دانشکده ریاضی

گروه ریاضی کاربردی

پایان‌نامه کارشناسی ارشد

سری‌های زمانی با تغییرات غیرنرمال و کاربردهای آن

سحر رفیعی دهبنه

استاد راهنما:

دکتر احمد نزاکتی رضازاده

اساتید مشاور:

دکتر حسین باغیشنی

دکتر محمدعلی مولایی

خرداد ۱۳۹۲

دانشگاه صنعتی شاهرود
دانشکده ریاضی

گروه ریاضی کاربردی

پایان‌نامه کارشناسی ارشد

سری‌های زمانی با تغییرات غیرنرمال و کاربردهای آن

سحر رفیعی دهبنه

استاد راهنما:

دکتر احمد نزاکتی رضازاده

اساتید مشاور:

دکتر حسین باغیشنی

دکتر محمد علی مولایی

پایان‌نامه ارشد جهت اخذ درجه کارشناسی ارشد

خرداد ۱۳۹۲

پروردگارا:

حکمت قدم هایی را که برایم برمی داری بر من آشکار کن،

تا دره‌ایی را که به سویم می گشایی، ندانسته نبندم

و درهایی که به رویم می بندی، به اصرار نگشایم .

پروردگارا:

نه میتوانم موهایشان را که در راه عزت من سفید شد، سیاه کنم و نه برای دستهای پینه بسته شان که ثمره تلاش برای افتخار من است، مرهمی دارم . پس توفیقم ده که هر لحظه شکر گزارشان باشم و ثانیه های عمرم را در عصای دست بودنشان بگذرانم .

تقدیم به پدر و مادر عزیزم .

خودایا:

حکمت قدمائی که مرا وسانی، روشن کون

تا ندسته، درائی که مر واز کونی، دنودم

و درائی که دودی، اصرار امرا واز نوکونم

خودایا:

نه تنم اوشن موکه در راه می عزت سفید بوس، سیاه بوکونم و نه اوشن دست که در راه می افتخار پینه دوست ، مرهم بنم . پس مر توفیق فده هر لحظه اوشن سگرگذار بم و ثانیه های می عمر اوشن دست عصا بم .

تقدیم کونم می پرومار عزیز

تقدیر و تشکر :

حال که به یاری پروردگار ، این دوره از تحصیلات خود را به پایان رساندم ، بر خود لازم می‌دانم از زحمات بی‌شائبه ی پدر و مادر عزیز و برادر و زن داداش نازنینم که همواره مشوق و پشتیبان من بوده‌اند و ادامه این راه را برایم هموار نمودند، تشکر کنم .

از اساتید گرامی جناب آقای دکتر احمد نراکی رضازاده که زحمت راهنمایی این پایان نامه را بر عهده داشتند و نیز جناب آقای دکتر حسین باغیشنی به عنوان استاد مشاور سپاسگذاری می نمایم .

در پایان از اعضای کمیته داوران جناب آقای دکتر محمد آرشی و آقای دکتر امید کریمی و همچنین جناب آقای سید رضا موسوی صمیمانه تشکر می کنم .

همچنین از تمامی دوستان و همه کسانی که مراد انجام این پایان نامه یاری کردند صمیمانه قدردانی می کنم و بهترین سرنوشت ها را برای آنان آرزو دارم .

سحر رفیعی دیبنه ۱۳۹۲

پیشگفتار :

رگرسیون چندکی توصیف کاملی از وابستگی توزیع شرطی Y به متغیر تبیینی X ارایه می‌دهد، در حالی‌که رگرسیون معمولی فقط وابستگی میانگین شرطی متغیر پاسخ Y به متغیر تبیینی X بررسی می‌کند. برآورد پارامترها در رگرسیون چندکی بر اساس تابع زیان نامتقارن برای جمله خطا است و مشابه برآورد پارامترها در رگرسیون کمترین توان‌های دوم خطا محاسبه می‌شود. ایده رگرسیون چندکی‌بیزی با استفاده از یک تابع درستنمایی بر اساس توزیع لاپلاس نامتقارن قرار داده شده است. مدل‌های رگرسیون چندکی دودویی و توبیت می‌توانند به عنوان رگرسیون چندکی خطی با پاسخ‌های پیوسته پنهان که به‌طور کامل مشاهده نشده‌اند، در نظر گرفته می‌شود. در حالت معمول سری‌زمانی را با جمله خطا نرمال در نظر گرفته شده است. در این پایان‌نامه از فرض غیرنرمال بودن جمله خطا استفاده نموده‌ایم و آن‌را به صورت نیمه‌پارامتری با تابع چندک نمایی در نظر گرفته‌ایم و سری‌زمانی با تغییرات غیر نرمال، در نظر گرفته شده است. روش مطرح شده با استفاده از یک سری داده‌های شبیه‌سازی و دو مجموعه واقعی شرح داده می‌شود.

واژگان کلیدی: رگرسیون چندک، روش بیزی، توزیع لاپلاس نامتقارن، پیشین ناسره، پسین سره، روش *MCMC*،

فهرست مطالب

فصل اول: مفاهیم پایه‌ای

فصل دوم: رگرسیون بیزی

فصل سوم: رگرسیون چندکی بیزی

فصل چهارم: رگرسیون چندکی دودویی و تویت

فصل پنجم: سری زمانی غیر نرمال

پیوست

فهرست شکل‌ها

فصل اول

مفاهیم پایه‌ای

۱,۱ مقدمه

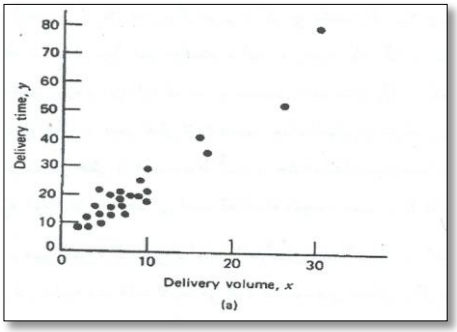
در این فصل مقدمه‌ای از رگرسیون و به‌خصوص رگرسیون چندکی و بعضی کاربردهای آن‌را بیان می‌کنیم.

دراین راستا مدل‌های رگرسیونی را برای حالت‌های پارامتری، نیمه‌پارامتری و ناپارامتری در نظر می‌گیریم.

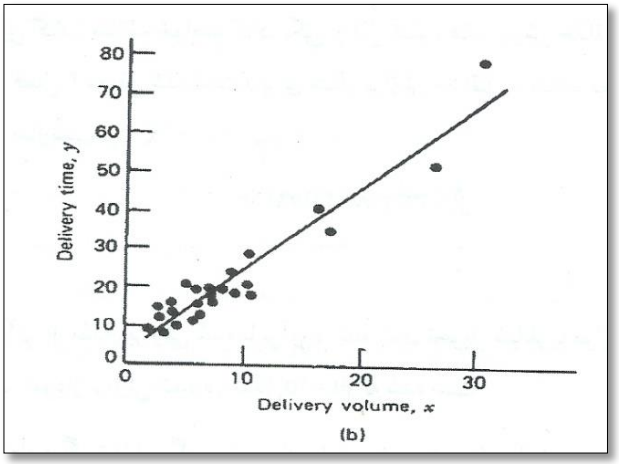
۲,۱ مدل‌های رگرسیونی

تحلیل رگرسیونی، روشی آماری برای بررسی و مدل‌بندی ارتباط بین متغیرها است. مدل‌های رگرسیونی در زمینه‌های متعددی از جمله مهندسی، فیزیک، اقتصاد، مدیریت، علوم زیستی و علوم اجتماعی کاربرد دارند. در حقیقت می‌توان گفت این مدل‌ها، یکی از پرکاربردترین مدل‌های آماری محسوب می‌شوند.

به‌عنوان مثالی برای درک تحلیل رگرسیونی، فرض کنید یک آماردان توسط یک سازنده بطری‌های نوشابه به منظور تحلیل و بررسی رابطه بین حجم محصول تحویلی و عملیات خدمت‌رسانی، برای ماشین-های فروش استخدام شده است. وی گمان می‌کند که زمان لازم برای تحویل محصولات به تعداد محصولات تحویل‌شده وابسته است. آماردان ۲۵ فروشنده جزئی را به تصادف بازدید نموده و زمان تحویل (برحسب دقیقه) و حجم محصول تحویل داده شده (برحسب مورد) را برای هر یک ثبت می‌نماید. پراکنش این مشاهدات در شکل ۱,۱ نمایش داده شده است. نمودار پراکنش ۱,۱ یک ارتباط خطی بین زمان تحویل و حجم تحویل داده شده را پیشنهاد می‌کند. شکل ۲,۱ این ارتباط خطی را نمایش می‌دهد.



شکل ۱,۱ نمودار پراکنش حجم و زمان تحویل محصولات



شکل ۲,۱ رابطه خطی حجم و زمان تحویل محصولات

اگر y و x به‌ترتیب نشان‌دهنده زمان تحویل و حجم تحویل داده شده باشند، معادله خط رگرسیونی به صورت زیر است:

$$y = \beta_0 + \beta_1 x, \qquad (1.1)$$

که در آن β_0 عرض از مبداء و β_1 شیب خط است. دقت کنید که داده‌ها دقیقاً روی خط قرار نمی‌گیرند. بنابراین رابطه (۱,۱) بایستی تعدیل شود. اختلاف بین مقدار مشاهده شده،y، و خط،$\beta_0 + \beta_1 x$، را خطای ε قرار دهند، یعنی ε متغیری است تصادفی که عدم برازش را بیان و ناتوانی مدل در برازش دقیق داده‌ها را اندازه‌گیری می‌کند. این خطا ممکن است ناشی از در نظر نگرفتن سایر متغیرهای مهم، خطاهای اندازه‌گیری و سایر عوامل روی زمان تحویل باشند. آنگاه مدل منطقی‌تر برای داده‌های زمان تحویل به صورت:

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (2.1)$$

خواهد بود. معادله (۲,۱)، یک مدل رگرسیون خطی نامیده می‌شود. x متغیر مستقل[۱] و y متغیر وابسته[۲] نامیده می‌شوند. این نام‌گذاری باعث می‌شود که این مفهوم با مفهوم استقلال آماری اشتباه گرفته شود.

[۱] *Independent variable*

[۲] *Dependent variable*

بنابراین معمولاً x را متغیر پیشگو[۱] یا متغیر تبیینی[۲] و y را متغیر پاسخ[۳] می‌نامند. چون در رابطه (۲,۱) فقط یک متغیر تبیینی وجود دارد، این مدل را مدل رگرسیون خطی ساده می‌نامند. در حالت کلی ممکن است متغیر پاسخ با k متغیر تبیینی $x_1, x_2, \dots, x_k$ ارتباط داشته باشد. بنابراین مدل (۲,۱) را می‌توان به صورت

$$y = \beta_\circ + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (3.1)$$

تعمیم داد. مدل (۳,۱) را رگرسیون خطی چندگانه می‌نامند.

یکی از اهداف مهم تحلیل رگرسیونی، برآورد پارامترهای مدل است. برآورد پارامتر و تعیین صورت مدل ریاضی نامیده می‌شود. یکی از روش‌های برآورد پارامترها، روش کمترین توان‌های دوم خطا[۴] است. برای مثال، برازش کمترین توان‌های دوم خطا به داده‌های زمان تحویل به صورت $\hat{y} = 3/3208 + 2/1762x$ است که در آن، $\hat{y}$ مقدار برازش‌شده یا برآوردشده زمان تحویل متناظر با حجم تحویلی x است. مدل برازش‌شده در شکل ۲,۱ نمایش داده شده است.

مدل رگرسیونی یک رابطه علت و معلولی[۵] را بین متغیرها نتیجه نمی‌دهد. چه بسا یک رابطه تجربی قوی بین دو یا چند متغیر وجود داشته باشد، اگر این رابطه نتواند بیانگر یک رابطه علت و معلولی بین متغیرهای تبیینی و متغیر پاسخ باشد، آنگاه برای داشتن یک ارتباط علت و معلولی بین متغیرهای تبیینی با متغیر پاسخ، بایستی پایه و مبنایی خارج از داده‌های نمونه‌ای داشته باشیم. به طور کلی تحلیل رگرسیونی می‌تواند در تصدیق یک رابطه علت و معلولی کمک کند، اما نمی‌تواند تنها پایه و مبنای چنین ادعایی باشد.

---

۱ *Predictor variable*

۲ *Regressor variable*

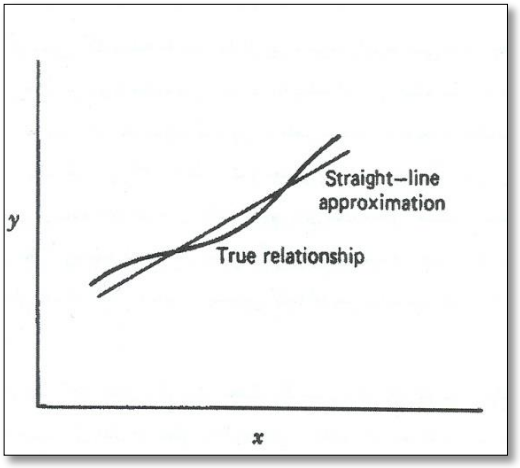
۳ *Response variable*

۴ *Least Sum of Square Errors*

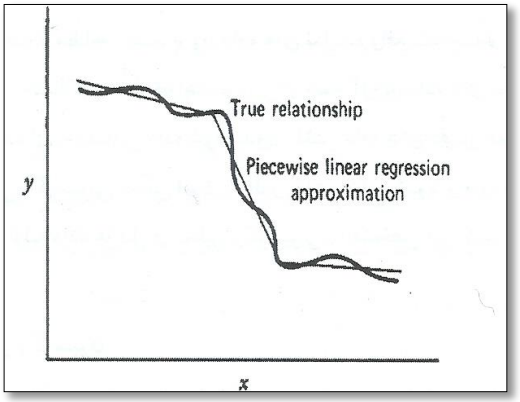
۵ رابطه‌ای که فقط برای آزمایش‌های تصادفی معتبر هست نه براساس مطالعه مشاهده‌ای

در اکثر کاربردهای رگرسیون خطی، معادله رگرسیون فقط یک تقریب از رابطه درست بین متغیرها است. برای مثال شکل ۳,۱ وضعیتی را نشان می‌دهد که در آن یک ارتباط نسبتاً پیچیده به وسیله یک معادله رگرسیون خطی تقریب زده شده است. گاهی یک تقریب پیچیده‌تر لازم است. به‌عنوان مثال در شکل ۴,۱ یک تابع رگرسیونی خطی تکه‌ای[۱] برای تقریب رابطه واقعی بین y و x مورد استفاده شده است.

به این نکته نیز توجه داشته باشید که، در حالت کلی، مدل‌های رگرسیونی تنها در ناحیه‌ای از متغیرهای تبیینی که شامل داده‌های مشاهده شده هستند، معتبرند.



شکل ۳,۱ تقریب خطی از یک ارتباط پیچیده



شکل ۴,۱ تقریب خطی تکه‌ای از یک ارتباط پیچیده

[۱] *Piecewise linear regression*

۳,۱ مفاهیم اساسی رگرسیون چندکی

رگرسیون چندکی، مدل آماری کامل‌تری از رگرسیون معمولی است و در حال حاضر کاربردهای وسیعی دارد. در این بخش ابتدا رگرسیون چندکی را معرفی و برخی از کاربردهای آن را مطرح می‌کنیم.

۱,۳,۱ معرفی چندک

فرض کنید متغیر تصادفی X دارای تابع توزیع $F_X(.)$ باشد، پارامتر Q_p را چندک مرتبه p برای $F(X)$ یا X می‌نامیم، هرگاه نامساوی دوطرفه زیر برقرار باشد:

$$P(X < Q_p) \le p \le P(X \le Q_p). \qquad (4.1)$$

معنی این نامساوی دوطرفه آن است که مقدار احتمال در فاصله $(-\infty, Q_p)$ حداکثر p و در فاصله $[-\infty, Q_p)$ حداقل p می‌باشد. برحسب اینکه $100p$، $10p$ یا $4p$ عدد صحیح باشد، Q_p را به ترتیب صدک، دهک یا چارک[۱] می‌نامند. مثلاً $Q_{0/23}$ را صدک بیست و سوم و $Q_{0/4}$ را دهک چهارم می‌نامند.

البته سه مقدار $Q_{0/75}, Q_{0/5}, Q_{0/25}$، به ترتیب چارک اول (یا چارک پایینی) چارک دوم (یا میانه) و چارک سوم (یا چارک بالایی) گفته می‌شود. اگر $F_X(.)$ پیوسته باشد، یعنی نمودار آن دارای خطوط افقی یا جهشی نباشد، آنگاه نامساوی (۴,۱) به تساوی $F(Q_p) = p$ تبدیل می‌شود و Q_p پاسخ یکتای معادله $F(Q_p) = \int_{-\infty}^{Q_p} f(x)dx = p$ خواهد بود.

مثال ۱: فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{سایر حالات} \end{cases}$$

آنگاه برای چارک اول Q_1 داریم:

$$F(Q_1) = \int_0^{Q_1} 2x dx = x^2|_0^{Q_1} = Q_1^2 = \frac{1}{4}.$$

[۱] *Quartile*

بنابراین با تکیه‌گاه $(0,1)$، $Q_1 = \frac{1}{2}$ است.■

برای هر متغیر تصادفی حقیقی مقدار X، که دارای تابع توزیع $F(x) = P(X \leq x)$ و برای‌هر>0 $p < 1$، داریم:

$$F^{-1}(p) = inf\{x: F(x) \geq p\}$$

که در آن، $F^{-1}(p)$ ، p امین چندک X می‌باشد.

توابع چندک را برای جمله خطا استفاده می‌کنیم که در فصل۳ مبنا جمله خطا بر روی تابع زیان با توزیع لاپلاس نامتقارن است به همین دلیل در ابتدا به معرفی تابع زیان می‌پردازیم.

تابع زیان تکه‌ای- خطی به صورت

$$\rho_p(u) = (u)\big(p - I(u<0)\big), \qquad (5.1)$$

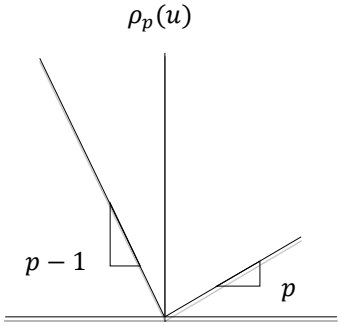
را در نظر بگیریم که در آن،

$$I_A(u) = \begin{cases} 1 & u \in A \\ 0 & u \notin A \end{cases}, \quad 0 \leq p \leq 1$$

نمودار تابع زیان تکه‌ای- خطی را در شکل۵,۱ می‌توانید ببینید.

آنگاه برای $p \in (0,1)$، $\hat{x}$ با می‌نیمم کردن امید ریاضی تابع زیان به‌دست می‌آید. عبارت زیر برآوردگر چندک تحت تابع زیان(۵,۱) می‌باشد

$$E\rho_p(X - \hat{x}) = (p-1)\int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + p\int_{\hat{x}}^{\infty} (x - \hat{x})dF(x). \qquad (6.1)$$



شکل ۵,۱ تابع زیان تکه‌ای- خطی

با مشتق‌گیری از رابطه (۶,۱) نسبت به $\hat{x}$ و متحد کردن آن با صفر، خواهیم داشت:

$$0 = (1-p)\int_{-\infty}^{\hat{x}} dF(x) - p\int_{\hat{x}}^{\infty} dF(x) = F(\hat{x}) - p.$$

چون F یکنواست، پس هر عضوی از $\{x: F(x) = p\}$ می‌نیمم کننده (۱,۹) است. جواب، وقتی یکتاست که $\hat{x} = F^{-1}(p)$. در غیر اینصورت p امین چندک شامل یک بازه است (کانکر[۱]،۲۰۰۵).

اگر تابع توزیع تجربی $F_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \le x)$ را با عنوان برآورد$F(x)$ در نظر بگیریم، آن‌گاه

$$\int \rho_p(x-\hat{x})dF_n(x) = n^{-1}\sum_{i=1}^{n} \rho_p(x_i - \hat{x}). \qquad (7.1)$$

می‌نیمم کننده (*7.1*)، p امین چندک نمونه است.

۲,۳,۱ ویژگی‌های تابع چندک

توابع چندک دارای ویژگی‌های بسیاری است که در زیر به چند مورد اشاره می‌نماییم ولی در این پایان نامه چون مبنا بر روی یک تابع چندک است از ویژگی‌های زیر استفاده نمی‌کنیم:

۱. مجموع توابع چندک، یک تابع چندک است.

۲. حاصل‌ضرب دو تابع چندک مثبت، تابع چندک است.

۳. اگر $Q_1(p)$ و $Q_2(p)$ دو تابع چندک باشند، آنگاه

$$Q(p) = wQ_1(p) + (1-w)Q_2(p), \quad 0 < w < 1$$

$Q(p)$ بین دو تابع چندک قرار می‌گیرد.

۴. اگر $Q(p)$ تابع چندک باشد، آنگاه $\lambda + \eta Q(p)$ نیز تابع چندک است.

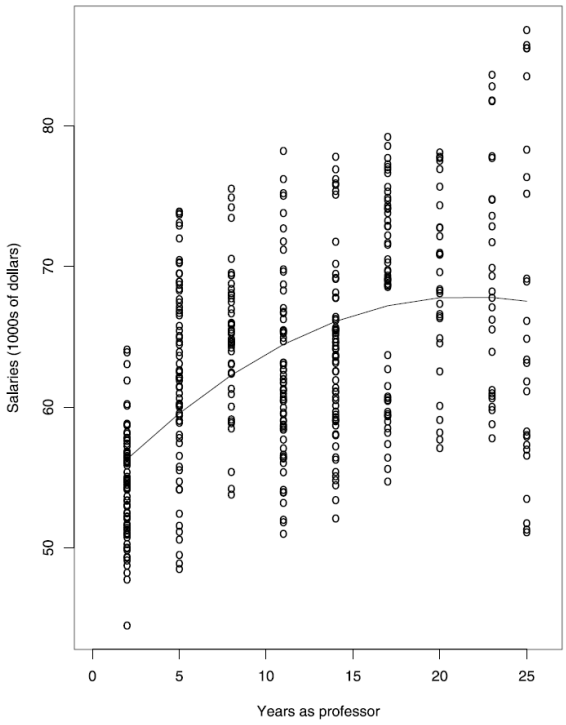
[۱] *Koenker*

۴,۱ از رگرسیون معمولی تا رگرسیون چندکی

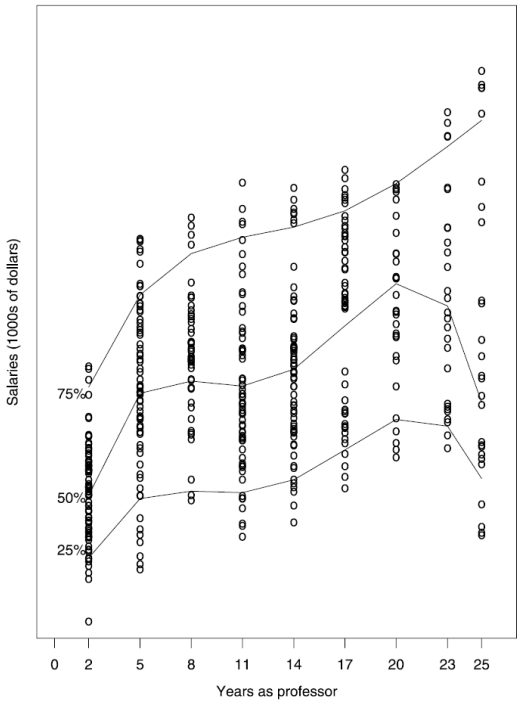
مدل رگرسیون معمولی، ارتباط بین متغیر پاسخ و بعضی متغیرهای تبیینی را مشخص می‌کند. این مدل، یکی از روش‌های آماری است که برای چندین دهه در تحقیقات کاربردی بکار گرفته می‌شود. برای مثال، چند سال پیش، دانشگاهی در آمریکا می‌خواست ارتباط بین دستمزد اساتید و تعداد سال‌های تدریس را به‌دست آورد. در این تحقیق، دانشگاه میزان حقوق ۴۵۹ استاد آمار و تعداد سال‌های تدریس را در طول مدت سال‌های ۱۹۸۰ تا ۱۹۹۰جمع آوری کرد. مدل رگرسیون خطی ساده بود

$$y = \boldsymbol{X}^T\boldsymbol{\beta} + \varepsilon,$$

که در آن، $\boldsymbol{X} = (1, x)^T$، $\boldsymbol{\beta} = (\beta_\circ, \beta_1)^T$ ،y دستمزد، x تعداد سال‌های تدریس و ε جمله خطا با پذیره نرمال بود. مدل‌های پیچیده‌تری، مانند مدل‌های رگرسیون چندجمله‌ای، نیز ممکن است برای این ارتباط بکار برده شوند. شکل۶,۱ پراکنش داده‌ها همراه با بهترین برازش رگرسیون درجه دوم را نشان می‌دهد. متاسفانه، منحنی شکل۶,۱ تصویر نامناسبی از توزیع دستمزد را نشان می‌دهد، به این معنی که تغییر در شکل توزیع دستمزد با سال‌های تدریس به صورت مناسب نشان داده نشده است. دلیل این امر آن است که مدل‌های برازش رگرسیون معمولی فقط ارتباط بین متوسط دستمزد و سال‌های تدریس را نشان می‌دهد. برای ارایه تصویر کامل‌تری از ارتباط بین دستمزد و سال‌های تدریس، در شکل ۷,۱ چندک‌های نمونه‌ای ۰/25 و ۰/5 و ۰/75 نمایش داده شده‌اند. منحنی‌های حاصل، منحنی‌های رگرسیون چندکی نامیده می‌شوند. به هر حال تغییر در شکل توزیع دستمزد، به وضوح در شکل ۷,۱ نمایش داده شده است. یک تقریب از توزیع احتمالی دستمزد کامل را می‌توان از خمیدگی رگرسیون چندک متناظر با مقدار برد p تولید کرد.



شکل ۶,۱: نمودار رگرسیون درجه دوم دستمزد و تعداد سالهای تدریس



شکل۷,۱: نمودار رگرسیون چندکی دستمزد و تعداد سال‌های تدریس برای سه مرتبه **۰/25** و **۰/5** و **۰/75**

۵,۱ از توزیع‌های چوله شرطی تا رگرسیون چندکی

نمودار (الف) شکل ۸,۱ پراکنش وزن در مقابل سن را برای یک نمونه ۴۰۱۱ تـایی از دختـران آمریکـایی (کول[۱]، ۱۹۸۸) نشان می‌دهد.

نمودار (ب) شکل۸,۱، تایید دیگری از ارتباط بین سن و وزن است، کـه چنـدین منحنـی رگرسـیون چندکی هموار با ۰/97, ۰/9, ۰/75, ۰/5, ۰/25, ۰/1, ۰/03 $p =$ را نمایش می‌دهـد. ملاحظـه مـی-کنید که توزیع‌های شرطی مربوط، دارای چولگی مثبت هستند. دو سوال جالب ممکن است پیش آید:

۱. وضعیت وزن به عنوان تابعی از سن چگونه است؟

۲. وضعیت وزن برحسب سن برای افراد سنگین وزن یا سبک وزن چگونه است؟

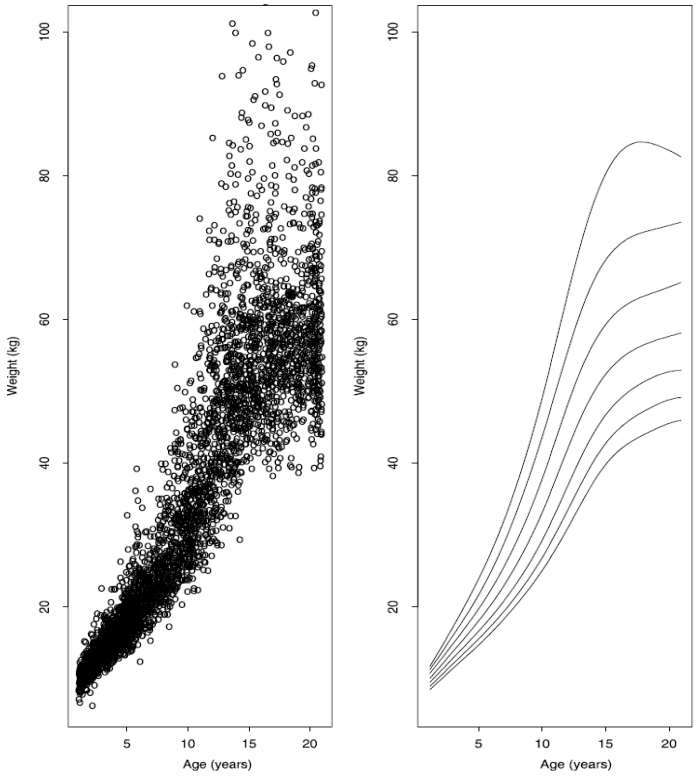
استفاده از رگرسیون، میانگین، پاسخ مناسبی برای سوال اول فراهم نمی‌آورد. از این‌رو، منحنی میانه برای متغیرهای تبیینی و پاسخ همیشه روشن نیست و گاهی وابسته به هدف تحلیل‌گـر اسـت. منحنـی میانـه، متناظر با منحنی رگرسیون چندکی میانی است که در شکل ۸,۱ (ب) نشان داده شده است. دخترانی کـه وزن آن‌ها رو یا بالای منحنی ۹۷ درصدی جامعه هستند، سـنگین وزن هسـتند. در ایـن‌صـورت منحنـی مناسب نمایش آن‌ها رگرسیون چندکی ۰/97= p است. به‌طور مشابه، منحنی رگرسیون چندکی $p = 0$، ۰/03 نشان‌دهنده ارتباط وزنی دخترهای سبک‌وزن با سن است. در بخش ۷,۱ نحوه برآورد ایـن منحنـی-های هموار را مطرح می‌کنیم.

در رگرسیون معمولاً فرض بر این است که متغیرهای تبیینی شامل تغییرات تصادفی نیسـتند، ولـی متغیرهای پاسخ تصادفی هستند. از نقطه نظر کاربردی، این مطلب اغلب صحیح نیست، اما به‌دلیـل اینکـه اگر متغیرهای تبیینی تصادفی باشند، شیوه برازش بسیار پیچیده می‌شود، برای اجتناب از این پیچیـدگی

[۱] *Cole*

معمولاً فرض می‌شود که تغییرات تصادفی هر یک از متغیرهای تبیینی در مقایسه با دامنه مشاهده شده، آن قدر کوچک است که می‌توان از آن صرفنظر کرد.

باید توجه داشت که در رگرسیون، رابطه بین متغیرهای تبیینی و پاسخ یک رابطه تابعی ریاضی نیست، چون اگر از x، مقدار y بدست آید از y مقدار x بدست نمی‌آید. نوع رابطه‌ای که ممکن است بین متغیرهای تبیینی و پاسخ برقرار باشد، متنوع است. اما معمولاً به‌دلیل سادگی و نیز وجود برخی مبانی نظری در مورد تقریب روابط غیرخطی یا خطی، به یافتن رابطه خطی علاقه‌مندی بیشتری نشان داده می‌شود.



(الف) (ب)

شکل۸,۱: نمودار وزن در برابر سن برای نمونه ۴۰۱۱ دختران آمریکایی. (الف) ارایه پراکنش

(ب) نمودار رگرسیون چندک از پایین به بالا $\boldsymbol{p = 0/03, 0/1, 0/25, 0/5, 0/75, 0/9, 0/97}$

۶,۱ کاربردهای رگرسیون چندکی

یو و استاندر[۱](۲۰۰۳) بر روی کاربردهای رگرسیون چندک بحث نموده‌اند، که در زیر به برخی از آن‌ها اشاره می‌نماییم.

۱,۶,۱ نمودارهای مرجع در پزشکی

نمودارهای مرجع یا صدکی، مجموعه‌ای از چندک‌های مفید در پزشکی را ارایه می‌دهند. این نمودارها به طور وسیع در تشخیص‌های اولیه پزشکی به منظور شناسایی موارد غیرمعمول، به‌کار برده می‌شوند، که مقدار بعضی از این اندازه‌گیری‌های خاص در دنباله‌های ابتدایی توزیع مرجع[۲]، مناسب می‌باشند. نیاز به منحنی‌های چندکی به جای یک بازه مرجع ساده، زمانی به‌وجود می‌آید که اندازه‌گیری‌ها (و به تبع آن بازه مرجعی) به شدت به متغیر تبیینی مثل سن وابسته باشند. کول و گرین[۳](۱۹۹۲) و روی استون و التمن[۴](۱۹۹۴) روی این مسأله بحث کرده‌اند. صدک‌های منتخب، معمولاً یک زیر مجموعه متقارن از $\{$۰/۰۳, ۰/۰۵, ۰/۰۱, ۰/۲۵, ۰/۵, ۰/۷۵, ۰/۹, ۰/۹۵, ۰/۹۷$\}$ هستند.

یک رهیافت بدیهی، استفاده از توزیع شرطی معلوم $F(Y|x)$، و برازش توزیع شرطی p امین منحنی چندکی متناظر با $q_p(x) = F^{-1}(p|x)$ است. اگر توزیع نرمال باشد، برآورد p امین منحنی چندک ساده است. اگر توزیع چوله باشد، که پذیره‌ای معمول است، اغلب از یک تبدیل نرمال‌ساز استفاده می‌شود. یکی از این تبدیلات، تبدیل باکس-کاکس[۵] است که در بخش ۲,۷,۱ به آن می‌پردازیم (کول، ۱۹۸۸؛ التمن ۱۹۹۰؛ روی‌استون و رایق[۶] ۲۰۰۰). برای حالاتی که تبدیلات نرمال‌ساز کارساز نیستند،

---

[۱] *Yu and Stander*

[۲] *Reference Distribution*

[۳] *Green*

[۴] *Royston and Altman*

[۵] *Box-Cox Transformation*

[۶] *Wrigh*

یک روش ناپارامتری[۱] یا نیمه‌پارامتری[۲] پیشنهاد می‌شوند (کول و گرین، ۱۹۹۲؛ هیگرتی و په‌په[۳] ۱۹۹۹).

۲,۶,۱ تحلیل بقا

یکی از کاربردهای رگرسیون چندکی در تحلیل بقا[۴]، مطالعه اثر یک متغیر تبیینی روی زمان بقای یک فرد است. یک متغیر تبیینی مشخص، ممکن است دارای اثرات متفاوت روی مخاطره پایین (مثل افراد خونسرد)، مخاطره متوسط (مثل دانشجویان)، و مخاطره بالای (مثل افرادی سیگاری) یک فرد باشد. این اثرات را می‌توان با در نظرگرفتن چندین تابع چندک از زمان بقا بررسی کرد (کانکر و گلینگ[۵]، ۲۰۰۱). شکل۹,۱ منحنی سه رگرسیون چندکی با $p = ۰/1,۰/5,۰/9$، بر اساس زمان بقا ۱۸۴ بیمار با متغیر تبیینی سن (بین ۱۲ تا ۶۴ سال)، از مطالعه جراحی پیوند قلب در استنفورد، می‌باشد (کروسلی و هو،[۶] ۱۹۷۷). برای جزییات بیشتر درباره رگرسیون میانگین سانسورشده یانگ[۷](۱۹۹۹) را ببینید.

مدل مخاطره متناسب کاکس[۸]، یکی از مدل‌های پرمصرف در تحلیل بقا محسوب می‌شود. مدل زمان شکست شتابنده[۹] (AFT) که مبنی بر مدل مخاطره متناسب کاکس ساخته می‌شود، لگاریتم زمان بقا را به عنوان یک تابع از متغیرهای تبیینی در نظر می‌گیرد (یانگ، ۱۹۹۹؛ کوتاس و گلفاند،[۱۰] ۲۰۰۱). این نوع مدل‌بندی به لحاظ تجربی مفید و مناسب است، چرا که تفسیر آن به راحتی انجام می‌شود.

فرض می‌کنیم مدل اساسی زمان‌های بقا برای $T_i , i = 1, \dots, n$ وابسته به متغیرهای تبیینی x_i

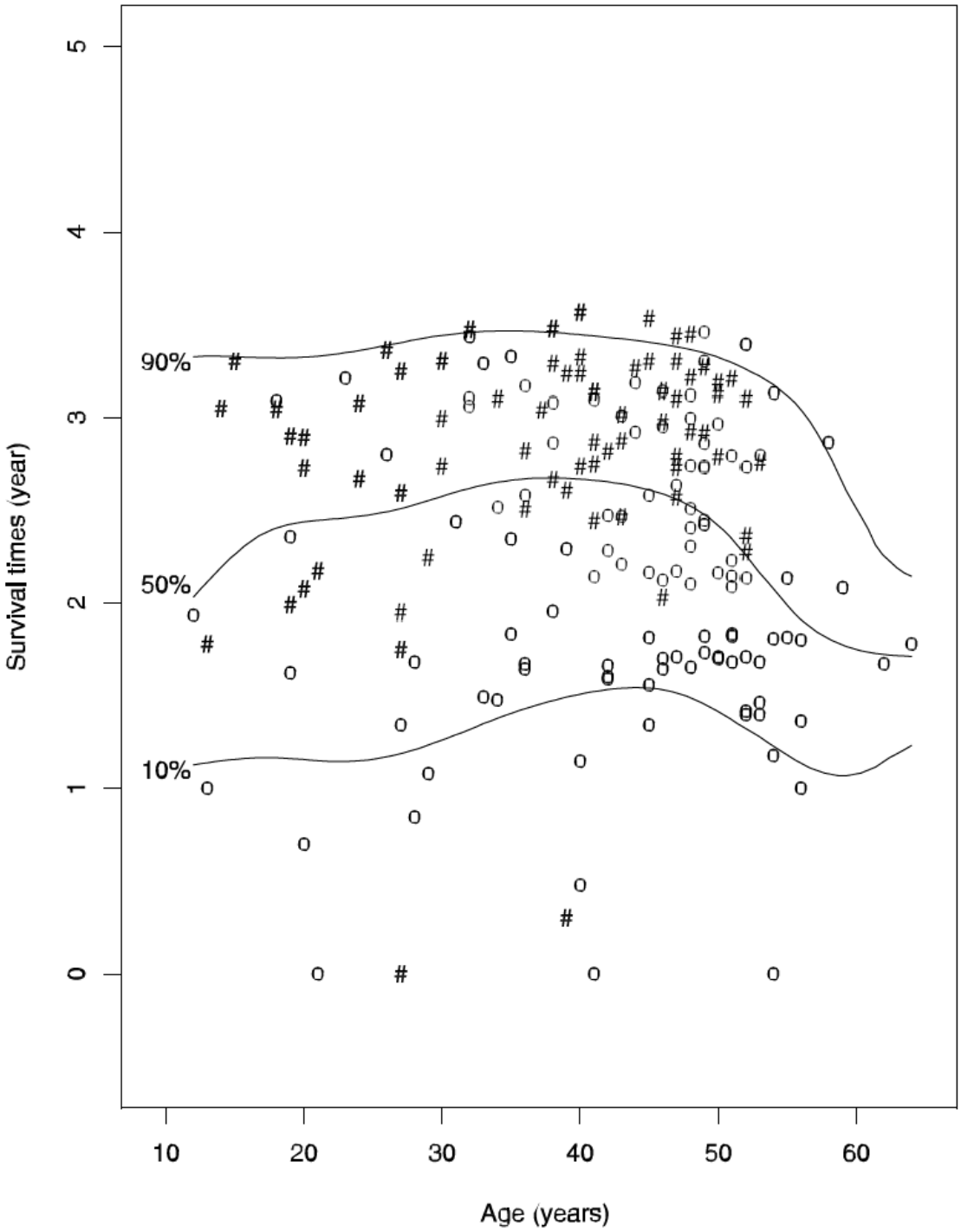
---

۱ *Non-Parametric*
۲ *Semi-parametric*
۳ *Heagerty and Pepe*
۴ *Survival Analysis*
۵ *Geling*
۶ *Crowley and Hu*
۷ *Yang*
۸ *Cox Proportional Hazard*
۹ *Accelerated Failure Time*
۱۰ *Kottas and Gelfand*

است، که در آن T_iها می‌توانند سانسور شده باشند. در صورت عدم وجود مشاهدات سانسور شده، طبیعی است تا زوج $\{T_i, x_i\}_{i=1}^{n}$ به صورت نمونه‌های دومتغیره مستقل و هم‌توزیع، مورد بررسی قرار گیرند. اگر iامین مشاهده سانسورشده باشد، در اینصورت Y_i را به‌جای T_i مشاهده می‌کنیم. تبدیل لگاریتم T_i، یک مدل AFT معمول را فراهم می‌کند که لگاریتم T_i را به صورت خطی روی x_i برازش می‌دهد. یعنی

$$log(T_i) = x_i{}^T\beta + \varepsilon_i,$$

که در آن، ε_i ها مستقل و هم‌توزیع با یک تابع توزیع نامعلوم می‌باشند. حال اگر در مدل به‌جای T_i، Y_i مشاهده شده باشد، در این‌صورت میانگین ε_i صفر نیست و بنابراین جمله عرض از مبدا شامل بردار β نخواهد بود. به همین دلیل، تحلیل رگرسیون میانگین، یک روش برازش خوب برای مدل AFT نمی‌باشد. با این وجود رگرسیون چندکی، که مدل‌های چندک‌های زمان بقا یا یک تبدیل یکنواخت وابسته به آن، مانند یک تابع از متغیرهای تبیینی را مدل‌بندی می‌کند، مناسب می‌باشد (یانگ، ۱۹۹۹).



شکل۹,۱: داده‌های جراحی قلب استنفورد با سه منحنی چندکی متناظر با $p = ۰/1,۰/5,۰/9$

: داده‌های سانسورشده o : داده‌های کامل

۳,۶,۱ اقتصاد

رگرسیون چندکی در مطالعه بازارهای مصرفی برای بررسی تاثیر یک متغیر تبیینی، که ممکن است مربوط به افرادی با گروه‌های درآمدی بالا، متوسط و پایین باشند، به‌کار برده شود. به‌طور مشابه، تغییرات نرخ‌های بهره ممکن است اثرات متفاوتی بر قیمت‌های سهام شرکت‌هایی که به گروه‌های با سود بالا، متوسط و پایین تعلق دارند، داشته باشند. رگرسیون چندکی به عنوان یک ابزار تحلیلی استاندارد، به‌ویژه برای مطالعه دستمزد و درآمد، در اقتصاد به کار برده می‌شود (بوچینسکی[۱]، ۱۹۹۸). این نوع رگرسیون، در مطالعه چگونگی توزیع درآمد در میان افراد جامعه نیز با اهمیت است. برای مثال در تعیین استراتژی‌های مالیاتی یا برای انجام سیاست‌های اجتماعی، می‌توان از آن استفاده کرد.

۴,۶,۱ مدل‌بندی پدیده‌های زیست‌محیطی

علم هیدرولوژی در مورد مدل‌بندی بارندگی و جریان آب رودخانه‌ها است. تامین آب شیرین، نیازمند برآورد احتمال وقوع خشکسالی جهت طراحی حجم مخازن سدها و نیز احتمال بارندگی‌های شدید جهت طراحی زه‌کشی سیلاب است. بنابراین مدل‌بندی دنباله توزیع و آگاهی از چندک‌های ابتدایی و انتهایی، مبحث اصلی هیدرولوژی آماری است. فرض کنید $q_p(x)$ تابع چندک برای متغیر هیدرولوژی، مانند ماکسیمم ارتفاع سالیانه آب باشد. بنابراین برای مقدار بزرگ p_0، مثلاً ۰/۹، احتمال وقوع یک مقداری بیشتر از $q_{p_0}(x)$ برابر $(1-p_0)$ خواهد بود. احتمال وقوع اولین مقدار بیشتر از $q_{p_0}(x)$ در سال k برابر با $P(K=k)=p_0^{k-1}(1-p_0)$ خواهد بود. میانگین متغیر K برابر $\frac{1}{(1-p_0)}$ است که دوره برگشت T با احتمال $(1-p_0)$ نامیده می‌شود. برای مثال، اگر T برابر صد سال باشد، پس $q_{0/99}(x)$ برابر مقداری است که با احتمال ۰/۹۹ سیلاب بعدی در ۱۰۰ سال آینده از مقدار حال تجاوز می‌کند.

[۱] *Buchinsky*

۵,۶,۱ نمایان سازی ناهم‌واریانسی

یک مسأله مهم در تحلیل داده‌ها، تشخیص وجود ناهم‌واریانسی[۱] است. ترسیم چندک‌ها، می‌تواند ابزار مهمی برای تشخیص باشد. این ترسیم نه تنها به شناخت ناهم‌پراشی کمک می‌کند، بلکه اثر مکانی، گسترش، و شکل توزیع Y به شرط $X = x$ را نیز فراهم می‌کند.

رگرسیون چندکی را می‌توان برای تعیین انحراف از پذیره‌های اولیه مدل $Y = x^T\beta + \varepsilon$ به‌کار برد. اگر توزیع ε به متغیر تبیینی x وابسته نباشد، تمام چندک‌های رگرسیونی موازی خواهند بود. برای مثال هفت منحنی چندکی برای داده‌های دختران آمریکایی در شکل ۸,۱ به وضوح موازی نیستند، که این نشان‌دهنده ناهم‌واریانسی است.

۷,۱ روش‌های برآورد

برای برآورد پارامترها در رگرسیون چندک، الگوریتم‌ها و روش‌های مختلفی پیشنهاد شده‌اند. در این بخش به چهار مورد از آن‌ها اشاره و در این پایان‌نامه از مدل رگرسیون چندکی پارامتری برای برآورد پارامترها استفاده می‌کنیم.

۱,۷,۱ مدل رگرسیون چندکی پارامتری

برای کمی کردن رابطه بین متغیر پاسخ $\boldsymbol{Y}$ و متغیر تبیینی $\boldsymbol{x}$، اغلب فرض می‌شود $E[\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}]$ به صورت ترکیب خطی ساده $\boldsymbol{x}^T\boldsymbol{\beta}$ باشد. به‌طور مشابه، مدل رگرسیون چندکی، وابستگی خطی چندک‌های $\boldsymbol{Y}$ به $\boldsymbol{x}$ را مشخص می‌کند. به عبارت دیگر، فرض می‌کنیم ارتباط بین p امین چندک $\boldsymbol{Y}$ و متغیرهای تبیینی $\boldsymbol{x}$ به صورت $q_p(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}$ باشد. با در نظر گرفتن مجموعه داده $\{x_i, y_i\}_{i=1}^n$ می‌توان برآورد پارامترهای $\boldsymbol{\beta}$ را با می‌نیمم کردن رابطه

[۱] *heteroscedasticity*

$$\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i{}^T \boldsymbol{\beta}),$$

به‌دست آورد. جواب صریحی برای ضرایب، تحت این مدل رگرسیونی وجود ندارد، زیرا تابع زیان در نقطه مبدا دارای مشتق نمی‌باشد.

۲,۷,۱ تبدیلات باکس کاکس مدل چندکی

در این زیربخش برآوردهای پارامترهای رگرسیون چندکی با کمک تبدیلات باکس کاکس[۱] را، که در بخش ۱,۶,۱ اشاره شد، توضیح می‌دهیم. فرض کنید تابع $g(\lambda; y)$ به صورت

$$g(\lambda; y) = \begin{cases} (y^\lambda - 1)/\lambda & \text{اگر } \lambda \neq ۰ \\ log(y) & \text{اگر } \lambda = ۰, \end{cases}$$

تعریف شود. همچنین فرض کنید $g^{-1}(\lambda; z)$ تابع معکوس $g(\lambda; y)$ نسبت به y باشد. برای یافتن توابع معکوس داریم:

$$(y^\lambda - 1)/\lambda = z \Longrightarrow \left(y^\lambda - 1\right) = \lambda z \Longrightarrow y = (1 + \lambda z)^{\frac{1}{\lambda}}$$
$$log(y) = z \Longrightarrow y = e^z.$$

بنابراین

$$g^{-1}(\lambda; z) = \begin{cases} (\lambda z + 1)^{1/\lambda} & \text{اگر } \lambda \neq ۰ \\ exp(z) & \text{اگر } \lambda = ۰, \end{cases} \qquad (8.1)$$

اگر p امین چندک $g(\lambda; Y)$، $q_p(g)$ باشد، آن‌گاه p امین چندک Y، $g^{-1}\{\lambda; q_p(g)\}$ می‌باشد. به‌ویژه اگر مقداری از λ وجود داشته باشد، به طوری که $g(\lambda; Y)$ و X به صورت نرمال توزیع شده باشند، پس با توجه به ویژگی توزیع نرمال می‌توان نشان داد که p امین چندک $g(\lambda; Y)$ به صورت خطی $\beta_۰ + \beta_1 x$ است که در آن، β_1 و $\beta_۰$ وابسته به p هستند. از این‌رو p امین چندک Y به‌وسیله $g^{-1}(\lambda; \beta_۰ + \beta_1 x)$ به‌دست می‌آید.

[۱] *Box-cox transformation*

مقداری از λ که متغیر Y را به نرمال تبدیل می‌کند، ممکن است وجود نداشته باشد. در این حالت، یک راه آن است که از تبدیلی، بر مبنای تابع زیان، برای پیدا کردن $\beta = (\beta_0, \beta_1)^T$ و λ به‌وسیله می‌نیمم کردن رابطه

$$\sum_{i=1}^{n} \rho_p\big(y_i - g^{-1}(\lambda; \beta_0 + \beta_1 x_i)\big),$$

استفاده شود (بوچینسکی، ۱۹۹۸).

تحت تبدیلات (*8.1*)، می‌توان فرض کرد که مقدار λ با تغییر متغیرهای تبیینی x، تغییر می‌کند و p امین چندک Y تحت تبدیلات باکس-کاکس، ممکن است توسط $g^{-1}(\lambda(x); \beta_0 + \beta_1 x_i)$ بررسی شود. به جای برآورد یک مقدار منفرد از λ، از برآورد منحنی $\lambda(x)$ می‌توان استفاده نمود. در این مورد، به یک برآورد هموار، بر اساس روش هسته[۱] یا اسپلاین هموار[۲] نیاز داریم. کول و گرین (۱۹۹۲) برای به‌دست آوردن برآورد درستنمایی تاوانیده[۳] $\lambda(x)$ روشی را معرفی کردند.

اگر $RSS(f)$ مجموع توان دوم خطای تاوانیده برای برازش f، به صورت زیر تعریف شود:

$$RSS(f, \alpha) = \sum_{i=1}^{n} (y_i - f(x_i))^2 + \alpha \int_a^b \big(f''(x)\big)^2 \, dx$$

آن‌گاه $\lambda(x)$ از ماکسیمم کردن رابطه

$$l(\lambda) - \alpha \int \lambda''(x)^2 dx,$$

برآورد می‌شود، که در آن $l(\lambda)$، لگاریتم تابع درستنمایی بوده و برابر است با

$$l(\lambda) = \sum_{i=1}^{n} \left\{ \lambda(x_i)\, log(y_i) - \frac{1}{2} z_i^2 \right\}, \qquad z_i = g(\lambda; y_i),$$

---

[۱] *Kernel*

[۲] *Smoothing spline*

[۳] *Penalized likelihood estimate*

و $\int \lambda''(x)^2 dx$ جریمه ناهمواری[۱] و α پارامتر هموارسازی می‌باشند.

۳,۷,۱ مدل رگرسیون چندکی ناپارامتری

از آنجایی‌که توزیع شرطی، جزء اصلی رگرسیون چندکی است، یو و جونز [۲](۱۹۹۸) و هال[۳] و همکاران (۱۹۹۹) روش‌هایی را برای برآورد توزیع شرطی مورد بررسی قرار دادند.

تعریف۷,۱: فرض کنیم که $x_1, \dots, x_n$ نمونه تصادفی از توزیعی با تابع چگالی احتمال f باشد، در این‌صورت برآوردگر هسته‌ای $f(\boldsymbol{x})$ به صورت زیر تعریف می‌شود:

$$\hat{f}_h(\boldsymbol{x}) = \hat{f}_h(x_1, \dots, x_n; x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

که $h >$ ۰ برآوردگر هموارسازی و $K(.)$ تابع هسته‌ای است که باید در سه شرط زیر صدق کند.

1. $\lim_{y\to\infty} |yK(y)| = ۰$
2. $\sup_{-\infty<y<\infty} |K(y)| < \infty$
3. $\int_{-\infty}^{+\infty} K(y)dy = 1$

برای مثال، $F(y|x)$ ممکن است، به صورت ناپارامتری، با فرم زیر برآورد شود:

$$\hat{F}(y|x) = \sum_{i=1}^{n} w_i(x) I(A_i), \qquad (9.1)$$

که در آن، $A_i = \{y_i < y\}$، w_i تابع وزن غیر صفر که به x_i وابسته است، و $\sum_{i=1}^{n} w_i(x) = 1$ است.

این نوع برآورد وزنی هسته برای توزیع شرطی با مشکلاتی همراه است. اول، این برآوردگر تابع توزیع نیست، زیرا نه تنها یکنوا نیست بلکه فقط مقادیر آن بین صفر و یک قرار می‌گیرند و هیچگاه به یک

[۱] *roughness penalty*

[۲] *Jones*

[۳] *Hall*

نمی‌رسند. به همین دلیل هال(۱۹۹۹) برآوردگر تصحیح‌شده مشهور ناداریا- واتسون[۱] را که دارای تابع توزیع شرطی به فرم

$$\hat{F}(y|x) = \frac{\sum_{i=1}^{n} p_i(x_i) K_h(x_i - x) I(y_i \le y)}{\sum_{i=1}^{n} p_i(x_i) K_h(x_i - x)},$$

است را پیشنهاد داد، که در آن، $K_h(x_i - x) = K(\frac{x_i - x}{h})$، h پارامتر برآوردگر هموارسازی، K تابع چگالی هسته $K_h(.)$، $\sum_{i=1}^{n} p_i(x_i) = 1$ و $p_i(x_i) \ge 0$ است.

دوم منحنی‌های چندکی بر اساس برآوردگرهای $F(y|x)$ ممکن است با یکدیگر تلاقی داشته باشند که نامطلوب است. این مسأله توسط شکل ۱۰,۱ نمایش داده‌شده است که در آن، برآورد رگرسیون چندکی هسته برای $\hat{F}(y|x)$، از هال و همکاران (۱۹۹۹)، برای مجموعه داده‌های جراحی قلب استنفورد نشان داده شده است. یو و جونز (۱۹۹۸) برای حل این مسأله، رهیافت هسته دوگان[۲] را به‌کار بردند. در این رهیافت، به برآوردگر هموارسازی برای y و x نیاز است. ایده اساسی آن است که تابع نشانگر $I(A_i)$ در رابطه (۹,۱) به وسیله یک تابع توزیع پیوسته $\Omega\{(y_i - y)/b\}$ جایگزین شود. که در آن،

$$\Omega(q) = \int_{-\infty}^{q} g(v) dv,$$

$\Omega(q)$ یک تابع توزیع با تابع چگالی $g(v)$، $b > 0$ برآوردگر هموارسازی در جهت y و $w_i(x)$ توابع وزن بر اساس هسته هستند. بنابراین داریم:

$$\hat{F}(y|x) = \sum_{i=1}^{n} w_i(x) \Omega\left(\frac{y_i - y}{b}\right).$$

اگر g دارای هسته چگالی یکنواخت باشد، $g(u) = 0/5 I(|u| \le 1)$، و

$$w_i(x) = \frac{K_h(x_i - x)}{\sum_{i=1}^{n} K_h(x_i - x)},$$

[۱] *Nadaraya-Watson*

[۲] *double-kernel*

و همچنین وزن‌ها با پبرآوردگر هموارسازی h در جهت x باشند، می‌توان نشان داد که $\frac{d}{dp}\hat{q}_p(x) > 0$. این نتیجه از آن‌جا ناشی می‌شود که $\hat{q}_p(x)$ تابعی یکنوا از p است و این‌که دنباله این روش از مسأله تلاقی، که در بالا اشاره شد، پیروی نمی‌کند. یو و جونز (۱۹۹۸) یک الگوریتم مکرر برای برآورد توابع چندک تحت رهیافت هسته دوگان معرفی کردند.

تحت تابع زیان مجموعه بالا، برآورد رگرسیون چندکی بر اساس هسته را می‌توان با می‌نیمم کردن تابع زیان وزنی هسته

$$\min_{a}\left\{\sum_{i=1}^{n}\rho_p(y_i-a)K_h(x_i-x)\right\},$$

به دست آورد که در آن، می‌نیمم‌کننده $\hat{a}=\hat{a}(x)$، برآوردگر رگرسیون چندکی p ام است. یک فرآیند کمترین توان‌های دوم باز وزن داده شده تکراری[۱] برای پیدا کردن $\hat{a}$ توسط یو و جونز (۱۹۹۸) معرفی شد.

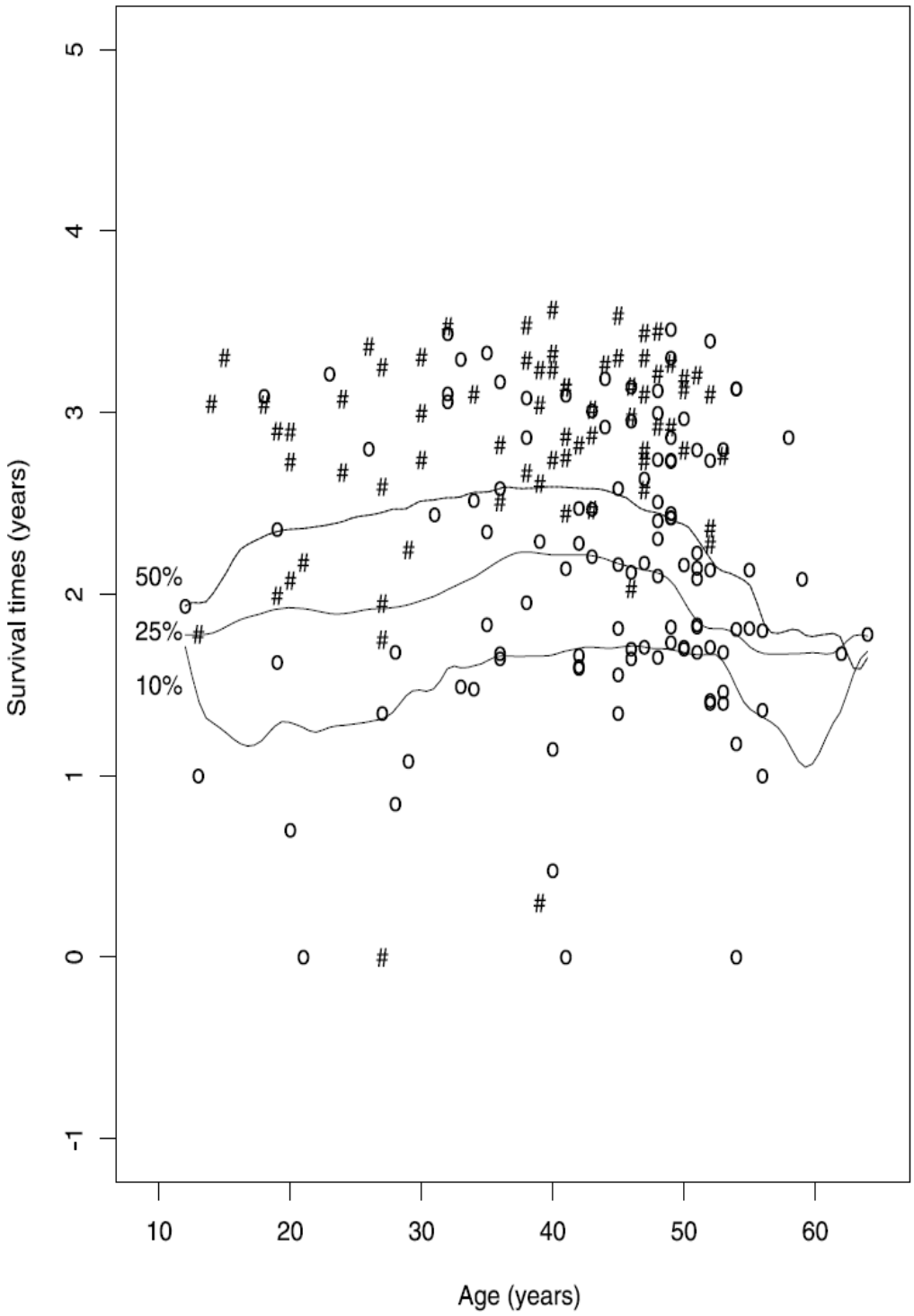
۴,۷,۱ مدل رگرسیون چندکی نیمه‌پارامتری

رهیافت برآورد درستنمایی تاوانیده که در بخش۲,۷,۱ شرح داده شد، مثالی از فن برازش نیمه‌پارامتری است که توسط کول و گرین (۱۹۹۲) مطرح شد. به‌طور کلی، در روش نیمه‌پارامتری در بعضی مواقع، اما نه در تمام موارد، فرض بر این است که متغیر پاسخ Y به متغیرهای تبیینی وابسته است. چنین مدلی با جایگزینی مدل خطی $\boldsymbol{x}^T\boldsymbol{\beta}+g(t)$ با $\boldsymbol{x}^T\boldsymbol{\beta}+g(t)$ به صورت $q_p=\boldsymbol{x}^T\boldsymbol{\beta}+g(t)$ تعیین می‌شود که در آن t و x متغیرهای تبیینی برای p امین چندکY و $\boldsymbol{\beta}$ و تابع g مجهول هستند و باید برآورد شوند. معمولاً برآورد g و $\boldsymbol{\beta}$ با می‌نیمم کردن

$$\sum_{i=1}^{n}\rho_p\{y_i-\boldsymbol{x}_i{}^T\boldsymbol{\beta}-g(t)\}+\alpha\int(g^{"}(t))^2dt\,,$$

به‌دست می‌آید کانکر و همکاران (۱۹۹۲) روشی را برای این منظور ارایه دادند.

[۱] *Iterated Reweighthed Least Square*


Survival times (years)
Age (years)
50%
25%
10%
-1
0
1
2
3
4
5
10
20
30
40
50
60

شکل۱۰,۱: داده‌های جراحی قلب استنفورد با سه منحنی رگرسیون چندک $\boldsymbol{p = ۰/\mathbf{1}, ۰/\mathbf{25}, ۰/\mathbf{5}}$ که با روش هسته به‌دست آمده‌اند

۸,۱ رگرسیون چندکی برای سری‌زمانی

در اکثر تحقیقات رگرسیون چندکی، فرض بر این است که متغیر پاسخ Y، با شرط مفروض بودن x مستقل است. اخیراً چندین محقق بر روی روش‌های مختلفی برای مدل‌بندی رگرسیون چندکی سری‌زمانی متمرکز شده‌اند. به عنوان مثال، روشی بر اساس برآورد توزیع شرطی، توسط کای[۱] (۲۰۰۲) و روشی بر اساس تابع زیان، توسط گانون[۲] (۲۰۰۳) ارایه شده است. در روش کای (۲۰۰۲) فرض شده است که سری‌زمانی $Y_i,\ i = 1, \dots, n$ به سری‌زمانی X_i توسط رابطه $Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i$ وابسته باشند، که در آن $\mu(X_i)$ تابع رگرسیونی و ε_i خطای مدل است. وابستگی $\sigma(X_i)$ به X_i به این معنی است که مدل دارای ناهمواریانسی است.

اولین روش، برآورد توزیع Y_i به شرط X_i توسط رهیافت بخش۳,۷,۱ است و سپس برآورد چندک شرطی توسط معکوس تابع توزیع شرطی است. در دومین روش، برای برآورد چندک شرطی از فرآیند اکیداً ایستای Z به شرط اطلاعات گذشته و حال استفاده می‌شود، چندک p ام Z از مسأله بهینه‌سازی زیر مشخص می‌شود:

$$q_p(x) = \underset{\theta \in R}{argmin}\{E[\rho_p(Z-\theta)|X = x]\}.$$

آن‌گاه برای برآورد از روش هسته بر اساس تابع زیان از بخش ۳,۷,۱ استفاده می‌شود.

[۱] *Cai*

[۲] *Gannoun*

فصل دوم

رگرسیون بیزی

۱,۲ مقدمه

روش‌های استنباط بیزی با روش‌های استنباط بسامدی[۱] متفاوت هستند. در روش بسامدی تنها منبع اطلاعاتی، داده‌های موجود هستند که در ساختن برآورد یا آزمون فرضیه پارامتر مورد توجه قرار می‌گیرند. در صورتی که در روش بیزی برآورد یا آزمون پارامتر از ترکیب داده‌های موجود با اطلاعاتی از گذشته، حاصل می‌گردد.

فرمول اساسی برای دخالت دادن اطلاعات گذشته در تحلیل آماری توسط توماس بیز[۲] در سال ۱۷۶۳ ارایه شد. اولین کاربرد آن برای مسائل رگرسیونی توسط هارولد جفری[۳] در سال ۱۹۲۹ ارایه شد.

۲,۲ توزیع‌های پیشین و پسین

توزیع پارامتر قبل از مشاهده هر داده‌ای، توزیع پیشین[۴] آن پارامتر نامیده می‌شود. توزیع شرطی پارامتر به شرط مشاهده داده‌ها، توزیع پسین[۵] نامیده می‌شود. اگر مقادیر مشاهده شده داده‌ها را در تابع احتمال شرطی یا تابع چگالی احتمال شرطی داده‌ها به شرط پارامتر، جایگذاری کنیم نتیجه این کار فقط تابعی از پارامتر است که آن را تابع درستنمایی می‌نامیم.

۱,۲,۲ تعیین توزیع پیشین

یک مسأله استنباط آماری را در نظر بگیرید که در آن قرار است مشاهدات از توزیعی با تابع چگالی احتمال یا تابع احتمال $f(x|\theta)$ انتخاب شوند، که در این‌جا θ پارامتری با مقادیر نامعلوم است. فرض می‌کنیم که مقدار نامعلوم θ در یک فضای پارامتر مانند Ω قرار دارد. در مسائل استنباط آماری سعی بر

---

۱ *Frequentist*

۲ *Bayes*

۳ *Jeffreys*

۴ *Prior Distribution*

۵ *Posterior Distribution*

این بوده است که بر اساس مشاهداتی از تابع چگالی احتمال یا تابع احتمال، $f(x|\theta)$ بتوان تعیین کرد که قرار گرفتن θ در کجای Ω محتمل‌تر است.

در بسیاری از مسائل، قبل از هر مشاهده‌ای از $f(x|\theta)$، آماردان یا آزمایشگر این توانایی را دارد که اطلاعات قبلی در مورد θ را به صورت یک توزیع احتمال برای θ در مجموعه Ω خلاصه کند. به عبارت دیگر، قبل از آن‌که داده‌های مربوط به آزمایش جمع آوری یا مشاهده شوند، دانش و تجربه قبلی آزمایشگر او را هدایت می‌کند تا تعیین کند قرارگرفتن پارامتر θ در کدام بخش Ω محتمل‌تر است.

۲,۲,۲ طبیعت جدال‌برانگیز توزیع پیشین

مفهوم توزیع پیشین در آمار بسیار جدال‌برانگیز است. بعضی از آماردانان عقیده دارند که توزیع پیشین را می‌توان در هر مسأله آماری برای θ در نظر گرفت. آن‌ها عقیده دارند که این توزیع، یک توزیع احتمال ذهنی[۱] است و اطلاع محقق و ذهنیت او را درباره اینکه کدام مقادیر θ محتمل‌تر هستند، نشان می‌دهد. آن‌ها همچنین عقیده دارند که توزیع پیشین هیچ تفاوتی با توزیع‌های دیگر که در سایر زمینه‌های آماری بکار می‌روند، ندارد و تمام قوانین نظریه احتمال برای توزیع پیشین نیز قابل طرح است. آماردانانی که این قبیل برداشت‌ها را باور دارند، جزء هواداران فلسفه بیزی به شمار می‌آیند.

سایر آماردانان عقیده دارند که در بسیاری از مسائل آماری صحبت کردن از توزیع احتمال درباره پارامتر θ مناسب نیست، زیرا مقدار واقعی θ اصلاً متغیر تصادفی نیست. آنچه که اتفاق افتاده است آن است که این مقدار واقعی برای محقق آشکار نیست. این آماردانان عقیده دارند زمانی می‌شود توزیع احتمالی برای θ در نظر گرفت که اطلاعات وسیعی درباره فراوانی‌های نسبی مقادیری که θ می‌تواند بگیرد از قبل موجود باشد. در این صورت است که دو آماردان متفاوت یک توزیع برای θ در نظر خواهند گرفت. به عنوان مثال، فرض کنید θ نسبت کالاهای معیوب تولید شده به وسیله یک دستگاه باشد که مقدار

[۱] *Subjective Distribution*

واقعی آن معلوم نیست. همچنین فرض کنید مسئولان این کارخانه تعداد زیادی از این کالاها را در گذشته تولید و نسبت کالاهای معیوب در گذشته را ثبت کرده باشند. در این صورت امکان تعیین توزیع پیشین برای θ، با استفاده از این سابقه، وجود دارد.

هر دو گروه آماردانان هم‌سخن هستند که اگر بتوان توزیع پیشین معقولی انتخاب کرد، آنگاه استفاده از روش‌هایی که در ادامه خواهیم گفت، مفید خواهد بود.

۳,۲,۲ توزیع پیشین ناسره

بسیار اتفاق می‌افتد که توزیع‌های پیشین متفاوت، اختلاف زیادی در توزیع‌های پسین ایجاد نمی‌کنند. از جمله شرایطی که ممکن است منجر به این نتیجه شوند، وجود حجم بالای داده‌ها و گستردگی بسیار زیاد توزیع‌های پیشین می‌باشند.

این نتیجه دو پیامد دارد. اول، اگر تعداد داده‌ها زیاد باشند، عدم توافق دو محقق روی یک توزیع پیشین، چندان اهمیتی ندارد. دوم، اگر برای محققین فرق زیادی نداشته باشد که چه توزیع پیشینی را انتخاب کنند، تمایل زیادی به صرف وقت برای تعیین آن نخواهند داشت. اگر توزیع پیشین تعیین نشود، آنگاه راهی برای تعیین توزیع شرطی پارامتر به شرط داده‌ها، وجود ندارد. تدبیری که در این حالت اندیشیده شده است، آن است که سعی می‌شود بر اساس برخی محاسبات، اطلاعات بیشتری از داده‌ها، نسبت به آنچه که از قبل موجود بود، استخراج شود. معمولاً این محاسبات بر اساس تابعی مانند $\xi(\theta)$ انجام می‌شود که در آن با تابع $\xi(\theta)$ مثل یک توزیع پیشین رفتار می‌شود اما حتی ممکن است $\int \xi(\theta)d\theta = \infty$، که به‌طور واضح تعریف تابع چگالی احتمال را نقض می‌کند. چنین پیشینی را توزیع پیشین ناسره[۱] می‌گویند.

[۱] *Improper prior*

۴،۲،۲ توزیع‌های پیشین پیوسته و گسسته

در برخی از مسائل، پارامتر θ حداکثر می‌تواند دنباله‌ای شمارا از مقادیر را اختیار کند. توزیع پیشین θ در این حالت یک توزیع گسسته خواهد بود. در این صورت $\xi(\theta)$ را تابع احتمال پیشین θ می‌نامیم. در مسائل دیگر، θ ممکن است بتواند مقادیر بازه‌ای از خط حقیقی را اختیار کند که در این صورت یک توزیع پیوسته برای θ در نظر گرفته می‌شود. در این صورت $\xi(\theta)$ را تابع چگالی احتمال پیشین θ می‌نامیم.

۵،۲،۲ توزیع‌های پیشین مزدوج

برای برخی از مدل‌های آماری بسیار رایج، خانواده‌ای از توزیع‌ها با این ویژگی وجود دارد که اگر توزیع پیشین را عضوی از آن خانواده در نظر بگیریم، توزیع پسین نیز عضوی از آن خانواده خواهد بود. چنین خانواده‌ای از توزیع‌ها را خانواده مزدوج می‌نامند.

برای روش‌های متفاوت مدل‌بندی آماری برای داده‌ها به شرط پارامتر، خانواده توزیع‌های مزدوج از پارامتر به‌دست می‌آید. به عنوان چند مثال، برای داده‌هایی که توزیع آن‌ها به آزمایش‌های برنولی وابسته است مثل دوجمله‌ای، هندسی، دوجمله‌ای منفی، خانواده مزدوج برای پارامتر احتمال پیروزی، توزیع‌های بتاست. برای داده‌های با توزیع وابسته به فرآیند پواسون مثل توزیع پواسون، گاما (با پارامتر شکل معلوم) و نمایی، خانواده مزدوج برای پارامتر نرخ، خانواده توزیع‌های گاماست. برای داده‌های با توزیع نرمال با واریانس معلوم، خانواده توزیع‌های مزدوج، خانواده نرمال است (پاشا، ۱۳۸۲).

۳،۲ روش بیزی

روش بیزی، یک روش استنباط آماری کلی است که تقریباً می‌توان برای تمام مسائل آماری از آن استفاده کرد. در این زیربخش، خلاصه‌ای از ویژگی‌های آن را، قبل از به کار بردن آن در مدل‌های رگرسیونی، بیان می‌کنیم. برای تحلیل مجموعه ای از داده‌ها، معمولاً یک مدل آماری برای آن در نظر می‌گیریم. فرض

کنید $\boldsymbol{y}$ بردار داده‌ها و $\boldsymbol{\theta}$ بردار پارامترهای مجهول را نمایش دهند. در مدل رگرسیون ساده، $\boldsymbol{y} = (y_1, \dots, y_n)$ بردار مشاهدات متغیر وابسته و $\boldsymbol{\theta} = (\alpha, \beta, \sigma)$ بردار پارمترهای مدل است. می‌خواهیم داده-های $\boldsymbol{y}$ را برای برآورد یا آزمون فرضیه درباره بعضی از پارامترها در $\boldsymbol{\theta}$ بکار بریم. مدل آماری بیان می‌کند که بردار داده‌های $\boldsymbol{y}$ که به عنوان یک متغیر تصادفی در نظر گرفته می‌شود، برای هر مقدار ثابت از بردار $\boldsymbol{\theta}$ دارای توزیع معینی است. مثلاً مدل رگرسیون خطی ساده از توزیع نرمال بیان می‌کند که برای مقادیر ثابت (α, β, σ)، توزیع y_i نرمال با میانگین $\alpha + \beta x_i$ و واریانس σ^2 می‌باشد و y_iها نیز از یکدیگر مستقل هستند. توزیع $\boldsymbol{y}$ برای یک مقدار ثابت $\boldsymbol{\theta}$ را توزیع شرطی y به شرط $\boldsymbol{\theta}$ می‌نامند. در روش بیزی باید برای $\boldsymbol{\theta}$ توزیعی نیز مشخص کنیم. قبل از مشاهده داده‌ها باید ارزیابی کنیم که درباره مقادیر احتمالی پارامترها چه باوری داریم و این باور را به صورت یک توزیع احتمال برای $\boldsymbol{\theta}$ بیان کنیم. در مرحله بعد، توزیع پیشین $\boldsymbol{\theta}$ و توزیع $\boldsymbol{y}$ به شرط $\boldsymbol{\theta}$ را ترکیب نموده، توزیع $\boldsymbol{\theta}$ به شرط $\boldsymbol{y}$ را به‌دست می‌آوریم. برای بردار داده‌های مشاهده شده $\boldsymbol{y}$، این توزیع را توزیع پسین $\boldsymbol{\theta}$ می‌نامند. در این روش اطلاعات پیشین درباره پارامترها را به وسیله داده‌های جاری، به‌روز[1] می‌کنیم تا اطلاعات پسین به‌دست آید.

فرمول بیز، ترکیبی از توزیع‌های $\boldsymbol{\theta}$ و توزیع $\boldsymbol{y}$ به شرط $\boldsymbol{\theta}$ را برای به‌دست آوردن توزیع پسین فراهم می‌کند. اغلب این توزیع‌ها را برحسب تابع چگالی احتمال آن‌ها، به ترتیب با $f(\boldsymbol{\theta})$، $f(\boldsymbol{y}|\boldsymbol{\theta})$ و $g(\boldsymbol{\theta}|\boldsymbol{y})$ نشان می‌دهند. فرمول بیز بیان می‌کند

$$g(\boldsymbol{\theta}|\boldsymbol{y}) = cf(\boldsymbol{\theta})f(\boldsymbol{y}|\boldsymbol{\theta}),$$

که در آن c، کمیتی است که به $\boldsymbol{\theta}$ بستگی ندارد.

برآورد بیزی یا آزمون‌های فرضیه بیزی درباره پارامترهای $\boldsymbol{\theta}$ از توزیع پسین $\boldsymbol{\theta}$ به‌دست می‌آیند. برای مثال، می‌توانیم برآورد $\boldsymbol{\theta}$ را با استفاده از میانگین توزیع پسین آن به‌دست آوریم.

[1] *Update*

۴,۲ برآورد بیزی خط رگرسیون

در مدل‌های رگرسیونی، توزیعی که به طور وسیعی بکار برده شده است توزیع نرمال می‌باشد. مدل رگرسیون خطی نرمال را در نظر بگیرید. با معلوم بودن پارامترهای α, β و σ بردار داده‌های $y = (y_1, \dots, y_n)$ دارای توزیع نرمال چندمتغیره است. در روش بیزی همچنین باید توزیع پیشین توأم پارامترهای α, β و σ را مشخص کنیم. به طور نظری، هر توزیعی را می‌توان به عنوان توزیع پیشین انتخاب کرد، اما به طور عملی اگر توزیع پیشینی را انتخاب کنیم که وقتی توزیع پسین را محاسبه می‌کنیم به طور مناسبی با توزیع بردار داده‌ها ترکیب شود، محاسبات آسان‌تر می‌شود.

توزیع پیشین ناآگاهی‌بخش[۱]: اطلاعات پیشین یا باوری که یک شخص درباره پارامترها دارد، بستگی به آن شخص دارد. فرض کنید هیچ اطلاع پیشین یا باوری درباره پارامترها نداریم. یعنی، قبل از مشاهده داده‌ها فرض کنید هیچ ایده‌ای در مورد این‌که α, β و σ چه مقادیر ممکنی دارند، به جز این‌که σ باید مثبت باشد، نداریم. از این‌رو نیاز به یک توزیع پیشین ناآگاهی‌بخش[۲] داریم که عدم آگاهی ما را بیان کند. برای مثال اگر در یک مسئله آزمون فرض ساده در مقابل ساده، احتمالات یکسان $\frac{1}{2}$ را به فرض صفر و مقابل نسبت دهیم آنگاه این توزیع پیشین، ناآگاهی بخش است.

یکی از روش‌های بدست آوردن توزیع پیشن ناآگاهی بخش توسط جفریز(۱۹۶۱) ارایه شده است، که در این روش چگالی پیشین را به صورت $f(\boldsymbol{\theta}) \propto \sqrt{|I(\boldsymbol{\theta})|}$ در نظر می‌گیرد که در آن $|I(\boldsymbol{\theta})|$ دترمینان ماتریس اطلاع فیشر می‌باشد.

[۱] *Noninformative Priors Distribution*

[۲] *Noninformative Priors Distribution*

فصل سوم

رگرسیون چندکی بیزی

۱,۳ مقدمه

رگرسیون چندکی یک رده از مدل‌هاست که بر مبنای نظریه مدل‌های رگرسیون ساخته می‌شود و می‌توان از آن برای سنجش اثر متغیرهای تبیینی روی توزیع متغیر پاسخ استفاده کرد. این نوع رگرسیون توسط

کانکر و باست[۱] (۱۹۷۸) معرفی شد. رگرسیون چندکی نه تنها یک روش مناسب برای اندازه‌گیری اثر متغیرهای تبیینی روی متغیر پاسخ است، بلکه می‌تواند به عنوان یک روش مناسب برای سنجش اثر متغیرهای تبیینی روی چندک‌های انتخاب شده از توزیع متغیر پاسخ برده شود. به همین دلیل است که این روش، به‌ویژه در تحقیقات کاربردی، استفاده عملی دارد. گر چه نظریه مجانبی برای رگرسیون چندکی به خوبی توسعه یافته است، اما برای نمونه‌های کوچک، توزیع‌های دقیق نامعلوم است. به همین دلیل استفاده‌کنندگان از این رگرسیون، برای توزیع نمونه‌های کوچک، متوسل به استفاده از روش خودگردان‌ساز[۲] شده‌اند (بوچینسکی، ۱۹۹۸).

در این‌جا نشان می‌دهیم که رگرسیون چندکی می‌تواند درون خانواده نرمال آمیخته مقیاسی قرار بگیرد و می‌توان از روش‌های نمونه‌گیری گیبز[۳] با داده‌افزایی[۴]، برای به‌دست آوردن توزیع‌های پسین حاشیه‌ای دقیق پارامترها استفاده کرد. برای محققینی که بخواهند ویژگی کوچک نمونه برآوردگرها را مطالعه کنند، توزیع‌های پسین دقیق ارزشمند هستند. مدل‌بندی بیزی رگرسیون چندکی، در فضای $\mathbb{R}$، L_1 است که در آن پارامترها بر اساس می‌نیمم کردن مجموع قدر مطلق انحرافات خطا، یعنی $\widehat{\theta} = arg\,\underset{\theta}{min}\sum_{i=1}^{n}|y_i - x_i^T\theta|$ برآورد می‌شوند. برای حالت خاصی که پارامتر چندک مساوی $\frac{1}{2}$ است، مسأله ساده‌شده و منحنی رگرسیونی، میانه چگالی شرطی $Y|x$ است. برازش رگرسیون L_1 معادل است با به‌دست آوردن برآوردگر ماکسیمم درستنمایی پارامترها تحت فرض توزیع لاپلاس برای جمله خطا (تزیوناس[۵]، ۲۰۰۳).

---

۱ *Bassett*

۲ *Bootstrap*

۳ *Gibbs Sampling*

۴ *Data Augmentation*

۵ *Tsionans*

در این بخش، برآورد توزیع‌های لاپلاس مکانی- مقیاسی را مورد استفاده قرار می‌دهد، زیرا قدرمطلق میانه یک برآورد ماکسیمم درستنمایی نااریب از پارامتر مکان است، اما برآوردگری با کمترین واریانس نیست. برای این منظور هارتر[۱] و همکاران (۱۹۷۹) برآوردهای سازوار تنومند[۲] را برای مکان و مقیاس پیشنهاد کردند. مدل‌های *ARMA* با خطاهای لاپلاس توسط دامسلت و الشراوی[۳] (۱۹۸۹) مورد بررسی قرار گرفتند. افرون[۴] (۱۹۸۶) نقش توزیع لاپلاس در رگرسیون خطی تعمیم یافته را مورد بررسی قرار داد و همچنین بالاکریشنان و کاتلر[۵] (۱۹۹۴) برآوردهای*ML* مربوط به توزیع لاپلاس در رگرسیون را مورد توجه قرار دادند. نظریه مجانبی و نسبت درستنمایی برای رگرسیونL_1 و آماره‌های چندگانه لاگرانژ توسط کانکر و باست (۱۹۸۲) توسعه داده شده است.

۲,۳ استنباط بیزی در رگرسیون چندکی

نظریه رگرسیون چندکی که توسط کانکر و باست (۱۹۷۸) معرفی شد، روی مدل خطی

$$y_i = \boldsymbol{x}_i{}^T\boldsymbol{\beta} + u_i, \quad i = 1, \dots, n \qquad (1.3)$$

در نظر گرفته شد که در آن، u_i ها خطاهای مستقل و هم توزیع، با تابع توزیعی متقارن حول صفر هستند. بنابراین p امین چندک رگرسیونی ($0 < p < 1$)، بر اساس تابع زیان قدرمطلق انحرافات به صورت

$$\min_{\beta}\left[\sum_{i|y_i \geq x_i{}^T\beta} p|y_i - x_i{}^T\beta| + \sum_{i|y_i < x_i{}^T\beta}(1-p)|y_i - x_i{}^T\beta|\right],$$

[۱] *Harter*

[۲] *Adaptive Robust Estimates*

[۳] *Damseleth and El-Shaarawi*

[۴] *Efron*

[۵] *Balakrishnan and Cutler*

به‌دست می‌آید. برای $p = \frac{1}{2}$ این برآوردگر معادل با رگرسیون L_1 است. یعنی برآوردگر رگرسیون چندکی، تعمیم‌یافته برآوردگر L_1 است. بنابراین اگر بخواهیم یک مدل‌بندی بیزی از رگرسیون چندکی ارایه دهیم، باید از رگرسیون L_1 استفاده کنیم.

در موارد بسیاری ممکن است در استنباط مجانبی مورد نظر، تردید داشته باشیم و استنباط نمونه کم مد نظر باشد. در این راستا، روش‌های خودگردان‌ساز یک انتخاب برای به‌دست آوردن خطاهای استاندارد با اندازه نمونه کوچک است (بوچینسکی، ۱۹۹۸؛ کول، ۱۹۹۲).

برای توسعه رهیافت بیزی، مدل خطی(۱,۳) را با چگالی

$$f(u_i) \propto \tau^{-1} exp\left[-\tau^{-1}|u_i|\{pI_{[0,\infty)}(u_i) + (1-p)I_{(-\infty,0)}(u_i)\}\right] \qquad (2.3)$$

برای جمله خطا در نظر می‌گیریم که در آن، τ پارامتر مقیاس است و ضریب تناسب، به دلیل وابستگی به مقدار معلوم p، بی‌اهمیت است.

توزیعی که در رابطه (۲,۳) بیان شد، از نوع لاپلاس چوله است. این چولگی ناشی از اختلاف پارامترهای p و $1-p$ به‌ترتیب برای مقادیر مثبت و منفی u می‌باشد. بنابراین تابع درستنمایی به صورت زیر می‌باشد (تزیوناس، ۲۰۰۳)

$$L_p(\beta, \tau; Y, X) \propto \prod_{i=1}^{n} f(u_i)$$

$$L_p(\beta, \tau; Y, X) \propto \tau^{-n} exp\left\{-\tau^{-1} \sum_{i=1}^{n} |y_i - x_i{}^T\beta|.\{pI_{[0,\infty)}(y_i - x_i{}^T\beta) + (1-p)I_{(-\infty,0]}(y_i - x_i{}^T\beta)\}\right\},$$

معادل با ماکزیمم کننده تابع درستنمایی برآوردگر رگرسیون چندکی β است. فرض می‌کنیم توزیع پیشین، $P(\beta, \tau) \propto \tau^{-1}$ باشد. بنابراین توزیع پسین به صورت

$$P_p(\beta, \tau|Y, X) \propto P(\beta, \tau) \times L_p(\beta, \tau; Y, X)$$

$$\propto \tau^{-(n+1)} exp\left\{-\tau^{-1}\sum_{i=1}^{n}|y_i - {x_i}^T\beta|.\{pI_{[0,\infty)}(y_i - {x_i}^T\beta) + (1-p)I_{(-\infty,0]}(y_i - {x_i}^T\beta)\}\right\}, \qquad (3.3)$$

حاصل می‌شود (تزیوناس، ۲۰۰۳). با انتگرال‌گیری از پارامتر مقیاس τ، توزیع پسین حاشیه‌ای ضرایب رگرسیون را به‌دست می‌آوریم.

اکنون فرض کنید

$$c = \sum_{i=1}^{n}|y_i - {x_i}^T\beta|.\{pI_{[0,\infty)}(y_i - {x_i}^T\beta) + (1-p)I_{(-\infty,0]}(y_i - {x_i}^T\beta)\}.$$

بنابراین باید از $\tau^{-(n+1)}exp\{-\tau^{-1}c\}$ نسبت به τ انتگرال بگیریم. در این راستا داریم

$$P_p(\beta|Y,X) = \int \tau^{-(n+1)}exp\{-\tau^{-1}c\}d\tau$$
$$= \frac{\Gamma(n)}{c^n}\int_0^{\infty}\tau^{-(n+1)}exp\{-\tau^{-1}c\}d\tau = \frac{\Gamma(n)}{c^n}.$$

بنابراین

$$P_p(\beta|Y,X) \propto \left\{\sum_{i=1}^{n}|y_i - {x_i}^T\beta|.\{pI_{[0,\infty)}(y_i - {x_i}^T\beta) + (1-p)I_{(-\infty,0]}(y_i - {x_i}^T\beta)\}\right\}^{-n}$$

این توزیع پسین را با استفاده از تکنیک‌های تحلیلی نمی‌توان مورد مطالعه قرار داد. بنابراین ناگزیر به استفاده از روش‌های عددی خواهیم بود.

نمایش نرمال آمیخته[۱] از توزیع لاپلاس، به‌ویژه، با تابع چگالی احتمال

$$f(u_i|\omega_i) \propto (\sigma^2\omega_i)^{-\frac{1}{2}}exp\left\{-\frac{{u_i}^2}{2\sigma^2\omega_i}\left[pI_{[0,\infty)}(u_i) + (1-p)I_{(-\infty,0]}(u_i)\right]\right\}, \qquad (4.3)$$

را برای پارامتر مقیاس جدید $\sigma = \sqrt{2\tau}$ بررسی می‌کنیم.

[۱] *Normal Mixture Representation*

توزیع نمایی استاندارد$P(\omega_i) = exp(-\omega_i)$ را در نظر بگیرید. بعد از انتگرال‌گیری نسبت به متغیر وزن‌های پنهان[۱] w، تابع چگالی دلخواه به‌دست می‌آید. اگر بردار پارامتر به‌وسیله متغیر وزن‌های پنهانω_i داده‌افزایی شود، چگالی پسین توام به صورت

$$P(\beta, \sigma, w|Y, X) \propto \sigma^{-(n+1)} \prod_{i=1}^{n} (\omega_i)^{\frac{-1}{2}} exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} \frac{(y_i - {x_i}^T\beta)^2}{\omega_i} \left[pI_{[0,\infty)}(y_i - {x_i}^T\beta) + (1-p)I_{(-\infty,0]}(y_i - {x_i}^T\beta)\right] - \omega_i\right\},$$

به‌دست می‌آید. استنباط بیزی بر اساس روش نمونه‌گیری گیبز (گلفاند و اسمیت[۲]، ۱۹۹۰؛ تانر و ونگ[۳]، ۱۹۸۷) انجام خواهد گرفت.

هدف از نمونه‌گیری گیبز، تولید داده‌های تصادفی از یک توزیع چندمتغیره، و تشکیل یک زنجیر مارکوف در فضای پارامتر می‌باشد، که این مستلزم دانستن اطلاعات در خصوص همه توزیع‌های شرطی است. تحت شرایط نظم (روبرتز[۴] و اسمیت، ۱۹۹۴) نمونه‌های تولید شده از نمونه‌گیری گیبز نمی‌توانند مستقل باشند، زیرا تولید اعداد متوالی وابسته‌اند.

اجرای موثر نمونه‌گیری گیبز، نیازمند توزیع‌های شرطی پسینی است که دارای صورت بسته بوده و تولید نمونه از آن‌ها راحت و آسان باشد.

می‌دانیم اگر $E(y_i) = \mu$ و $var(y_i) = \sigma^2$، در این‌صورت

$$\frac{(y_i - \mu)^2}{\sigma^2} \sim \chi^2(1),$$

و توزیع پسین شرطیσ^2 به صورت

[۱] *Latent Weights Variable*

[۲] *Gelfand and Smith*

[۳] *Tanner and Wong*

[۴] *Roberts*

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i - {x_i}^T\beta)^2}{\omega_i}|(\beta, W, Y, X)\sim\chi^2(n),$$

به‌دست می‌آید. توزیع پسین شرطی هر وزن ناهم‌پراشی ω_i به صورت

$$P(\omega_i|\beta,\sigma,Y,X) \propto {\omega_i}^{\frac{-1}{2}} exp\{-q_i(\beta,\sigma).{\omega_i}^{-1} - \omega_i\},$$

خواهد بود که در آن، $q_i(\beta,\sigma) = \frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2$ انتخاب اعداد تصادفی از توزیع شرطی متغیر پنهان ω_i بر اساس روش نمونه‌گیری پذیرش بهینه از یک توزیع گاما(فرناندز و استیل[1]، ۱۹۹۸) است. توزیع پسین شرطی پارامترهای رگرسیون β نیز به صورت

$$P(\beta|\sigma,W,Y,X) \propto exp\left\{\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{(y_i - {x_i}^T\beta)^2}{\omega_i}\left[pI_{[\circ,\infty)}(y_i - {x_i}^T\beta) + (1-p)I_{(-\infty,\circ]}(y_i - {x_i}^T\beta)\right] - \omega_i\right\},$$

است.

۳,۳ رگرسیون چندکی بیزی

در این بخش، ایده رگرسیون چندکی بیزی با استفاده از یک تابع درستنمایی که بر اساس توزیع لاپلاس نامتقارن[2] بنا شده است، معرفی می‌شود. علاوه بر این، نشان داده می‌شود که صرفنظر از توزیع اصلی داده‌ها، استفاده از توزیع لاپلاس نامتقارن راهی موثر و طبیعی برای مدل‌بندی رگرسیون چندکی بیزی است. همچنین توضیح خواهیم داد که پیشین‌های یکنواخت ناسره برای پارامترهای نامعلوم مدل، یک توزیع پسین توام سره را نتیجه می‌دهد.

نظریه کلاسیک مدل‌های خطی، نظریه‌ای برای مدل‌های میانگین شرطی،$E(Y|x)$، است. البته در بسیاری از کاربردها خارج شدن از این رده از مدل‌ها می‌تواند مفید باشد. به‌تدریج مشخص شده است که رگرسیون چندکی، روشی جامع برای تحلیل آماری مدل‌های پاسخ خطی و غیرخطی است.

[1] *Fernandez and Steel*

[2] *Asymmetric Laplace Distribution*

رگرسیون چندکی، تمرکز تکنیک‌های کمترین توان‌های دوم برای برآورد توابع میانگین شرطی را با یک روش عمومی برای برآورد خانواده‌های توابع چندک شرطی، تکمیل می‌کند. این امر انعطاف‌پذیری روش‌های رگرسیون پارامتری و ناپارامتری را تا حد زیادی توسعه می‌دهد.

در مدل‌های خطی، یک روش ساده برای برآورد چندک شرطی توسط کانکر و باست (۱۹۷۸) ارایه شده است. مدل خطی استاندارد به صورت

$$y_t = \mu(\boldsymbol{x}_t) + \varepsilon_t, \qquad (5.3)$$

می‌باشد، که در آن، $\mu(\boldsymbol{x}_t)$ میانگین y_t به شرط بردار $\boldsymbol{x}_t$ و ε_t جمله خطا با میانگین صفر و واریانس ثابت است. مشخص کردن توزیع ε_t ضروری نیست. در مدل (۵,۳) برای بردار پارامترها $\boldsymbol{\beta}$ می‌توان نوشت $\mu(\boldsymbol{x}_t) = \boldsymbol{x}_t^T\boldsymbol{\beta}$. p امین چندک $(0< p < 1)$، ε_t مقداری مثل q_p، است که $P(\varepsilon_t < q_p) = p$. p امین چندک از y_t به شرط $\boldsymbol{x_t}$ نیز به سادگی به صورت

$$q_p(y_t|\boldsymbol{x_t}) = \boldsymbol{x}_t^T\boldsymbol{\beta}(p),$$

بیان می‌شود که در آن، $\boldsymbol{\beta}(p)$ بردار ضرایب وابسته به p می‌باشد.

$\widehat{\boldsymbol{\beta}}(p)$، p امین $(0< p < 1)$ چندک رگرسیونی، می‌تواند جواب معادله

$$\min_{\boldsymbol{\beta}} \sum_t \rho_p(y_t - \boldsymbol{x}_t^T\boldsymbol{\beta}),$$

باشد، که در آن تابع زیان به صورت

$$\rho_p(u) = u.\big(p - I(u <0)\big), \qquad (6.3)$$

تعریف می‌شود. به طور معادل رابطه (۶,۳) را می‌توان به صورت

$$\rho_p(u) = u.\big(pI(u >0) - (1-p)I(u <0)\big),$$

و یا

$$\rho_p(u) = \frac{|u| + (2p-1)u}{2},$$

نیز نوشت.

برخلاف حالت معمول که از تابع زیان درجه دوم برای برآورد رگرسیون میانگین استفاده می‌شود (هوبر[1]، ۱۹۸۱)، در برآورد رگرسیون چندکی، از رده مخصوصی از توابع زیان که دارای ویژگی‌های تنومند هستند استفاده می‌کند. برای مثال، در حضور نقاط پرت در مشاهدات، رگرسیون شرطی میانه (بر اساس قدر مطلق) نیرومندتر از رگرسیون میانگین شرطی است.

بکارگیری استنباط بیزی در مدل‌های خطی تعمیم‌یافته، امروزه امری نسبتاً متداول است. استفاده از روش‌های $MCMC$[2] (پیوست الف) برای به‌دست آوردن توزیع پسین، حتی در برخی وضعیت‌های پیچیده، استنباط بیزی را خیلی مفید و جذاب جلوه می‌دهد. برخلاف روش‌های متعارف در استنباط بیزی، توزیع پسین برای کل پارامترها قابل محاسبه است و با استفاده از آن می‌توان برآوردگرهای مناسبی برای پارامترها به‌دست آورد.

یو و موید[3] (۲۰۰۱) یک روش بیزی برای رگرسیون چندکی ارایه دادند که کاملاً متفاوت از مطالبی است که قبل از آن در سایر منابع آورده شده است. در این روش، بدون توجه به توزیع واقعی داده‌ها، از استنباط بیزی برای رگرسیون چندکی، با استفاده از تابع درستنمایی مبتنی بر توزیع لاپلاس نامتقارن، استفاده می‌شود. به طور کلی، ممکن است هر پیشینی انتخاب شود، حتی پیشین‌های یکنواخت ناسره. در این حالت نشان داده می‌شود توزیع پسین توأم، سره می‌باشد.

یو و موید (۲۰۰۱) نشان دادند که می‌نیمم کردن تابع زیان رابطه (۶,۳)، دقیقاً معادل با ماکسیمم کردن تابع درستنمایی است که از ترکیب چگالی‌های مستقل لاپلاس نامتقارن به‌دست می‌آید. برای این منظور، ابتدا ویژگی توزیع لاپلاس نامتقارن را یادآوری می‌کنیم.

---

[1] *Huber*

[2] *Markov Chain Monte Carlo*

[3] *Moyeed*

تعریف: متغیر تصادفی U دارای توزیع لاپلاس نامتقارن است، اگر تابع چگالی احتمال آن به صورت

$$f_p(u) = p(1-p)exp\{-\rho_p(u)\}, \qquad (7.3)$$

باشد که در آن، $0 < p < 1$ و $\rho_p(u)$ تابعی است که در رابطه (۶,۳) تعریف شد.

هنگامی‌که $p = \frac{1}{2}$ باشد، رابطه (۷,۳) به صورت $\frac{1}{4}exp(\frac{-|u|}{2})$ تبدیل می‌شود که همان تابع چگالی توزیع لاپلاس متقارن استاندارد است. برای سایر مقادیرp، چگالی رابطه (۷,۳) نامتقارن است. میانگین U برابر $\frac{1-2p}{p(1-p)}$ و فقط برای $p < \frac{1}{2}$ مثبت بوده و واریانس U برابر $\frac{1-2p+2p^2}{p^2(1-p)^2}$ است.

با توجه به تابع چگالی توزیع لاپلاس نامتقارن داریم

$$f_p(u) = \int_{-\infty}^{\circ} p(1-p)e^{-u(p-1)}\,du + \int_{\circ}^{\infty} p(1-p)e^{-up}\,du = 1,$$

$$E(u) = p(1-p)\int_{-\infty}^{\circ} ue^{-u(p-1)}\,du + p(1-p)\int_{\circ}^{\infty} ue^{-up}\,du$$

که با انتگرال‌گیری جز به جز و در نظر گرفتن $\int_{-\infty}^{\circ} e^{-u(p-1)}\,du = dv$ برای جمله اول و $\int_{\circ}^{\infty} ue^{-up}\,du = dv$ برای جمله دوم داریم:

$$E(u) = p(1-p)\left[\frac{-ue^{-u(p-1)}}{p-1} + \int_{-\infty}^{\circ}\frac{e^{-u(p-1)}}{p-1}du\right] + p(1-p)\left[\frac{-ue^{-up}}{p} + \int_{\circ}^{\infty}\frac{e^{-up}}{p}du\right]$$

$$= p(1-p)\left[\frac{-e^{-u(p-1)}}{(p-1)^2}|_{-\infty}^{\circ}\right] + p(1-p)\left[\frac{-e^{-up}}{p^2}|_{\circ}^{+\infty}\right] = \frac{-p}{1-p} + \frac{1-p}{p}$$

$$= \frac{1-2p}{p(1-p)}$$

همچنین برای واریانس داریم از انتگرال‌گیری جز به جز استفاده می‌کنیم.

$$E(u^2) = p(1-p)\int_{-\infty}^{\circ} u^2e^{-u(p-1)}\,du + p(1-p)\int_{\circ}^{\infty} u^2e^{-up}\,du$$

که همان فرضیات قبلی را در نظر می‌گیریم

$$E(u^2) = p(1-p)\left[\frac{-u^2e^{-u(p-1)}}{p-1} + \frac{2}{p-1}\int_{-\infty}^{\circ} ue^{-u(p-1)}\,du\right]$$
$$+ p(1-p)\left[\frac{-u^2e^{-up}}{p} + \frac{2}{p}\int_{\circ}^{\infty} ue^{-up}\,du\right]$$

با استفاده از انتگرال امید ریاضی داریم

$$E(u^2) = p(1-p)\left[\frac{2}{p-1}\left(\frac{1}{(p-1)^2}\right)\right] + p(1-p)\left[\frac{2}{p}\left(\frac{1}{p^2}\right)\right] = \frac{-2p}{(1-p)^2} + \frac{2(1-p)}{p^2}$$
$$= \frac{6p^2-6p+2}{p^2(1-p)^2}$$

بنابراین

$$var(u) = E(u^2) - (E(u))^2 = \frac{2p^2-2p+1}{p^2(1-p)^2}.$$

در صورتی که p به صفر یا یک نزدیک شود، این واریانس به سرعت افزایش می‌یابد.

می‌توان پارامترهای مقیاس و مکان،σ و μ، را به ترتیب در چگالی (۷,۳) اضافه کرد، که منجر به تابع چگالی احتمال به صورت زیر می‌شود.

$$f_p(u;\mu,\sigma) = \frac{p(1-p)}{\sigma} exp\left\{-\rho_p\left(\frac{u-\mu}{\sigma}\right)\right\},$$

را به‌دست آورد.

در مدل خطی تعمیم‌یافته معمولی،که برای داده‌های غیرنرمال استفاده می‌شوند، برآورد پارامترهای نامعلوم رگرسیونی $\boldsymbol{\beta}$ تحت مفروضاتی به‌دست می‌آیند که عبارتند از:

۱. متغیرهای تصادفی Y_i، $i = 1, \dots, n$، به شرط $\boldsymbol{x}$، مستقل و دارای توزیع $f(y;\mu_i)$ هستند که μ_i ها با رابطه $\mu_i = E[Y_i|\boldsymbol{x_i}]$ تعریف می‌شوند.

۲. برای تابع پیوند[۱] معلوم g، $g(\mu_i) = \boldsymbol{x_i}^T\boldsymbol{\beta}$

[۱] *Link Function*

در این ساختار شناخته‌شده، یک توزیع نرمال برای y_iها با یک تابع پیوند همانی، همان مدل خطی کلاسیک است و برآوردگر $\boldsymbol{\beta}$ تحت تابع زیان درجه دوم به‌دست می‌آید.

در صورتی‌که به جای میانگین شرطی $E[Y_i|\boldsymbol{x_i}]$، علاقه‌مند به چندک شرطی، $q_p(y_i|\boldsymbol{x_i})$، باشیم هنوز هم می‌توانیم این مسأله را در قالب مدل خطی تعمیم‌یافته، بررسی کنیم. این بررسی با در نظر گرفتن فرض‌های زیر صورت می‌گیرد:

۱. $f(y,\mu_i)$، یک تابع چگالی لاپلاس نامتقارن است.

۲. برای هر $0< p <1$، $g(\mu_i) = \boldsymbol{x}_i{}^T\boldsymbol{\beta}_{(p)} = q_p(y_i|\boldsymbol{x_i})$.

برای انجام استنباط بیزی، گرچه توزیع پیشین مزدوج استاندارد برای پارامترهای رگرسیون چندکی در دسترس نیست، اما می‌توان از روش‌های نمونه‌گیری $MCMC$ برای محاسبه توزیع پسین پارامترهای نامعلوم، با توجه به هر پیشینی که در نظر گرفته شود، استفاده کرد (یو و موید، ۲۰۰۱). این روش، توزیع‌های پسین توأم و حاشیه‌ای تمام پارامترهای نامعلوم را مشخص می‌کند.

اگر توزیع پیشین $\boldsymbol{\beta}$ را $p(\boldsymbol{\beta})$ در نظر بگیریم، بر اساس مشاهدات $\boldsymbol{y} = (y_1, \dots, y_n)$ ، تابع درستنمایی $L(\boldsymbol{y}|\boldsymbol{\beta})$ قابل محاسبه بوده و توزیع پسین به صورت

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto L(\boldsymbol{y}|\boldsymbol{\beta})P(\boldsymbol{\beta}), \qquad (8.3)$$

نتیجه می‌شود. می‌توان تابع درستنمایی را بر اساس توزیع لاپلاس نامتقارن و پارامتر مکان $\mu_i = \boldsymbol{x}_i{}^T\boldsymbol{\beta}$ به صورت

$$L(\boldsymbol{y}|\boldsymbol{\beta}) = p^n(1-p)^n exp\left\{-\sum_i \rho_p(y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta})\right\},$$

تابعی که رابطه بین پیش بینی خطی و میانگین(متوسط) تابع توزیع را بیان می‌کند.

نوشت. از لحاظ نظری، می‌توان از هر توزیع پیشین در رابطه (۸,۳) استفاده کرد. اما در نبود اطلاعات واقع‌بینانه، می‌توان از توزیع پیشین یکنواخت ناسره برای تمام اعضای $\boldsymbol{\beta}$ استفاده کرد، زیرا در توزیع یکنواخت احتمال وجود مقادیر مختلف پارامترها با هم یکسان است. در این‌صورت، توزیع پسین توأم حاصل، متناسب با تابع درستنمایی است.

۴,۳ پیشین‌های ناسره برای پارامترها

قضیه و لم زیر نشان می‌دهند که اگر توزیع پیشین $\boldsymbol{\beta}$ یکنواخت ناسره انتخاب شود، توزیع پسین توأم سره می‌باشد.

لم ۱,۳: فرض کنید بردار $\boldsymbol{\beta}$ فقط یک مولفه داشته باشد، در این‌صورت p امین چندک شرطی به صورت $q_p(y|x) = \beta(p)$ می‌باشد، که برای سادگی، می‌نویسیم $\beta = \beta(p)$. حال اگر فرض کنیم $p(\beta) \propto 1$، آن‌گاه

$$0 < \int L(y|\beta)d\beta < \infty.$$

اثبات: بدون کاستن از کلیت، فرض می‌کنیم مشاهدات $\{y_j\}_{j=1}^{n}$ به صورت $y_1 \leq y_2 \leq \cdots \leq y_n$ مرتب شده باشند، همچنین فرض می‌کنیم n و p معلوم باشند. در این‌صورت

$$\int_{-\infty}^{+\infty} L(y|\beta)d\beta = p^n(1-p)^n \int_{-\infty}^{+\infty} exp\left[-\sum_{j=1}^{n}(y_j-\beta)\{p - I(y_j \leq \beta)\}\right]d\beta$$

از آنجایی‌که مشاهدات y_j صعودی هستند، پس از y_n که بزرگترین مشاهده است استفاده می‌کنیم. داریم:

$$\int_{-\infty}^{+\infty} L(y|\beta)d\beta \geq p^n(1-p)^n \int_{y_n}^{+\infty} exp\left[-\sum_{j=1}^{n}(y_j-\beta)\{p-I(y_j\leq\beta)\}\right]d\beta$$

$$= p^n(1-p)^n \int_{y_n}^{+\infty} exp\left[(1-p)\sum_{j=1}^{n}(y_j-\beta)\right]d\beta$$

$$= \frac{p^n(1-p)^n}{(n(1-p))} exp\left\{(1-p)\sum_{j=1}^{n}(y_j-y_n)\right\}$$

$$= \frac{p^n(1-p)^{n-1}}{n} exp\left\{(1-p)\sum_{j=1}^{n-1}(y_j-y_n)\right\}$$

$\geq$۰.

در این‌قسمت با تغییر متغیر در نظر می‌گیریم $(y_j-\beta)\{p-I(y_j\leq\beta)\}=x_j$. با استفاده مکرر از نامساوی کوشی- شوارتز داریم:

$$\int_{-\infty}^{+\infty}\prod_{j=1}^{n} exp[-(y_j-\beta)\{p-I(y_j\leq\beta)\}]\,d\beta = \int e^{-x_1}e^{-x_2}\dots e^{-x_n}\,d\boldsymbol{x}$$

$$= \left[\left(\int e^{-x_1}e^{-x_2}\dots e^{-x_n}\,d\boldsymbol{x}\right)^2\right]^{\frac{1}{2}} \leq \left(\int e^{-2x_1}\,dx_1\right)^{\frac{1}{2}}\left(\int e^{-2x_2}\dots e^{-2x_n}\,d\boldsymbol{x}_{-1}\right)^{\frac{1}{2}}$$

$$= \left[\int e^{-2x_1}dx_1\right]^{\frac{1}{2}}\left[\left(\int e^{-2x_2}\dots e^{-2x_n}\,d\boldsymbol{x}_{-1}\right)^2\right]^{\frac{1}{4}}$$

$$\leq \left[\int e^{-2x_1}dx_1\right]^{\frac{1}{2}}\left(\int e^{-4x_2}\,dx_2\right)^{\frac{1}{4}}\left(\int e^{-4x_3}\dots e^{-4x_n}\,d\boldsymbol{x}_{-1,2}\right)^{\frac{1}{4}}$$

$$= \left[\int e^{-2x_1}dx_1\right]^{\frac{1}{2}}\left[\int e^{-4x_2}\,dx_2\right]^{\frac{1}{4}}\left[\left(\int e^{-4x_3}\dots e^{-4x_n}\,d\boldsymbol{x}_{-1,2}\right)^2\right]^{\frac{1}{8}}$$

$$\leq \left[\int e^{-2x_1}dx_1\right]^{\frac{1}{2}}\left[\int e^{-4x_2}\,dx_2\right]^{\frac{1}{4}}\left[\left(\int e^{-8x_3}\,dx_3\right)\left(\int e^{-8x_4}\dots e^{-8x_n}\,dx_{-1,2,3}\right)\right]^{\frac{1}{8}}.$$

با ادامه این روند، تا مرحله *n-2*، داریم:

$$\int e^{-x_1}e^{-x_2}\dots e^{-x_n}\,d\boldsymbol{x}$$

$$\leq \left[\int e^{-2x_1}dx_1\right]^{\frac{1}{2}}\dots\left[\int e^{-2^{n-2}x_{n-2}}dx_{n-2}\right]^{\frac{1}{2^{n-2}}}\left[\left(\int e^{-2^{n-2}x_{n-1}}e^{-2^{n-2}x_n}\,dx_{n-1}\right)^2\right]^{\frac{1}{2^{n-1}}}$$

$$\leq \left[\int e^{-2x_1}dx_1\right]^{\frac{1}{2}}\dots\left[\int e^{-2^{n-2}x_{n-2}}dx_{n-2}\right]^{\frac{1}{2^{n-2}}}\left(\int e^{-2^{n-1}x_{n-1}}\,dx_{n-1}\right)^{\frac{1}{2^{n-1}}}\left(\int e^{-2^{n-1}x_n}\,dx_n\right)^{\frac{1}{2^{n-1}}}$$

$$=\prod_{j=1}^{n-1}\left[\int e^{-2^j x_j}\,dx_j\right]^{\frac{1}{2^j}}\left[\int e^{-2^{n-1}x_n}\,dx_n\right]^{\frac{1}{2^{n-1}}}.$$

بنابراین، می‌توان نوشت

$$\int_{-\infty}^{+\infty}L(y|\beta)d\beta = p^n(1-p)^n\int_{-\infty}^{+\infty}\prod_{j=1}^{n}exp\left[-(y_j-\beta)\{p-I(y_j\leq\beta)\}\right]d\beta$$

$$\leq\ p^n(1-p)^n\prod_{j=1}^{n-1}\left(\int_{-\infty}^{+\infty}exp\left[-2^j(y_j-\beta)\{p-I(y_j\leq\beta)\}\right]d\beta\right)^{\frac{1}{2^j}}$$

$$\times\left(\int_{-\infty}^{+\infty}exp[-2^{n-1}(y_n-\beta)\{p-I(y_n\leq\beta)\}]\,d\beta\right)^{\frac{1}{2^{n-1}}}.$$

بنابراین

$$\int_{-\infty}^{+\infty} L(y|\beta)d\beta \le p^n(1-p)^n$$

$$\times \prod_{j=1}^{n-1}\left(\int_{-\infty}^{y_j} exp\left[-2^j p\left(y_j-\beta\right)\right]d\beta + \int_{y_j}^{+\infty} exp\left[2^j(1-p)\left(y_j-\beta\right)\right]d\beta\right)^{\frac{1}{2^j}}$$

$$\times \left(\int_{-\infty}^{y_n} exp[-2^{n-1}p(y_n-\beta)]\,d\beta + \int_{y_n}^{+\infty} exp[2^{n-1}(1-p)(y_n-\beta)]d\beta\right)^{\frac{1}{2^{n-1}}}$$

$$= p^n(1-p)^n \times \prod_{j=1}^{n-1}\left(\frac{1}{2^j p}+\frac{1}{2^j(1-p)}\right)^{\frac{1}{2^j}} \times \left(\frac{1}{2^{n-1}p}+\frac{1}{2^{n-1}(1-p)}\right)^{\frac{1}{2^{n-1}}}$$

$$= p^n(1-p)^n \times \prod_{j=1}^{n-1}\left(\frac{1}{2^j p(1-p)}\right)^{\frac{1}{2^j}} \times \left(\frac{1}{2^{n-1}p(1-p)}\right)^{\frac{1}{2^{n-1}}}$$

$$= p^n(1-p)^n \times \prod_{j=1}^{n-1}\frac{1}{\frac{j}{2^{2^j}}} \times \frac{1}{2^{\frac{n-1}{2^{n-1}}}} \times \frac{1}{(p(1-p))^{\sum_{j=1}^{n-1}\frac{1}{2^j}}} \times \frac{1}{(p(1-p))^{\frac{1}{2^{n-1}}}} < \infty.$$

از آن‌جایی که

$$\sum_{j=1}^{n-1}\frac{1}{2^j} = \frac{2^{n-1}-1}{2^{n-1}} < \infty, \qquad \prod_{j=1}^{n-1}\frac{1}{\frac{j}{2^{2^j}}} = \frac{1}{2^{\frac{2^n-(n+1)}{2^{n-1}}}} < \infty.$$

$$p(1-p) \le \frac{1}{4}$$

بنابراین

$$\int L(y|\beta)d\beta \le p^{n-1}(1-p)^{n-1} \times \frac{1}{2^{\frac{2^n-2}{2^{n-1}}}} < \infty.$$

و لم ۱,۳ ثابت می‌شود.□

اکنون با استفاده از لم ۱,۳، می‌توان قضیه *1.3* بیان و اثبات کرد.

قضیه*1.3*: اگر تابع درستنمایی، لاپلاس نامتقارن باشد و $P(\boldsymbol{\beta}) \propto 1$، آن‌گاه توزیع پسین $\boldsymbol{\beta}$، $\pi(\boldsymbol{\beta}|\boldsymbol{y})$، سره است. یعنی

$$0 < \int \pi(\boldsymbol{\beta}|\boldsymbol{y}) d\boldsymbol{\beta} < \infty,$$

و به طور معادل

$$0 < \int L(\boldsymbol{y}|\boldsymbol{\beta}) P(\boldsymbol{\beta}) d\boldsymbol{\beta} < \infty.$$

اثبات: فرض کنید $\boldsymbol{x}_{\boldsymbol{j}_{-q}}^T$ و $\boldsymbol{\beta}_{-q}$ نشان‌دهنده $\boldsymbol{x}_j^T$ و $\boldsymbol{\beta}$ بدون q امین مولفه باشند. بدون کاستن از کلیت، فرض کنید که x_{qj} نشان‌دهنده j امین مشاهده از مولفه q باشد و برای $j = 1, \dots, n$، $x_{qj} > 0$ و $v_j(\boldsymbol{\beta})$ نشان‌دهنده وزن متغیر پاسخ نسبت به j امین مشاهده از مولفه q باشد، به طوری که $v_j(\boldsymbol{\beta}) = \frac{(y_j - \boldsymbol{x}_{j_{-q}}^T \boldsymbol{\beta}_{-q})}{x_{qj}}$. همچنین فرض کنید $v_1(\boldsymbol{\beta}) \le v_2(\boldsymbol{\beta}) \le \cdots \le v_n(\boldsymbol{\beta})$. ابتدا نشان می‌دهیم $y_j \le \boldsymbol{x}_j^T \boldsymbol{\beta}$، اگر و فقط اگر $v_j(\boldsymbol{\beta}) \le \beta_q$ و $y_j - \boldsymbol{x}_j^T \boldsymbol{\beta} = (v_j(\boldsymbol{\beta}) - \beta_q) x_{qj}$.

فرض کنید $y_j \le \boldsymbol{x}_j^T \boldsymbol{\beta}$، در این‌صورت

$$v_j(\boldsymbol{\beta}) x_{qj} + \boldsymbol{x}_{\boldsymbol{j}_{-q}}^T \boldsymbol{\beta}_{-q} = y_j \le \boldsymbol{x}_j^T \boldsymbol{\beta}$$

پس $v_j(\boldsymbol{\beta}) x_{qj} \le \boldsymbol{x}_j^T \boldsymbol{\beta} - \boldsymbol{x}_{\boldsymbol{j}_{-q}}^T \boldsymbol{\beta}_{-q}$ می‌شود. بنابراین

$$v_j(\boldsymbol{\beta}) x_{qj} \le x_{qj} \beta_q \Rightarrow v_j(\boldsymbol{\beta}) \le \beta_q.$$

می‌شود و

$$v_j(\boldsymbol{\beta}) x_{qj} = y_j - \boldsymbol{x}_{\boldsymbol{j}_{-q}}^T \boldsymbol{\beta}_{-q}$$

پس $v_j(\boldsymbol{\beta}) x_{qj} = y_j - \boldsymbol{x}_j^T \boldsymbol{\beta} + x_{qj} \beta_q$ می‌شود و

$$(v_j(\boldsymbol{\beta}) - \beta_q)x_{qj} = y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}.$$

است. بنابراین اگر $p(\boldsymbol{\beta}) \propto 1$، آن‌گاه

$$\int_{-\infty}^{+\infty} L(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta} = p^n(1-p)^n \int_{-\infty}^{+\infty} exp\left[-\sum_{j=1}^{n}(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta})\{p - I(y_j \le \boldsymbol{x}_j^T\boldsymbol{\beta})\}\right] d\boldsymbol{\beta}$$

$$= p^n(1-p)^n \int_{-\infty}^{+\infty} exp\left[-\sum_{j=1}^{n} x_{qj}(v_j(\boldsymbol{\beta}) - \beta_q)\{p - I(v_j(\boldsymbol{\beta}) \le \beta_q)\}\right] d\boldsymbol{\beta}$$

$$\ge p^n(1-p)^n \int d\boldsymbol{\beta}_{-q} \int_{v_n(\boldsymbol{\beta})}^{+\infty} exp\left[(1-p)\sum_{j=1}^{n} x_{qj}(v_j(\boldsymbol{\beta}) - \beta_q)\right] d\beta_q$$

$$= p^n(1-p)^n \frac{1}{(1-p)\sum_{j=1}^{n} x_{qj}}$$

$$\times \int exp\left\{-(1-p)\sum_{j=1}^{n} x_{qj}\left(v_n(\boldsymbol{\beta}) - v_j(\boldsymbol{\beta})\right)\right\} d\boldsymbol{\beta}_{-q}.$$

عبارت $\sum_{j=1}^{n} x_{qj}(v_n(\boldsymbol{\beta}) - v_j(\boldsymbol{\beta}))$ یک تابع خطی از $\boldsymbol{\beta}_{-q}$ است، و به کمک لم ۱,۳ می‌بینیم که $\int L(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta} > 0$. با استفاده از نامساوی کوشی- شوارتز، داریم:

$$\int_{-\infty}^{+\infty} L(y|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta} = p^n(1-p)^n \int_{-\infty}^{+\infty} \prod_{j=1}^{n} exp\left[-(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta})\{p - I(y_j \le \boldsymbol{x}_j^T\boldsymbol{\beta})\}\right] d\boldsymbol{\beta}$$

$$\le p^n(1-p)^n \prod_{j=1}^{n-1}\left(\int_{-\infty}^{+\infty} exp\left[-2^j(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta})\{p - I(y_j \le \boldsymbol{x}_j^T\boldsymbol{\beta})\}\right] d\boldsymbol{\beta}\right)^{\frac{1}{2^j}}$$

$$\times \left(\int_{-\infty}^{+\infty} exp\left[-2^{n-1}(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta})\{p - I(y_n \le \boldsymbol{x}_j^T\boldsymbol{\beta})\}\right] d\boldsymbol{\beta}\right)^{\frac{1}{2^{n-1}}}.$$

برای هر $1 \le j \le n$، فرض کنید $x_{qj} \neq 0$ و قرار دهید

$$I = \int_{-\infty}^{+\infty} exp\left[-2^j(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta})\{p - I(y_j \le \boldsymbol{x}_j^T\boldsymbol{\beta})\}\right] d\boldsymbol{\beta}.$$

اگر قرار دهیم$(\boldsymbol{\gamma}^T = (\boldsymbol{\gamma}_{-q}^T, \gamma_q) = (\boldsymbol{\beta}_{-q}^T, \boldsymbol{x}_j^T\boldsymbol{\beta})$ و تبدیل ژاکوبین $\left|\frac{d\boldsymbol{\beta}}{d\boldsymbol{\gamma}}\right| = \frac{1}{|x_{qj}|}$ را در نظر بگیریم، آن‌گاه

$$I = \frac{1}{|x_{qj}|}\int_{-\infty}^{+\infty} exp\left[-2^j(y_j - \gamma_q)\{p - I(y_j \le \gamma_q)\}\right] d\boldsymbol{\gamma},$$

و با استفاده از لم ۱,۳

$$\int_{-\infty}^{+\infty} L(y|\boldsymbol{\beta})d\boldsymbol{\beta} = p^n(1-p)^n \int_{-\infty}^{+\infty} \prod_{j=1}^{n} \frac{1}{|x_{qj}|} exp\left[-(y_j - \gamma_q)\{p - I(y_j \le \gamma_q)\}\right] d\boldsymbol{\gamma}$$

$$\le p^n(1-p)^n \prod_{j=1}^{n-1} \left(\frac{1}{|x_{qj}|}\int_{-\infty}^{+\infty} exp\left[-2^j(y_j - \gamma_q)\{p - I(y_j \le \gamma_q)\}\right] d\boldsymbol{\gamma}\right)^{\frac{1}{2^j}}$$

$$\times \left(\frac{1}{|x_{qn}|}\int_{-\infty}^{+\infty} exp\left[-2^{n-1}(y_n - \gamma_q)\{p - I(y_n \le \gamma_q)\}\right] d\boldsymbol{\gamma}\right)^{\frac{1}{2^{n-1}}}.$$

بنابراین داریم

$$\int_{-\infty}^{+\infty} L(y|\boldsymbol{\beta})d\boldsymbol{\gamma} \le p^n(1-p)^n$$
$$\times \prod_{j=1}^{n-1}\left(\frac{1}{|x_{qj}|}\int_{-\infty}^{y_j} exp\left[-2^j p(y_j-\gamma_q)\right]d\boldsymbol{\gamma} + \int_{y_j}^{+\infty} exp\left[2^j(1-p)(y_j-\gamma_q)\right]d\boldsymbol{\gamma}\right)^{\frac{1}{2^j}}$$
$$\times\left(\frac{1}{|x_{qn}|}\int_{-\infty}^{y_n} exp\left[-2^{n-1}p(y_n-\gamma_q)\right]d\boldsymbol{\gamma}\right.$$
$$\left.+\int_{y_n}^{+\infty} exp\left[2^{n-1}(1-p)(y_n-\gamma_q)\right]d\boldsymbol{\gamma}\right)^{\frac{1}{2^{n-1}}}$$

$$\le p^n(1-p)^n \times \prod_{j=1}^{n-1}\left(\int_{-\infty}^{y_j} exp\left[-2^j p(y_j-\beta)\right]d\boldsymbol{\gamma} + \int_{y_j}^{+\infty} exp\left[2^j(1-p)(y_j-\beta)\right]d\boldsymbol{\gamma}\right)^{\frac{1}{2^j}}$$
$$\times\left(\int_{-\infty}^{y_n} exp[-2^{n-1}p(y_n-\beta)]\,d\boldsymbol{\gamma} + \int_{y_n}^{+\infty} exp[2^{n-1}(1-p)(y_n-\beta)]d\boldsymbol{\gamma}\right)^{\frac{1}{2^{n-1}}}$$
$$= p^n(1-p)^n \times \prod_{j=1}^{n-1}\left(\frac{1}{2^j p}+\frac{1}{2^j(1-p)}\right)^{\frac{1}{2^j}} \times \left(\frac{1}{2^{n-1}p}+\frac{1}{2^{n-1}(1-p)}\right)^{\frac{1}{2^{n-1}}}$$
$$= p^n(1-p)^n \times \prod_{j=1}^{n-1}\left(\frac{1}{2^j p(1-p)}\right)^{\frac{1}{2^j}} \times \left(\frac{1}{2^{n-1}p(1-p)}\right)^{\frac{1}{2^{n-1}}}$$
$$= p^n(1-p)^n \times \prod_{j=1}^{n-1}\frac{1}{\frac{j}{2^{2^j}}} \times \frac{1}{2^{\frac{n-1}{2^{n-1}}}} \times \frac{1}{(p(1-p))^{\sum_{j=1}^{n-1}\frac{1}{2^j}}} \times \frac{1}{(p(1-p))^{\frac{1}{2^{n-1}}}} < \infty$$

اثبات کامل می‌شود. □

در عمل می‌توان فرض کرد که اجزاء $\boldsymbol{\beta}$ دارای توزیع پیشین یکنواخت ناسره مستقل هستند که این حالت خاص قضیه۱,۳ خواهد شد.

فصل چهارم

رگرسیون چندکی دودویی و توییت

۱,۴ مقدمه

کار اصلی کانکر و باست (۱۹۷۸) روی رگرسیون چندکی به تدریج به عنوان یک رهیافت جامع برای تحلیل آماری مدل‌های با پاسخ خطی و غیرخطی، در حال گسترش است. اخیراً، محققان بررسی و انجام رگرسیون چندکی از دیدگاه بیزی را شروع کرده‌اند. مدل رگرسیون چندکی بیزی، برای داده‌های مستقل،

اولین بار توسط والکر و مالیک[۱] (۱۹۹۹) معرفی شد و تدی و کوتاس[۲](۲۰۱۰) رهیافت آن‌ها را برای رگرسیون چندک توبیت[۳] گسترش دادند. همچنین گریسی و بوتای[۴] (۲۰۰۷) و ریچ[۵] و همکاران (۲۰۱۰)، رگرسیون چندکی بیزی را برای داده‌های خوشه‌ای مطرح کردند. به طور کلی، یک رهیافت بیزی نیاز به مشخص کردن رگرسیون چندکی برای تعیین تابع درستنمایی و یک رهیافت مشترک که از بکارگیری توزیع خطا لاپلاس نامتقارن است، دارد.

اولین بار کانکر و ماچادو[۶] (۱۹۹۹) ارتباط بین توزیع لاپلاس نامتقارن و رگرسیون چندکی را به‌دست آوردند. این یافته‌ها توسط یو و موید (۲۰۰۱) به کار برده شد و رگرسیون چندکی بیزی با اعمال توزیع لاپلاس نامتقارن روی جملات خطا در مدل خطی، مطرح شد. آن‌ها الگوریتم *MCMC* قدم زدن تصادفی متروپولیس-هستینگس[۷] را برای تولید نمونه از توزیع پسین به‌کار گرفتند. گریسی و بوتای (۲۰۰۷) نیز بکارگیری توزیع‌های لاپلاس نامتفارن را برای تحلیل داده‌های طول جغرافیایی در رگرسیون چندکی مطرح کردند. توجیه نظری بکارگیری توزیع لاپلاس نامتقارن را در کومجیر[۸](۲۰۰۵) می‌توان دید.

در رگرسیون چندکی، وقتی که بردار متغیرهای تبیینی شامل متغیرهای زیادی باشد، آن‌گاه انتخاب متغیر برای بهبود دقت مدل برازش شده، لازم می‌شود. در این موارد آماردانان معمولاً از روش‌های منظم‌سازی[۹] مانند لاسو [۱۰](وانگ[۱] و همکاران ۲۰۰۵) و لاسو سازوار[۲] (وو و لیو ۲۰۰۹)، برای انتخاب مدل

---

۱ *Walker and Mallick*
۲ *Taddy and Kottas*
۳ *Tobit Quantile Regression*
٤ *Geraci and Bottai*
٥ *Riech*
٦ *Machado*
۷ *Random Walk Metropolis-Hastings*
۸ *Komjuer*
۹ *Regularization Methods*
۱۰ *Least absolute shrinkage and selection operator*

در رگرسیون چندکی استفاده می‌کنند. در این روش‌ها، برخی از ضرایب رگرسیونی بطور خودکار به سمت صفر منقبض می‌شوند.

رید[۳] و همکاران (۲۰۱۰) یک رهیافت بیزی را با بکارگیری روش انتخاب متغیر با جستجوی تصادفی[۴] پیشنهاد دادند.

جی و لین و ژانگ[۵](۲۰۱۲) از روش انتخاب متغیر با جستجوی تصادفی بیزی در چارچوب رگرسیون چندکی دودویی و توبیت استفاده کردند. ما این فصل را با استفاده از همین فرض دنبال می‌کنیم.

۲,۴ مدل سلسله‌مراتبی بیزی

در طول دهه گذشته، رگرسیون‌های چندکی دودویی[۶] و توبیت توجه زیادی را در آمار کاربردی و نظری به خود جلب کرده‌اند. اولین بار مانسکی[۷] (۱۹۸۵) برآوردگر رگرسیون چندکی دودویی نیمه‌پارامتری را معرفی کرد و برآوردگر ماکسیمم امتیاز[۸] را پیشنهاد داد.

مدل‌های رگرسیون چندکی دودویی و توبیت هر دو می‌توانند به عنوان رگرسیون چندکی خطی با پاسخ‌های پیوسته پنهان که به‌طور کامل مشاهده نشده‌اند، در نظر گرفته شوند. مدل مورد نظر به صورت زیر است:

$$y_i^* = \boldsymbol{x}_i^T \boldsymbol{\beta} + u_i \quad \text{و} \quad y_i = g(y_i^*),\ i = 1, \dots, n. \qquad (1.4)$$

---

۱ *Wang*

۲ *adaptive lasso*

۳ *Reed*

٤ *Stochastic Search Variable Selection*

٥ *Ji and Lin and Zhang*

٦ *Binary Quantile Regression*

۷ *Manski*

۸ *Maximum score estimator*

که در آن، $g(.)$ تابع پیوند و y_i پاسخ مشاهده شده از i امین گزاره‌ای است که توسط پاسخ مشاهده نشده پنهان y_i^* تعیین شده است. یک مثال ساده از این روش، وزن بدن است که برای آن یک آستانه در نظر می‌گیریم. اگر وزن کمتر از آستانه باشد y_i را صفر قرار می‌دهیم و اگر بیشتر از آن باشد یک قرار می‌دهیم. $\boldsymbol{x}_i$ بردار q بعدی متغیرهای تبیینی، $\boldsymbol{\beta}$ بردار q بعدی پارامترها و u_i یک متغیر اختلال[۱] تصادفی است.

دو مورد خاص از مدل (۱,۴)، رگرسیون چندکی دودویی (مانسکی، ۱۹۸۵) و رگرسیون چندکی توبیت (پاول[۲]، ۱۹۸۶) هستند که به ترتیب دارای توابع پیوند $g(y_i^*) = I\{y_i^* > 0\}$ و $g(y_i^*) = max(y_i^*, c)$، برای c ثابت معلوم، می‌باشند.

روش انتخاب متغیر با جستجوی تصادفی، اولین بار برای انتخاب متغیر در رگرسیون خطی بیزی، توسط جورج و مک کولاج[۳] (۱۹۹۳)، پیشنهاد شد. از این روش در گستره وسیعی از مدل‌ها، از جمله مدل‌های خطی تعمیم‌یافته (وانگ و ژرژ، ۲۰۰۳)، مدل‌های توافقی (اسمیت و کوهن[۴]، ۱۹۹۶)، تحلیل رابطه ژنتیکی (یی[۵] و همکاران، ۲۰۱۰)، مدل‌های سلسله‌مراتبی و گروه‌بندی مقید[۶] (فارکومنی[۷]، ۲۰۱۰)، و رگرسیون چندکی خطی (رید و همکاران، ۲۰۱۰) استفاده می‌شود.

---

۱ *Disturbance*

۲ *Powell*

۳ *George and McCulloch*

۴ *Smith and Kohn*

۵ *Yi*

۶ *Grouping Constrained Model*

۷ *Farcomeni*

در روش انتخاب متغیر با جستجوی تصادفی، هر مولفه از بردار پارامترهای $\boldsymbol{\beta}$ به عنوان یک ترکیب از دو توزیع نرمال مرکزی‌شده[1] با واریانس‌های متفاوت (جورج و مک کولاج، ۱۹۹۳) یا از ترکیب یک توزیع نرمال مرکزی‌شده و یک توزیع تباهیده در صفر (کو[2] و مالیک، ۱۹۹۸) مدل‌بندی می‌شود.

یک برآوردگر شهودی برای p امین چندک مدل رگرسیون چندکی دودویی و توبیت که توسط مانسکی (۱۹۸۵) و پاول (۱۹۸۶) ارایه شده است، به صورت زیر می‌باشد

$$\widehat{\boldsymbol{\beta}}_p = arg\,\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - g(\boldsymbol{x}_i^T\boldsymbol{\beta})),$$

است که در آن، $g(\boldsymbol{x}_i^T\boldsymbol{\beta}) = I\{\boldsymbol{x}_i^T\boldsymbol{\beta} > 0\}$ برای رگرسیون چندکی دودویی، $g(\boldsymbol{x}_i^T\boldsymbol{\beta}) = max(\boldsymbol{x}_i^T\boldsymbol{\beta}, c)$ برای رگرسیون چندکی توبیت است و $\rho_p(.)$ تابع زیان (۵,۱) می‌باشد.

در این بخش، رگرسیون چندکی بیزی را به عنوان یک رگرسیون خطی با جمله خطایی از توزیع لاپلاس نامتقارن مورد بررسی قرار می‌دهیم. فرض می‌کنیم جمله خطا u_i در مدل (۱,۴) از توزیع لاپلاس نامتقارن به صورت زیر پیروی می‌کند.

$$f_p(u) = p(1-p)exp\big(-\rho_p(u)\big)$$

توجه کنید که

$$\int_{-\infty}^{0} f_p(u)du = p\,(1-p)\int_{-\infty}^{0} e^{u(1-p)}\,du = p.$$

یعنی p امین چندک، برای حالت متقارن ($p = 0/5$) صفر است. برای مدل رگرسیون چندکی دودویی و توبیت، y_i جایگزین y_i^* شده و چون تابع پیوند $g(.)$ یکنواست، در صورتی که $\boldsymbol{x}_i^T\boldsymbol{\beta}$، p امین چندک y_i^* باشد، $g(\boldsymbol{x}_i^T\boldsymbol{\beta})$، p امین چندک $y_i = g(y_i^*)$ خواهد بود.

طبق مطالعات یو و استاندر (۲۰۰۷)، تابع درستنمایی مدل (۱,۴) به صورت

[1] *Centered Normal*

[2] *Kuo*

$$L(\boldsymbol{y}|\boldsymbol{\beta}) = p^n(1-p)^n exp\left\{-\sum_{i=1}^{n}\rho_p(y_i - g(\boldsymbol{x}_i^T\boldsymbol{\beta}))\right\}, \qquad (2.4)$$

است. می‌نیمم کردن رابطه (۲,۴)، معادل ماکسیمم کردن تابع درستنمایی مدل رگرسیونی با فرض خطای لاپلاس نامتقارن در مدل (۱,۴) است. زیرا اگر بخواهیم $\sum_{i=1}^{n}\rho_p(y_i - g(\boldsymbol{x}_i^T\boldsymbol{\beta}))$ می‌نیمم شود، کافی است همین عبارت در توان تابع نمایی دارای علامت منفی باشد یک پیشین متداول در این مدل‌ها، مدل پیشین *spike–slab* است که توسط محققان بسیاری بکار گرفته می‌شود. در ادامه آن‌را توضیح می‌دهیم.

(ایشواران و رائو[۱]، ۲۰۰۵) فرض کنید $(y_1, \dots, y_n)$، n پاسخ مستقل، $\boldsymbol{x}_i = \left(x_{i,1}, \dots, x_{i,K}\right)^T$ متغیر تبیینی K بعدی و $\beta = (\beta_1, \dots, \beta_K)^T$ بردار پارامترها باشد. مدل *spike-slab* به صورت یک مدل بیزی تعریف می‌شود که از مدل‌بندی سلسله‌مراتبی زیر پیروی می‌کند:

$$Y_i|X_i, \beta, \sigma^2 \sim N(x_i^T\beta, \sigma^2), \qquad i = 1, \dots, n,$$
$$(\beta|\gamma) \sim N(\circ, \boldsymbol{\Gamma}),$$
$$\gamma \sim \pi(d\gamma),$$
$$\sigma^2 \sim \mu(d\sigma^2),$$

که در آن، $\circ$ یک بردار K بعدی از صفر است، $\boldsymbol{\Gamma}$ ماتریس قطری $K \times K$ با عناصر $\gamma_1, \dots, \gamma_K$ است، π اندازه پیشین برای $\gamma = (\gamma_1, \dots, \gamma_K)^T$ است و μ اندازه پیشین برای σ^2 است. علاوه بر این، فرض می‌کنیم برای $k = 1, \dots, K$، $\pi\{\gamma_k > \circ\} = 1$ و $\pi\{\sigma^2 > \circ\} = 1$ که این نشان‌دهنده سره بودن توزیع پیشین‌ها می‌باشد.

لمپر[۲](۱۹۷۱) و میشل و بیوچمپ[۳](۱۹۸۸) از پیشگامان روش *spike-slab* هستند. عبارت *spike-slab* برای پیشین β در فرمول‌بندی سلسله‌مراتبی بکار برده می‌شود. β_kها دو به دو مستقل بود و از

[۱] *Ishwaran and Rao*

[۲] *Lempers*

[۳] *Mitchell and Beauchamp*

توزیع ترکیبی دو نقطه، که از توزیع تخت یکنواخت (*slab*) و توزیع تباهیده در صفر (*spike*) ساخته شده است.

در مدل (۱,۴) بررسی شمول و عدم شمول متغیرهای تبیینی که q بعدی می‌باشند و هر یک از بعدها دارای دو حالت می‌باشند، بنابراین 2^q، مدل ممکن می‌سازند. فرض کنید $\gamma_j, j=1,\dots,q$ متغیرهای نشانگری باشند، به طوری که $\gamma_j = 1$ به این معناست که j امین متغیر تبیینی در مدل رگرسیون وجود دارد و $\gamma_j = 0$ یعنی در مدل وجود ندارد. مدل رگرسیون برای i امین عضو، $i = 1, \dots, n$، به صورت

$$y_i^* = \sum_{j=1}^{q} \beta_j \gamma_j x_{ij} + u_i, \qquad (3.4)$$

است که در آن، u_i دارای توزیع لاپلاس نامتقارن است.

کوزومی و کوبایشی[۱] (۲۰۱۱) یک الگوریتم نمونه‌گیری گیبز معرفی کردند که از نمایش توزیع لاپلاس نامتقارن به صورت ترکیبی از توزیع نرمال و نمایی، استفاده می‌کند.

لم زیر این صورت نمایش را بیان می‌کند.

لم ۱,۴: فرض کنید z یک متغیر تصادفی نمایی با میانگین یک و ξ متغیر تصادفی نرمال استاندارد باشد. برای $p \in (0,1)$، متغیر تصادفی $u = \theta z + \tau\sqrt{z}\,\xi$ از توزیع لاپلاس نامتقارن پیروی می‌کند، به طوری که

$$\theta = \frac{1-2p}{p(1-p)}, \qquad \tau^2 = \frac{2}{p(1-p)}.$$

اثبات: فرض می‌کنیم متغیر تصادفی u دارای تابع چگالی احتمال به صورت زیر باشد:

$$f_p(u) = p(1-p)\, exp\{-\rho_p(u)\}, \qquad \rho_p(u) = u\{p - I(u<0)\},$$

و همچنین

[۱] *Kozumi and Kobayashi*

$$f(u) = \int_{-\infty}^{\circ} p(1-p)e^{-u(p-1)}\,du + \int_{\circ}^{\infty} p(1-p)e^{-up}\,du = 1.$$

آن‌گاه تابع مشخصه u، به صورت زیر به دست می‌آید:

$$\begin{aligned}
\varphi(u) &= E\left[e^{itu}\right] \\
&= \int_{-\infty}^{+\infty} e^{itu} f_p(u)du \\
&= \int_{-\infty}^{\circ} p(1-p)e^{itu+(1-p)u}du + \int_{\circ}^{\infty} p(1-p)e^{itu-pu}du \\
&= p(1-p)\left\{\frac{1}{it+(1-p)} + \frac{1}{p-it}\right\} \\
&= p(1-p)\left\{\frac{1}{t^2 - i(1-2p)t + p(1-p)}\right\} \\
&= \left\{\frac{1}{p(1-p)}t^2 - i\frac{1-2p}{p(1-p)}t + 1\right\}^{-1},
\end{aligned}$$

که در آن، $i^2 = -1$.

فرض کنید z یک متغیر نمایی استاندارد و ξ یک متغیر نرمال استاندارد باشد. تابع مشخصه $\acute{u} = \theta z + \tau\sqrt{z}\,\xi$ به صورت زیر بیان می‌شود:

$$\begin{aligned}
\phi(\acute{u}) &= E\left[e^{it(\theta z + \tau\sqrt{z}\,\xi)}\right] \\
&= \int_{\circ}^{\infty} e^{it\theta z} E\left[e^{it\tau\sqrt{z}\,\xi}\right] e^{-z}dz
\end{aligned}$$

از آنجایی‌که ξ از توزیع نرمال استاندارد پیروی می‌کند، داریم:

$$\begin{aligned}
E\left[e^{it\tau\sqrt{z}\,\xi}\right] &= \int_{-\infty}^{+\infty}\left(e^{it\tau\sqrt{z}\,\xi}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{\xi^2}{2}}\right)d\xi \\
&= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\left(\frac{\xi^2 - 2it\tau\sqrt{z}\,\xi}{2}\right)}d\xi \\
&= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2\tau^2 z}e^{-\left(\frac{\xi^2 - 2it\tau\sqrt{z}\,\xi - t^2\tau^2 z}{2}\right)}d\xi
\end{aligned}$$

$$= e^{-\frac{1}{2}t^2\tau^2 z}.$$

بنابراین، تابع مشخصه $\acute{u}$ به صورت زیر تبدیل می‌شود

$$\phi(t) = \int_{\circ}^{\infty} e^{-z\left(1+\frac{1}{2}t^2\tau^2 - i\theta t\right)} dz$$

$$= \left(1 + \frac{1}{2}t^2\tau^2 - i\theta t\right)^{-1}.$$

این دو تابع مشخصه وقتی هم ارزند که

$$\theta = \frac{1-2p}{p(1-p)}, \qquad \tau^2 = \frac{2}{p(1-p)}.$$

با توجه به توزیع لاپلاس نامتقارن می‌توان نتیجه گرفت

$$E(u) = \frac{1-2p}{p(1-p)}$$

$$, \qquad var(u) = E(u^2) - (E(u))^2 = \frac{2p^2 - 2p + 1}{p^2(1-p)^2}$$

اگر بخواهیم مقدار θ و τ^2 را محاسبه کنیم، از رابطه $u = \theta z + \tau\sqrt{z}\,\xi$ استفاده می‌کنیم. داریم:

$$E(u) = \theta E(z) + \tau E\left(\sqrt{z}\right)E\left(\xi\right) = \theta = \frac{1-2p}{p(1-p)}$$

$$E(u^2) = \theta^2 E(z^2) + \tau^2 E\left(\xi^2\right)E(z) + 2\theta\tau E\left(z\sqrt{z}\right)E\left(\xi\right) = 2\theta^2 + \tau^2$$

$$\frac{6p^2 - 6p + 2}{p^2(1-p)^2} = 2\left(\frac{1-2p}{p(1-p)}\right)^2 + \tau^2$$

$$\tau^2 = \frac{2}{p(1-p)}. \qquad \square$$

با استفاده از لم ۱,۴ مدل ۳,۴ را بازنویسی می‌کنیم:

$$y_i^* = \sum_{j=1}^{n} \beta_j \gamma_j x_{ij} + \theta z_i + \tau\sqrt{z_i}\,\xi_i,$$

که در آن، $z_i \sim Exp(1)$ و $\xi_i \sim N(\circ, 1)$.

برای $\boldsymbol{\beta}$ از پیشین‌های *spike-slab* استفاده می‌کنیم. فرض کنید پیشین حاشیه‌ای $\boldsymbol{\beta}$ دارای $N(\beta_0, D_0)$ است. بنابراین $\beta_j = 0$ است، اگر $\gamma_j = 0$ شود و $\beta_\gamma \sim N(\beta_{0\gamma}, D_{0\gamma})$ باشد که زیرنویس γ نمایش دهنده زیرمجموعه شاخص دار توسط γ_j های غیر صفر است. (نماد X_j بطور مشابه تعریف شده است.) آن‌گاه نشانگر γ_j دارای توزیع پیشین مستقل برنولی با پارامتر احتمال p_j هستند. فرض کنید $\boldsymbol{y}^* = (y_1^*, \dots, y_n^*)^T$، $X = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)^T$، $\Sigma = diag(\boldsymbol{z})$ و $\boldsymbol{z} = (z_1, \dots, z_n)^T$. بنابراین مدل سلسله-مراتبی به صورت زیر خواهد بود:

$$y_i = g(y_i^*), \quad i = 1, \dots, n,$$
$$\boldsymbol{y}^* | \boldsymbol{\beta}, \boldsymbol{z}, \boldsymbol{\gamma} \sim N_n\left(\boldsymbol{X}_\gamma \beta_\gamma + \theta \boldsymbol{z}, \tau^2 \Sigma\right),$$
$$\boldsymbol{\beta} | \boldsymbol{\gamma} \sim N_q\left(\beta_{0\gamma}, D_{0\gamma}\right),$$
$$\boldsymbol{\gamma} \sim \prod_{j=1}^{q} \pi_j^{\gamma_j} (1 - \pi_j)^{1-\gamma_j},$$
$$\boldsymbol{z} \sim \prod_{i=1}^{n} exp(-z_i),$$

در این مدل، $\boldsymbol{\gamma}$ و $\boldsymbol{z}$ دو به دو مستقل هستند. پارامتر احتمال π_j نشان‌دهنده پیشین نامعلوم، روی همه متغیرها بجز jامین متغیر تبیینی وجود دارد. وقتی اطلاعات پیشین کافی نباشد، آنگاه اغلب برای هر j، مقدار $\pi_j = 0/5$ را قرار می‌دهیم.

۳,۴ توزیع شرطی کامل برای نمونه‌گیری گیبز

هدف اصلی، از تحلیلهای ما به‌دست آوردن احتمال‌های پسین برای $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)$، $\pi(\boldsymbol{\gamma}|\boldsymbol{y})$، می‌باشد. توزیع پسین برای $\boldsymbol{\gamma}$، وابسته به داده‌ها و اطلاعات پیشین است.

آلبرت و چیب[1] (۱۹۹۳) از نمونه‌گیری گیبز برای مدل‌های رگرسیون دودویی ساده استفاده کردند.

[1] *Albert and Chib*

در این زیر بخش، متغیرهای پنهان $y_1^*, \dots, y_n^*$ با استفاده از الگوریتم نمونه‌گیری گیبز و بر اساس توزیع‌های شرطی کامل $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{z}$ شبیه سازی می‌شوند.

در ابتدا، توزیع شرطی کامل y_i^* را برای رگرسیون چندکی دودویی در نظر می‌گیریم. تحت توزیع‌های پیشین مشخص شده، توزیع شرطی کامل y_i^* را به صورت ترکیبی از دو توزیع نرمال بریده شده در نظر می‌گیریم که آن‌را به صورت قسمت‌های مثبت و منفی مشاهده می‌کنیم:

$$y_i^*|y_i, \boldsymbol{\beta}, x_i, \boldsymbol{z}, \boldsymbol{\gamma} \sim \begin{cases} N\left(\boldsymbol{x}_i^T \beta_\gamma + \theta z_i, \tau^2 z_i\right) I(y_i^* > 0), & if \quad y_i = 1 \\ N\left(\boldsymbol{x}_i^T \beta_\gamma + \theta z_i, \tau^2 z_i\right) I(y_i^* \leq 0). & \text{سایر حالات} \end{cases} \qquad (4.4)$$

برای رگرسیون توبیت، توزیع شرطی کامل y_i^* را مانند بالا در دو قسمت بزرگتر یا کوچکتر از c در نظر می‌گیریم:

$$y_i^*|y_i, \beta, x_i, z, \gamma \sim \begin{cases} \delta(y_i) & if \quad y_i > c \\ N\left(\boldsymbol{x}_i^T \beta_\gamma + \theta z_i, \tau^2 z_i\right) I(y_i^* \leq c). & \text{سایر حالات} \end{cases} \qquad (5.4)$$

به طوری‌که $\delta(y_i)$ نشان‌دهنده توزیع تباهیده در y_i است. در واقع اگر $y_i > c$، آن‌گاه $y_i = y_i^*$ ولی برای $y_i < c$، توزیع نرمال را در نظر می‌گیریم.

چون متغیر پاسخ پنهان $\boldsymbol{y}^*$ بر اساس توزیع‌های شرطی کامل $\boldsymbol{\beta}$، $\boldsymbol{z}$ و $\boldsymbol{\gamma}$ محاسبه شده است، بنابراین توزیع‌های شرطی کامل تنها به مشاهدات پاسخ $\boldsymbol{y}$ وابسته نیستند و مشابه رگرسیون چندکی توبیت و رگرسیون چندکی دودویی می‌باشند.

توزیع‌های شرطی کامل برای ضرایب رگرسیونی به صورت زیر است

$$\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{\gamma}, \boldsymbol{z} \sim N_q\left(\boldsymbol{\mu}_\gamma, B_\gamma\right), \qquad (6.4)$$

که در آن

$$\boldsymbol{B}_\gamma^{-1} = \tau^{-2} X_\gamma^T \Sigma^{-1} X_\gamma + D_{0\gamma}^{-1},$$
$$\boldsymbol{\mu}_\gamma = \boldsymbol{B}_\gamma \{\tau^{-2} X_\gamma^T \Sigma^{-1} (\boldsymbol{y}^* - \theta z) + D_{0\gamma}^{-1} \boldsymbol{\beta}_{0\gamma}\},$$

و $\Sigma = diag(z)$.

برای اثبات (۹,۴) داریم

$$f(\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{y}^*,\boldsymbol{\gamma},\boldsymbol{z}) = f(\boldsymbol{\gamma})f(\boldsymbol{z}|\boldsymbol{\gamma})f(\boldsymbol{\beta}|\boldsymbol{\gamma},\boldsymbol{z})f(\boldsymbol{y}^*|\boldsymbol{\beta},\boldsymbol{z},\boldsymbol{\gamma}).$$

از آنجایی‌که $\boldsymbol{z},\boldsymbol{\gamma}$ از هم مستقلند، بنابراین

$$f(\boldsymbol{z}|\boldsymbol{\gamma}) = \frac{f(\boldsymbol{\gamma},\boldsymbol{z})}{f(\boldsymbol{\gamma})} = \frac{f(\boldsymbol{z})f(\boldsymbol{\gamma})}{f(\boldsymbol{\gamma})} = f(\boldsymbol{z}),$$

$$f(\boldsymbol{\beta}|\boldsymbol{\gamma},\boldsymbol{z}) = \frac{f(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{z})}{f(\boldsymbol{\gamma},\boldsymbol{z})} = \frac{f(\boldsymbol{\gamma})f(\boldsymbol{\beta}|\boldsymbol{\gamma})f(\boldsymbol{z}|\boldsymbol{\gamma},\boldsymbol{\beta})}{f(\boldsymbol{z})f(\boldsymbol{\gamma})} = \frac{f(\boldsymbol{\gamma})f(\boldsymbol{\beta}|\boldsymbol{\gamma})f(\boldsymbol{z})}{f(\boldsymbol{z})f(\boldsymbol{\gamma})} = f(\boldsymbol{\beta}|\boldsymbol{\gamma}).$$

و از طرفی

$$g(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{y}^*,\boldsymbol{z},\boldsymbol{\gamma}) = \frac{f(\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{y}^*,\boldsymbol{\gamma},\boldsymbol{z})}{\int f(\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{y}^*,\boldsymbol{\gamma},\boldsymbol{z})d\boldsymbol{\beta}}.$$

بنابراین

$$f(\boldsymbol{\beta},\boldsymbol{y},\boldsymbol{y}^*,\boldsymbol{\gamma},\boldsymbol{z}) = \left[\prod_{j=1}^{q}\pi_j^{\gamma_j}(1-\pi_j)^{1-\gamma_j}\right]$$

$$\times\left[\frac{1}{(2\pi)^{\frac{q}{2}}\left|D_{\circ\gamma}\right|^{\frac{1}{2}}}exp\left\{\frac{-1}{2}(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})^T\left(D_{\circ\gamma}\right)^{-1}(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})\right\}\right]$$

$$\times\left[\prod_{i=1}^{n}exp(-z_i)\right]\left[\frac{exp\left[\frac{-1}{2}\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)\right]}{(2\pi)^{n/2}\left|\tau^2\Sigma\right|^{1/2}}\right]$$

$$=\left[\prod_{j=1}^{q}\pi_j^{\gamma_j}(1-\pi_j)^{1-\gamma_j}\right]\left[\prod_{i=1}^{n}exp(-z_i)\right]\left[(2\pi)^{-n/2}\left|\tau^2\Sigma\right|^{-1/2}(2\pi)^{\frac{-q}{2}}\left|D_{\circ\gamma}\right|^{-1/2}\right]$$

$$\times exp\left\{\frac{-1}{2}\left[(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})^T D_{\circ\gamma}^{-1}(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})\right.\right.$$

$$\left.\left.+\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)\right]\right\}.$$

بنابراین با اطلاعات به‌دست آمده، داریم:

$$g(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{y}^*,\boldsymbol{z},\boldsymbol{\gamma}) =$$

$$\frac{exp\left\{\frac{-1}{2}\left[(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})^T D_{\circ\gamma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_{\circ\gamma})+\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)\right]\right\}}{\int exp\left\{\frac{-1}{2}\left[(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})^T D_{\circ\gamma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_{\circ\gamma})+\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)\right]\right\}d\boldsymbol{\beta}_\gamma}$$

$$\propto exp\left\{\frac{-1}{2}\left[(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})^T D_{\circ\gamma}^{-1}(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})+\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)\right]\right\}.$$

برای حل عبارت بالا، آن‌را به صورت زیر تجزیه می‌کنیم:

$$\begin{aligned}(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma})^T D_{\circ\gamma}^{-1}\left(\boldsymbol{\beta}_\gamma-\boldsymbol{\beta}_{\circ\gamma}\right)&=\boldsymbol{\beta}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}-\boldsymbol{\beta}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}-\boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}+\boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}\\&=\boldsymbol{\beta}_\gamma{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_\gamma-2\boldsymbol{\beta}_\gamma{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}+\boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}\,.\end{aligned}$$

و

$$\begin{aligned}&\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^*-\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\theta\boldsymbol{z}\right)\\&=\tau^{-2}\left[\boldsymbol{y}^{*T}\Sigma^{-1}\boldsymbol{y}^*-\boldsymbol{y}^{*T}\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma-\boldsymbol{y}^{*T}\Sigma^{-1}\theta\boldsymbol{z}-\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{y}^*+\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma\right.\\&\qquad\left.+\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\theta\boldsymbol{z}-\theta\boldsymbol{z}^{\boldsymbol{T}}\Sigma^{-1}\boldsymbol{y}^*+\theta\boldsymbol{z}^{\boldsymbol{T}}\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma+\theta\boldsymbol{z}^{\boldsymbol{T}}\Sigma^{-1}\theta\boldsymbol{z}\right]\\&=\tau^{-2}\left[\boldsymbol{y}^{*T}\Sigma^{-1}\boldsymbol{y}^*-2\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{y}^*+\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma+2\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\theta\boldsymbol{z}\right.\\&\qquad\left.-2\theta\boldsymbol{z}^{\boldsymbol{T}}\Sigma^{-1}\boldsymbol{y}^*+\theta^2\boldsymbol{z}^{\boldsymbol{T}}\Sigma^{-1}\boldsymbol{z}\right].\end{aligned}$$

بنابراین عبارت داخل قسمت نمایی به صورت

$$\begin{aligned}&\tau^{-2}\left[\boldsymbol{y}^{*T}\Sigma^{-1}\boldsymbol{y}^*-2\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{y}^*+\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma+2\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\theta\boldsymbol{z}-2\theta\boldsymbol{z}^{\boldsymbol{T}}\Sigma^{-1}\boldsymbol{y}^*\right.\\&\qquad\left.+\theta^2\boldsymbol{z}^{\boldsymbol{T}}\Sigma^{-1}\boldsymbol{z}\right]+\boldsymbol{\beta}_\gamma{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_\gamma-2\boldsymbol{\beta}_\gamma{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}+\boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}\\&=\boldsymbol{\beta}_\gamma{}^T\left[\tau^{-2}\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma+D_{\circ\gamma}^{-1}\right]\boldsymbol{\beta}_\gamma-2\boldsymbol{\beta}_\gamma{}^T\left[\tau^{-2}\boldsymbol{X}_\gamma{}^T\Sigma^{-1}(\boldsymbol{y}^*-\theta\boldsymbol{z})+D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}\right]\\&\qquad+\tau^{-2}(\boldsymbol{y}^*-2\theta\boldsymbol{z})^T\Sigma^{-1}\boldsymbol{y}^*+\theta^2\tau^{-2}\boldsymbol{z}^{\boldsymbol{T}}\Sigma^{-1}\boldsymbol{z}+\boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}\end{aligned}$$

تبدیل می‌شود. اگر در نظر بگیریم $g\sim N_p(A,C)$ باشد، در این‌صورت

$$f(g)=(2\pi)^{\frac{p}{2}}|C|^{-\frac{1}{2}}exp[(F-A)^T C^{-1}(F-A)].$$

حال اگر قسمت نمایی را در نظر بگیریم، داریم:

$$F^T C^{-1}F-F^T C^{-1}A-A^T C^{-1}F+A^T C^{-1}A$$

$$= F^T C^{-1} F - 2F^T C^{-1} A + A^T C^{-1} A$$

که در واقع

$$A = C\left[\tau^{-2}\boldsymbol{X}_\gamma{}^T\Sigma^{-1}(\boldsymbol{y}^* - \theta\boldsymbol{z}) + C^{-1} = \left[\tau^{-2}\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma + D_{\circ\gamma}^{-1}\right], \quad F = \boldsymbol{\beta}_\gamma.$$

$\left[, D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}\right]$ بنابراین کافی است ثابت کنیم

$$A^T C^{-1} A = \tau^{-2}(\boldsymbol{y}^* - 2\theta\boldsymbol{z})^T\Sigma^{-1}\boldsymbol{y}^* + \theta^2\tau^{-2}\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{z} + \boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}$$

و یا معادل به طور

$$C^{-1}A = \left[\tau^{-2}\boldsymbol{X}_\gamma{}^T\Sigma^{-1}(\boldsymbol{y}^* - \theta\boldsymbol{z}) + D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}\right]$$

$$C^{-1}A = \left[\tau^{-2}\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma + D_{\circ\gamma}^{-1}\right]A = \tau^{-2}\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma A + D_{\circ\gamma}^{-1}A$$

بنابراین $\boldsymbol{\beta}_{\circ\gamma} = A$ و $\boldsymbol{X}_\gamma A = (\boldsymbol{y}^* - \theta\boldsymbol{z})$.

$$A^T C^{-1} A = \boldsymbol{\beta}_{\circ\gamma}{}^T\left[\tau^{-2}\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma + D_{\circ\gamma}^{-1}\right]\boldsymbol{\beta}_{\circ\gamma} = \tau^{-2}\boldsymbol{\beta}_{\circ\gamma}{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_{\circ\gamma} + \boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}$$

$$= \tau^{-2}(\boldsymbol{y}^* - \theta\boldsymbol{z})^T\Sigma^{-1}(\boldsymbol{y}^* - \theta\boldsymbol{z}) + \boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}$$

$$= \tau^{-2}\left[\boldsymbol{y}^{*T}\Sigma^{-1}\boldsymbol{y}^* - \boldsymbol{y}^{*T}\Sigma^{-1}\theta\boldsymbol{z} - \theta\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{y}^* + \theta^2\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{z}\right] + \boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}$$

$$= \tau^{-2}\left[\boldsymbol{y}^{*T}\Sigma^{-1}\boldsymbol{y}^* - 2\theta\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{y}^* + \theta^2\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{z}\right] + \boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}$$

$$= \tau^{-2}(\boldsymbol{y}^* - 2\theta\boldsymbol{z})^T\Sigma^{-1}\boldsymbol{y}^* + \theta^2\tau^{-2}\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{z} + \boldsymbol{\beta}_{\circ\gamma}{}^T D_{\circ\gamma}^{-1}\boldsymbol{\beta}_{\circ\gamma}.$$

فرض کنید $\boldsymbol{\gamma}_{-i} = (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_q)$. متغیر ویژه پاسخ $\boldsymbol{\gamma}$ می‌تواند مولفه‌ای از توزیع شرطی $\gamma_i|y^*, \beta, z, \gamma_{-i}$ باشد که هر کدام از γ_i دارای توزیع برنولی با پارامترهای احتمال

$$P(\gamma_i = 1|\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{\beta}, \boldsymbol{z}, \boldsymbol{\gamma}_{-i}) = \frac{c_i}{c_i + d_i}$$

می‌باشند. در آن

$$c_i = f(\boldsymbol{y}^*|\boldsymbol{\beta}, \gamma_i = 1, \boldsymbol{z}, \boldsymbol{\gamma}_{-i}) f(\boldsymbol{\beta}|\gamma_i = 1, \boldsymbol{\gamma}_{-i}, \boldsymbol{z}) f(\gamma_i = 1, \boldsymbol{\gamma}_{-i}|\boldsymbol{z})$$

می‌باشد. از آنجایی‌که $\boldsymbol{\gamma}$ و $\boldsymbol{z}$ دوبه دو مستقل هستند، بنابراین

$$f(\gamma_i = 1, \boldsymbol{\gamma}_{-i}|\boldsymbol{z}) = f(\gamma_i = 1, \boldsymbol{\gamma}_{-i}) f(\boldsymbol{z})$$

$$d_i = f(\boldsymbol{y}^*|\boldsymbol{\beta}, \gamma_i = \circ, \boldsymbol{z}, \boldsymbol{\gamma}_{-i}) f(\boldsymbol{\beta}|\gamma_i = \circ, \boldsymbol{\gamma}_{-i}) f(\gamma_i = \circ, \boldsymbol{\gamma}_{-i}|\boldsymbol{z})$$

و

$$f(\gamma_i = 0, \boldsymbol{\gamma}_{-i}|\boldsymbol{z}) = f(\gamma_i = 0, \boldsymbol{\gamma}_{-i}) f(\boldsymbol{z})$$

در ادامه نشان می‌دهیم که شرایط روی $\boldsymbol{\beta}, \boldsymbol{y}^*, \boldsymbol{\gamma}, z_i$ بطور دو طرفه مستقل هستند. یعنی

$$z_i|\boldsymbol{\beta}, \boldsymbol{y}^*, \boldsymbol{\gamma} \sim GIG(\frac{1}{2}, \hat{\lambda}_i, \hat{\eta}_i)$$

است. به طوری که، $\hat{\lambda}_i = (y_i^* - \boldsymbol{x}_i^T \boldsymbol{\beta}_\gamma)^2/\tau^2$ ،$\hat{\eta}_i = 2 + \theta^2/\tau^2$ و $GIG(v, a, b)$ نشان‌دهنده توزیع گاوسی معکوس تعمیم‌یافته با تابع چگالی احتمال زیر است

$$f(x|v, a, b) = \frac{(b/a)^{v/2}}{2K_v(\sqrt{ab})} x^{v-1} exp\left\{-\frac{(ax^{-1} + bx)}{2}\right\}, \quad x > 0$$

به طوری که $K_v(.)$ تابع بسل اصلاح‌شده نوع سوم با مرتبه v است (بارندورف- نلسون[1] و شفرد[2]، ۲۰۰۱) و به صورت زیر تعریف می‌شود:

$$K_v(x) \sim \begin{cases} -log(x) & \text{اگر } v = 0 \\ 2^{|v|-1}\Gamma(|v|)x^{-|v|}. & \text{اگر } v \neq 0 \end{cases}$$

بنابراین باید نشان دهیم

$$\begin{aligned} f(z_i|\boldsymbol{\beta}, \boldsymbol{y}^*, \boldsymbol{\gamma}) &= \frac{\left(\hat{\eta}_i/\hat{\lambda}_i\right)^{1/4}}{\sqrt{2\pi}\left(\hat{\eta}_i\hat{\lambda}_i\right)^{-1/4}} {z_i}^{-1/2} exp\left\{-\frac{(\hat{\lambda}_i {z_i}^{-1} + \hat{\eta}_i z_i)}{2}\right\} \\ &= \frac{\sqrt{\hat{\eta}_i}}{\sqrt{2\pi z_i}} exp\left\{-\frac{(\hat{\lambda}_i + \hat{\eta}_i {z_i}^2)}{2z_i}\right\} \\ &= \frac{\sqrt{2\tau^2 + \theta^2}}{\tau\sqrt{2\pi z_i}} exp\left\{-\frac{(y_i^* - \boldsymbol{x}_i^T\boldsymbol{\beta}_\gamma)^2 + (2\tau^2 + \theta^2){z_i}^2}{2\tau^2 z_i}\right\} \end{aligned}$$

برای اثبات این‌که $z_i|\boldsymbol{\beta}, \boldsymbol{y}^*, \boldsymbol{\gamma}$ دارای توزیع گاوسی معکوس تعمیم‌یافته است، ابتدا توزیع $\boldsymbol{z}|\boldsymbol{\beta}, \boldsymbol{y}^*, \boldsymbol{\gamma}$ را به‌دست می‌آوریم و پس از آنجایی‌که $\boldsymbol{z}$ شامل n تا z_i است، داریم

[1] *Barndorf-Nielsen*

[2] *Shepherd*

$$g(\boldsymbol{z}|\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma}) = \prod_{i=1}^{n} h(z_i|\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma}).$$

دوباره رابطه $f(\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma},\boldsymbol{z}) = f(\boldsymbol{\gamma})f(\boldsymbol{z}|\boldsymbol{\gamma})f(\boldsymbol{\beta}|\boldsymbol{\gamma},\boldsymbol{z})f(\boldsymbol{y}^*|\boldsymbol{\beta},\boldsymbol{z},\boldsymbol{\gamma})$ را در نظر بگیرید.

از آنجایی‌که $\boldsymbol{z},\boldsymbol{\gamma}$ از هم مستقلند، داریم

$$g(\boldsymbol{z}|\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma}) = \frac{f(\boldsymbol{z},\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma})}{\int f(\boldsymbol{z},\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma})\,d\boldsymbol{z}}$$

$$= \frac{\dfrac{exp\left[\frac{-1}{2}\left(\boldsymbol{y}^* - \boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - \theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^* - \boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - \theta\boldsymbol{z}\right) - \sum_{i=1}^{n} z_i\right]}{\left|\tau^2\Sigma\right|^{1/2}}}{\int \dfrac{exp\left[\frac{-1}{2}\left(\boldsymbol{y}^* - \boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - \theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^* - \boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - \theta\boldsymbol{z}\right) - \sum_{i=1}^{n} z_i\right]}{\left|\tau^2\Sigma\right|^{1/2}} d\boldsymbol{z}}$$

$$\propto \frac{exp\left[-\frac{1}{2}\left(\boldsymbol{y}^* - \boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - \theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^* - \boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - \theta\boldsymbol{z}\right) - \sum_{i=1}^{n} z_i\right]}{\left|\tau^2\Sigma\right|^{1/2}}.$$

برای حل عبارت بالا آن‌را به صورت زیر تجزیه می‌کنیم:

$$\left(\boldsymbol{y}^* - \boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - \theta\boldsymbol{z}\right)^T\left(\tau^2\Sigma\right)^{-1}\left(\boldsymbol{y}^* - \boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - \theta\boldsymbol{z}\right)$$

$$= \tau^{-2}\left[\boldsymbol{y}^{*T}\Sigma^{-1}\boldsymbol{y}^* - 2\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{y}^* + \boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma + 2\theta\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - 2\theta\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{y}^* + \theta^2\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{z}\right]$$

بنابراین عبارت داخل قسمت تابع نمایی به صورت زیر بازنویسی می‌شود:

$$-\frac{1}{2}\tau^{-2}\left[\boldsymbol{y}^{*T}\Sigma^{-1}\boldsymbol{y}^* - 2\boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{y}^* + \boldsymbol{\beta}_\gamma{}^T\boldsymbol{X}_\gamma{}^T\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma + 2\theta\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{X}_\gamma\boldsymbol{\beta}_\gamma - 2\theta\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{y}^* + \theta^2\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{z} + 2\tau^2\sum_{i=1}^{n} z_i\right]$$

و در نتیجه آن را می‌توان به صورت زیر نوشت:

$$-\frac{1}{2}\tau^{-2}\left[\frac{y_i^{*2}}{z_i}-2\frac{y_i^*\boldsymbol{x_i}^T\boldsymbol{\beta}_\gamma}{z_i}+\frac{\left(\boldsymbol{x_i}^T\boldsymbol{\beta}_\gamma\right)^2}{z_i}+2\theta\frac{z_i\boldsymbol{x_i}^T\boldsymbol{\beta}_\gamma}{z_i}-2\theta\frac{y_i^*z_i}{z_i}+\theta^2\frac{{z_i}^2}{z_i}+2\tau^2z_i\right].$$

بنابراین با استفاده از رابطه

$$g(\boldsymbol{z}|\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma})=\prod_{i=1}^{n}h(z_i|\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma}),$$

می‌توان نوشت:

$$h(z_i|\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma})\propto\frac{exp\left(\tau^2\theta\left(y_i^*-\boldsymbol{x_i}^T\boldsymbol{\beta}_\gamma\right)\right)}{\tau\sqrt{z_i}}exp\left\{-\frac{1}{2}\tau^{-2}\left[\frac{\left(y_i^*-\boldsymbol{x_i}^T\boldsymbol{\beta}_\gamma\right)^2}{z_i}+\frac{\theta^2{z_i}^2+2\tau^2{z_i}^2}{z_i}\right]\right\}.$$

نمونه‌گیری از یک توزیع گاوسی معکوس تعمیم‌یافته، بیشتر توسط تبدیل می‌تواند کارا شود. نشان می‌دهیم که $z_i^{-1}|\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma}$ از توزیع گاوسی معکوس با تابع چگالی احتمال زیر پیروی می‌کند.

$$f(z_i^{-1}|\boldsymbol{\beta},\boldsymbol{y}^*,\boldsymbol{\gamma})=\frac{\left(\frac{\hat{\eta}_i}{\hat{\lambda}_i}\right)^{-\frac{1}{4}}}{2K_{-\frac{1}{2}}(\sqrt{\hat{\eta}_i\hat{\lambda}_i})}{z_i}^{-\frac{3}{2}}exp\left\{-\frac{\hat{\lambda}_i+\hat{\eta}_i{z_i}^2}{2z_i}\right\}$$

$$=\frac{\left(\hat{\eta}_i\hat{\lambda}_i\right)^{\frac{1}{4}}}{\left(\frac{\hat{\eta}_i}{\hat{\lambda}_i}\right)^{\frac{1}{4}}\sqrt{2\pi}}{z_i}^{-\frac{3}{2}}exp\left\{-\frac{\hat{\lambda}_i+\hat{\eta}_i{z_i}^2}{2z_i}\right\}=\frac{\sqrt{\hat{\lambda}_i}}{\sqrt{2\pi{z_i}^3}}exp\left\{-\frac{\hat{\lambda}_i+\hat{\eta}_i{z_i}^2}{2z_i}\right\}$$

گاوسی معکوس، نوع خاصی از گاوسی معکوس تعمیم‌یافته است وقتی که $v=\frac{-1}{2}$ باشد. فرم استاندارد گاوسی معکوس به صورت زیر می‌باشد:

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

با در نظر گرفتن $\frac{\acute{\lambda}(x-\acute{\mu})^2}{2x(\acute{\mu})^2}=\frac{\acute{\lambda}(x^2+\acute{\mu}^2)}{2x(\acute{\mu})^2}-\frac{\acute{\lambda}}{\acute{\mu}}$، داریم

$$f\left(x|\acute{\mu},\acute{\lambda}\right)=\sqrt{\frac{\acute{\lambda}}{2\pi x^3}}ex\,p\left(\frac{\acute{\lambda}}{\acute{\mu}}\right)exp\left\{-\frac{\acute{\lambda}(x^2+\acute{\mu}^2)}{2x(\acute{\mu})^2}\right\}.$$

از مقایسه این فرمول با فرمول بالا داریم $\acute{\lambda} = \hat{\lambda}_i$ و $exp(\frac{\acute{\lambda}}{\acute{\mu}})$ را به عنوان جمله ثابت در نظر می‌گیریم، بنابراین

$$\acute{\lambda}(\acute{1} + \frac{x^2}{\acute{\mu}^2}) = \hat{\lambda}_i \left(1 + \frac{\hat{\eta}_i}{\hat{\lambda}_i} x^2\right)$$

در این‌صورت

$$\acute{\mu} = \sqrt{\frac{\hat{\lambda}_i}{\hat{\eta}_i}}.$$

می‌شود.

فصل پنجم

سری زمانی غیر نرمال

۱,۵ مقدمه

بسیاری از سری‌های زمانی اقتصادی و مالی، غیر نرمال هستند. در این فصل، رهیافت بیزی مدل‌های سری‌زمانی اتورگرسیو[۱] را با توابع چندک پیشنهاد می‌دهیم که این رهیافت پارامتری است. در این فصل،

[۱]. *Autoregressive*

رهیافت پارامتری را با رهیافت نیمه‌پارامتری مقایسه شده است. مطالعات شبیه‌سازی و کاربرد سری‌زمانی با داده‌های واقعی را در این فصل مورد مطالعه قرار داده‌ایم.

۲,۵ مدل

یک مدل سری‌زمانی اتورگرسیو مرتبه k به صورت زیر تعریف می‌شود:

$$y_t = a_0 + a_1 y_{t-1} + \cdots + a_k y_{t-k} + \varepsilon_t, \quad (۱.۵)$$

در آن، ε_t دارای میانگین صفر و واریانس σ^2 است. مدل (۱,۵) برای سری‌های زمانی نرمال، مناسب است.

برای سری‌های زمانی غیرنرمال، یکی از مشکلات مدل (۱,۵) پیدا کردن توزیع مناسب برای جمله خطا(ε_t) است. کانکر (۲۰۰۵) روش نیمه‌پارامتری را برای حل این مشکل پیشنهاد داد. در این روش با برآورد یک دنباله از چندک‌های شرطی y_t، تابع توزیع شرطی y_t را می‌توان تقریب زد. این مدل را از آنجا نیمه پارامتری می‌گویند که p امین چندک y_t به شرط $Y_{t-1} = (y_1, y_2, \dots, y_{t-1})$ را می‌توان به صورت

$$q^p_{y_t|Y_{t-1}} = a^p_0 + a^p_1 y_{t-1} + \cdots + a^p_k y_{t-k}, \quad 0 < p < 1, \quad (2.5)$$

بیان کرد که در آن، $a^p_0, a^p_1, \dots, a^p_k$ پارامترهای مدل وابسته به p هستند.

برای سری‌زمانی $y_1, y_2, \dots, y_{t-1}$، می‌توان پارامترها را با می‌نیمم کردن تابع هزینه زیر برآورد کرد:

$$\min_{\boldsymbol{\beta}^p} \sum_{t=k+1}^{n} \rho_p(u_t),$$

که در آن

$$\rho_p(u_t) = u_t.\left(p - I_{[u_t<0]}\right).$$

برای مدل (۱,۵) داریم:

$$u_t = y_t - a^p_0 - a^p_1 y_{t-1} - \cdots - a^p_k y_{t-k}, \quad t = k+1, \dots, n.$$

مدل‌های متفاوتی برای حل این مسأله بهینه‌سازی پیشنهاد شده است. برای مثال، کانکر و دی‌اوری[۱] (۱۹۹۴) و کانکر و دی‌اوری (۱۹۸۷) را ببینید. یو و موید (۲۰۰۱) یک رهیافت بیزی را برای برآورد پارامترها پیشنهاد دادند که در فصل۳ به آن پرداخته شد.

چون توزیع جمله خطا مشخص نیست و نیازی به انتخاب یک توزیع برای جمله خطا ،ε_t، نیست این رهیافت را نیمه پارامتری گویند. روش دیگر تعیین توزیع برای جمله خطا، بکارگیری یک رهیافت پارامتری است که این روش برای اولین دفعه توسط گیل کریست[۲] (۲۰۰۰) پیشنهاد شده است. در این حالت تابع چندک شرطی y_t به صورت

$$Q_{y_t}(p|Y_{t-1}) = a_\circ + a_1 y_{t-1} + \cdots + a_k y_{t-k} + \eta Q(p,\gamma), \qquad (3.5)$$

در نظر گرفته می‌شود که در آن $i = \circ, 1, \ldots, k$ ،a_i ،η وγ پارامترهای مدل هستند و $Q(p,\gamma)$ تابع توزیع چندک و p مقادیر چندک است.

در این قسمت به تفاوت‌های بین مدل‌های (۱,۵)، (۲,۵) و (۳,۵) اشاره می‌کنیم.

در واقع مدل (1.5) کار برآورد میانگین شرطی y_t و واریانس باقیماندهσ^2 را انجام می‌دهد. در حالی که مدل (۲,۵) چندک شرطی p ام ازy_t برای $p \in (\circ, 1)$ را برآورد می‌کند. بنابراین اگر یک دنباله از چندک‌های برآورد شده داشته باشیم، آنگاه همه توابع توزیع شرطی y_t توسط بکارگیری چندک شرطی برآورد شده می‌توانند برآورد شوند.

مدل (۱,۵) برای سری‌های‌زمانی غیرنرمال مناسب نیست چون مشخص کردن توزیع جمله خطا، ε_t، مشکل است. مدل (۲,۵) به انتخاب توزیع برای ε_t وابسته نیست، اما در قسمت شبیه‌سازی مشاهده شد که چندک برآورد شده ممکن است سازگار نباشد و مدل (۳,۵) به $Q(p,\gamma)$ که توسط توزیع جمله

[۱]. *D'Orey*

[۲]. *Gilchrist*

خطا تعریف می‌شود، وابسته است، اما به‌طور طبیعی چندک‌های برآورد شده سازگارند و انتخاب $Q(p,\gamma)$ به دلیل ویژگی تابع چندک خیلی منعطف‌تر از مدل (۱,۵) است، یعنی تابع چندک‌های مختلفی را می‌توان در نظر گرفت.

هر تابع چندک دارای یک توزیع می‌باشد و با استفاده از ویژگی توابع چندک که در فصل اول اشاره شده است می‌توانیم مدل سره (۳,۵) را برای سری زمانی مشاهده کنیم که برای آن با سری‌های غیر نرمال که منعطف توزیع شده‌اند سرو کار داریم. گیل کریست(۲۰۰۰) چندین روش را برای برآورد پارامتر مدل (۳,۵) پیشنهاد داده است ولی به دلیل دشواری روش برآورد ماکسیمم درستنمایی، او روی این روش تمرکز نکرده است که در واقع انگیزه‌ای برای کارهای جدید شود.

۳,۵ مطالعه روش *MCMC*

فرض کنید $y_1,\dots,y_n$ سری‌زمانی مشاهده شده از مدل (۳,۵) و $\beta=(a_\circ,\dots,a_k,\eta,\gamma)$ باشند. بنابراین تابع درستنمایی شرطی $y_{k+1},\dots,y_n$ با شرط $Y_k=(y_1,\dots,y_k)^T$، به صورت زیر است:

$$L(y_{k+1},\dots,y_n|Y_k,\beta)=f(y_{k+1}|Y_k,\beta)\times f\left(y_{k+2}\middle|\underbrace{y_{k+1},Y_k}_{Y_{k+1}},\beta\right)\times\dots\times f\left(y_n\middle|\underbrace{y_{n-1},Y_{n-2}}_{Y_{n-1}},\beta\right).$$

معکوس تابع چندک، تابع توزیع است و معکوس تابع توزیع یک تابع چندک می‌شود که در آن p چندک است، یعنی

$$Q(p)=F^{-1}(p)\quad,\quad F(p)=Q^{-1}(p)$$

حال اگر در نظر بگیریم $F(p)=Q^{-1}(p)$ آنگاه داریم

$$\frac{\partial F(p)}{\partial p}=\frac{1}{\frac{\partial Q(p)}{\partial p}}\Longrightarrow f(p)=\frac{1}{\frac{\partial Q(p)}{\partial p}}$$

$$f(y_{k+1}|Y_k,\beta)=\frac{1}{\frac{\partial Q_{y_{k+1}}(p,\gamma)}{\partial p}\Big|_{p=p_t}}$$

بنابراین

$$L(y_{k+1},\dots,y_n|Y_k,\beta)=\prod_{t=k+1}^{n} f(y_t|Y_{t-1},\beta)=\prod_{t=k+1}^{n}\frac{1}{\frac{\partial Q_{y_t}(p,\gamma)}{\partial p}\Big|_{p=p_t}}$$

است، که $t=k+1,\dots,n$ ، p_t در رابطه زیر صدق می کند

$$y_t=a_\circ+a_1y_{t-1}+\cdots+a_ky_{t-k}+\eta Q(p_t,\gamma) \qquad (4.5)$$

در حالت عادی، معادله (۴,۵) به حل عددی نیاز دارد.

فرض کنید $\pi(\beta)$ تابع چگالی پیشین پارامترها باشد. آن‌گاه تابع چگالی پسین β به صورت زیر است:

$$\pi(\beta|Y_n)\propto L(y_{k+1},\dots,y_n|Y_k,\beta)\pi(\beta),\quad \beta\in\Omega$$

که Ω فضای پارامتر است. اگر $\pi(\beta|Y_n)$ به‌طور مناسب بر روی فضای Ω تعریف شده باشد، آن‌گاه روش $MCMC$ می‌تواند برای برآورد پارامترهای مدل، توسعه داده شود. برای مثال به چیب و گرینبرگ[۱] (۱۹۹۵) را ببینید. در این پایان نامه الگوریتم $MCMC$ قدم زدن تصادفی متروپولیس-هستینگس به صورت زیر طراحی شده است.

اگر فرض کنیم β مقدار واقعی پارامترها و $p_t,\ t=k+1,\dots,n$ احتمال‌های متناظر با β مشخص شده باشند و $\acute{\beta}$ مقدار پیشنهاد شده و $\acute{p}_t$ احتمال‌های مشخص شده با مقدار پیشنهادی $\acute{\beta}$ باشند، آنگاه نمونه قدم زدن تصادفی $MCMC$ برای مدل به صورت زیر می‌تواند طراحی شود:

a) تولید مقدار پیشنهاد شده $\acute{\beta}$ از طریق چگالی پیشنهادی $g(\beta,\acute{\beta})$ به طوری که $\acute{\beta}\in\Omega$.

b) معادله (۴,۵) برای $\acute{p}_t$ حل می کنیم.

c) پذیرش مقدار پیشنهادی تولید شده با احتمال $min\ (AB,1)$ که در آن

$$A=\frac{\pi(\acute{\beta}|Y_n)}{\pi(\beta|Y_n)}\ ,\qquad B=\frac{g(\acute{\beta},\beta)\ /\int_\Omega\ g(\acute{\beta},\beta)d\beta}{g(\beta,\acute{\beta})\ /\int_\Omega\ g(\beta,\acute{\beta})\,d\acute{\beta}}.$$

[۱]. *Greenberg*

اگر نمونه به اندازه کافی بزرگ باشد، آن‌گاه انتظار داریم توزیع زنجیر مارکوف شبیه‌سازی شده به توزیع پسین پارامترها همگرا شود.

توجه کنید که اگر تکیه‌گاه g نیز Ω باشد، آنگاه انتگرال موجود در B یک می‌شود. در این پایان‌نامه فرض شده که مدل شبیه سازی برای برآورد همه انتگرال‌ها در B شامل می‌شوند.

مثال ۱,۵.حالت خاص: برای مثال فرض کنید $Q(p,\gamma)$ تابع چندک نمایی باشد. بنابراین

$$F_X(x) = 1 - e^{-\lambda x} \quad , \quad F_X(x) = p$$
$$1 - e^{-\lambda x} = p \quad \Rightarrow \quad e^{-\lambda x} = 1 - p$$
$$-\lambda x = ln(1-p) \quad \Rightarrow x = -\frac{1}{\lambda} ln(1-p)$$
$$Q(p,\gamma) = -\frac{1}{\lambda} ln(1-p), \qquad ۰ \leq p < 1$$

با فرض این‌که $Q(p,\gamma)$ نمایی است، مدل (۴,۵) را می‌توان بازنویسی کرد.

$$Q_{y_t}(p|Y_{t-1}) = a_۰ + a_1 y_{t-1} + \cdots + a_k y_{t-k} - \eta \frac{ln(1-p)}{\lambda}$$
$$= a_۰ + a_1 y_{t-1} + \cdots + a_k y_{t-k} - \frac{ln(1-p)}{\gamma}, \qquad (5.۵)$$

که در آن، $\gamma = \lambda/\eta$ و $\beta = (a_۰, \dots, a_k, \gamma)$.

بنابراین درستنمایی شرطی $y_t,\ t = k+1, \dots, n$ به صورت زیر است.

$$Q_{y_t}(p,\gamma) = -\frac{\eta}{\lambda} ln(1-p) \quad , \quad \frac{\partial Q_{y_t}(p,\gamma)}{\partial p} = -\frac{\eta}{\lambda}\left(-\frac{1}{1-p}\right) = \frac{\eta}{\lambda(1-p)}$$
$$\left.\frac{\partial Q_{y_t}(p,\gamma)}{\partial p}\right|_{p=p_t} = \frac{\eta}{\lambda(1-p_t)}.$$

$$L(y_{k+1},\dots,y_n|Y_k,\beta)=\prod_{t=k+1}^{n} f(y_t|Y_{t-1},\beta)=\prod_{t=k+1}^{n}\frac{1}{\frac{\partial Q_{y_t}(p,\gamma)}{\partial p}\Big|_{p=p_t}}=\prod_{t=k+1}^{n}\frac{\lambda}{\eta}(1-p_t)$$

$$=\prod_{t=k+1}^{n}\gamma(1-p_t).$$

و p_t جوابی از معادله زیر است:

$$y_t=a_\circ+a_1y_{t-1}+\cdots+a_ky_{t-k}-\frac{ln(1-p_t)}{\gamma}\qquad(6.5)$$

در این مورد خاص p_t می‌تواند جواب دقیق معادله بالا باشد.

فرض کنید تابع چگالی پیشین پارامترها به صورت زیر باشد:

$$\pi(\beta)=\prod_{i=\circ}^{k}\pi(a_i)\pi(\gamma)$$

$$\pi(a_i)\sim N\left(\circ,\tilde{\sigma}_{a_i}^2\right),\quad \pi(\gamma)=\alpha e^{-\alpha\gamma}.$$

آن‌گاه توزیع پسین به خوبی روی $\Omega=\Omega_1\times\Omega_2$ تعریف می‌شود که در آن،

$$\Omega_1=\{(a_\circ,\dots,a_k)|a_\circ+a_1y_{t-1}+\cdots+a_ky_{t-k}\le y_t,\quad t=k+1,\dots,n\},$$

$$\Omega_2=(\circ,\infty).$$

نمونه‌های تولید شده از الگوریتم *MCMC* می‌توانند برای برآورد پارامترهای مدل بکار برده شوند، که انتخاب توزیع پیشنهادی انعطاف‌پذیر باشد در این‌جا فرض می‌کنیم:

$$g\left(\beta,\acute{\beta}\right)=\prod_{i=\circ}^{k}g_i(a_i,\acute{a}_i)g_{k+1}(\gamma,\acute{\gamma})$$

که در آن $g_i(a_i,\acute{a}_i)\sim N(a_i,\sigma_{a_i}^2)$ و $g_{k+1}(\gamma,\acute{\gamma})\sim N(\gamma,\sigma_\gamma^2)$.

از آنجایی‌که برای $\pi(a_i)$ هر تابع چگالی سره را می توان در نظر گرفت، چون تابع درستنمایی برای هر $t=k+1,\dots,n$, $\circ\le p_t<1$ همواره کمتر از γ^{n-k} است. برای $\pi(\gamma)$ نیاز به انتخاب یک توزیعی که توزیع پسین سره بسازد، داریم.

حال سوالی که این‌جا مطرح می‌شود این است که چرا تابع درستنمایی همواره کمتر از γ^{n-k} است؟ در جواب باید گفت چون برای تابع درستنمایی داشتیم که $\prod_{t=k+1}^{n}\gamma(1-p_t)$ و از آنجایی‌که چندک‌ها همواره بین صفر و یک هستند، پس $0 < (1-p_t) < 1$ و حاصل‌ضرب بالا همواره کمتر از γ^{n-k} می‌شود.

۴,۵ کاربردها

در این بخش، مدل اتورگرسیو غیرنرمال را با یک مثال شبیه‌سازی شده و دو مثال واقعی شرح می‌دهیم. اساس مثال واقعی روی مجموعه داده‌های آب دریاچه هورون[۱] و سهام داکس[۲] آلمان است.

برای مثال در قسمت شبیه‌سازی فرض شده $k = 2$ باشد. آن‌گاه داریم:

$$Q_{y_t}(p|y_{t-1}) = a_0 + a_1 y_{t-1} + a_2 y_{t-2} - \frac{ln(1-p)}{\gamma} \qquad (7.5)$$

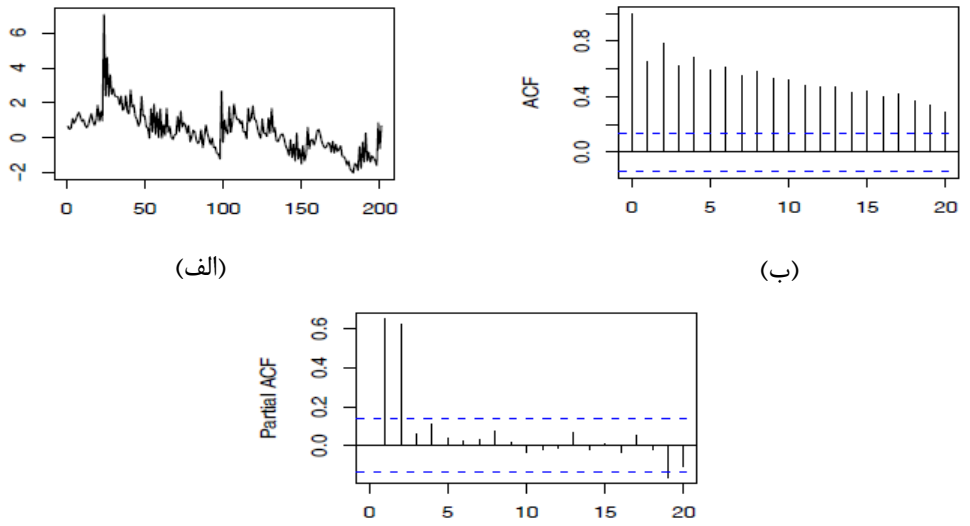
در ابتدا فرض شده $a_0 = -0/6$، $a_1 = 0/3$، $a_2 = 0/6$، $\gamma = 1/6$ و $y_1 = y_2 = 0$ باشند. با این فرضیات، سری‌زمانی به طول ۹۲۰۰ از مدل (۷,۵) را شبیه‌سازی شده است و سپس با حذف نمودن ۹۰۰۰ داده اول این سری، سری‌زمانی به طول ۲۰۰ از مدل خواهند داشت. در این‌جا فرض شده اطلاعات پیشین را با مقادیر $\tilde{\sigma}_{a_i}, i = 0,1,2$ و α مشخص کنند، که به صورت تصادفی $\tilde{\sigma}_{a_i} = 10$ و $\alpha = 0/5$ انتخاب شوند. برای مقادیر دیگر ما دارای نتایج خیلی مشابه هستیم. در این‌جا مقادیر آغازین $(\min_{1\le t\le n} y_t, 0, 0, \gamma_0)$ انتخاب شده است، که γ_0 نمونه تصادفی از توزیع پیشین γ است که این نشان می‌دهد که مقادیر آغازین در تکیه‌گاه Ω توزیع پسین قرار دارند. در این‌جا برای شبیه سازی انجام شده، مقادیر ابتدایی $(-2/0\ 11, 0, 0, 0/11)$ می‌شوند.

نمودار سری‌زمانی داده‌های شبیه‌سازی شده با نمودارهای تابع خود همبستگی[۱] (acf) و تابع خود همبستگی جزیی[۲] $(pacf)$ در شکل زیر آمده است.

---

۱. *Lake Huron*

۲. *CAX*



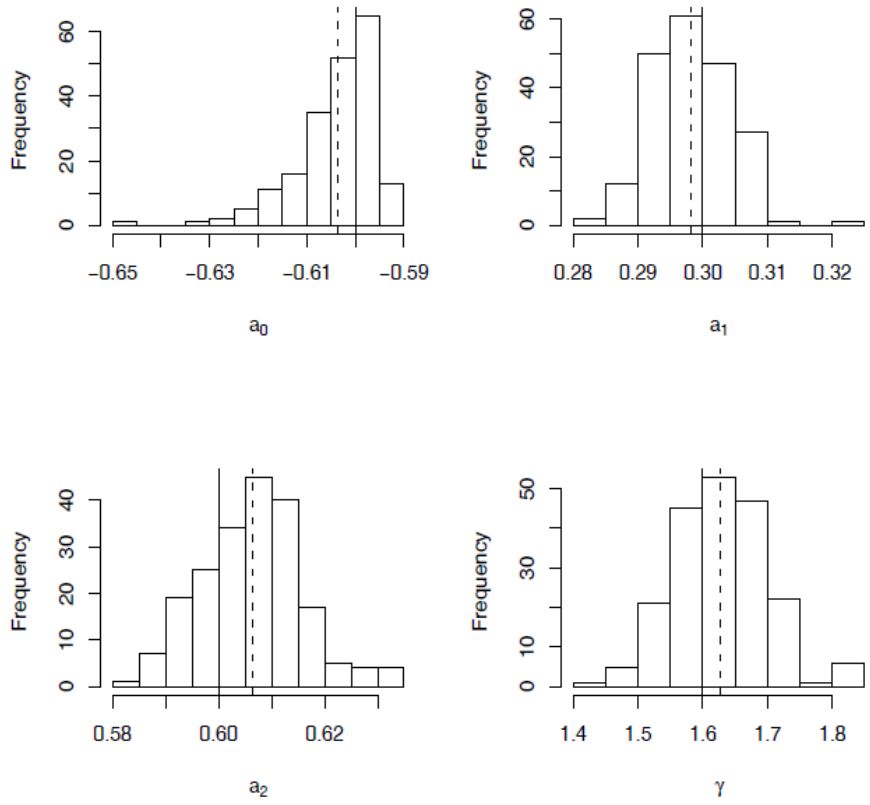
شکل۵,۱(الف) نمودار سری‌زمانی سری شبیه‌سازی (ج)(ب) نمودار تابع خودهمبستگی سری شبیه‌سازی . (ج) نمودار تابع خودهمبستگی جزیی سری شبیه‌سازی.

کای(۲۰۰۹) برای برآورد پارامترهای مدل از الگوریتم *MCMC* قدم زدن تصادفی متروپولیس-هستینگس استفاده کرده است. برای این‌کار در ابتدا یک نمونه ۲۰۰۰۰ داده‌ای را با استفاده از اطلاعات پیشین داده شده اجرا کرده است، به این صورت که با استفاده از β که مقادیر واقعی هستند، $\acute{\beta}$ را از توزیع نرمال با میانگین عناصر β داده شده اجرا نمود، سپس با داغیدن ۱۰۰۰۰ داده اول از کل نمونه، نمونه‌های باقیمانده را در هر ۵۰ مرحله ذخیره نمود، به طوری که نمونه‌های جمع‌آوری شده برای هر پارامتر دارای طول ۲۰۰ باشند، یعنی داده‌های $1,51,101, \ldots, 9951$ را انتخاب می‌شوند.

[۱] . *AutoCorrelation Function*

[۲] . *Partial AutoCorrelation Function*

شکل ۲,۵ بافت نگار[۱] نمونه‌های جمع آوری شده است از این الگوریتم را نشان می‌دهد، که خطوط عمودی بسته نشان دهنده وضعیت مقادیر واقعی پارامتر و خطوط عمودی خط تیره برآورد بیزی پارامترها را نشان می‌دهد. مشاهده می‌شود که مقادیر واقعی پارامتر به خوبی بین پسین حاشیه‌ای قرار دارند.



شکل۲,۵ بافت نگارها نمونه‌های جمع‌آوری شده برای هر پارامتر در مطالعه شبیه‌سازی

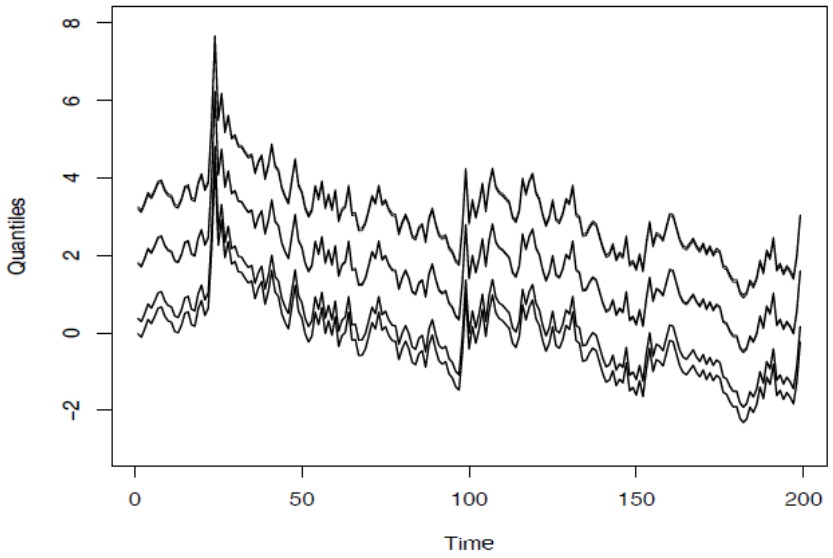
بنابراین مدل برازش شده به صورت

$$Q_{y_t}(p|Y_{t-1}) = -\circ/6\circ 27 + \circ/2958 y_{t-1} + \circ/6\circ 66 y_{t-2} - \frac{ln(1-p)}{1/6125} \qquad (8.5)$$

می‌شود. ملاحظه می‌فرمایید که مدل برازش شده خیلی مشابه مدل واقعی است. برای تاثیر چندک‌ها در سری زمانی مورد نظر به شکل 3.5 که چندک‌های شرطی مدل واقعی (۷,۵) و مدل برازش شده (۸,۵) با

[۱] . *Histogram*

چندک‌های که به ترتیب از پایین به بالا p =۰/۰ 5,۰/5,۰/95,۰/995 به‌دست آمده را نشان می‌دهد، در نظر گرفته می‌شود. واضح است که توزیع شرطی سری‌های زمانی در هر نقطه زمانی دارای چولگی به راست است و این چیزی است که انتظار داشتیم.



شکل۳,۵ چندک‌های شرطی مدل واقعی (۷,۵) و مدل برازش شده (۸,۵) که به‌ترتیب از پایین به بالا

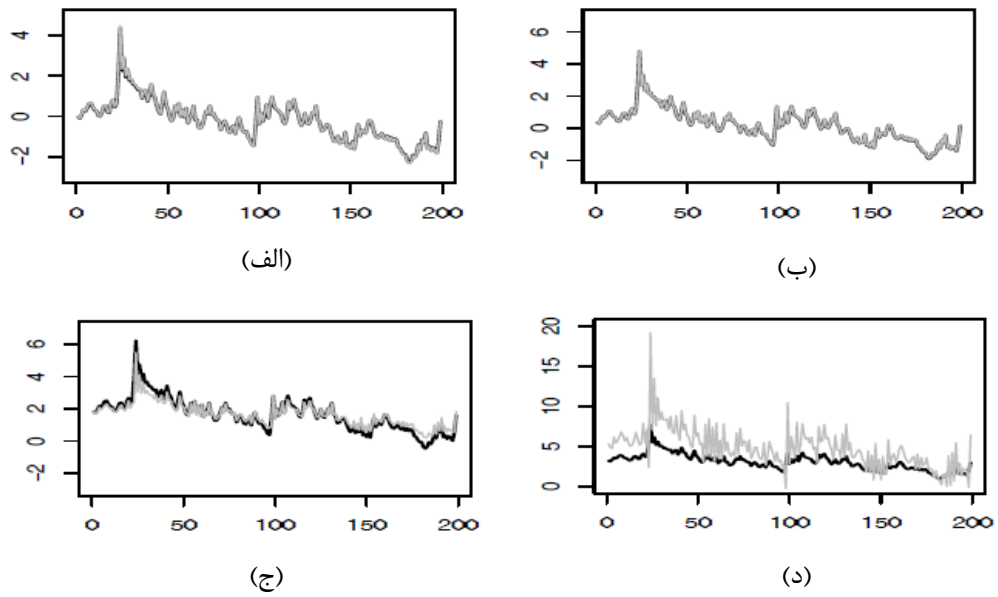
$\boldsymbol{p}$ =۰/۰ **5,۰/5,۰/95,۰/995**

متناظر با مدل(7.5)، از مدل نیمه پارامتری که توسط کانکر (۲۰۰۵) پیشنهاد شده است و به صورت زیر می‌باشد استفاده شده است:

$$q^{p}_{y_t|Y_{t-1}} = a^{p}_{\circ} + a^{p}_{1} y_{t-1} + a^{p}_{2} y_{t-2} \qquad (9.5)$$

همانطور که مشاهده می‌شود در آن پارامترها به مقادیر p وابسته هستند. از آن‌جایی که این مدل، وابستگی به جمله خطا ندارد، می‌توان از مدل نیمه‌پارامتری (۹,۵) در سری‌زمانی شبیه سازی شده مشابه نیز استفاده نمود.

کای(۲۰۰۹) از نرم افزار آماری R برای برازش کردن p امین چندک شرطی سری‌زمانی شبیه‌سازی شده استفاده کرد که در آن $p =$ ۰/۰ 5,۰/5,۰/95,۰/995 است. شکل۴,۵ چندک‌های شرطی مدل واقعی (۷,۵) (منحنی مشکی) و مدل نیمه پارامتری (۹,۵) (منحنی روشن‌تر) نشان می‌دهد.



شکل۴,۵ چندک‌های شرطی از مدل ۷,۵ (منحنی مشکی) و مدل برازش‌شده ۹,۵ (منحنی روشن‌تر) متناظر با (الف)$p =$۰ (ب) $p =$۰/5 (ج) $p =$۰/95 (د) $p =$۰/995

مشاهده می‌نمایید که وقتی p خیلی کوچک باشد، چندک‌های شرطی دو مدل خیلی مشابه هستند. و وقتی که p خیلی نزدیک یک شود، تفاوت بین دو چندک‌های شرطی خیلی بزرگ می‌شود. در واقع این یکی از مشکلات رهیافت نیمه پارامتری است، زیرا وقتی p رهیافت‌هایش به صفر یا یک نزدیک شود داده-ها قادر به برآورد کردن پارامترها نیستند، بیشتر چندک‌های برآورد شده نامطلوب هستند.

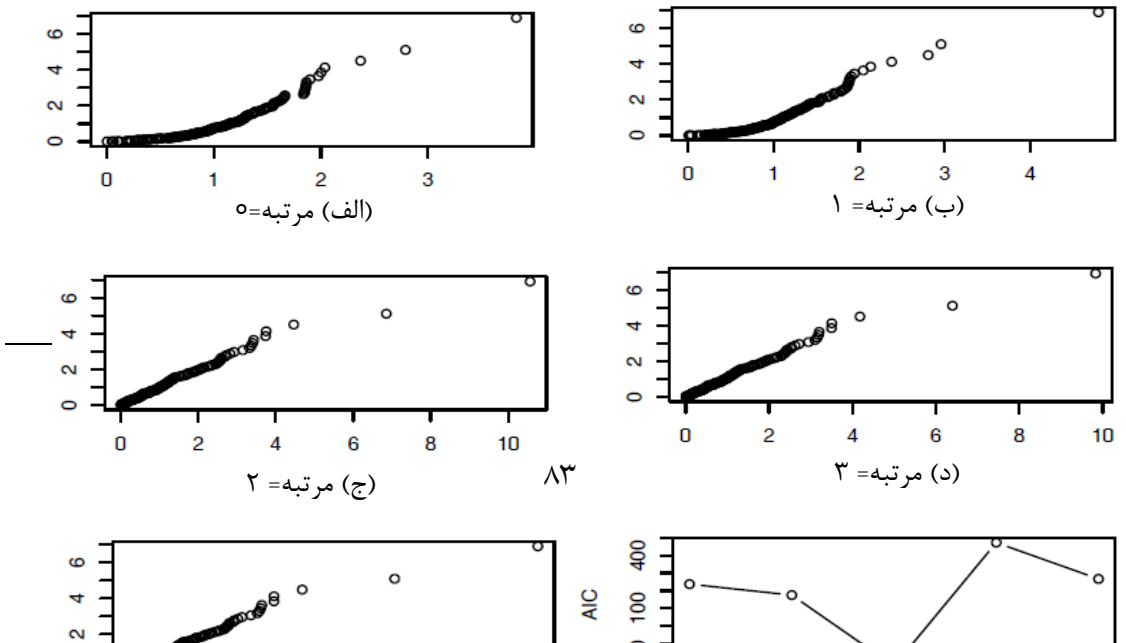
مدل برازش شده (۵,۵) را برای مرتبه‌های $k = ۰,1,2,3,4$ مشابه داده‌های شبیه‌سازی شده در نظر گرفته‌شده است تا مشاهده شود که آیا با این مرتبه‌ها قادر به تشخیص مدل واقعی هست یا نه. اگر مدل برازش شده مناسب باشد، آنگاه انتظار خواهیم داشت که باقیمانده‌ها استاندارد شده

$$e_t = \gamma(y_t - a_۰ - a_1 y_{t-1} - \cdots - a_k y_{t-k}), \quad t = k+1, \ldots, n$$

از توزیع نمایی با نرخ واحد تعریف شده به صورت $Q(p) = -ln(1-p)$ پیروی کند.

بنابراین باید انتظار داشت که نقاط روی نمودار چندک‌های نمونه، از باقیمانده‌های استاندارد شده e_t، در مقابل چندک‌های توزیع نمایی با نرخ واحد به طور تقریبی در میان خط راست قرار دارد.

شکل ۵,۵ نمودار چندک در مقابل چندک ($qq\ plot$[۱]) برای مدل‌های برازش شده مختلف نشان می‌دهد. مشاهده می‌شود که برای این مجموعه داده‌ها مدل‌های با مرتبه $k = ۰,1$ از مدل‌های با مرتبه $k = 2,3,4$ بدتر اجرا می‌شوند. به هر حال برای مدل‌های با $k \geq 2$ نمودارهای چندک-چندک خیلی مشابه به نظر می‌رسند. با بررسی مقادیر محاسبه شده معیار آکایک[۲] (AIC) توسط بکارگیری برآورد بیزی پارامترها برای هر مدل برازش شده و با توجه به شکل۵,۵ (ی) مشاهده شده است که مدل مرتبه۲ دارای کمترین مقدار آکایک نسبت به هر مدل دیگر است و نشان‌دهنده این موضوع است که در حقیقت مرتبه واقعی مدل را از ابتدا درست در نظر گرفته بود.

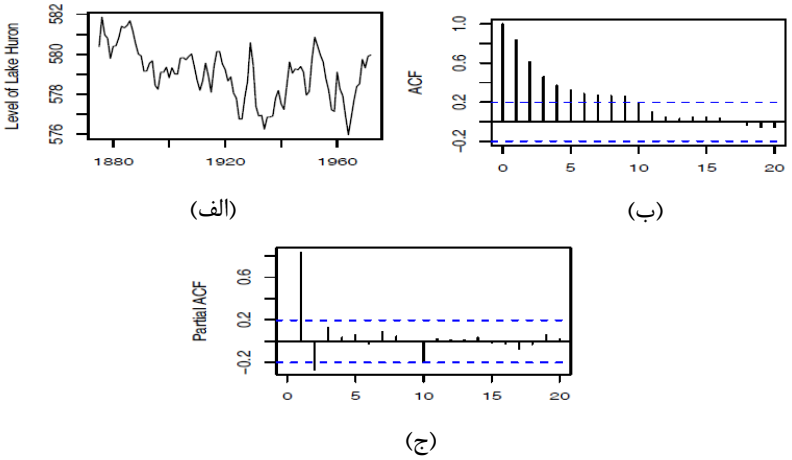


شکل۵,۵ (الف)-(ه) نمودار چندک-چندک و (ی) مقادیر آکاییک مدل‌های برازش شده از مطالعات شبیه‌سازی

۱,۴,۵ کاربردهای واقعی سری‌زمانی

داده‌های سطح آب دریاچه هورون طی سال‌های ۱۹۷۲-۱۸۷۵

در این بخش از داده‌های مربوط به سطح آب دریاچه هورون که طی سال‌های ۱۹۷۲-۱۸۷۵ اندازه‌گیری شده است، استفاده می‌کنیم. نمودار سری زمانی و نمودار تابع خودهمبستگی و تابع خودهمبستگی جزیی را در شکل ۵,۶ نشان داده‌ایم. مشاهده شد که سری زمانی دریاچه هورون خیلی مشابه ساختار همبستگی شکل ۱,۵ می‌باشد.



(الف) (ب)

(ج)

شکل۶,۵ (الف) نمودار سری‌زمانی سری دریاچه هورون . (ب) نمودار تابع خودهمبستگی سری زمانی دریاچه هورون . (ج) نمودار تابع خودهمبستگی جزیی سری زمانی دریاچه هورون.

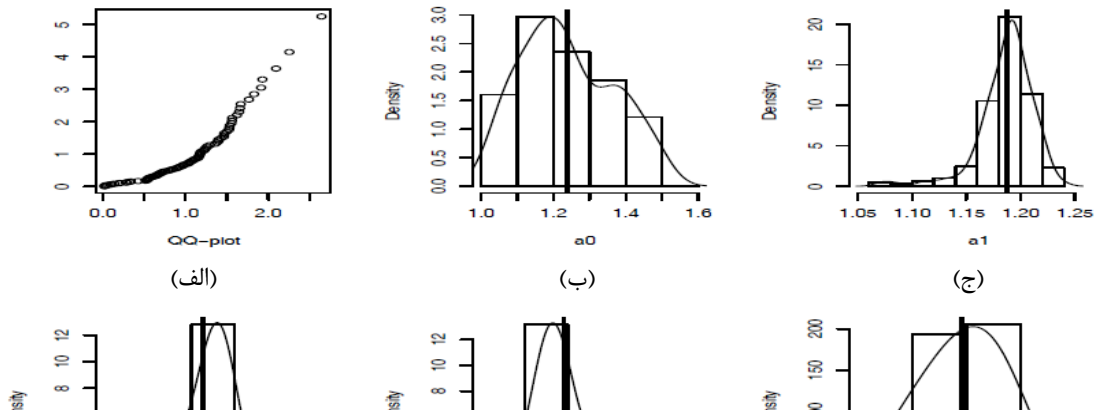
کای(۲۰۰۹) در ابتدا یک دنباله از مدل برای سری زمانی با مرتبه $k=1,2,3,4$ برازش نمود و برای هر مدل برازش شده، نمونه ۵۵۰۰۰۰ مرحله‌ای را اجرا کرد و سپس با داغیدن ۱۵۰۰۰۰ نمونه اول، نمونه‌های باقیمانده را در هر ۱۰داده ذخیره نمود، یعنی برای هر پارامتر ۴۰۰۰۰ داده داشت. در این مرحله برای اینکه مدل صحیح شناسایی شود مقادیر آکایک را بررسی می‌شوند و نمودار چندک-چندک مدل‌های برازش شده رسم شده است. مشاهده شد که آکایک و نمودار چندک-چندک، هر دو، مدل با مرتبه ۳ را پیشنهاد می‌دهندکه دارای بهترین مدل می‌باشند. بنابراین فقط نمودار بافت نگار نمونه‌های مجموعه بندی شده و چندک-چندک باقیمانده استاندارد در شکل ۷,۵ نشان داده شده است. منحنی پیوسته، نمودار چگالی و خط عمودی محل قرار گرفتن مقادیر پارامتر برآورد شده که میانگین نمونه است را نشان می‌دهد. مدل برازش شده به صورت

$$Q_{y_t}(p|Y_{t-1}) = 1/238 + 1/187y_{t-1} - ۰/537y_{t-2} + ۰/345y_{t-3} - \frac{log(1-p)}{۰/767} \qquad (10.5)$$

است و مدل نیمه پارامتری متناظر با آن به صورت

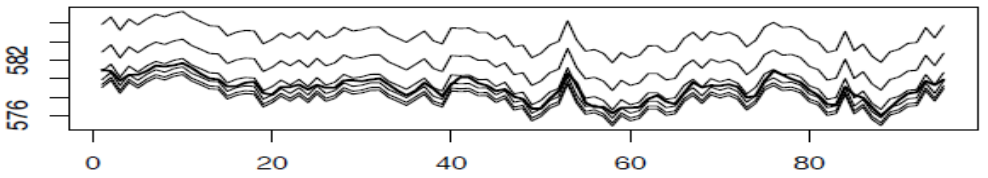
$$q^p_{y_t|Y_{t-1}} = a^p_۰ + a^p_1 y_{t-1} + a^p_2 y_{t-2} + a^p_3 y_{t-3} \qquad (11.5)$$

است. چندک‌های شرطی برازش شده برای $p =$ ۰/۰ 5,۰/25,۰/5,۰/75,۰/95,۰/995 در شکل ۸,۵ نشان داده شده است و انتظار داریم که رفتار ناسازگار روی چندک‌های شرطی برآورد شده مدل(۱۱,۵) دوباره اتفاق بیافتد.

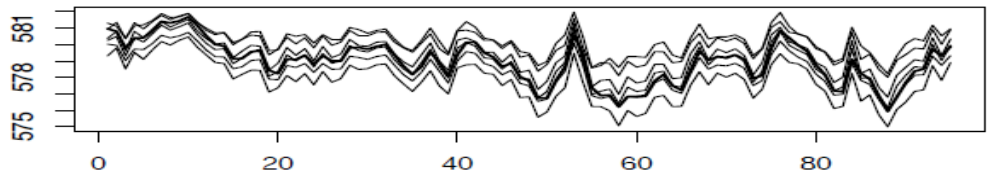


(الف) (ب) (ج)

شکل ۷,۵ (الف) نمودار چندک-چندک باقیمانده‌های مدل (ب)-(ی) نمودارهای چگالی بافت نگار(منحنی پیوسته) و مقادیر پارامتر برآورد شده(خط‌های عمودی تیره) نمونه‌های دسته بندی شده



(الف)



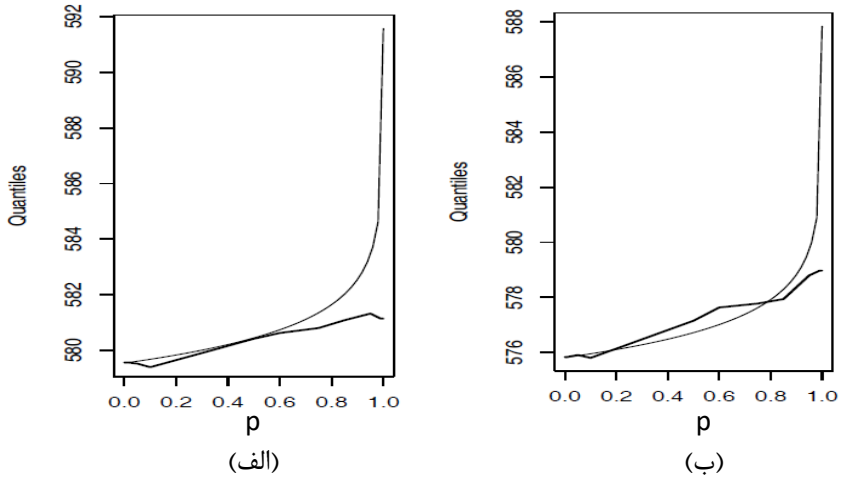
(ب)

شکل۸,۵ چندک‌های شرطی برازش شده برای(از پایین به بالا) **۵,۰/25,۰/5,۰/75,۰/95,۰/995** p =۰/۰ از (الف) مدل (۱۰,۵) و (ب) مدل (۱۱,۵) که منحنی‌های تیره‌تر سری زمانی دریاچه هورون هستند

توجه کنید که نمودار چندک-چندک شکل ۷,۵ برخی شک‌ها روی کیفیت مدل برازش شده را نشان می-دهد. به هر حال در مقایسه با رهیافت نیمه پارامتری، مدل (۱۰,۵) همچنان اجرای بهتری دارد و این برای تحقیقات بیشتر تائید کننده است.

برای اطلاعات بیشتر کای(۲۰۰۹) تابع چندک شرطی برآورد شده y_7 و y_{33} را برای مدل برازش شده (۱۰,۵) و مدل نیمه پارامتری (۱۱,۵) در نظر گرفت که این دو نقطه بطور تصادفی انتخاب شده‌اند و می‌توانیم تابع چندک شرطی y_t را برای هر t مشابهی برآورد کنیم. بنابراین شرطها به ترتیب روی Y_6و Y_{32} برآورد تابع چندک y_7 و y_{33} در نظر گرفته شده است و نمودار تابع چندک در شکل ۹,۵ نشان داده شده است که منحنی کلفت‌تر تابع چندک توسط مدل(۱۱,۵) و منحنی نازک‌تر توسط مدل (۱۰,۵) بدست آمده اند.

.



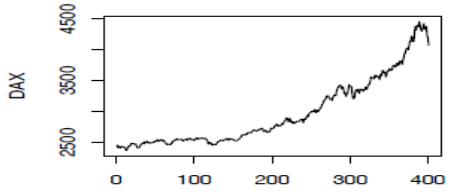
شکل۹,۵ توابع چندک شرطی برآورد شده (الف)y_7 و (بy_{33}) ،که منحنی باریک‌تر از مدل (۱۰,۵) و منحنی کلفت‌تر از مدل (۱۱,۵) بدست آمده‌اند

شاخص سهام اروپا

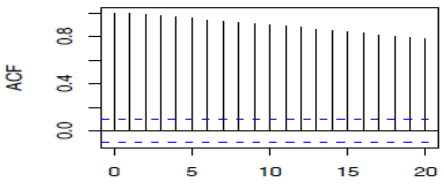
بررسی قیمت روزانه شاخص سهام در اروپا طی شش ماه دوم سال۱۹۹۱ تا شش ماه دوم سال ۱۹۹۸ جزء داده‌های استاندارد در نرم افزار R می‌باشد. کای(۲۰۰۹) به‌طور ویژه به بررسی داده‌های شاخص داکس آلمان، که رایج‌ترین معیار برای اندازه گیری برگشت سود توسط ارزش سهام در بورس اوراق بهادار فرانکفورت است، پرداخته است. این شاخص مبتنی بر عملکردشان است بدان معنی که هر گونه سود و حوادث دیگر در محاسبه نهایی شاخص تاثیرگذار است. برای بررسی شاخص این سهام طی سال‌های ۱۹۹۶-۱۹۹۵ سری‌زمانی به طول ۴۰۰ در نظر گرفته شده که در شکل ۱۰,۵ رسم شده است. دنباله‌ای از مدل (۵,۵) به این سری‌زمانی برازش نموده و مشاهده شد که $k=4$ بهترین مدل برازش شده است. شکل۱۱,۵ نمودار چندک-چندک از باقیمانده‌ها و بافت نگار نمونه‌های جمع‌آوری شده از روش $MCMC$ نشان می‌دهند، که خطوط عمودی نشان‌دهنده وضعیت مقادیر پارامتر برآوردشده هستند. بنابراین مدل برازش شده به صورت زیر است:

$$Q_{y_t}(p|Y_t) = ۰/۷۵۶ + ۰/۹۴۴y_{t-1} + ۰/۳۴۳y_{t-2} + ۰/۲۵y_{t-3} - ۰/۵۷۱y_{t-4} - \frac{log(1-p)}{۰/۰\,11} \qquad (12.5)$$

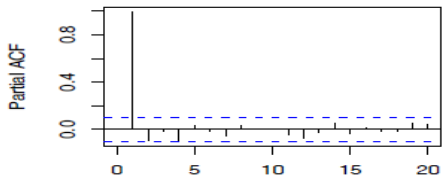
که چندک‌های شرطی برازش شده در شکل ۱۲,۵ نشان‌داده‌شده است. بر اساس مدل برازش شده (۱۲,۵) توانستند تابع چندک تبیینی یک مرحله جلوتر Y_{400} را با شرط y_{401} به‌دست آورند.



(الف)

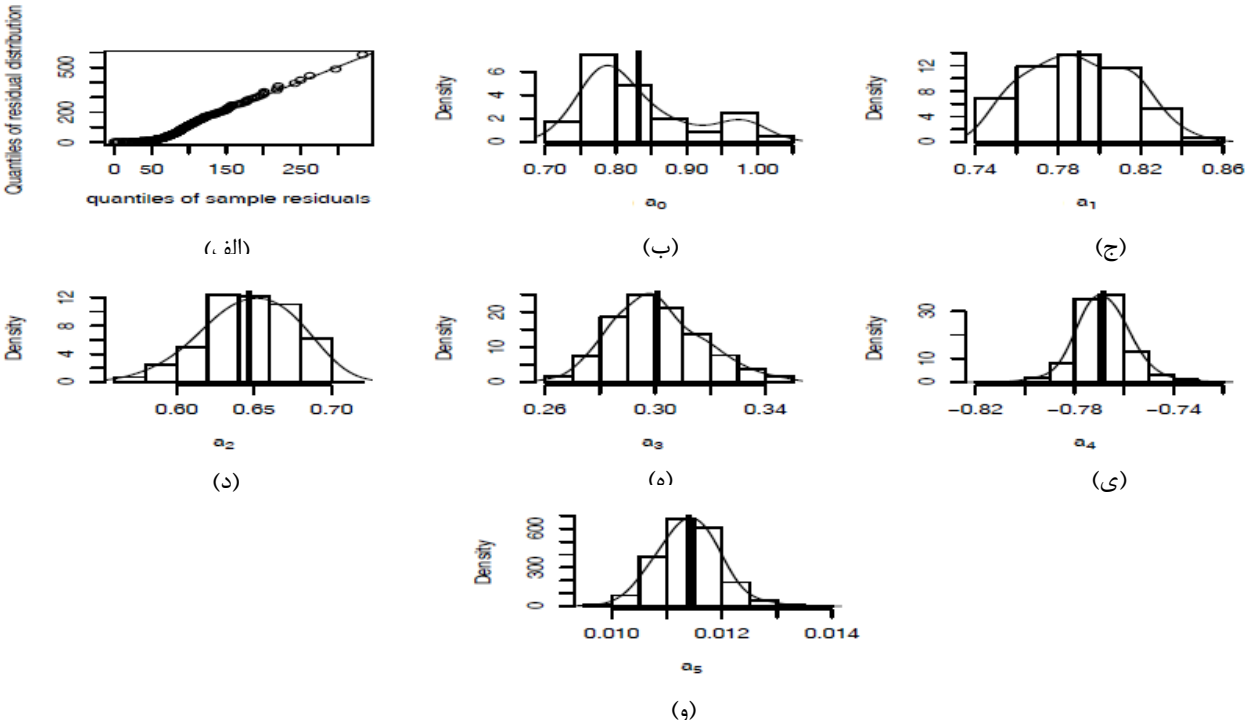


(ب)

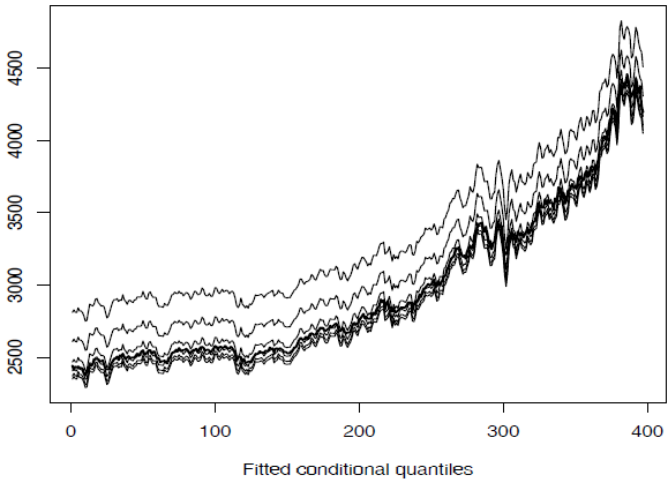


(ج)

شکل۱۰,۵ (الف) نمودار سری‌زمانی از سری داکس آلمان (ب) نمودار تابع خودهمبستگی از سری داکس آلمان (ج) نمودار تابع خودهمبستگی جزیی از سری داکس آلمان



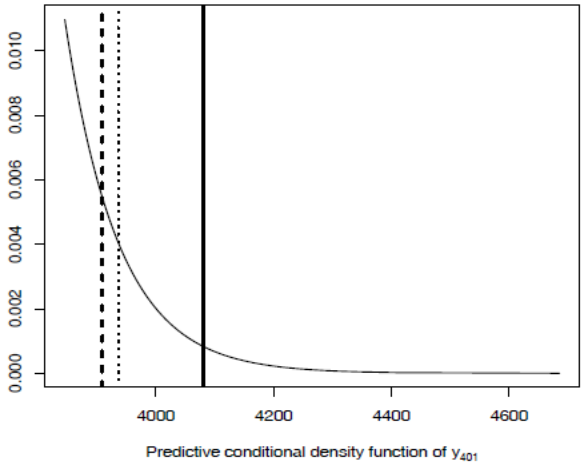
شکل۱۱,۵ (الف) نمودار چندک-چندک باقیمانده‌های مدل(۱۲,۵). (ب)-(و) نمودارهای چگالی بافت نگار (منحنی‌های پیوسته) و مقادیر پارامترهای برآورد شده (خط‌های عمودی) از نمونه‌های جمع‌آوری شده



شکل۱۲,۵ چندک‌های شرطی برازش شده (منحنی روشن‌تر) به ترتیب از پایین به بالا

$\boldsymbol{p}$ =۰/۰ **5,۰/25,۰/5,۰/75,۰/95,۰/995** ۰/۰ و منحنی مشکی سری‌زمانی واقعی

شکل ۱۳,۵ تابع چگالی شرطی y_{401} را نشان می‌دهد که در آن خط عمودی پیوسته وضعیت مقادیر مشاهده شده واقعی و خط‌های خط تیره و نقطه چین به ترتیب متناظر با میانگین پیش بینی شده و میانه پیش بینی شده مدل برازش شده(۱۲,۵) را نشان می‌دهند.



شکل ۱۳,۵ تابع چگالی شرطی پیشگویانه $\boldsymbol{y_{401}}$

در این پایان‌نامه، از رهیافت بیزی برای مدل‌بندی سری‌زمانی خود برگشت غیرنرمال با تابع چندک استفاده شده است که این رهیافت با رهیافت نیمه پارامتری قابل مقایسه بود که با مطالعه شبیه‌سازی و سری‌زمانی واقعی، به خوب بودن عملکرد آن پی بردیم.

از این روش در آینده می‌توان برای سری‌های زمانی غیر خطی و حالتی که پارامتر مقیاس از مدل به‌مقادیر سری‌های زمانی نیز بستگی دارد، توسعه داده شوند. به علاوه روش‌های پیش‌بینی رو به جلو مرحله چندگانه از مدل (۳,۵) به توسعه در آینده نیاز دارند، روش جدید نیز به توسعه سری‌های زمانی چندمتغیره نیاز دارند.

پیوست

پیوست الف: روش مونت کارلو با زنجیره‌ی مارکف *MCMC*

برای تقریب کمیت

$$\theta = E(X) = \int_S h(x)f(x)dx$$

به روش مونت کارلو، با استخراج مشاهدات مستقل و هم توزیع، از چگالی $f(x)$ و استفاده از $\hat{\theta}_{MC} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$ کمیت θ را تقریب می‌زنیم. اما استخراج نمونه از چگالی $f(x)$ مشکل اساسی روش مونـت کارلو می‌باشد. برای این کار دو روش معکوس تابع توزیع و روش پذیرش–رد را می‌توان معرفی کرد. حـال به دنبال روش قدرتمندی برای استخراج نمونه هستیم که محدودیتهای کمتری داشته باشـد. در زنجیـر-های مارکوف ارگودیک، می‌توان الگوریتمی را معرفی کرد که به کمک آن از چگالی هدف $f(x)$ با تکیه‌گاه S نمونه استخراج کرد. برای این کار، لازم بود که یک زنجیره‌ی مارکوف ارگودیک با توزیع ایستای حـدی یکتای $f(x)$ داشته باشیم بطوریکه از هر سطر ماتریس احتمال انتقالات در حالـت S شـمارا و یـا از هـر چگالی $k(x,y)$، برای $x \in S$ ثابت، در حالت S ناشمارا بتوان نمونه تولید کرد. به علاوه زنجیره‌ی مارکوف تحویل ناپذیر، نامتناوب و زمان برگشت پذیر ارگودیک است. حال مسأله این است که چگونه می‌توان یک زنجیره‌ی مارکوف تحویل ناپذیر، نامتناوب و زمان برگشت پذیر با توزیع ایستای حدی یکتای $f(x)$ بـر S ساخت به طوریکه از سطرهای ماتریس احتمال انتقالات و یا چگالی‌های هسـته احتمـال انتقـالات بتـوان نمونه تولید کرد.

در واقع روش مونت کارلو با زنجیرهای مارکوف $(MCMC)$، چنین زنجیره‌ای را برای اسـتخراج یـک نمونه‌ی تقریبی از $f(x)$ ساخته (بخش زنجیره‌ی مارکوف روش) و با استخراج نمونـه، روش مونـت کـارلو (بخش مونت کارلو روش) را برای تقریب کمیت‌ها به کار می‌برد. مارکوف نشان داد که قانون اعداد بـزرگ در حالت وابستگی مارکوفی نمونه، باز هم برقرار است. بنـابراین در بخـش مونـت کـارلو روش، چنانچـه از تقریبی بودن توزیع کناری مشاهدات به‌دست آمده چشم پوشی و فرض کنیم که مشاهدات مـا وابسـته‌ی مارکوفی باشند و توزیع کناری آنها $f(x)$ است آنگاه با پشتوانه‌ی قانون اعداد بزرگ می‌تـوان روش مونـت کارلو را برای تقریب کمیت‌ها به کار برد. لذا بخش مونت کارلو روش با پذیرش یک تقریب سر راست است.

اما مسأله اصلی ساختن زنجیره‌ی مارکوف تحویل ناپذیر، نامتناوب و زمان برگشت پذیر با چگالی ایستای حدی یکتای $f(x)$، یعنی بخش زنجیره‌ی مارکوف $MCMC$ می‌باشد.

برای تولید چنین زنجیره‌ای دو الگوریتم ارایه شده است: الگوریتم متروپلیس- هستینگز و نمونه گیری گیبز

الگوریتم متروپولیس-هستینگس

این الگوریتم با انتخاب یک هسته‌ی احتمال انتقالات $g(y|x)$ بر S، با داشتن مشاهده $X^{(t)}$، پس از سه گام، مشاهدی $X^{(t+1)}$ را به ما می‌دهد. بدین منظور به یک مقدار اولیه برای به‌دست آوردن سایر مشاهدات نیاز داریم:

در زمان t=۰، $X^{(t)} = x^{(0)}$ را از توزیع ابتدایی دلخواه $\Pi^{(0)}$ با چگالی $\pi^{(0)}(x)$، که بر S تعریف شده است، با این شرط که $f(x^{(0)}) >$۰، تولید کند. البته می‌توان $x^{(0)} \in S$ را با شرط $f\left(x^{(0)}\right) >$۰، به دلخواه انتخاب کرد که معادل است با انتخاب یک توزیع بر S با چگالی

$$\pi^{(0)}\left(x^{(0)}\right) = 1 \;;\; \forall y \in S, y \neq x^{(0)} \;:\; \pi^{(0)}(y) =0$$

حال با داشتن $X^{(t)} = x^{(t)}$، الگوریتم مقدار $X^{(t+1)}$ را بصورت زیر تولید می‌کند:

۱. مقدار پیشنهادی X^* را از چگالی $g(.|x^{(t)})$ تولید می‌کنیم.

۲. نسبت متروپلیس- هستینگز را که بصورت $R(u,v) = \frac{f(v)g(u|v)}{f(u)g(v|u)}$ تعریف شده است، را برای $(x^{(t)}, X^*)$ محاسبه می‌کنیم، یعنی $R(x^{(t)}, X^*)$.

۳. $X^{(t+1)}$ را بصورت زیر انتخاب می‌کنیم:

- با احتمال $min\{1, R(x^{(t)}, X^*)\}$; $X^{(t+1)} = X^*$
- با احتمال $1 - min\{1, R(x^{(t)}, X^*)\}$; $X^{(t+1)} = x^{(t)}$

این الگوریتم را تا زمانی که تعداد مشاهدات مورد نیاز به‌دست آید، تکرار می‌کنیم. لذا این الگوریتم با انتخاب هسته‌ی احتمال انتقالات $g(y|x)$ بر S، و مقدار اولیه‌ی با احتمال $x^{(0)}$ یک زنجیر ما می‌دهد.

توجه

- مقدار با احتمال X^* را که گام یک تولید می‌شود، مقدار پیشنهادی می‌گوییم چون در گام سوم پذیرفته یا رد می‌شود.
- چگالی شرطی $g(y|x)$ را که آن مقادیر پیشنهادی X^* را با داشتن $x^{(t)}$ تولید می‌کنیم، چگالی پیشنهادی می‌گوییم.
- نسبت متروپلیس- هستینگز $R(x^{(t)}, X^*)$ همواره تعریف می‌شود. زیرا مقدار پیشنهادی $X^* = x^*$ تنها وقتی تولید می‌شود که $f(x^{(t)}) > 0$ و $g(x^{(*)}|x^{(t)}) > 0$.
- مارکوف بودن زنجیر به‌دست آمده با این روش بدیهی است زیرا $X^{(t+1)}$ تنها به $X^{(t)}$ بستگی دارد.
- تحویل ناپذیری و نامتناوب بودن زنجیر به انتخاب چگالی پیشنهادی $g(y|x)$ بستگی دارد.

بنابراین باید $g(y|x)$ طوری انتخاب شود که در نهایت زنجیره‌ی به‌دست آمده از الگوریتم تحویل ناپذیر و نامتناوب باشد.

- زمان برگشت پذیری زنجیر همواره برقرار است.

در الگوریتم متروپولیس-هستینگس مسأله پیدا کردن $g(y|x)$ ایی است

اولاً برای هر $x \in S$ ثابت قابل نمونه گیری باشد

ثانیاً باعث شود زنجیره‌ی حاصل از الگوریتم ارگودیک باشد.

علاوه بر اینها، ملاک‌های دیگری را نیز برای انتخاب $g(y|x)$ مناسب باید مد نظر قرار داد. واضح است که مشاهدات حاصل از الگوریتم باید نمونه‌ی خوبی از جامعه‌ی $f(x)$ باشند. به عنوان مثال زنجیرای که بیشتر مقادیر پیشنهادی را رد می‌کند، منجر به مشاهداتی از $f(x)$ می‌شود که خیلی از آنها تکراری و

ثابت هستند و بنابراین این نمونه نمی‌تواند پراکندگی جامعه‌ی $f(x)$ را به خـوبی نشـان دهـد. همچنـین زنجیرای که همه‌ی مقادیر پیشنهادی را بپذیرد، منجر به مشاهداتی از $f(x)$ می‌شود کـه بسـیار پراکنـده بوده و این در سرعت همگرایی توزیع مشاهدات به توزیع حدی $f(x)$ تاثیر گذاشته، باعث می‌شـود بـرای به‌دست آوردن یک نمونه‌ی قابل قبول، مشاهدات بسیار زیادی را بـا الگـوریتم تولیـد کنـیم. لـذا انتخـاب $g(y|x)$ مناسب در الگوریتم متروپولیس-هستینگس حیاتی است. اجراهای عملی این الگوریتم نشان داده که با انتخاب $g(y|x)$ مناسب، نتایج بسیار رضایت بخش‌اند.

در حالت کلی نمی‌توان قاعده‌ای برای انتخاب $g(y|x)$ مناسب ارایه کرد. بلکه باید با توجه به ساختار چگالی$f(x)$ و تکیه‌گاه آن S، و با در نظر گرفتن شرط‌ها و ملاک‌های انتخاب $g(y|x)$ مناسب، $g(y|x)$ را انتخاب کرد. بهترین کار دسته‌بندی مسائلی است که با روش $MCMC$ در پی حل آنهـا هسـتیم. در آمـار بیشترین کاربرد این روش در استنباط بیزی است. وقتی که می‌خواهیم میانگین پسـین، میانـه‌ی پسـین، احتمال‌های پسین و ... را محاسبه کنیم، اما در حل معمول انتگرال حاصـل از تعریـف ایـن کمیـت‌هـا، بـا مشکل روبروهستیم. استفاده از زنجیرهای مستقل[1] حداقل در این مورد خاص بسیار کارساز است.

الگوریتم متروپلیس- هستینگز و زنجیرهای مستقل

در الگوریتم متروپلیس- هستینگز، $g(y|x)$ طوری انتخاب شود که داشته باشـیم: $g(y|x) = g(y)$، کـه در آن $g(x)$ چگالی یک توزیع بر S است، در اینصورت هر مقدار پیشنهادیX^* مستقل از مشاهده‌ی قبلی $x^{(t)}$ از $g(x)$ تولید می‌شود. لذا زنجیره‌ی حاصل از الگوریتم مستقل می‌باشد.

با این انتخاب، نسبت متروپلیس- هستینگز به فرم زیر است:

$$R\left(x^{(t)}, X^*\right) = \frac{f(X^*)g(x^{(t)})}{f(x^{(t)})g(X^*)}$$

[1]. *Independent Chains*

به علاوه اگر برای هر $x \in S$ که $f(x) >0$، داشته باشیم: $g(x) >0$ آنگاه زنجیر حتماً تحویل ناپـذیر و نامتناوب بوده، لذا ارگودیک با چگالی ایستای حدی یکتای $f(x)$ می‌باشد. بنابراین در این حالـت دیگـر نیازی به چک کردن تحویل ناپذیری و نامتناوب بودن زنجیره‌ی حاصل از الگوریتم نبوده، بلکـه کافیسـت شرط فوق در مورد $g(x)$ صدق کند.

الگوریتم متروپلیس- هستینگز و زنجیر قدم زدن تصادفی

در الگوریتم متروپلیس- هستینگز، $g(y|x)$ طوری انتخاب شود که داشته باشیم: $g(y|x) = g(y-x)$، که در آن $g(y-x)$ چگالی یک توزیع بر S است. اگر چگالی g حول صفر متقارن باشد، زنجیـر مـارکوف متناظر با آن، یک فرآیند قدم زدن تصادفی است.

با این انتخاب، نسبت متروپلیس- هستینگز به فرم زیر است:

$$R\left(x^{(t)}, X^*\right) = \frac{f(X^*)}{f(x^{(t)})}$$

احتمال پذیرش به g وابسته نیست. بـه ایـن معنـی کـه بـرای یـک جفـت مفـروض$(x^{(t)}, X^*)$، احتمـال پذیرش X^* از همه توزیع‌ها با هم یکسان است، اما انتخاب gهای مختلـف منجـر بـه دامنـه‌هـای مختلـف مقادیر برای X^*ها و در نتیجه نرخ‌های پذیرش متفاوت خواهد شد.

نمونه گیری گیبز

برای تقریب $\theta = E(X) = \int_S h(x)f(x)dx$ به روش مونت کارلو وقتی که با روش‌هـای معمـول نتـوان نمونه‌ای از $f(x)$ استخراج کرد، الگوریتم متروپلیس- هستینگز روشی عمـومی بـرای تولیـد یـک زنجیـر ارگودیک با چگالی ایستای حدی یکتای $f(x)$ می‌باشد که تحقق‌های این زنجیـر را مـی‌تـوان بـه عنـوان نمونه‌ای تقریبی از $f(x)$ فرض کرد. اما اجرای این الگوریتم زمانی امکان پذیر است که هسـته‌ی احتمـال

انتقالات (چگالی پیشنهادی) مناسبی را برای تولید یک مشاهده به شرط داشـتن مشـاهده قبلـی، داشـته باشیم. اشاره شد که بطور کلی نمی‌توان قاعده‌ای برای انتخاب توزیع پیشنهادی ارایه کرد.

بنابراین بهتر است بر اساس مسائلی که بـه روش *MCMC* مـی‌خـواهیم حـل کنـیم، الگـوریتم‌هـای مخصوصی را طرح کرده و زنجیرهای تولید شده با این الگوریتم‌ها را دسته‌بندی کنـیم. مـثلاً در الگـوریتم متروپلیس- هستینگز، یک دسته‌بندی بر اساس انواع زنجیرهایی که می‌توان با این الگـوریتم تولیـد کـرد، زنجیرهای مستقل است.

حال با طرح مسأله‌ای، الگوریتم دیگری را برای حل این مسأله طرح می‌کنیم:

فرض کنیم $\boldsymbol{X} = (X_1, \dots, X_p) \sim f(x)$، $\boldsymbol{X}_{-i}$ را تعریف می‌کنیم:

$$\boldsymbol{X}_{-i} = \left(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p\right) \; ; \; i = 1, \dots, p$$

(بردار $\boldsymbol{X} = (X_1, \dots, X_p)$ با حذف مولفه‌ی iام). همچنین برای $i = 1, \dots, p$، توزیع شرطی $X_i|\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$ با چگالی $f(X_i|x_{-i})$ را در نظر می‌گیریم.

می‌خواهیم $\theta = E(X) = \int_S h(x)f(x)dx$ را به روش مونت کارلو تقریب بزنیم، اما بـه روش‌هـای معمول نمی‌توانیم از $f(x)$ نمونه استخراج کنیم. چنانچه برای $i = 1, \dots, p$، $f(x_i|\boldsymbol{x}_{-i})$ طوری باشد کـه با داشتن $\boldsymbol{x}_{-i}$ بتوان از آن نمونه تولیـد کـرد، آنگـاه بـا انتخـاب وضـعیت اولیـه‌ی $x^{(0)}$، الگـوریتم زیـر بـا داشتن $\boldsymbol{X}^{(t)}$، $\boldsymbol{X}^{(t+1)}$ را می‌توان تولید کرد:

۱) یک ترکیب از مولفه‌های $\boldsymbol{x}^{(t)}$ را انتخاب می‌کنیم.

۲) برای $i = 1, \dots, p$، X_i^* را از $f(x_i|\boldsymbol{x}_{-i}^{(t)})$ تولید کرده، تا $\boldsymbol{X}^* = (X_1^*, \dots, X_p^*)$ به‌دست آید.

۳) قرار می‌دهیم: $\boldsymbol{X}^{(t+1)} = \boldsymbol{X}^*$

توجه

- این الگوریتم را گیبز برای تقریب یک انتگرال در فیزیک پیشنهاد کرده است و به همـین خاطر به نمونه گیری گیبز یا الگوریتم گیبز مشهور است.
- زنجیرای که با این الگوریتم به‌دست می‌آید، $\{\boldsymbol{X}^{(t)}\}_{t\geq 0}$، مارکوف است زیـرا $\boldsymbol{X}^{(t+1)}$ تنهـا به $\boldsymbol{X}^{(t)}$ وابسته است.
- در گام دوم به جای اینکه $\boldsymbol{X}^*$ از $f(\boldsymbol{x}|\boldsymbol{x}^{(t)})$ تولید شود، با تولید p مولفه‌ی X_i^* بردار $\boldsymbol{X}^*$ به‌دست می‌آید. برای اینکه ترتیب تولید مولفه‌ها بر تولید $\boldsymbol{X}^*$ تاثیر نداشـته باشـد، در گـام اول از بین $p!$ جایگشت ممکن از مولفه‌های $\boldsymbol{X}^{(t)}$ یکی را انتخـاب و گـام دوم را بـر اسـاس ایـن ترتیـب انتخاب شده اجرا می‌کنیم. به همین دلیل در برخی از موارد ابتـدا یـک توزیـع احتمـال را بـر $p!$ جایگشت مولفه‌های $\boldsymbol{X}^{(t)}$ در نظر گرفته، گام اول را چنین بیان می‌کنند: یک جایگشت را از توزیع تعریف شده بر جایگشت‌های مولفه‌های $\boldsymbol{X}^{(t)}$ تولید می‌کنیم. سپس گام دوم را با این ترتیب تولید شده اجرا می‌کنند.
- اجرای گام دوم بر اساس ترتیب انتخاب شده در گام اول را یک سیکل می‌گوییم.
- برای یک سیکل خاص، در i امـین مرحلـه‌ی سـیکل الگـوریتم بـا داشـتن $\boldsymbol{X}^{(t)}$، مقـدار پیشنهادی $\boldsymbol{X}^* = (x_1^{(t)}, \dots, X_i^*, \dots, x_p^{(t)})$ را از چگالی پیشنهادی $g_i(\boldsymbol{x}^*|\boldsymbol{x}^{(t)})$ تولید می‌کند. که در آن:

$$g_i(\boldsymbol{x}^*|\boldsymbol{x}^{(t)}) = \begin{cases} f\left(\boldsymbol{x}_i^*\middle|\boldsymbol{x}_{-i}^{(t)}\right) \;;\; \boldsymbol{x}_{-i}^* = \boldsymbol{x}_{-i}^{(t)} \\ 0 \quad ; \quad \text{سایر نقاط} \end{cases}$$

در این حالت $R(\boldsymbol{x}^{(t)}, \boldsymbol{x}^*) = \frac{f(x^*)g(x^{(t)}|x^*)}{f(x^{(t)})g(x^*|x^{(t)})} \equiv 1$، یعنی مقدار پیشنهادی $\boldsymbol{X}^*$ پذیرفته می‌شـود. بنابراین هر سیکل از p الگوریتم متروپلیس- هستینگز تشکیل شده است و می‌توان الگوریتم فوق را حالت خاصی از متروپلیس- هستینگز در نظر گرفت.

- می‌دانیم که $f(x_i|\boldsymbol{x}_{-i}) = c_i f(\boldsymbol{x})$، که در آن $c_i^{-1} = \int f(\boldsymbol{x}_{-i})d\boldsymbol{x}_{-i}$. چنانچه در این الگوریتم $f(x_i|\boldsymbol{x}_{-i})$ برای برخی از i ها نرمال نشده باشد[1] یعنی : c_i مجهول باشد، چون هر سیکل در الگوریتم متشکل از p الگوریتم متروپلیس- هستینگز است، مشکلی برای اجرای الگوریتم نخواهیم داشت.

- برای بهبود عملکرد این الگوریتم، در یک سیکل وقتی که مولفه‌ی اول پیشنهادی X_1^* با داشتن $\boldsymbol{X}^{(t)}$ تولید شد، مولفه‌ی دوم پیشنهادی X_2^* را از $g_i(\boldsymbol{x}^*|x_1^{(t+1)}, x_2^{(t)}, \dots, x_p^{(t)})$ به جای $g_i(x^*|x_1^{(t)}, x_2^{(t)}, \dots, x_p^{(t)})$ تولید کرده و به همین ترتیب برای سایر مولفه‌ها تا کامل شدن سیکل از مقدار جدید $x_i^{(t+1)}$ به جای $x_i^{(t)}$ استفاده می‌کنیم.

- زنجیره‌ی مارکوفی که با این الگوریتم تولید می‌شود تحویل ناپذیر، نامتناوب و زمان برگشت پذیر و در نتیجه ارگودیک بوده، بنابراین این الگوریتم نیز یکی از ابزارهای روش *MCMC* است

[1] تابع $f: \Re \to \Re$ تابع چگالی است اگر ۱) $\forall x : f(x) \geq 0$ ۲) $\int_{\Re} f(x)dx = 1$. اگر f در شرط اول صدق کند و بدانیم $\int_{\Re} f(x)dx < \infty$، آنگاه f را یک چگالی نرمال نشده می‌گوییم

منابع

ساموئل کارلین، هووارد ام. تیلور؛(۱۳۷۲). نخستین درس در فرآیندهای تصادفی؛ ترجمه: علی‌اکبر عالم زاده، عین اله پاشا. **ناشر: موسسه نشر علوم نوین صفار**

موریس دگروت، مارک اسکرویش،(۱۳۸۲) احتمال و آمار، ترجمه: پاشا عین اله، **انتشارات مبتکران. ۲ج**

داگلاس مونتگمری، الیزابت پک؛(۱۳۸۶)، مقدمه‌ای بر تحلیل رگرسیون خطی؛ ترجمه: سید ابراهیم رضوی پاریزی. **انتشارات دانشگاه شهید باهنر کرمان**

Altman, N. S., (1990). *Kernel smoothing of data with correlated errors.* ***J. Am. Statist. Ass.,* 85, 749-759.**

Albert, J. H., Chib, S., (1993). *Bayesian analysis of binary and polychotomous response data.* ***J. Am. Statist. Ass.,* 88, 669–679.**

Balakrishnan, N., Cutler, C. D., (1994). *Maximum Likelihood Estimation of the Laplace Parameters Based on TypeII Censored Samples Volume.* ***Springer Verlag, New York.***

Barndorf-Nielsen, O., Shepherd, N., (2001). *Non-Gaussian Orensien-Uhlenbeck-based models and some of their uses in financial economics.* ***J.R. Statist. Soc., B* 63, *part*2, *pp.*167-241**

Birkes, D., Dodge, Y., (1993). *Alternative Methods of Regression.* ***Wiley and Sonc New York.***

Buchinsky, M., (1998). *Recent Advances in Quantile Regression Models.* ***J. Hurn. Res.,* 33, 88-126.**

Cai, Z., (2002). *Regression quantiles for time series.* ***Econometr. Theory,* 18, 169–192.**

Cai, Y., (2009). *Autoregression with Non-Gaussian Innovations.* ***J. Time Series Econometrics., Vol*.1, *Issue* 2.**

Chen, R., Tsay, R. S., (1991). *On the ergodicity of TAR*(1*) processes.* ***Annals of Applied Probability*1, 613–34.**

Chibs, S., Greenberg, E., (1995). *Understanding the Metropolis-Hasting Algorithm.* ***American Statistican,* 49, 327-335.**

Cole, T. J., (1988). *Fitting smoothed centile curves to reference data* (*with discussion*). ***J. R. Statist. Soc. A,* 151, 385-418.**

Cole, T. J., Green, P. J., (1992). *Smoothing reference centile curves: the LMS Method and penalized likelihood*. ***Statist*. *Med.* 11, 1305-1319.**

Crowley, J., Hu, M., (1977). *Covariance analysis of heart transplant data.* ***J. Am. Statist. Ass., 72,27–36.***

Damsleth, A. H., El-Shaarawi, E., (1989). *ARMA Models with Double-Exponentially Distributed Noise.* ***J. R Statist Soc. B,* 51, 61-69.**

Efron, B., (1986). *Double Exponential Families and their use in Generalized Linear Regression.* ***J.Am. Statist. Ass.,* 81, 709-721.**

Farcomeni, A., (2010). *Bayesian constrained variable selection.* ***Statistica Sinica* 20, 1043–1062.**

Fernandez, C., Steel, M., (1998). *On bayesian modeling of fat tails and skewness.* ***J. Am. Statist. Ass., 93, 359-371.***

Gannoun, A., Saracco, J., Yu, K., (2003). *Nonparametric prediction by conditional median and quantiles.* ***J. Statist. Planng Inf.***

Gelfand, A. E., Smith, A. F. M., (1990). *Sampling based approaches to calculating marginal densities.* ***J. Am. Statist.Ass., Vol. 85, No. 410. pp. 398-409.***

George, E.L., McCulloch, R.E., (1993). *Variable selection via Gibbs sampling.* ***J. Am . Statist .Ass., 88, 881–889.***

Geraci, M., Bottai, M., (2007). *Quantile regression for longitudinal data using the asymmetric laplace distribution.* ***Biostatistics* 8, 140–154.**

Gilchrist, W. G., (2000). *Statistical modelling with quantile functions.* ***Chapman & Hall/ CRC.***

Givens, G., Hoeting. J., (2005). *Computational statistics.*

Hall, P.,W olff,R. C. L., Yao, Q ., (1999). *Methods for estimating a conditional distribution.* ***J. Am. Statist. Ass.,* 94,154–163.**

Harter, H. L., Moore, A. H., Curry, T. F., (1979). *Adaptive robust estimation of location and Scale parameters of symmetric populations.* ***Computations in Statistics- Theory and Methods* 8, 1473-1492.**

Heagerty, P. J., Pepe, M. S., (1999). *Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children.* ***Appl. Statist.,* 48, 533–551.**

Huber, P. J., (1981). *Robust statistics.* ***Wiley, New York.***

Ishwaran, h., Rao, J. S., (2005). *Spike and slab variable selection frequentist and Bayesian strategies.* ***Institute of Mathematical Statistics. Vol.*33, *No.* 2,730-744.**

Ji, Y., Lin, N., Zhang, B., (2012). *Model selection in binary and tobit quantile regression sing the Gibbs sampler.* ***Computational Statistics and Data Analysis* 56, 827-839.**

Koenker, R., Bassett, G. W., (1978). *Regression for quantile regression.* ***Econometrica.* 46, 33-50.**

Koenker, R., Bassett,G .W., (1982). *Robust tests for heteroscedasticity based on quantile regression.* ***Econometrica.* 50, 43-62**

Koenker, R., D'Orey, V., (1987). *Computing regression quantiles.* ***J . Applied Statistics,* 36, 383–393.**

Koenker, R., D'Orey, V., (1994). *A remark on Algorithm AS*229*: Computing dual regression quantiles and regression rank scores.* ***J . Applied Statistics,* 43,*414-410.***

Koenker, R., Geling, R., (2001). *Reappraising medfly longevity a quantile regression survival analysis.* ***J. Am. Statist. Ass.,* 96,*4*58–468.**

Koenker, R., Machado, J. A. F., (1999). *Goodness of fit and related inference processes for quantile regression.* ***J. Am. Statist. Ass.,* 94, 1296–1310.**

Koenker, R., (2005). *quantile regression.* ***Cambridge University Press.***

Komujer, I., (2005). *Quasi-maximum likelihood estimation for conditional quantiles.* ***J. Econometrics* 128, 137–164.**

Kottas, A., Gelfand, A. E., (2001). *Bayesian semiparametric median regression model.* ***J. Amer. Statist. Ass.*****, 96, 1458-1468.**

Kozumi, H., Kobayashi, G., (2011). *Gibbs sampling methods for Bayesian quantile regression.* ***J. Statistical Computation and Simulation.***

Kuo, L., Mallick, B., (1998). *Variable selection for regression models.* ***Sankhya B*****60, 65–81.**

Lempers, F. B. (1971). *Posterior Probabilities of Alternative Linear Models.* ***Rotterdam Univ. Press.***

Manski, C.F., (1985). *Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator.* ***J. Econometrics*** **27,313–333.**

Mitchell, T. J. and Beauchamp, J. J. (1988). *Bayesian variable selection in linear regression (with discussion).* ***J. Amer. Statist. Assoc.*** **83 1023–1036.**

Petruccelli, J., Woolford, S. W., (1984). *A threshold AR*(1) *model.* ***J. Applied Probability*** **21, 270–86.**

Powell, J., (1986). *Censored regres**si**on quantiles.* ***J. Econometrics*** **32, 143–155.**

Reed, C., Dunson, D.B., Yu, K., (2010). *Bayesian quantile regression model selection using stochastic search. Tech. Rep.,* ***Brunel Univ****.*

Reich, B.J., Bondell, H.D., Wang, H.X., (2010). *Flexible Bayesian quantile regression for independent and clustered data.* ***Biostatistics*** **11, 337–352***.*

Roberts, C. O., Smith, A. F. M., (1994). *Simple Conditions for the convergence of the gibbs sampler and hastings-metropolis algorithms.* ***Stochastic Processes and their Applications.*** **49, 207-216.**

Royston, P., Altman, D. G., (1994). *Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling with discussion.* ***J . Appl. Statist.*** **43, 429-467.**

Royston, P., Wright, E. M., (2000). *Goodness-of-fit statistics for age-specific reference intervals.* ***Statist. Med.*****,19,2943–2962.**

Smith, M., Kohn, R.,(1996). *Nonparametric regression using Bayesian variable selection.* ***J. Econometrics*** **75, 317–343.**

Taddy, M.A., Kottas, A., (2010). *A Bayesian nonparametric approach to inference for quantile regression.* ***J. Business and Economic Statistics*** **28,357–369.**

Tanner, M. A., Wong, W. H., (1987). *The calculation of posterior distributions (with discussion)* . ***J. Amer. Statist. Assoc.*** **82, 528-550.**

Tong, H., Lim, K. S. (1980). *Threshold autoregression, limit cycles and cyclical data* (*with discussion*). ***J. Royal Staistical Society, Series B*** **42, 245–92.**

Tsionans, E. G., (2003). *Bayesian quantile inference.* ***J. Statist. Computation and Simulation.*** **73,** ***No.*****9,659-674**.

Walker, S.G., Mallick, B.K., (1999). *A Bayesian semiparametric accelerated failure time model.* ***Biometrics*** **55, 477–483.**

Wang, H.S., Li, G., Jiang, G., (2005). *Robust regression shrinkage and consistent variable selection through the LAD-lasso.* ***J. Business and Economic Statistics*** **25, 347–355.**

Yang, S., (1999). *Censored median regression using weighted empirical survival and hazard functions.* ***J. Am. Statist. Ass.,*** **94,137–145.**

Yi, N., George, V., Allison, D.B., (2003). *Stochastic search variable selection for identifying multiple quantitative trait loci.* ***Genetics*** **164, 1129–1138.**

Yu, K., Jones, M. C., (1998). *Local linear regression quantile estimation.* ***J. Am. Statist. Ass.,*** **93,228–238**.

Yu, K., Lu, Z., Stander, J., (2003). *Quantile regression, applications and current research areas.* ***J. R. Statist. Soc. Ser. D*** **(*The Statistication*) 52** ***Issue*** **3** ***Page,*** **331-350**

Yu, K., Moyeed, R., (2001). *Bayesian quantile regression.* ***Statistics and Probability Letters.*** **54, 437-447.**

Yu, K., Stander, J.,(2007). *Bayesian analysis of a tobit quantile regression model.* ***J. Econometrics*** **137, 260–276.**

Abstract

Quantile regression a complete description of conditional dependence of Y on the explanatory variables. In ordinary regression model the dependence of conditional mean of response variable Y on explanatory variable X is checked. Parameters estimation in quantile regression is based on an asymmetric loss function and calculates in the same way of least square methods. The idea of Bayesian quantile regression employing a likelihood function that is based on the asymmetric laplace distribution. Binary and tobit quantile regression can both be viewed as linear quantile regression with a latent continuous response which is incompletely observed. Normally the time series of error is considered normal. In this thesis the assumption of non gausian of the error and have used it as a semi-parametric exponential function category considered when we set abnormal changes proposed and the simulation data sets describe building and two real series.

Keywords: Quantile Regression, Bayesian Inference, Asymmetric Laplace Distribution, Improper Prior, Proper Posterior, Markov Chain Monte Carlo Methods.

Shahrood University of Technology

Faculty of Mathematical Sciences

M.Sc Thesis

Times series with Non-Gaussian and Applications Innovations

Sahar Rafiey Dehbaneh

Supervisor:

Dr.Ahmad Nezakati Rezazadeh

Advisors:

Dr.Hosein Baghishani

Dr.Mohammad Ali Molaee

June 2013