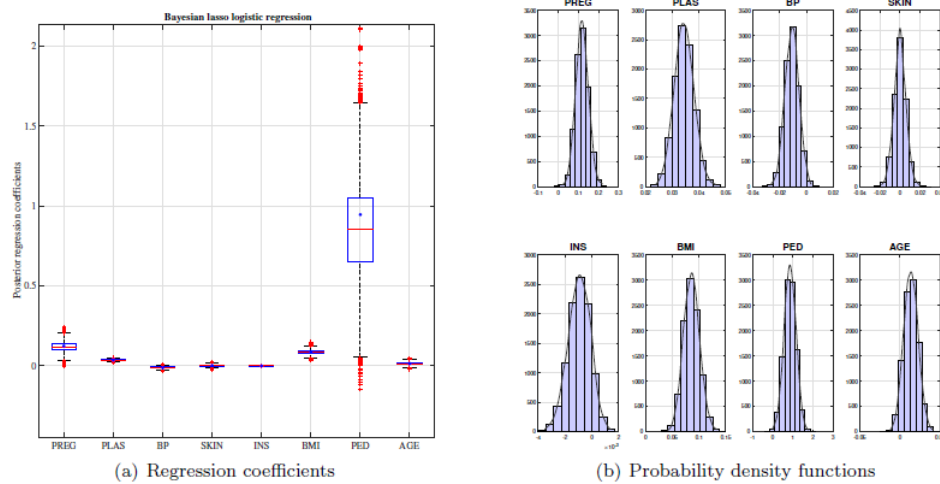
● AGE – سن (سال)

در این قسمت از مدل رگرسیون لجستیک برای پیشگویی در دیدگاه بیزی استفاده می‌کنیم. لازم به ذکر است که پاسخ در این جا یک متغیر دودوئی است که نشان می‌دهد که فرد دارای بیماری دیابت است یا خیر.

جدول ۱.۴: پیشگویی پارامترها بر اساس مدل لجستیک از دیدگاه بیزی برای داده‌های دیابت

پارامتر	میانگین ضرایب	انحراف استاندارد ضرایب	ناحیه اعتبار ۹۵٪		tStat	رتبه	ESS
PREG	۱/۱۲۴۵۷	۰/۰۳۵۹۹۸	۱/۰۵۶۶۲	۱/۱۹۷۶۴	۳/۷۱۹	۳ **	۹۵/۲
PLAS	۱/۰۳۴۸۵	۰/۰۰۳۷۷	۱/۰۲۷۷۹	۱/۰۴۲۵۷	۹/۳۷۴	۱ **	۹۲/۵
BP	۰/۹۸۹۳۲	۰/۰۰۵۰۳	۰/۹۷۹۵۶	۱۹۹۹۲۸	-۲/۰۹۹	۵ **	۹۷/۶
SKIN	۰/۹۹۹۹۴	۰/۰۰۶۱۳	۰/۹۸۸۱۵	۱/۰۱۲۱۷	-۰/۰۱۱	۷ **	۹۰/۸
INS	۰/۹۹۹۰۸	۰/۰۰۰۸۵	۰/۹۹۷۳۳	۱/۰۰۰۶۶	-۱/۰۸۳	۷ *	۱۰۰/۰
BMI	۱/۰۸۹۶۷	۰/۰۱۶۳۱	۱/۰۵۹۲۱	۱/۱۲۳۱۶	۵/۷۸۸	۲ **	۹۶/۰
PED	۲/۳۵۵۵۴	۰/۷۴۹۷۲	۱/۳۴۹۲۵	۴/۲۸۸۱۵	۲/۸۸۵	۴ **	۹۳/۶
AGE	۱/۰۱۳۶۴	۰/۰۰۸۶۹	۰/۹۹۶۹۴	۱/۰۳۲۱۰	۱/۵۲۲	۶ *	۹۴/۴
CONS	۰/۰۰۰۲۶	۰/۰۰۰۲۴	۰/۰۰۰۰۶	۰/۰۰۱۰۱	-۱۱/۵۶۱		

شکل ۱.۴ نمودار جعبه‌ای از نمونه‌های پسین ضرایب رگرسیونی و توابع چگالی احتمالی پسین شرطی را نشان می‌دهد. در این مثال، رتبه‌بندی رگرسیون لاسوی بیزی برای پاسخ لجستیک که بیشترین متغیر مهم در سطح گلوکز پلاسما دارد، که به دنبال BMI، تعداد حاملگی‌ها، سابقه خانوادگی دیابت و دیاستولیک فشار خون است. با توجه به خروجی سه متغیر سن، سطح سرم انسولین و ضخامت پوست مهم تلقی نمی‌شود.



شکل ۱.۴: آنالیز داده‌های دیابت برای لاسوی بی‌زی با پاسخ لجستیک

۶.۴ نتیجه‌گیری و پیشنهادات برای آینده تحقیق

با توجه به آنچه گفته شد و این که داشتن اطلاعات پیشین کمک می‌کند تا بتوانیم برآورد درستی از پارامترها داشته باشیم و زمانی که تعداد پارامترها زیاد است از روش‌های جریمه کردن استفاده می‌شود. در این بحث برای انجام عمل جریمه، روش‌های لاسوی بی‌زی و نعل اسبی بی‌زی اشاره شده است و توانستیم با استفاده از این دو رهیافت به نتایج قابل قبولی برسیم. برای آینده تحقیق می‌توانیم از روشی به نام برآورد نعل اسبی پلاس از دیدگاه بی‌زی به عنوان یکی از روش‌های رگرسیون جریمه شده بی‌زی استفاده کنیم.

مراجع

- [۱] آرست، م. (۱۳۹۵). مقایسه رفتار برخی برآوردگرهای انقباضی بریج در مدل رگسیون چندگانه، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۲] برزوئی بیدگلی، م. (۱۳۹۳). بررسی رفتار برآوردگر جک نایف ریج، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۳] بلوری کلوزی، س. (۱۳۹۵). تحلیل پاسخ‌های دو سطحی نامتعادل با مدل رگسیون مقادیر فرین تعمیم‌یافته، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۴] بدخشان، م. (۱۳۹۷). مدل‌های بقای فضایی بیزی در چهارچوب مخاطره‌های متناسب پارامتری، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۵] رجب‌لو، س. (۱۳۹۷). رگسیون جریمه شده در بعد بالا با تابع جریمه لاسوی مشتق‌پذیر، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۶] رحمانی، م. (۱۳۹۷). استنباط مبتنی بر درست‌نمایی در رگسیون ریج برای داده‌های طولی، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۷] عجم، م. (۱۳۹۳). تحلیل بیزی مدل‌های رگرسیونی بتا با پاسخ‌های وابسته، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۸] عابدین‌پور لیاردمه، ل. (۱۳۹۶). یک مدل منعطف برای پیش‌گویی فضایی پاسخ‌های نرخ به کمک توزیع بتا-دوجمله‌ای، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۹] نوروزی‌راد، م. (۱۳۹۶). برآوردگرهای بهبودیافته در برخی مدل‌های رگرسیونی خطی جریمه شده، رساله دکتری، دانشکده ریاضی، دانشگاه صنعتی شاهرود.
- [۱۰] ولیپور پاشا کلایی، م. (۱۳۹۲). مدل‌بندی داده‌های نرخ و نسبت با رگسیون بتا، پایان نامه کارشناسی ارشد، دانشکده ریاضی، دانشگاه صنعتی شاهرود.

- [11] Agresti, A. (1996). *An Introduction to Categorical Data Analysis*, Wiley, New York.
- [12] Andrews, D. F., and Mallows, C. L. (1974), Scale mixtures of normal distributions, *Journal of the Royal Statistical Society, Ser. B*, **36**, 99–102.
- [13] Armagan, A. Dunson, D. and Lee. J. (2013). Generalized double Pareto shrinkage, *Statistica Sinica*, **23**(1): 119-143.
- [14] Atkinson, A. C. (1982), The simulation of generalized inverse Gaussian and hyperbolic random variables, *SIAM Journal on Scientific and Statistical Computing*, **3**, 502–515.
- [15] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Function With Formulas, Graphs, and Mathematical Tables*.
- [16] Bae, K., and Mallick, B. K. (2004), *Gene selection using a two-level hierarchical Bayesian Model*, *Bioinformatics*, **20**, 3423–3430.
- [17] Berger, J. O. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8**, 716-61.
- [18] Barndor-Nielsen, Kent, J. and Sorensen, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review*, **50**(2):145-159.
- [19] Bhattacharya, A. Chakraborty, A. and Mallick, B. K. (2016). *Fast sampling with Gaussian scale-mixture priors in high-dimensional regression*, arXiv:1506.04778.
- [20] Breheny, P. and Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Annals of Applied Statistics*, **5**(1):232-253.
- [21] Carlin. B. P. Polson. N. G. (1991). Inference for nonconjugate Bayesian models using the Gibbs sampler, *Can. J. Statist.* **19**, 399-405.
- [22] Casella, G. (2001), *Empirical Bayes Gibbs Sampling*, *Biostatistics*, **2**, 485– 500.
- [23] Chhikara, R. S., and Folks, L. (1989), *Inverse Gaussian Distribution: Theory, Methodology, and Applications*, New York: Marcel Dekker.
- [24] Carvalho, C. M. Scott, J. G. (2009). Objective Bayesian model selection in Gaussian graphical models. *Biometrika*, **96**, 497-512.

- [25] Clarke, B. Barron, A. (1990). Information-theoretic asymptotics of Bayes methods. *IEEE Trans. Info. Theory*, **36**, 453-71.
- [26] Carlos, M. Carvalho, Nicholas, Polson, G. (2010). *The horseshoe estimator for sparse signals*.
- [27] Carvalho, C. M. Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, **97**(2):465-480
- [28] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm*, Journal of the Royal Statistical Society. Series B (Methodological), **39**(1), 1-38.
- [29] Figueiredo, M. (2003). Adaptive sparseness for supervised learning. *IEEE Trans. Pat. Anal. Mach. Intel*, **25**, 1150-9.
- [30] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**(456):1348-1360.
- [31] Friedman, J. Hastie, T. and Tibshirani, R. (2010). Regularized paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**(1).
- [32] Fruhwirth-Schnatter, S. and Fruhwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data Analysis*, 51
- [33] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), Least Angle Regression, *The Annals of Statistics*, **32**, 407–499.
- [34] 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
- [35] Figueiredo, M. A. T. (2003), Adaptive Sparseness for Supervised Learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**, 1150– 1159.
- [36] Frank, I. E., and Friedman, J. H. (1993), A Statistical View of Some Chemometrics Regression Tools, *Technometrics*, **35**, 109–135.
- [37] Fu, W. J. (1998), Penalized Regressions: The Bridge versus the Lasso, *Journal of Computational and Graphical Statistics*, **7**, 397–416.
- [38] Gneiting, T. (1997), Normal Scale Mixtures and Dual Probability Densities, *Journal of Statistical Computation and Simulation*, **59**, 375–384.

- [39] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**(3):515-533.
- [40] Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical Science*, **7**(4):473-483
- [41] Gramacy, R. B. and Polson, N. G. (2012). Simulation-based regularized logistic regression. *Bayesian Analysis*, **7**(3).
- [42] Griffin, J. E. and P. J. Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, **5**(1):171-188
- [43] Hastie, T., Tibshirani, Inference, and Prediction, New York: *Springer- Verlag*.
- [44] Hans, C. M. (2009). Bayesian lasso regression. *Biometrika*, **96**, 835-45.
- [45] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: bised estimation for non-orthogonal problems. *Technometrics*, **12**, 55-67.
- [46] Holmes, C. C and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, **1**(1):145-168
- [47] Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society (Series B)*, **34**(1):1-41.
- [48] Lichman, M. (2013). UCI machine learning repository.
- [49] Makalic, E. and Schmidt, D. F. (2011). A simple Bayesian algorithm for feature ranking in high dimensional regression problems. In 24th Australasian Joint Conference on Advances in Artificial Intelligence(AIA 2011), volume 7106 of Lecture Notes in Artificial Intelligence, pages 223-230, Perth, Australia, December 5-8.
- [50] McCullagh, P. Nelder, J.A. (1989). *Generalized Linear Models, 2nd ed.*, Chapman and Hall, London.
- [51] Meyer, J.P., Stanley, D.J., Herscovitch, L. Topolnytsky, L. (2002). Affective, Continuance, and Normative Commitment to the Organization: A Metaanalysis of Antecedents, Correlates, and Consequences, *Journal of Vocational Behavior*, **61**, 20–52
- [52] Makalic, E. Schmidt, D. F. (2016). *High-Dimensional Bayesian Regularized Regression with the bayesreg Package*.

- [53] Meng, X. and Van Dyk, D. (1997). *The EM Algorithm — An Old Folk-Song Sung to a Fast New Tune*, Journal of the Royal Statistical Society. Series B (Methodological), **59**(3), 511-567.
- [54] Nelder, J.A. Wedderburn, R.W.M. (1972), Generalized Linear Models, *Journal of the Royal Statistical Society, A*, **135**, 370–384.
- [55] Jørgensen, B. (1987), Exponential Dispersion Models, *Journal of the Royal Statistical Society, Ser. B*, **49**, 127–162.
- [56] Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. New York: Oxford University Press.
- [57] Johnstone, I. Silverman, B. W. (2004). Needles and straw in haystacks: empirical-Bayes estimates of possibly Sparse sequences. *Ann. Statist.*, **32**, 1594-649.
- [58] Knight, K., and Fu, W. (2000), Asymptotics for Lasso-Type Estimators, *The Annals of Statistics*, **28**, 1356–1378.
- [59] Osborne, M. R., Presnell, B., and Turlach, B. A. (2000a), A New Approach to Variable Selection in Least Squares Problems, *IMA Journal of Numerical Analysis*, **20**, 389–404.
- [60] Park, T. Casella, G. (2008). The Bayesian lasso. *J. Am. Statist. Assoc.* **103**, 681-6.
- [61] Pericchi, L. R. Smith, A. (1992). Exact and approximate posterior moments for a normal location parameter. *J. R. Statist. Soc. B*, **54**, 793-804.
- [62] Polson, N. G. (1991). A representation of the posterior mean for a location model. *Biometrika*, **78**, 426-30.
- [63] Polson, N. G. and Scott, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In *Bayesian Statistics*, volume 9.
- [64] Polson, N. G. Scott, J. G. and Windle, J. (2013). Bayesian inference for logistic models using Polya-gamma latent variables. **108**(504):1339-1349.
- [65] Rosset, S., and Zhu, J. (2004), Discussion of Least Angle Regression, by B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *The Annals of Statistics*, **32**, 469–475.
- [66] Rue, H. (2001). Fast sampling of Gaussian markov random fields. *Journal of the Royal Statistical Society (Series B)*, **63**(2):325-338.
- [67] Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer.

-
- [68] Scott, J. G. Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Statist. Plan. Infer*, **136**, 2144-62.
- [69] Scott, J. G. Carvalho, C. M. (2008). Feature-inclusion stochastic search for Gaussian graphical models. *J. Comp. Graph. Statist*, **17**, 790-808.
- [70] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. In *3rd Berkeley Symp. Math. Statist. Prob.*, vol. 1.
- [71] Strawderman, W. (1971). proper Bayes minimax estimator of multivariate normal mean. *Ann. Statist.* **42**, 385-8.
- [72] Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Ser. B*, **58**, 267–288.
- [73] Tipping, M. (2001). Sparse Bayesian learning and the relevance vector machine. *J. Math. Learn. Res.* **1**, 211-44.
- [74] Tibshirani, R. (2011) Regression shrinkage and selection via the lasso: a retrospective. *J. R. Statist. Soc. B*, **73**:273-282.
- [75] West, M. (1984), Outlier Models and Prior Distributions in Bayesian Linear Regression, *Journal of the Royal Statistical Society, Ser. B*, **46**, 431–439.
- [76] Windle, J. Polson, N. G. and Scott, J. G. (2014). Sampling Polya-gamma random variates: alternate and approximate techniques.
- [77] Wang, D. Q., Chukova, S. and Lai, C. D. (2004). On the relationship between regression analysis and mathematical programming. *Journal of Applied Mathematica and Decision Sciences* , **8**, 131-140.
- [78] Wakefield, J. (2013). *Bayesian and Frequentist Regression Methods*.
- [79] Xu, Z. Schmidt, D. F. Makalic, E. Qian, G. and Hopper, J. L. (2016). Bayesian grouped horseshoe regression with application to additive models. In *29th Australasian Joint Conference on Artificial Intelligence*, Hobart, Australia.
- [80] Yuan, M., and Lin, Y. (2005), Efficient Empirical Bayes Variable Selection and Estimation in Linear Models, *Journal of the American Statistical Association*, **100**, 1215–1225.
- [81] Yao, Y. (2004). *Statistical Applications of Linear Programming for Feature Selection Via Regularization Methods*. Ph. D. dissertation, The Ohio State University.

- [82] Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, **38**(2):894-942.
- [83] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society (Series B)*, **67**(2):301-320.

پیوست آ

در این بخش یک سری مفاهیم و تعاریف تکمیلی را در رابطه با مطالب ذکر شده در فصل‌های گذشته، تنها به منظور افزایش آگاهی خواننده ارائه می‌کنیم. مطالب این بخش تماماً برگرفته از پایان‌نامه عجم (۱۳۹۳) است. اگر چه که در ساختار پایان‌نامه خاص از همه مطالب این قسمت استفاده نشده اما به اعتقاد نویسندگان این مجموعه هم چون مطالب گفته شده در عجم (۱۳۹۳) بسیار جامع و متصل است برای فهم کامل موضوع مورد بحث، تمامی موارد آن بند به بند آورده شده است.

۱.۱ روش نمونه‌گیری MCMC

هدف اصلی در روش‌های انتگرال‌گیری مونت کارلویی، ارائه تقریبی مناسب برای انتگرال‌های از نوع $\tau = \int g(x)dx$ می‌باشد. این انتگرال را می‌توان به صورت

$$\tau = E_f[h(X)] = \int h(x)f(x)dx,$$

نوشت که در آن، $f(\cdot)$ یک تابع چگالی و $h(\cdot)$ تابعی است که متناسب با $g(\cdot)$ تعریف می‌شود. حل مونت کارلویی این انتگرال، نیازمند تولید نمونه X_1, \dots, X_m از $f(x)$ می‌باشد. با محاسبه میانگین نمونه‌ای $\bar{h}_m = \frac{1}{m} \sum_{i=1}^m h(X_i)$ و بنابر قانون قوی اعداد بزرگ، چنان‌چه $m \rightarrow \infty$ ، همگرایی به مقدار انتگرال، τ ، تضمین می‌شود، یعنی

$$\bar{h}_m \xrightarrow{p} E_f[h(X)]$$

روش‌های MCMC ، شامل روش‌هایی است که برای انتگرال‌گیری به روش مونت کارلویی، اغلب در مسائل مربوط به استنباط بیزی، مورد استفاده قرار می‌گیرند. در یک مدل بیزی، با شرط مفروض بودن چگالی پیشین، $\pi(\theta)$ و چگالی توام (x, θ) ، $f(x, \theta)$ ، مساله انتگرال‌گیری به امید ریاضی شرطی تابعی مانند $g(\theta)$ نسبت به توزیع پسین $\pi(\theta | x)$ تبدیل می‌شود، یعنی

$$E[g(\theta | x)] = \int_{\Theta} g(\theta)\pi(\theta | x)d\theta$$

روش‌های انتگرال‌گیری MCMC ، به ویژه الگوریتم‌های متروپولیس-هستینگس (متروپولیس و همکاران، ۱۹۵۳ و هستینگز، ۱۹۷۰) و نمونه‌گیر گیبز (گمان و گمان، ۱۹۸۴ و گل‌فاند و اسمیت، ۱۹۹۰) به عنوان ابزار بسیار مفید برای تحلیل مدل‌های آماری پیچیده پدید آمده‌اند. روش‌های معمول نمونه‌گیری MCMC ، از جمله متروپولیس-هستینگس دارای مشکلات متعددی نظیر تشخیص همگرایی زنجیر و زمان طولانی محاسبات می‌باشند. اما نمونه‌گیر گیبز این گونه مسائل را به دنباله‌ای از مسائل ساده‌تر تقسیم می‌کند و اجرای آن ساده‌تر است. البته ممکن است این دنباله‌ها زمان زیادی برای همگرایی نیاز داشته باشند ولی با این وجود، این الگوریتم مفید و جالب‌توجه است، زیرا این اجازه را می‌دهد که یک نمونه از توزیع پسین توام تولید کنیم. در ادامه دو الگوریتم متروپولیس-هستینگس و نمونه‌گیر گیبز را به‌طور مختصر شرح می‌دهیم. قبل از تشریح این الگوریتم‌ها، چند تعریف را بیان می‌کنیم.

تعریف ۱.۱.۱. (فرآیند تصادفی). اگر t متغیر شاخص (که معمولاً بیانگر متغیر زمان است) و $X^{(t)}$ متغیر متناسب با t باشد، آن‌گاه یک فرآیند تصادفی، مجموعه‌ای از متغیرهای تصادفی $\{X^{(t)} : t \in T\}$ یک متغیر تصادفی است که در آن $X^{(t)}$ به ازای هر $t \in T$ یک متغیر تصادفی است و T مجموعه اندیس‌گذار فرآیند نامیده می‌شود.

تعریف ۲.۱.۱. (زنجیر مارکوف). فرآیند تصادفی $\{X^{(t)}\}$ را زنجیر مارکوف می‌گوییم، هرگاه

$$\begin{aligned} P(X^{(i+1)} = a | X^{(i)}, X^{(i-1)}, \dots, X^{(0)}) &= P(X^{(i+1)} = a | X^{(i)}) \\ &= K(X^{(i)}, X^{(i+1)}) \end{aligned}$$

که در آن $K(., .)$ را هسته انتقال یا هسته مارکوف می‌نامند بنابراین یک زنجیر مارکوف، دنباله‌ای از متغیرهای تصادفی است که با فرض معلوم بودن $X^{(i)}$ ، که حالت فرآیند در زمان فعلی i ام است، متغیر $X^{(i+1)}$ که حالت بعدی فرآیند است مستقل از حالت‌های گذشته، یعنی $X^{(i-1)}, \dots, X^{(0)}$ می‌باشد. از آن‌جا که نمونه بعدی در هر زنجیر فقط به نمونه قبلی آن وابسته است، این عدم وابستگی به گذشته اجازه می‌دهد تا زنجیرهای مارکوف برای ساده‌سازی مسائل پیچیده مورد استفاده قرار گیرند.

تعریف ۳.۱.۱. (زنجیر مارکوف مانا). زنجیر مارکوف $\{X^{(t)}\}$ یک زنجیر مارکوف مانا است، هرگاه دارای یک توزیع مانای $f(.)$ باشد به طوری که اگر $X^{(t)} \sim f(.)$ آنگاه $X^{(t+1)} \sim f(.)$

الگوریتم متروپولیس - هستینگز

الگوریتم متروپولیس - هستینگز مثالی از مجموعه روش‌های نمونه‌گیری MCMC است. این الگوریتم، با شرط داشتن توزیع پسین توام $\pi(\theta | \mathbf{x})$ ، مجموعه‌ای از متغیرهای تصادفی $\{\theta^{(t)}\}$ ، برای $t = 1, 2, \dots$ ، را به طور پی‌در پی تولید می‌کند. برای این نمونه‌ها، این الگوریتم، ما را ملزم به تعیین چگالی پیشنهادی $p(\theta^{(t+1)}, \theta^{(t)})$ می‌کند. این تابع، تابع چگالی احتمال θ در زمان $t + 1$ ، با توجه به مقدار آن در زمان t است. بنابراین عملکرد الگوریتم متروپولیس - هستینگز به صورت زیر بیان می‌شود:

۱. نمونه $\theta^{(t+1)} \sim p(\theta^{(t+1)}, \theta^{(t)})$ را تولید کن.

۲. قرار بده

$$\theta^t = \begin{cases} \theta^{(t+1)} & \rho(\theta^{(t)}, \theta^{(t+1)}), \\ \theta^{(t)} & 1 - \rho(\theta^{(t)}, \theta^{(t+1)}), \end{cases}$$

به طوری که

$$\rho(\theta^{(t)}, \theta^{(t+1)}) = \min \left\{ 1, \frac{f(\mathbf{x} | \theta^{(t+1)}) \pi(\theta^{(t+1)}) p(\theta^{(t)}, \theta^{(t+1)})}{f(\mathbf{x} | \theta^{(t)}) \pi(\theta^{(t)}) p(\theta^{(t+1)}, \theta^{(t)})} \right\},$$

که به آن احتمال پذیرش متروپولیس - هستینگز می‌گویند. اجرای یک الگوریتم متروپولیس - هستینگز با مشکلاتی مواجه است. در مواردی ممکن است تعیین چگالی پیشنهادی دشوار باشد. یک روش جایگزینی، کاوش در همسایگی مقدار جاری زنجیر می‌باشد. اجرای این ایده، شبیه‌سازی θ در زمان $t + 1$ از رابطه $\theta^{(t+1)} = \theta^{(t)} + \varepsilon$ است، که در آن ε یک متغیر تصادفی مستقل از θ با میانگین صفر است. با شرط متقارن بودن توزیع ε ، و در نتیجه توزیع پیشنهادی، یعنی $p(\theta^{(t+1)}, \theta^{(t)}) = p(\theta^{(t)}, \theta^{(t+1)})$ ، الگوریتم را الگوریتم متروپولیس - هستینگز قدم زدن تصادفی می‌نامند و احتمال پذیرش آن به صورت

$$\rho(\theta^{(t)}, \theta^{(t+1)}) = \min \left\{ 1, \frac{f(\mathbf{x} | \theta^{(t+1)}) \pi(\theta^{(t+1)})}{f(\mathbf{x} | \theta^{(t)}) \pi(\theta^{(t)})} \right\},$$

تعریف خواهد شد. برای آگاهی بیشتر و عمیق‌تر جزئیات این الگوریتم و الگوریتم‌های مشابه به رابرت و کسلا (۲۰۰۴) مراجعه کنید.

نمونه‌گیری گیبز

بیشترین کاربرد نمونه‌گیر گیبز در مدل‌های بیزی، زمانی است که توزیع‌های شرطی کامل پارامترهای مورد علاقه وجود دارند. برای تولید نمونه از توزیع توام $\pi(\theta | \mathbf{x})$ ، که در آن

آوریم. این چگالی شرطی عبارتند از

$$\begin{aligned} \pi(\theta_1 | \mathbf{x}, \theta_2, \theta_3, \dots, \theta_B), \\ \pi(\theta_2 | \mathbf{x}, \theta_1, \theta_3, \dots, \theta_B), \\ \vdots \\ \pi(\theta_B | \mathbf{x}, \theta_1, \theta_2, \dots, \theta_{B-1}). \end{aligned}$$

نمونه‌گیر گیبز با فرض داشتن $\theta^{(t)} = (\theta_1^{(t)}, \dots, \theta_B^{(t)})$ در مرحله t ام، $t = 1, 2, \dots$ ، برای تولید یک زنجیر مارکوف از توزیع پسین توام $\pi(\theta | \mathbf{x})$ ، دنباله‌ای از نمونه‌ها را به صورت زیر تولید می‌کند

$$\begin{aligned} 1 : \theta_1^{(t+1)} &\sim \pi(\theta_1 | \mathbf{x}, \theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_B^{(t)}), \\ 2 : \theta_2^{(t+1)} &\sim \pi(\theta_2 | \mathbf{x}, \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_B^{(t)}), \\ &\vdots \\ B : \theta_B^{(t+1)} &\sim \pi(\theta_B | \mathbf{x}, \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{B-1}^{(t+1)}). \end{aligned}$$

ثابت شده است (رابرت و کسلا ۲۰۰۴) که نمونه تولید شده توسط الگوریتم بالا، نمونه‌ای پسین توام $\pi(\theta | \mathbf{x})$ است.

همگرایی یک زنجیر مارکوف

در کنار استفاده از روش‌های نمونه‌گیری MCMC در استنباط بیزی، دو چالش مهم تشخیص همگرایی و ویژگی‌های آمیختگی^۱ زنجیر و زمان طولانی محاسبات مطرح هستند. برای شناسایی همگرایی زنجیرهای MCMC، معیارهای مختلفی پیشنهاد شده‌اند. اما ابتدا همگرایی یک زنجیر مارکوف را تشریح می‌کنیم.

همان‌طور که قبلاً اشاره شده، در روش‌های انتگرال‌گیری مونت‌کارلویی، مساله انتگرال‌گیری به مساله یافتن راهی برای تولید نمونه $f(\cdot)$ تبدیل می‌شود. در بسیاری از موارد، تولید نمونه مستقیم و مستقل از چگالی هدف $f(\cdot)$ مشکل است. در این موارد می‌توان با استفاده از یک زنجیر مارکوف مانا که توزیع مانای آن $f(\cdot)$ است، به طور غیر مستقیم یک نمونه تولید کرد. اما از آنجایی که نمونه تولید شده از زنجیر مارکوف وابسته است، کیفیت یک نمونه مستقل با حجمی مشابه را ندارد. کیفیت چنین نمونه‌ای را می‌توان از طریق کمیتی به نام حجم نمونه موثر^۲ (ESS) سنجید (گلمن و همکاران ۱۹۹۸، رابرت و کسلا ۲۰۰۴). هر چه مقدار

^۱Mixing Properties

^۲Effective Sample Size

ESS محاسبه شده، برای هر پارامتر به تعداد نمونه‌های تولیدشده نزدیکتر باشد، نمونه‌های تولیدشده از کیفیت بالاتری برخوردار هستند.

تعریف ۴.۱.آ. اگر n تعداد نمونه‌های تولیدشده از زنجیر باشد، آن‌گاه حجم نمونه موثر برای پارامتر مورد نظر به صورت

$$ESS = \frac{n}{1 + 2 \sum_{k=1}^{\infty} \rho_k}$$

محاسبه می‌شود که در آن ρ_k مقدار خودهمبستگی در گام k است. در مواقع ESS کمیتی است که تعداد نمونه‌های مستقل به‌دست آمده از زنجیر را تخمین می‌زند.

در مباحث نظری زنجیرهای مارکوف، اگر زنجیر دارای یک توزیع مانای $f(\cdot)$ باشد، می‌توان نمونه‌ها را از آن تولید کرد. در نتیجه، در یک زنجیر، هسته مارکوف K و توزیع مانای $f(\cdot)$ در رابطه زیر صدق می‌کنند (رابرت و کسلا ۲۰۰۴)

$$\int_{\mathcal{X}} K(x, y) f(x) dx = f(y).$$

وجود یک توزیع مانا، ویژگی نافروکاستنی^۳ را بر هسته K تحمیل می‌کند. یعنی، K ، صرفنظر از مقدار اولیه $X^{(0)}$ ، اجازه حرکت بر روی تمام وضعیت‌های فضای حالت زنجیر را خواهد داشت. از دیگر نتایج وجود یک توزیع مانا در زنجیرهای مارکوف، بازگشتی^۴ بودن زنجیر است. بازگشتی بودن به این معنی است که زنجیر می‌تواند به هر مجموعه دلخواه، به تعداد نامتناهی بازگردد. در یک زنجیر بازگشتی، توزیع مانای یک توزیع حدی^۵ است. یعنی توزیع $X^{(t)}$ ، به ازای هر مقدار اولیه $X^{(0)}$ ، به توزیع مانای $f(\cdot)$ همگرا می‌شود. این ویژگی، ارگودیک بودن نیز نامیده می‌شود. بنابراین، بنابر قضیه ارگودیک، همگرایی به صورت زیر حاصل می‌شود

$$\frac{1}{T} \sum_{t=1}^T h(X^{(t)}) \xrightarrow{P} E_f(h(X)).$$

با توجه به مطالب فوق می‌توان گفت، رهیافت MCMC برای نمونه‌گیری از چگالی هدف $f(\cdot)$ ، یک زنجیر مارکوف $X^{(t)}$ ، با مقدار اولیه $X^{(0)}$ ، را با استفاده از یک هسته مارکوف K با توزیع مانای $f(\cdot)$ تولید می‌کند و سپس برای یک مقدار به اندازه کافی بزرگ T_0 ، نمونه‌های $X^{(T_0)}, X^{(T_0+1)}, \dots$ را تولید می‌کند که از توزیع $f(\cdot)$ گرفته شده‌اند. بنابراین، بررسی همگرایی زنجیر به توزیع مانای $f(\cdot)$ ، یکی از مباحث مهمی است که در روش‌های نمونه‌گیری MCMC مطرح می‌شود.

شبه‌سازی از یک زنجیر MCMC به دو بخش تقسیم می‌شود: پیش از همگرایی زنجیر و بعد از همگرایی. بخش پیش از همگرایی که به دوره داغیدن^۶ مشهور است، به عنوان بخشی

^۳ Irreducibility

^۴ Recurrency

^۵ Limiting Distribution

^۶ Burn-in

از زنجیر است که در آن نمونه‌های تولیدشده را نمی‌توان به عنوان یک معرف خوب از توزیع پسین در نظر گرفت. به عبارت ساده‌تر، نمونه‌های $X^{(0)}, X^{(1)}, \dots, X^{(T_0-1)}$ باید نادیده گرفته شوند و $X^{(T_0)}$ به عنوان تحقق از توزیع $f(\cdot)$ در نظر گرفته می‌شود. پیدا کردن دوره داغیدن و حذف این نمونه‌ها، باعث صرفه‌جویی در زمان (هزینه محاسباتی) خواهد شد. در بخش بعد از همگرایی، نمونه‌های (وابسته) $X^{(T_0)}, X^{(T_0+1)}, \dots$ تولید شده از $f(\cdot)$ برای استنباط در مورد کمیت‌های مورد علاقه استفاده می‌شوند. این امر، نظارت بر همگرایی را با مقایسه استنباط‌های ساخته شده از چند دنباله تولید شده مستقل با نقاط شروع متفاوت، پیشنهاد می‌کند (گلمن و روبین ۱۹۹۲).

حال بحثی که در این جا مطرح می‌شود این است که چگونه تشخیص دهیم یک زنجیر همگراست. زیرا اغنتظار داریم زنجیر تولیدشده در نهایت به توزیع مانایی که مدنظر است، همگرا شود. کاربران MCMC، همگرایی را با پیاده کردن روش‌های تشخیص همگرایی روی خروجی تولیدشده از نمونه‌گیری، بررسی می‌کنند. روش‌های متعددی برای بررسی همگرایی زنجیر MCMC وجود دارند. از جمله این روش‌ها می‌توان به معیارهای همگرایی گلمن و روبین (۱۹۹۲)، جی‌وک (۱۹۹۲)، هیدل برگر و ولچ (۱۹۸۳) اشاره کرد. برای یک زنجیر مشخص، این معیارها توسط توابعی از بسته‌های *lattice* (سرکار ۲۰۰۸) و *coda* (پلامر و همکاران ۲۰۰۶) موجود در نرم‌افزار R قابل محاسبه هستند. ما در اینجا دو مورد از معروف‌ترین معیارها را معرفی می‌کنیم.

۲.آ الگوریتم EM

الگوریتم EM یک نوع الگوریتم تکراری بوده که برای محاسبه برآورد ML یا برآورد ماکزیمم درست‌نمایی محدود شده^۷ REML در حضور داده‌های گمشده یا غیر قابل مشاهده استفاده می‌شود. الگوریتم EM توسط دمپستر و همکاران (۱۹۷۷) در مقاله‌ای پایه‌ای تدوین و منتشر شد. استفاده از الگوریتم EM در این دهه‌های اخیر به شدت افزایش پیدا کرده و ارائه شد. یک مشکل الگوریتم EM، در این است که به کندی همگرا می‌شود اگر چه پیشنهادها برای سرعت بخشیدن به روند همگرایی در این الگوریتم وجود دارد. مینگ و ون‌دیک (۱۹۷۷) استراتژی را برای سرعت بخشیدن به این الگوریتم با حفظ سادگی و ثبات آن پیشنهاد کرده‌اند.

چهار چوب نظری

الگوریتم EM روشی است تکراری که برای پیدا کردن ماکزیمم درست‌نمایی از یک مدل آماری وقتی که وابستگی به متغیرهای نامعلوم قابل مشاهده نیست استفاده می‌شود. چهارچوب‌های

^۷Restricted maximum likelihood

الگوریتم EM در دو مرحله اساسی دنبال می‌شود. مرحله اول مرحله برآورد است. در این مرحله تابع Q را که نشان‌دهنده امید ریاضی لگاریتم درست‌نمایی است تعریف می‌کنیم. این مرحله تابع Q را برآورد پارامتر ارزیابی می‌کند. دومین مرحله، مرحله ماکزیم‌سازی نامیده می‌شود. این مرحله درست‌نمایی را به حداکثر می‌رساند.

چون اثرات تصادفی متغیرها غیرقابل مشاهده هستند فرض کنیم داده گمشده بوده و پاسخ‌های داده مشاهده شده باشند. در این صورت داده‌ها را می‌توان به صورت زیر نشان داد

$$(y, b) = \{(y_i, b_i), \quad i = 1, 2, \dots, m\}$$

k (تعداد تکرارها را نشان می‌دهد و تا زمانی که به همگرایی نرسیدیم فرآیند تکراری ادامه دارد، پس انتخاب اولین مجموعه از مقادیر ابتدایی الگوریتم EM به صورت زیر ادامه می‌یابد (وو، ۲۰۱۰)

- مرحله E: در تکرار k ام این مرحله لگاریتم درست‌نمایی شرطی داده‌های کامل و داده‌های مشاهده شده و برآورد پارامتر محاسبه می‌شود. در این حالت Q به صورت امید ریاضی شرطی تعریف می‌شود

$$Q(\theta|\theta^{(k)}) = E[\log L(\theta|Y, b)|Y, \theta^{(k)}]$$

- مرحله M: این مرحله بیشترین امید ریاضی شرطی $Q(\theta|\theta^{(k)})$ را در مرحله E، با توجه به رابطه پارامت نامعلوم، برای یافتن پارامترهای بروز شده پیدا می‌کند به طوری که داشته باشیم

$$Q(\theta^{(k+1)}|\theta^{(k)}) \geq Q(\theta|\theta^{(k)})$$

این روند بارها و بارها تکرار می‌شود تا همگرایی رخ دهد این نکته مهم را هم باید توجه داشت که الگوریتم EM همگرایی در نقطه ماکزیمم دارد. نقاط شروع مهم هستند و باید نقاط مختلف را امتحان کرد تا به نقطه‌ای برسیم که الگوریتم در آن نقطه به ماکزیمم همگرایی کلی خود نزدیک شود. در ابتدا شاید به نقاط عجیبی برخورد کنیم جای تعجب ندارد چون الگوریتم EM به کندی همگرا می‌شود.

پیوست ب

توزیع‌های آماری

در این قسمت ابتدا توزیع‌های آماری مهم به کار رفته در این مجموعه آورده شده است اما در انتهای آن چند قضیه آورده شده، که ممکن است به عنوان توزیع‌های آماری ربطی نداشته باشد اما از آنها به مراتب استفاده شده و امکان تقسیم آن‌ها و ارائه به صورت یک بخش مجزا وجد نداشت.

ب.۱ توزیع‌های تک متغیره

توزیع بتا

می‌گوییم متغیر تصادفی X دارای توزیع بتا با پارامترهای a و b است و تابع چگالی آن به صورت زیر می‌باشد

$$f(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \quad 0 < x < 1$$

که در آن $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ و a و b پارامترهای آزاد می‌باشند. در حالت از نماد $X \sim B(a, b)$ استفاده می‌کنیم

توزیع نیم کوشی

گوییم متغیر تصادفی X دارای توزیع نیم کوشی با پارامتر b است اگر تابع چگالی آن به صورت زیر باشد

$$f(x, b) = \frac{2}{\pi b} \left(\frac{b^2}{x^2 + b^2} \right)$$

در این حالت از نماد $X \sim C^+(0, b)$ استفاده می‌کنیم

پیشین جفریز

فرض کنید متغیر (بردار) X دارای توزیعی باشد که تابع چگالی آن به صورت $f_\theta(\cdot)$ بوده که در آن $\theta \in \Theta \subseteq \mathbb{R}$ است. جفریز (۱۹۶۱) یک توزیع پیشین ناآگاهی به صورت زیر ارائه کرد

$$\pi(\theta) \propto (I(\theta))^{-\frac{1}{2}}$$

که در آن $I(\theta)$ میزان اطلاع فیشر در نمونه تصادفی است و از رابطه زیر محاسبه می‌شود

$$I(\theta) = E \left(\frac{\partial}{\partial \theta} \ln f_\theta(X) \right)^2 = -E \left(\frac{\partial^2}{\partial \theta^2} \ln f_\theta(X) \right)$$

اگر $\theta \in \Theta \subseteq \mathbb{R}^k$ باشد، پیشین جفریز به صورت زیر تعیین می‌شود

$$\pi(\theta) \propto |I(\theta)|^{-\frac{1}{2}}$$

که در آن $I(\theta)$ ماتریس $k \times k$ اطلاع فیشر است که دارای (i, j) امین مولفه به صورت زیر بوده

$$\begin{aligned} I_{ij}(\theta) &= E \left(\frac{\partial}{\partial \theta_i} \ln f_\theta(X) \frac{\partial}{\partial \theta_j} \ln f_\theta(X) \right) \\ &= -E \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f_\theta(X) \right) \quad i, j = 1, \dots, k \end{aligned}$$

و منظور از $|I(\theta)|$ ، دترمینان ماتریس $I(\theta)$ است.

توزیع نرمال معکوس

می‌گوییم متغیر تصادفی X دارای توزیع نرمال معکوس با پارامترهای μ و λ است اگر تابع چگالی آن به صورت زیر باشد

$$f(x) = \sqrt{\frac{\lambda}{2\pi}} x^{-\frac{3}{2}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2(\mu)^2 x}\right\} \quad x > 0$$

است که در این حالت از نماد $X \sim IG(\mu, \lambda)$ استفاده می‌کنیم.

توزیع گامای معکوس

می‌گوییم متغیر تصادفی X دارای توزیع گامای معکوس با پارامترهای α و β است اگر تابع چگالی آن به صورت زیر باشد توزیع گامای معکوس به صورت زیر است

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right) \quad \alpha > 0, \beta > 0 \quad x > 0$$

که در آن α پارامتر شکل و β پارامتر مقیاس می‌باشد. در این حالت از نماد $X \sim IGamma(\alpha, \beta)$ استفاده می‌کنیم.

توزیع پولیا-گاما

گوییم متغیر تصادفی X دارای توزیع پولیا-گاما با پارامترهای b و c است اگر تابع چگالی توزیع پولیا-گاما به صورت زیر باشد

$$f(x|a, b) = \left\{ \cosh^b\left(\frac{c}{\sqrt{x}}\right) \right\} \frac{\sqrt{x}^{b-1}}{\Gamma(b)} \sum_{n=0}^{\infty} (-1)^n \frac{\Gamma(n+b)}{\Gamma(n+1)} \frac{(\sqrt{x}n+b)}{\sqrt{\pi x^3}} \exp\left\{ -\frac{(\sqrt{x}n+b)^2}{\lambda x} - \frac{c^2}{2} x \right\}$$

که در آن پارامترهای $b > 0$ و $c \in \mathbb{R}$ می‌باشد. در این حالت از نماد $X \sim PG(b, c)$ استفاده می‌کنیم.

چگالی نرمال آمیخته

گوییم متغیر تصادفی Z دارای توزیع آمیخته مقیاسی نرمال با پارامترهای μ و σ^2 است اگر تابع چگالی آن به صورت زیر باشد

$$\pi_Z(z | \mu, \sigma^2) = \int_0^\infty N(z | \mu, g(\lambda)\sigma^2) \pi_\lambda(\lambda) d\lambda,$$

که در آن $N(\cdot | \mu, \sigma^2)$ مشخص کننده توزیع نرمال با میانگین $\mu \in \mathbb{R}$ و واریانس σ^2 ، $g(\lambda)$ نیز یک تابع مثبت از پارامتر آمیخته λ و $\pi_\lambda(\cdot)$ یک تابع چگالی روی \mathbb{R}^+ است که به چگالی آمیخته معروف است.

ب.۲ توزیع آمیخته مقیاسی نرمال چند متغیره

توزیع نرمال n - متغیره

بردار تصادفی $\mathbf{X} = (X_1, \dots, X_n)$ دارای توزیع نرمال n متغیره است با نماد $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ نمایش می‌دهند، هرگاه تابع چگالی احتمال آن به صورت زیر باشد

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T (\boldsymbol{\Sigma})^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad x \in \mathbb{R}^n$$

که در آن بردار میانگین و $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ و ماتریس $n \times n$ متقارن و معین مثبت $\boldsymbol{\Sigma} = (\sigma_{ij})$ با درایه‌های $\sigma_{ij} = Cov(X_i, X_j)$ است

توزیع آمیخته مقیاسی نرمال n - متغیره

گوییم متغیر تصادفی Z دارای توزیع آمیخته مقیاسی نرمال با پارامترهای $\boldsymbol{\mu}$ و σ^2 است اگر تابع چگالی آن به صورت زیر باشد

$$\pi_Z(z | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_0^\infty N_n(z | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \pi_\lambda(\lambda) d\lambda,$$

که در آن $N(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ مشخص کننده توزیع نرمال با میانگین $\boldsymbol{\mu} \in \mathbb{R}^n$ و واریانس - کوواریانس $\boldsymbol{\Sigma}$ نیز یک ماتریس معین مثبت است و پارامتر آمیخته λ و $\pi_\lambda(\cdot)$ یک تابع چگالی روی \mathbb{R}^+ است که به چگالی آمیخته معروف است.

ب.۳ قضایای حدی و احتمالاتی

قضیه ب.۳.۱. اگر \mathbf{A} یک ماتریس $n \times n$ متقارن و خودتوان با رتبه r باشد، آن‌گاه

$$\text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A}) = r$$

تعریف ب.۳.۱. (تجزیه طیفی) ماتریس \mathbf{A} را می‌توان به مقادیر و بردارهای ویژه‌اش به صورت زیر تجزیه کرد

$$\mathbf{A} = \boldsymbol{\Gamma} \boldsymbol{\Lambda} \boldsymbol{\Gamma}^T$$

که در آن $\boldsymbol{\Lambda}$ یک ماتریس قطری با مقادیر ویژه ماتریس \mathbf{A} و $\boldsymbol{\Gamma}$ ماتریس متعامدی است که ستون‌های آن بردارهای ویژه متناظر با مقادیر ویژه \mathbf{A} می‌باشد.

تعریف ۲.۳.ب. (معیار اطلاع انحراف). معیار اطلاع انحراف^۱ (DIC) در مسائل مربوط به انتخاب مدل بیزی توزیع پسین توسط شبیه‌سازی MCMC به دست آمده‌اند، مناسب می‌باشد. این معیار به عنوان تعمیمی از AIC برای انتخاب مدل در چهارچوب بیزی در نظر گرفته می‌شود.

اگر انحراف را به صورت $D(\theta) = -2 \ln(L(y | \theta))$ و تعداد پارامترهای موثر مدل به صورت $pD = \bar{D} - D(\bar{\theta})$ تعریف می‌کنیم، که در آن $\bar{D} = E(D(\theta))$ ، $D(\bar{\theta})$ و $\bar{\theta} = E(\theta | y)$ باشند، آن‌گاه DIC برابر است با

$$\begin{aligned} DIC &= pD + D \\ &= D(\bar{\theta}) + pD. \end{aligned}$$

در بین چندین مدل بیزی نامزد، مدلی که کمترین DIC را داشته باشد، انتخاب می‌شود.

قضیه ۲.۳.ب. (قانون ضعیف اعداد بزرگ) اگر Y_1, \dots, Y_n متغیرهای تصادفی مستقل و هم‌توزیع با میانگین $\mu = E(Y)$ به طوری که $E(|Y|) < \infty$ باشد در این صورت

$$\bar{Y}_n \xrightarrow{p} \mu.$$

قضیه ۳.۳.ب. (قانون قوی اعداد بزرگ) اگر Y_1, \dots, Y_n متغیرهای مستقل و هم‌توزیع با میانگین $\mu = E(Y)$ به طوری که $E(|Y|) < \infty$ باشد در این صورت

$$\bar{Y}_n \xrightarrow{a.s.} \mu.$$

قضیه ۴.۳.ب. (قضیه حد مرکزی) اگر Y_1, \dots, Y_n متغیرهای مستقل و هم‌توزیع با میانگین $\mu = E(Y)$ و واریانس σ^2 (به طوری که $E(Y^2) < \infty$) در این صورت

$$\sqrt{n}(\bar{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

قضیه ۵.۳.ب. (استلاسی) فرض کنید $A_n \xrightarrow{p} a$ و $B_n \xrightarrow{p} b$ ، برای ثابت‌های a و b و داریم $Y_n \xrightarrow{d} Y$

$$A_n Y_n + B_n \xrightarrow{d} aY + b.$$

^۱Deviance Information Criterion

پیوست پ

گزیده‌ای از برنامه R

شکل ۱.۲

```
rm(list=ls())
library(pls)
library(lars)
library(MASS)
library(monomvn)
data(diabetes)
diabetes
attach(diabetes)
dim(diabetes)
x
y
reg.ols <- regress(x, y)
reg.ols
reg.las <- regress(x, y, method="lasso")
reg.las
reg.blas <- blasso(x, y)
```



```
reg.blas
plot(reg.blas, burnin=200)
points(drop(reg.las$b), col=2, pch=20)
points(drop(reg.ols$b), col=3, pch=18)
legend("topleft", c("blasso-map", "lasso", "lsr"),
+col=c(2,2,3), pch=c(21,20,18))
plot(reg.blas, burnin=200, which="m")

rm(list=ls())
theta<-seq(-3,3,0.1)
theta
double.exponential<-function(theta){
d<-exp(-abs(theta))*1/(2)
return(d)
}
double.exponential(theta)

Cauchy<-function(theta){
c<- 1/(pi*(1+theta^2))
return(c)
}
Cauchy(theta)

horseshoe<-function(theta){
K<-(1/(2*pi^3))^1/2
xx<-(K/2)*log(1+(1/theta^2))
ss<-K*log(1+(2/theta^2))
return(ss)
return(xx)
}
horseshoe(theta)

plot(theta,double.exponential(theta),ylab="density",type="l",col="blue")
lines(theta,Cauchy(theta),col="red")
```

```
lines(theta,horseshoe(theta),ylab="density",type="l",col="black")
```

شکل ۱.۳

```
rm(list=ls())
```

```
k<-seq(0,1,0.01)
```

```
k
```

```
Double.exponential<-function(kappa){
```

```
a<-k^(-2)
```

```
b<-exp(-1/(2*k))
```

```
c<-a*b
```

```
return(c)
```

```
}
```

```
Double.exponential(k)
```

```
Cauchy<-function(kappa){
```

```
aa<-k^(-1/2)
```

```
bb<-(1-k)^(-3/2)
```

```
dd<-2/(1-k)
```

```
cc<-exp(-k/dd)
```

```
d<-aa*bb*cc
```

```
return(d)
```

```
}
```

```
Cauchy(k)
```

```
Strawderman.Berger<-function(ka){
```

```
f<-k^(-1/2)
```

```
return(f)
```

```
}
```

```
Strawderman.Berger(k)
```

```
Normal.exponential.gamma<-function(k){
```

```
t<-4
```

```
g<-k^(t-1)
```

```
return(g)
```

```
}
```

```
Normal.exponential.gamma(k)
```

```
#plot(k,Normal.exponential.gamma(k),main="Normal.exponential.gamma(c=4,d=1)",type="l",col="b
```

```
Normal.exponential.gamma<-function(k){
t<-1/4
g<-k^(t-1)
return(g)
}
Normal.exponential.gamma(k)
Normal.Jeffreys<-function(k){
h<-k^(-1)
i<-(1-k)^(-1)
oo<-h*i
return(oo)
}
Normal.Jeffreys(k)
Horseshoe<-function(k){
m<-k^(-1/2)
n<-(1-k)^(-1/2)
o<-m*n
return(o)
}
Horseshoe(k)
#plot(k,Horseshoe(k),main="Horseshoe",type="l",col="black")

par(mfrow=c(3,3))
plot(k,Double.exponential(k),main="Double.exponential",type="l",col="black")
plot(k,Cauchy(k),main="Cauchy",type="l",col="black")
plot(k,Strawderman.Berger(k),main="Strawderman.Berger",type="l",col="black")
plot(k,Normal.exponential.gamma(k),main="Normal.exponential.gamma(c=4,d=1)",
type="l",col="black")
plot(k,Normal.exponential.gamma(k),main="Normal.exponential.gamma(c=1/4,d=1)",
type="l",col="black")
plot(k,Normal.Jeffreys(k),main="Normal.Jeffreys",type="l",col="black")
plot(k,Horseshoe(k),main="Horseshoe",type="l",col="black")
```

```
rm(list=ls())
library(horseshoe)
X1 <- rnorm(600)
X=matrix(X1,ncol=30)
beta=matrix(c(rep(2,5),rep(0,25)),nr=30)
e1=rnorm(20)
e=matrix(e1,nr=20)
y<- X %*% (beta)+(e)
res <- horseshoe(y, X, method.tau = "halfCauchy", method.sigma = "Jeffreys")
boxplot(res)
```

جدول ١.٤

```
clear;
load data/pima.mat
[beta, beta0, retval] = bayesreg(X,y,'binomial','lasso', 'displayor', true, ...
'nsamples',1e4,'burnin',1e4,'thin',5,'varnames',varnames);
```

Abstract

Some regression models have a large number of explanatory variables, where, it is suggested to use penalized methods for estimation. Among these methods, the LASSO simultaneously selects the variables and estimates the parameters of the regression model. Now, from Bayesian view point, we can use Bayesian penalized methods. In this modeling, the prior distribution is usually chosen in such a way that the resultant is similar to one of the penalty estimators in classical models. In this dissertation, we study the Bayesian penalized regression and examine Bayesian LASSO and Bayesian horseshoe approaches in more details. For performance evaluation of the Bayesian penalized regression, we analyze two diabetes data sets.

key words: Bayesian penalized regression, Bayesian LASSO, Bayesian horseshoe , diabetes data , Normal scale mixture distribution , Half- Cauchy distribution.



Shahrood University of Technology

Faculty of Mathematical Sciences

MSc Thesis in: Mathematical statistics

Bayesian penalized regression

By: Seyed Ali Asghar Rahmati

Supervisor

Dr. Mohammad Arashi

Janvay 2019