
بسم الله الرحمن الرحيم

دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد آمار

استنباط بیزی در مدل‌های رگرسیون گشتاوری فضایی

نگارنده: نسیم اندخس

استاد راهنما

دکتر حسین باغیشنی

بهمن ۱۳۹۷

تقدیم به

همه کسانی که در راه رسیدن به هدف‌شان سختی بسیاری کشیده اند
و در این راه ناامید نشده اند...

هرگاه از سختی کاری ترسیدی در برابر آن سر سختی نشان ده
رامت می شود...

«امام علی (ع)»

سپاس گزاری...

خدایا!
تو، چگونه زیستن را به من بیاموز، چگونه مردن را خود خواهم آموخت.

به من توفیق تلاش در شکست، صبر در نومیدی، رفتن بی همراه، جهاد بی سلاح، کار بی پاداش، فداکاری در سکوت، مذهب بی عوام، عظمت بی نام، خدمت بی نان، ایمان بی ریا، خوبی بی نمود، گستاخی بی خامی، قناعت بی غرور، عشق بی هوس، تنهایی در انبوه جمعیت و دوست داشتن بی آنکه دوست بداند، روزی کن...

از استاد با کمالات و شایسته و دلسوز؛ جناب آقای دکتر حسین باغیشنی که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و بدون حضور ایشان اتمام این رساله برایم ناممکن بود، نهایت تشکر و امتنان را دارم و برای آن بزرگوار آرزوی سلامتی و کامیابی از درگاه خداوند تبارک و تعالی می نمایم.

از داوران جناب آقای دکتر داود شاهسونی و جناب آقای دکتر محمدرضا ربیعی که زحمت مطالعه و داوری پایان نامه این جانب را تقبل کردند صمیمانه کمال تشکر را دارم.

همچنین از پدر و مادر عزیزم و برادر مهربانم، که آرامش روحی و آسایش فکری فراهم نمودند تا با حمایت های همه جانبه در محیطی مطلوب، مراتب تحصیلی و نیز پایان نامه درسی را به نحو احسن به اتمام برسانم، سپاسگزاری می نمایم.

نسیم اندخس
بهمن ۱۳۹۷

تعهد نامه

این‌جانب **نسیم اندخس** دانشجوی کارشناسی ارشد رشته **آمار** دانشکده **علوم ریاضی** دانشگاه صنعتی شاهرود، نویسنده پایان‌نامه با عنوان **استنباط بیزی در مدل‌های رگرسیون گشتاوری فضایی** ، تحت راهنمایی دکتر **حسین باغیشنی** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط این‌جانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام ‘‘ دانشگاه صنعتی شاهرود’’ یا “ Shahrood University of Technology “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آن‌ها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

نسیم اندخس
بهمن ۱۳۹۷

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

امروزه مدل‌های رگرسیونی متنوعی فراتر از رگرسیون میانگین مورد توجه قرار گرفته‌اند که هدف آن‌ها بررسی طیف وسیعی از ویژگی‌های توزیع شرطی پاسخ به شرط متغیرهای تبیینی (نه فقط میانگین) است. به‌عنوان نمونه، ممکن است هدف بررسی تاثیر متغیرهای رگرسیونی بر دم توزیع پاسخ باشد. مثلا در حوزه مطالعات سلامت کودکان، پزشکان علاقه‌مند هستند تا عوامل موثر بر سوء تغذیه کودکان را شناسایی کنند؛ یا در مطالعات زلزله‌شناسی، پژوهشگران به‌دنبال شناخت شرایطی هستند که سهمگین‌ترین زلزله رخ می‌دهد. مدل‌های معمول برای رسیدن به این هدف، مدل‌های رگرسیون چندکی هستند. چالش اصلی استنباط آماری در مدل‌های رگرسیون چندکی، در برازش آن‌ها به داده‌های تحت مطالعه است که نیازمند بهینه‌سازی تابع هدف بر اساس برنامه‌ریزی خطی است. با توجه به مشکلات محاسباتی مدل‌های رگرسیونی چندکی، رده مدل‌های رگرسیون گشتاوری جانشین مناسبی برای هدف مورد اشاره هستند که به‌ویژه برای پاسخ‌های فضایی کمتر مورد توجه و استفاده قرار گرفته‌اند. در این پایان‌نامه، مدل رگرسیون گشتاوری را برای پاسخ‌های فضایی معرفی کرده و از یک رهیافت بیزی برای برازش و استنباط مدل استفاده می‌کنیم. همچنین مدل‌بندی توابع غیرخطی متغیرهای تبیینی توسط اسپلاین‌های جریمه‌ای و انتخاب متغیر با اثرات خطی توسط پیشین‌های میخ و تخته مد نظر هستند.

کلمات کلیدی: رگرسیون چندکی، رگرسیون گشتاوری، میدان تصادفی مارکوفی گاوسی، تنظیم‌سازی بیزی، اسپلاین‌های جریمه‌ای، پیشین میخ و تخته.

پیش‌گفتار

امروزه کاربردهای بسیاری وجود دارند که در آن‌ها هدف از به‌کارگیری یک مدل رگرسیونی بررسی تاثیر متغیرهای تبیینی نه فقط بر میانگین بلکه بر سایر ویژگی‌های توزیع پاسخ مانند میانه، یا چندک‌ها است. به‌عنوان نمونه، ممکن است علاقه‌مند باشیم بدانیم درآمد یک خانوار چه تابعی از هزینه ماهیانه برای سرانه خوراک خانوار است؛ و به‌طور ویژه این مساله برای خانوارهای کم‌درآمد (دم پایین) که توان مالی ضعیفی دارند، برای یک ارگان دولتی مربوط به رفاه اجتماعی آن جامعه، مد نظر است. در چنین مساله‌ای هدف اصلی بررسی مقادیر فرین یا همان دم توزیع شرطی پاسخ، به شرط متغیرهای تبیینی، است.

مدل‌های معمول برای رسیدن به این هدف، مدل‌های رگرسیون چندکی هستند. ایده رگرسیون چندکی توسط کوئنکر و باست در سال ۱۹۷۸ مطرح شد. چالش اصلی مدل‌های رگرسیون چندکی برازش آن‌ها است که نیازمند بهینه‌سازی یک تابع هدف مشتق‌ناپذیر است و باید بر اساس برنامه‌ریزی خطی انجام شود. با توجه به این چالش محاسباتی مدل‌های رگرسیون چندکی، مدل‌های رگرسیون گشتاوری توسط نیوی و پاول ۱۹۸۷ معرفی شدند که در سال‌های اخیر مورد توجه قرار گرفته‌اند. در این پایان‌نامه، یک مدل بیزی رگرسیون گشتاوری را در حضور اثرات غیرخطی متغیرهای تبیینی و اثر فضایی معرفی می‌کنیم. به‌طور خلاصه، ساختار این پایان‌نامه به صورت زیر است:

- در فصل اول، تعاریف و مفاهیم اولیه را ارایه می‌کنیم.
- در فصل دوم، مدل‌های رگرسیون گشتاوری و چندکی را معرفی کرده و به مقایسه آن‌ها می‌پردازیم.
- در فصل سوم، مدل بیزی رگرسیون گشتاوری را، با تمرکز بر اثرات فضایی و اثرات غیرخطی متغیرهای تبیینی، معرفی می‌کنیم و نحوه استنباط بیزی مدل را با کمک الگوریتم‌های مونت کارلوی زنجیر مارکوفی بیان می‌کنیم.
- در فصل چهارم، با دو مثال شبیه‌سازی و دو مثال واقعی به بررسی عملکرد مدل پیشنهادی می‌پردازیم.
- در پیوست نیز گزیده‌ای از برنامه‌های نوشته‌شده برای بازتولید نتایج ارایه‌شده در پایان‌نامه که در محیط R نوشته شده‌اند، آمده است.

لیست مقالات مستخرج از پایان‌نامه

۱. اندخس، ن.، باغیشنی، ح. (۱۳۹۷). ’’رگرسیون گشتاوری و چندکی’’، اولین سمینار دانشجویی آمار ایران، دانشگاه تربیت مدرس.

فهرست مطالب

فهرست تصاویر

فهرست جداول

فصل ۱

تعاریف و مفاهیم اولیه

۱.۱ مقدمه

رگرسیون[۱] یکی از معمول‌ترین و پرکاربردترین مدل‌های آماری در بسیاری از زمینه‌های کاربردی است. مدل‌های کلاسیک رگرسیون، معمولا میانگین توزیع شرطی پاسخ را هدف قرار می‌دهند، به‌طوری‌که میانگین شرطی $E(y|x)$ را به‌عنوان تابع رگرسیونی می‌شناسند.

از آن‌جا که میانگین یکی از معیارهای تمرکز است، آگاهی از آن به تنهایی نمی‌تواند اطلاعات کاملی از شکل توزیع به همراه داشته باشد، پس رگرسیون میانگین کلاسیک ممکن است نتواند اطلاعات کافی درباره شکل توزیع تصادفی تحت مطالعه را در سطوح مختلف به‌دست آورد. از طرفی، مسائلی وجود دارند که تمرکز آن‌ها بر سایر ویژگی‌های توزیع قرار دارد، مانند میانه، چندک‌ها و گشتاورها؛ به‌عنوان مثال در حوزه سلامت کودکان، پزشکان علاقه‌مند هستند تا عوامل مؤثر بر سوءتغذیه کودکان را شناسایی کنند؛ یا در مطالعات زلزله شناسی، پژوهشگران به دنبال شناخت شرایطی هستند که سهمگین‌ترین زلزله رخ می‌دهد.

در این موارد تعمیم‌های دیگری از مدل‌های رگرسیونی مانند مدل‌های رگرسیون چندکی[۲] و گشتاوری[۳] معرفی شده‌اند (ریج و همکاران، ۲۰۱۰ و نیب، ۲۰۱۳). در این پایان‌نامه، با

---

[۱] Regression

[۲] Quantile Regression

[۳] Expectile Regression

تمرکز بر مدل‌های رگرسیون گشتاوری به مدل‌بندی بیزی آن‌ها برای تحلیل داده‌های فضایی و غیرفضایی می‌پردازیم. در کنار این هدف اصلی، مدل‌بندی ناپارامتری اثرات رگرسیونی و انتخاب متغیر بیزی که به تنظیم‌سازی نیز معروف است، مد نظر ما قرار گرفته است.

۲.۱ مقدمه ای بر رگرسیون

واژه رگرسیون از نظر لغوی به معنی برگشت است اما در آمار، به‌عنوان «بازگشت به یک مقدار متوسط یا میانگین» استفاده می‌شود. در سال ۱۸۷۷، فرانسیس گالتون، مقاله‌ای درباره بازگشت به میانگین منتشر کرد و در آن بیان کرد که متوسط قد پسران دارای پدران قد بلند، کوتاه‌تر از قد پدران‌شان است. همین‌طور متوسط قد پسران دارای پدران کوتاه قد، بلندتر از قد پدران‌شان است. به این ترتیب، گالتون پدیده بازگشت به میانگین را در داده‌های قد پدران و پسران تأیید کرد.

گالتون، مفهومی زیست شناختی به رگرسیون نسبت داده بود، اما کارل پیرسون تحقیقات گالتون را برای مفاهیم آماری توسعه داد. گالتون، برای تأکید بر پدیده «بازگشت به میانگین» از تحلیل رگرسیون استفاده کرد، اما امروزه تحلیل رگرسیونی و تکنیک‌های آماری مرتبط با آن برای بررسی و مدل‌سازی ارتباط بین متغیرها استفاده می‌شود.

۱.۲.۱ رگرسیون خطی

ساده‌ترین رابطه رگرسیونی بین دو متغیر، رگرسیون خطی است. در این رگرسیون، متغیر پاسخ y، یک ترکیب خطی از پارامترها است. زمانی از رگرسیون متغیر پاسخ y بر متغیر تبیینی x صحبت می‌شود، که بین متغیرها همبستگی وجود داشته باشد. همبستگی بین متغیرهای x و y، بدین معنی است که تغییر در متغیر تبیینی x، چقدر بر متغیر پاسخ y تأثیر می‌گذارد.

در یک تحلیل رگرسیونی، در ابتدا فرض می‌شود بین دو متغیر ارتباطی وجود دارد. در رگرسیون خطی، این ارتباط باید به‌صورت یک خط بین دو متغیر وجود داشته باشد. با رسم نمودار پراکنش، این نوع ارتباط قابل بررسی است. در صورتی که نمودار نشان‌دهنده این باشد که داده‌ها تقریبا در امتداد یک خط مستقیم پراکنده شده‌اند، در این حالت این ارتباط به‌صورت $y = \beta_0 + \beta_1 x$ نشان داده می‌شود که در آن β_0 عرض از مبدأ و β_1 شیب خط است. از آن‌جایی که ممکن است در اندازه‌گیری متغیرها یا شرایط محیطی آزمایش مقداری خطا رخ دهد، معادله اولیه به‌صورت زیر بازنویسی می‌شود:

$$y = \beta_0 + \beta_1 x + \epsilon. \qquad (۱.۱)$$

معادله (۱.۱) یک معادله رگرسیون خطی ساده نامیده می‌شود که ϵ خطای تصادفی است. معمولا فرض می‌شود مجموع خطاها و در نتیجه میانگین آن‌ها، صفر است، مستقل از یکدیگر

هستند و واریانس آن‌ها ثابت است. چنان‌چه $\mathrm{E}(\epsilon) = ۰$ باشد، می‌توان معادله (۱.۱) را به منزله یک رگرسیون میانگین دانست که در آن میانگین شرطی پاسخ به‌صورت یک رابطه خطی مدل می‌شود:

$$\mathrm{E}(y|x) = \beta_۰ + \beta_۱ x.$$

این مدل بیان می‌کند که میانگین‌های y در سطوح مختلف متغیر تبیینی x در امتداد یک خط راست قرار دارند. به‌عبارت دیگر متغیر y در هر سطح از متغیر تبیینی دارای توزیعی است که میانگین‌های این توزیع‌ها روی یک خط راست جای گرفته‌اند.

رده مدل‌های خطی در کنار پرکاربرد بودن آن‌ها، در مسائل متعددی موردنظر نیستند. به‌ویژه زمانی‌که متغیرهای پاسخ ناگاوسی و همبسته هستند. در این موارد، رده مدل‌های آمیخته خطی تعمیم‌یافته[۴] (GLMMs) انتخاب معمول است که در ادامه معرفی می‌کنیم.

۳.۱ مدل‌های خطی تعمیم‌یافته

رده مدل‌های خطی تعمیم‌یافته[۵] (GLMs) اولین بار توسط نلدر و ودربرن (۱۹۷۲) معرفی شدند. این مدل‌ها این امکان را فراهم می‌آورند که رگرسیون معمولی به مواردی که متغیر پاسخ نرمال نیست، توسعه داده شود. در مدل‌های خطی کلاسیک فرض بر این است که جمله خطای متغیرهای تصادفی مستقل و دارای توزیع نرمال هستند؛ در حالی‌که در مدل‌های خطی تعمیم‌یافته با استفاده از یک تابع پیوند[۶] بین میانگین پاسخ‌ها و متغیرهای تبیینی ارتباط خطی برقرار می‌شود. واضح است که مدل‌های خطی تعمیم‌یافته رده گسترده‌ای از مدل‌های آماری را فراهم می‌آورند که به‌عنوان تعمیمی از مدل‌های خطی کلاسیک شمرده می‌شوند. مک‌کولاگ و نلدر (۱۹۸۹)، فارمیر و توتز (۱۹۹۱) و مک‌کولا و سیرل (۲۰۰۱) مبانی مدل‌های خطی تعمیم‌یافته و استنباط در آن‌ها را تبیین کرده‌اند. در مدل‌های خطی تعمیم‌یافته توزیع متغیرهای پاسخ عضوی از خانواده توزیع‌های نمایی با تابع چگالی احتمال به‌صورت

$$f(y|\theta, \phi) = \exp\{[y\theta - b(\theta)]/a(\phi) + c(y, \phi)\}$$

است، که در آن θ پارامتر کانونی، ϕ پارامتر پراکندگی و $b(\cdot)$، $a(\cdot)$ و $c(\cdot)$ توابعی معلوم هستند. همچنین تابع $b(\cdot)$ به‌گونه‌ای است که $\mu = \mathrm{E}(Y) = \dfrac{\partial b(\theta)}{\partial \theta}$ و $\mathrm{Var}(Y) = a(\phi)\dfrac{\partial^۲ b(\theta)}{\partial \theta^۲}$.

هر گاه متغیر پاسخ را با y_i، $i = ۱, ۲, \cdots, n$، نشان دهیم، آن‌گاه با پذیره استقلال داده‌ها، مدل رگرسیونی در یک GLM به‌صورت

$$g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$$

---

[۴] Generalized Linear Mixed Models

[۵] Generalized Linear Model

[۶] Link Function

بیان می‌شود، که در آن $\boldsymbol{x}_i$ها، متغیرهای تبیینی با بعد p، $\boldsymbol{\beta}$ بردار ضرایب رگرسیونی با بعد p و $g(\cdot)$ تابع پیوند را نشان می‌دهند. بنابراین یک مدل خطی تعمیم‌یافته دارای سه مؤلفه است:

۱. مؤلفه تصادفی که به‌وسیله متغیر پاسخ y و توزیع احتمال آن مشخص می‌شود. این توزیع عضوی از خانواده توزیع‌های نمایی است.

۲. مؤلفه روند که شامل متغیرهای تبیینی است و به پیش‌گوی خطی[۷] معروف است.

۳. یک تابع پیوند که رابطه بین ترکیبی خطی از مؤلفه روند و $\mathrm{E}(y)$ را تعیین می‌کند.

اگر تابع پیوند $g(\cdot)$ به‌صورت $\eta_i = g(\mu_i) = \mu_i$ باشد، تابع پیوند را تابع همانی و اگر $g(\cdot)$ به‌صورت $\eta_i = g(\mu_i) = \theta_i$ باشد، آن را یک تابع پیوند کانونی می‌نامند.

۱.۳.۱ روش کمترین توان‌های دوم وزنی تکراری

پیش‌گوی خطی زیر را در نظر بگیرید:

$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} = \sum_{j=۱}^{p} x_{ij}\beta_j \qquad i = ۱, ۲, \cdots, n$$

که در آن η_i پیش‌گوی خطی برای مشاهده iام متغیر پاسخ است. فرض کنید که پیش‌گوی خطی η_i به $\mathrm{E}(y_i) = \mu_i$ با یک تابع پیوند مشتق‌پذیر به‌صورت

$$g(\mu_i) = \eta_i, \qquad i = ۱, ۲, \cdots, n$$

مرتبط باشد. اگر سهم هر مشاهده در تابع درستنمایی را با $\ell_i(\boldsymbol{\beta})$ نشان دهیم، تابع درستنمایی پارامترهای رگرسیونی به‌صورت

$$\ell(\boldsymbol{\beta}) = \sum_{i=۱}^{n} \ell_i(\boldsymbol{\beta}) \tag{۲.۱}$$

است. برای محاسبه برآوردگرهای درستنمایی ماکسیمم $\boldsymbol{\beta}$ باید تابع درستنمایی (۲.۱) را نسبت به $\boldsymbol{\beta}$ ماکسیمم کنیم. مشتق تابع (۲.۱) نسبت به پارامترهای β_j با استفاده از مشتق‌های زنجیری از روابط زیر محاسبه می‌شود:

$$\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=۱}^{n} \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}. \tag{۳.۱}$$

که در آن

$$\frac{\partial \ell}{\partial \theta_i} = y_i + b'(\theta_i) = \frac{y_i - \mu_i}{\phi}$$

[۷] Linear Predictor

و داریم

$$\begin{aligned}
\mu_i &= -b'(\theta_i) \\
\frac{\partial \mu_i}{\partial \theta_i} &= -b''(\theta_i) = \mathrm{Var}(\mu_\mathrm{i}) \\
\frac{\partial \theta_i}{\partial \mu_i} &= \frac{1}{\mathrm{Var}(\mu_\mathrm{i})} \\
\frac{\partial \eta_i}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j}\{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_j x_{ij} + \cdots + \beta_k x_{ik}\} \\
&= x_{ij}
\end{aligned}$$

با توجه به روابط بالا خواهیم داشت (اگرستی، ۲۰۱۳)

$$\frac{\partial \ell_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\phi \mathrm{Var}(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \tag{۴.۱}$$

$$= \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} \tag{۵.۱}$$

و با مساوی قرار دادن مشتق (۴.۱) برای همه مشاهدات برابر صفر، داریم

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0. \tag{۶.۱}$$

معادله امتیاز (۶.۱) به‌عنوان تابعی از β غیرخطی است و برای حل آن باید از روش‌های عددی استفاده کرد.

تابع درستنمایی در مدل‌های خطی تعمیم‌یافته ماتریس کوواریانس مجانبی بردار پارامترها را تعیین می‌کند. ماتریس اطلاع برابر است با

$$\mathrm{E}\left(\frac{\partial^2 \ell_i}{\partial \beta_j \partial \beta_h}\right) = -\mathrm{E}\left[\left(\frac{\partial \ell_i}{\partial \beta_j}\right)^2\right] = -\frac{x_{ij}x_{ih}}{\mathrm{Var}(y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.$$

در نتیجه

$$\mathrm{E}\left(\frac{-\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta_h}\right) = \sum_{i=1}^{n} \frac{x_{ij}x_{ih}}{\mathrm{Var}(y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

و نمایش ماتریسی برای تمام ترکیب‌های (i,h) به‌صورت زیر است:

$$I = X'WX$$

که در آن X ماتریس طرح حاصل از متغیرهای تبیینی است و برای هر $i = 1,2,\cdots,n$ داریم

$$W = \mathrm{diag}(\mathrm{w}_1, \cdots, \mathrm{w}_\mathrm{n}), \qquad \mathrm{w_i} = \frac{(\partial \mu_\mathrm{i}/\partial \eta_\mathrm{i})^2}{\mathrm{Var}(\mathrm{y_i})}.$$

بنابراین ماتریس کوواریانس مجانبی برابر است با

$$\hat{I}^{-1} = (X'\hat{W}X)^{-1}$$

که $\hat{W}$ همان ماتریس W است که با $\hat{\beta}$ مقداردهی شده است.

ماتریس I در برآورد پارامترها به روش درستنمایی ماکسیمم در مدل‌های خطی تعمیم‌یافته به‌کار می‌رود. این برآوردها با استفاده از یک روش تکراری به نام الگوریتم فیشر[۸] به‌دست می‌آیند.

الگوریتم فیشر

الگوریتم فیشر برای برآورد پارامترهای مدل رگرسیون خطی تعمیم‌یافته در معادلات غیرخطی به‌کار می‌رود. به‌دلیل آن‌که معادله‌های برآورد پارامترها غیرخطی هستند، استفاده از روش‌های عددی یک راه‌حل مناسب برای به‌دست آوردن برآورد پارامترها است. روش به‌دست آوردن برآوردها با این الگوریتم به شرح زیر است (اگرستی، ۲۰۱۳):

فرض کنید $\boldsymbol{\beta}^{(t)}$، t ــ امین تقریب از برآورد درستنمایی ماکسیمم $\hat{\boldsymbol{\beta}}$ باشد. آن‌گاه فرمول الگوریتم فیشر عبارت است از:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + (X'WX)^{-1}\boldsymbol{g}^{(t)} \qquad (۷.۱)$$

که با توجه به رابطه (۴.۱) بردار مشتق اول $\boldsymbol{\beta}^{(t)}$ عبارتست از

$$\boldsymbol{g}^{(t)}(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial(\boldsymbol{\beta})}.$$

برای مقادیر آغازین این الگوریتم می‌توان از بردارهای

$$\hat{\boldsymbol{\beta}}^{(0)} = (X'X)^{-1}X'\boldsymbol{y}$$

یا

$$\hat{\boldsymbol{\beta}}^{(0)} = (X'X)^{-1}X'\boldsymbol{g}(\boldsymbol{y})$$

استفاده کرد. با تکرار معادله (۷.۱) برآوردهای مختلفی برای $\boldsymbol{\beta}$ نتیجه می‌شود تا برآوردها همگرا شوند. بعد از همگرایی، برآورد مرحله آخر در واقع تابع درستنمایی را بیشینه می‌کند. الگوریتم فیشر برای برآورد پارامترها در مدل‌های خطی تعمیم‌یافته در نرم‌افزارهای آماری پیاده‌سازی شده‌اند (تابع glm را در R ببینید).

استفاده از الگوریتم فیشر برای یافتن برآوردگرهای درستنمایی ماکسیمم در واقع نوعی استفاده متوالی از روش کمترین توان‌های دوم موزون[۹] است. با توجه به اهمیت این موضوع و استفاده از آن در فصل‌های بعد، این الگوریتم را به اختصار شرح می‌دهیم.

اگر مدل خطی $Y = X\boldsymbol{\beta} + \epsilon$ در نظر گرفته شود که در آن X ماتریس طرح حاصل از مشاهدات برای بردار متغیرهای تبیینی با بعد $n \times p$ است و ماتریس کوواریانس ϵ برابر با V

[۸] Fisher Algorithm

[۹] Weighted Least Squares

باشد، برآوردگر کمترین توان‌های دوم موزون $\boldsymbol{\beta}$ به‌صورت زیر است:

$$\hat{\boldsymbol{\beta}} = (X'V^{-1}X)X'V^{-1}Y.$$

با استفاده از $I = X'WX$ و اعضای قطری w_i معادله زیر حاصل می‌شود:

$$I^{(t)}\boldsymbol{\beta}^{(t)} + \boldsymbol{g}^{(t)} = X'W^{(t)}Y^{(t)}$$

که اعضای $Y^{(t)}$ به شکل زیر هستند:

$$\begin{aligned} Y_i^{(t)} &= \sum_j x_{ij}\beta_j^{(t)} + (y_i - \mu_i^{(t)})\frac{\partial \eta_i^{(t)}}{\partial \mu_i^{(t)}} \\ &= \eta_i^{(t)} + (y_i - \mu_i^{(t)})\frac{\partial \eta_i^{(t)}}{\partial \mu_i^{(t)}} \end{aligned}$$

و الگوریتم فیشر به شکل زیر تبدیل می‌شود:

$$(X'W^{(t)}X)\boldsymbol{\beta}^{(t+1)} = X'W^{(t)}Y^{(t)}.$$

این معادله، معادله‌ای نرمال برای برازش مدل خطی با متغیر پاسخ $Y^{(t)}$، ماتریس X و معکوس ماتریس کوواریانس $W^{(t)}$، با استفاده از کمترین توان‌های دوم موزون است و به روش کمترین توان‌های دوم بازموزون تکراری[۱۰] معروف است زیرا در هر گام دو بار وزن‌دهی انجام می‌گیرد. جواب معادله نیز به شکل زیر است:

$$\boldsymbol{\beta}^{(t+1)} = (X'W^{(t)}X)^{-1}X'W^{(t)}Y^{(t)}.$$

در این رابطه، Y صورت خطی‌شده تابع پیوند g است که در y ارزیابی می‌شود، یعنی

$$g(y_i) \approx g(\mu_i) + (y_i - \mu_i)g'(\mu_i) = \eta_i + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) Y_i.$$

در گام t الگوریتم، متغیر پاسخ اصلاح‌شده Y دارای عضوهایی است که مؤلفه iام آن با $Y_i^{(t)}$ تقریب زده می‌شود. در این گام، $Y^{(t)}$ با وزن $W^{(t)}$ رگرسیون می‌شود تا برآورد جدید $\boldsymbol{\beta}^{(t+1)}$ به‌دست آید. از این برآورد یک پیش‌گوی خطی جدید $\eta^{(t+1)} = \boldsymbol{x}'\boldsymbol{\beta}^{(t+1)}$ و یک متغیر پاسخ اصلاح‌شده $Y^{(t+1)}$ برای مرحله بعدی حاصل می‌شود.

۲.۳.۱ مدل‌های آمیخته خطی تعمیم‌یافته

وقتی داده‌ها مستقل نباشند، از GLMMs استفاده می‌شود که شامل متغیرهای تصادفی پنهان برای بیان اثرات تصادفی است (اسکابنبرگر و گاتوی، ۲۰۰۵). به‌عبارت دیگر، در این مدل‌ها پذیره استقلال به استقلال شرطی تعدیل و همبستگی بین آن‌ها با اضافه کردن اثرات تصادفی

[۱۰] Iteratively Re-weighted Least Squares

از طریق متغیرهای پنهان به مدل وارد می‌شود. فرض کنید $\boldsymbol{y} = (y_1, \cdots, y_n)'$ بردار متغیرهای پاسخ، X ماتریسی $n \times p$ از متغیرهای تبیینی مربوط به اثرات ثابت و $\boldsymbol{u} = (u_1, \cdots, u_q)'$ بردار اثرات تصادفی با بعد q و $Z = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n)'$ ماتریس $n \times q$ از متغیرهای تبیینی متناظر با اثرات تصادفی یا متغیرهای پنهان باشند. در این صورت، فرض می‌شود متغیرهای پاسخ مستقل شرطی و از یک خانواده توزیع‌های نمایی استخراج شده‌اند، به‌طوری‌که بین میانگین شرطی پاسخ و متغیرهای پنهان ارتباط خطی به‌صورت

$$g(\mathrm{E}[\boldsymbol{y}|X, Z]) = X\boldsymbol{\beta} + Z\boldsymbol{u}$$

وجود دارد. متغیرهای پنهان در مدل‌های آمیخته خطی تعمیم‌یافته می‌توانند بیانگر همبستگی فضایی، خوشه‌بندی یا صورت‌های دیگر از وابستگی بین داده‌ها باشد. اغلب فرض می‌شود متغیرهای پنهان دارای توزیع نرمال چندمتغیره با میانگین صفر و ماتریس واریانس Σ_θ هستند که $\boldsymbol{\theta}$ بردار پارامترهای مربوط به همبستگی است.

۴.۱ رهیافت‌های استنباط آماری

مجموعه روش‌های آماری که برای به‌دست آوردن دانش، در مورد ویژگی‌های مورد نظر داده‌ها به‌کار می‌روند را استنباط آماری می‌گویند. کسب اطلاعات، بر اساس داده‌های مشاهده‌شده در مورد مسئله مورد بررسی مهم‌ترین هدف استنباط آماری است. استنباط آماری به‌طور عمومی دو رهیافت اصلی را شامل می‌شود:

- رهیافت کلاسیک[۱۱] ، که به‌طور عمده مبتنی بر روش درستنمایی[۱۲] است.
- رهیافت بیزی[۱۳]

۱.۴.۱ رهیافت کلاسیک استنباط آماری

در آمار کلاسیک، پارامترهای مدل کمیت‌هایی ثابت ولی معمولا نامعلوم فرض می‌شوند. در این رهیافت، احتمال یک پیشامد برابر با فراوانی نسبی وقوع آن پیشامد تعبیر می‌شود. این بدین معناست که احتمال برای پیشامدهای قابل تکرار در نظر گرفته می‌شود که در آن عدم حتمیت به دلیل تصادفی بودن پیشامدها تعریف می‌شود. در صورتی‌که عدم حتمیت وقوع پیشامد به دلیل عدم آگاهی در مورد آن پیشامد باشد، نمی‌توان خواص احتمالی پارامترها را به‌دست آورد. بنابراین اصل تکرارپذیری[۱۴] در این رهیافت یکی از اصول پایه‌ای محسوب

[۱۱] Classical Approach
[۱۲] Likelihood-Based
[۱۳] Bayesian
[۱۴] Repeated Measure Principle

می‌شود. به‌عنوان مثال، در رهیافت کلاسیک، اگر یک فاصله اطمینان ۹۵% برای میانگین نامعلوم جامعه، $[-۱,۱]$ باشد، به‌این معنی نیست که احتمال این‌که میانگین در این فاصله قرار بگیرد، برابر ۹۵٪ است؛ بلکه بدین معنی است که اگر به‌دفعات زیادی نمونه‌گیری کنید و فاصله اطمینان ۹۵% بسازید، انتظار می‌رود در ۹۵% موارد این فاصله‌ها شامل میانگین واقعی باشند.

روش درستنمایی ماکسیمم

یکی از روش‌های معمول برای برآورد پارامترهای نامعلوم مدل‌های پارامتری آماری، روش برآورد درستنمایی ماکسیمم [۱۵] نام دارد که به اختصار به‌عنوان MLE یاد می‌شود. این روش شاید متداول‌ترین روش کلاسیک در بین آماردانان باشد که نخستین بار توسط گاوس (۱۸۲۱) به‌کار گرفته شد و پس از آن به صورت گسترده‌تری در سال ۱۹۲۵ توسط فیشر به‌عنوان جانشینی برای روش گشتاوری مورد استفاده قرار گرفت. برای معرفی این روش، ابتدا تابع درستنمایی را تعریف می‌کنیم.

تعریف ۱.۴.۱. (تابع درستنمایی). برای هر مقدار داده‌شده $X = \boldsymbol{x}$، تابع درستنمایی $\boldsymbol{x}$ را تابع چگالی احتمال توأم (تابع احتمال توأم) $\boldsymbol{x}$، یعنی $f(\boldsymbol{x}|\theta)$، تعریف می‌کنیم که به‌صورت تابعی از پارامتر θ در نظر گرفته می‌شود و آن را با نماد $\mathrm{L}(\theta)$ نمایش می‌دهیم. بنابراین

$$\mathrm{L}(\theta) = f(\boldsymbol{x}|\theta).$$

ذکر چند نکته در ادامه ضروری است:

- تابع درستنمایی $\mathrm{L}(\theta)$ لزوما نسبت به θ مشتق‌پذیر نیست.
- تابع درستنمایی $\mathrm{L}(\theta)$ لزوما یک تابع چگالی احتمال (تابع احتمال) نیست، بلکه متناسب با تابع چگالی احتمال است.
- اگر $X_۱, X_۲, \cdots, X_n$ یک نمونه تصادفی nتایی از توزیعی از خانواده چگالی‌های

$$\{f(x|\theta) : \theta \in \Theta\}$$

باشند، آن‌گاه

$$\mathrm{L}(\theta) = \prod_{i=۱}^{n} f(x_i|\theta)$$

تعریف ۲.۴.۱. (برآوردگر درستنمایی ماکسیمم). اگر $\delta(x)$ برآوردگری برای θ باشد، به‌طوری‌که

[۱۵] Maximum Liklihood Estimation

الف)

$$\mathrm{P}_\theta(\delta(X) \in \Theta) = ۱ \qquad \forall \theta \subseteq \mathbb{R}^p$$

ب)

$$\mathrm{L}(\delta(X) \geq \mathrm{L}(\theta)) \qquad \forall \theta \subseteq \mathbb{R}^p$$

آن‌گاه $\delta(X)$ به‌عنوان برآوردگر درستنمایی ماکسیمم θ تعریف می‌شود. معمولا برآوردگر درستنمایی ماکسیمم θ را با $\hat{\theta}$ نمایش می‌دهند و لزوما منحصر به‌فرد نیست. همچنین بر اساس تعریف $\hat{\theta}$ (پارسیان، ۱۳۹۱)

$$\mathrm{L}(\hat{\theta}) = \sup_{\theta \in \Theta} \mathrm{L}(\theta).$$

۲.۴.۱ رهیافت بیزی استنباط آماری

یک رهیافت کاربردی و منعطف در استنباط آماری، رهیافت بیزی است. در رهیافت بیزی پارامتر واقعی و نامعلوم θ، به‌عنوان تحققی از یک توزیع احتمالی به نام توزیع پیشین در نظر گرفته می‌شود. این در حالی است که آماردان‌های کلاسیک بر این باور هستند که پارامتر نامعلوم را به دشواری می‌توان به‌صورت یک متغیر تصادفی به حساب آورد. همچنین تعیین توزیع احتمالی برای چنین متغیری نیز به نظر و عقیده آماردان‌ها بستگی دارد. آماردان‌های بیزی بر این باور هستند که با ترکیب اطلاعات شخصی توسط توزیع پیشین و اطلاعات نهفته در داده‌ها توسط تابع درستنمایی، می‌توان به واقعیت نزدیک شد. اگر در استنباط بیزی، میزان اطلاعات حاصل از داده‌ها افزایش یابد، یعنی به‌طور مثال اگر حجم نمونه زیاد شود، اثر توزیع پیشین ناچیز می‌شود. پس انتظار بر این است که برآوردگرهای بیزی و کلاسیک، عملکرد یکسانی داشته باشند. اما اگر نمونه کوچک باشد، ممکن است اختلاف چشمگیری حاصل شود. در این صورت استنباط بیزی می‌تواند بسیار قوی‌تر از استنباط کلاسیک ظاهر شود. به‌علاوه در این دیدگاه می‌توان با استفاده از داده‌ها، نسخه به‌روزتری از توزیع پیشین[۱۶] ساخت و توزیع احتمال جدید را که توزیع پسین[۱۷] نام دارد، به‌دست آورد.

پایه دیدگاه بیزی استنباط آماری به قضیه بیز برمی‌گردد که اولین بار توماس بیز در سال ۱۷۶۳ مطرح شد.

قضیه ۱.۴.۱. (قضیه بیز). فرض کنید $A_۱, A_۲, \cdots, A_n$ یک افراز از فضای نمونه و E یک پیشامد دلخواه باشد. در این صورت برای هر $i = ۱, \cdots, n$ می‌توان نوشت

$$\mathrm{P}(A_i|E) = \frac{\mathrm{P}(E|A_i)\mathrm{P}(A_i)}{\sum_{i=۱}^{n} \mathrm{P}(E|A_i)\mathrm{P}(A_i)}.$$

[۱۶] Prior Distribution

[۱۷] Posterior Distribution

این قضیه یک روش برای به‌روز رساندن شانس A_i از $\mathrm{P}(A_i)$ به $\mathrm{P}(A_i|E)$، وقتی E مشاهده‌شده است، می‌باشد. $\mathrm{P}(A_i)$ را احتمال پیشین و $\mathrm{P}(A_i|E)$ را احتمال پسین می‌نامند (جفریز، ۱۹۳۶).

این قضیه برای تابع چگالی به‌صورت زیر بازنویسی می‌شود:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_\Theta f(x|\theta)\pi(\theta)d\theta}$$

که در آن $\pi(\theta)$ تابع چگالی احتمال توزیع پیشین و $\pi(\theta|x)$ تابع چگالی احتمال توزیع پسین θ است. از آن‌جا که مقدار $\int_\Theta f(x|\theta)\pi(\theta)d\theta$ مستقل از θ بوده و حاصل آن یک عدد ثابت است، در محاسبه $\pi(\theta|x)$ معمولا از محاسبه انتگرال خودداری کرده و توزیع پسین را به‌صورت زیر بیان می‌کنند:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta).$$

توزیع پیشین

توزیع پیشین θ را با $\pi(\theta)$ نشان می‌دهند که بیان‌کننده اطلاعات پیشین آماردانان درباره مقدار نامعلوم θ است. به بیان دیگر، قبل از مشاهده و گردآوری هر گونه داده، اطلاعات و دانسته‌های قبلی آماردان، وی را متقاعد می‌کند بر این باور باشد که شانس قرار گرفتن مقادیر θ در فضای Θ چگونه است. فرض می‌شود چنین باورهایی می‌توانند در قالب یک توزیع بیان شوند. بنابراین، مجموعه همه این اطلاعات در قالب یک توزیع که همان توزیع پیشین است، خلاصه می‌شوند.

تعریف ۳.۴.۱. (توزیع پیشین سره[۱۸]). توزیع با تابع چگالی احتمال $\pi(\theta)$ را سره گویند، هرگاه

$$\int_\Theta \pi(\theta)d(\theta) < \infty.$$

تعریف ۴.۴.۱. (توزیع پیشین ناسره[۱۹]). توزیع پیشین با تابع چگالی $\pi(\theta)$ را ناسره گویند، هرگاه

$$\int_\Theta \pi(\theta)d\theta = +\infty.$$

اگر توزیع پیشین سره باشد، توزیع پسین حتما سره خواهد بود؛ اما اگر ناسره باشد، ممکن است توزیع پسین سره شود. توجه داشته باشید زمانی می‌توان از توزیع پیشین ناسره برای θ استفاده کرد که توزیع پسین به‌دست آمده سره باشد. در غیر این صورت استنباط آماری غیرممکن خواهد بود. پس اگر در مدل بیزی از یک توزیع پیشین ناسره استفاده شود، بررسی سره بودن پسین به یک مسأله کلیدی و مهم تبدیل می‌شود.

تعریف ۵.۴.۱. (توزیع پیشین مزدوج[۲۰]). توزیع پیشین با تابع چگالی احتمال $\pi(\theta)$ را برای مدل با تابع چگالی احتمال $f(x|\theta)$ مزدوج گویند، هرگاه توزیع پسین آن متعلق به خانواده توزیع پیشین $\pi(\theta)$ باشد (رابرت، ۲۰۰۷).

---

[۱۸] proper

[۱۹] Improper

[۲۰] Conjugate Prior

تعریف ۶.۴.۱. (توزیع پیشین ناآگاهی‌بخش[۲۱]). یکی از مهم‌ترین توزیع‌های پیشین، پیشین ناآگاهی‌بخش است و در مواردی که اطلاع چندانی در مورد پارامتر موجود نباشد، مورد استفاده قرار می‌گیرد. در این حالت آماردان سعی می‌کند احتمال یکسانی را به همه پیشامدهای ممکن تخصیص دهد. پیشین ناآگاهی‌بخش به‌صورت $\pi(\theta) \propto K$، ($K > ۰$) تعریف می‌شود که انتخاب K اختیاری است. با انتخاب پیشین ناآگاهی‌بخش توزیع پسین به صورت

$$\pi(\theta|x) \propto Kf(x|\theta) = KL(\theta) \qquad (۸.۱)$$

بیان می‌شود. اگر متغیر مورد بررسی روی دامنه فضای پارامتر نامحدود باشد، در آن صورت تخصیص یک مقدار ثابت یکنواخت به‌عنوان توزیع پیشین، توزیع پیشین ناسره ایجاد می‌کند. نحوه انتخاب K در رابطه (۸.۱) به روش‌های مختلفی صورت می‌پذیرد (برگر و برنارد، ۱۹۹۲). به‌عنوان مثال، می‌توان به پیشین ناآگاهی‌بخش یکنواخت [۲۲] و پیشین ناآگاهی‌بخش جفریز[۲۳] اشاره کرد.

پیشین ناآگاهی‌بخش یکنواخت: برای نخستین بار لاپلاس (۱۸۱۲) با انتخاب $K = ۱$ پیشین ناآگاهی‌بخش را به‌صورت $\pi(\theta) \propto ۱$ به‌کار برد که به پیشین ناآگاهی‌بخش یکنواخت معروف است.

پیشین ناآگاهی‌بخش جفریز: جفریز (۱۹۴۶) توزیع پیشین ناآگاهی‌بخش را که بر اساس ماتریس اطلاع فیشر[۲۴] پایه‌ریزی شده بود، ارائه داد. اگر

$$I(\theta) = (I_{ij}(\theta)) \qquad i, j = ۱, \cdots p$$

ماتریس اطلاع فیشر باشد، که در آن

$$I_{ij}(\theta) = -\mathrm{E}_\theta \left[\frac{\partial^۲}{\partial\theta_i \partial\theta_j} \log f(x|\theta) \right]$$

آن‌گاه توزیع پیشین جفریز به صورت

$$\pi(\theta) \propto \sqrt{detI(\theta)}$$

تعریف می‌شود، که منظور از det دترمینان ماتریس است.

توزیع پسین

در ساده‌ترین مسائل آماری، محاسبه توزیع پسین مستلزم محاسبه انتگرال‌های چندگانه است؛ اما انجام بسیاری از انتگرال‌های چندگانه مشکل است و روش‌های متعارف برای حل آن‌ها وجود

[۲۱] Noninformative Prior
[۲۲] Noninformative Uniform Prior
[۲۳] Noninformative Jeffreys Prior
[۲۴] Fisher Information

ندارد و تنها راه، حل تقریبی آن‌هاست. یک رده کلی از روش‌های تقریب توزیع‌های پسین، روش‌های مبتنی بر نمونه‌گیری[۲۵] مانند روش‌های رد و پذیرش[۲۶] (رابرت و کسلا، ۲۰۱۰)، نمونه‌گیری نقاط مهم[۲۷] و نمونه‌گیری مونت کارلوی زنجیر مارکوفی[۲۸](MCMC) هستند. با استفاده از این روش‌ها، برای انجام استنباط‌های بیزی بر اساس توزیع پسین، دیگر نیازی به بهینه‌سازی توابع پیچیده و محاسبه گشتاورهای توزیع‌های پیچیده (انتگرال‌گیری‌های پیچیده) نداریم. در ادامه برخی از این روش‌ها را معرفی می‌کنیم.

روش‌های تقریب توزیع پسین مبتنی بر نمونه‌گیری

محاسبه توابع چگالی نامعلوم در بسیاری از کاربردها مانند مدل‌های بیزی پیچیده، تنها به کمک شبیه‌سازی ممکن بوده و در این میان روش‌های مونت کارلویی بیشترین سهم را دارند. روش‌های مونت کارلویی متکی بر نمونه‌گیری مکرر به منظور دستیابی به نتایج محاسباتی هستند. به‌طور کلی، روش‌هایی که در آن از اعداد تصادفی برای شبیه‌سازی استفاده می‌شود، روش‌های شبیه‌سازی مونت کارلویی نامیده می‌شوند. کلمه مونت کارلو در حقیقت برگرفته از مراکز تفریحی مشهور در بندر مونت کارلو در کشور فرانسه است، که در آن‌ها برای تصادفی کردن بازی‌ها از اعداد تصادفی استفاده می‌شد (رابرت و کسلا، ۲۰۰۴). استفاده گسترده از این روش از سال ۱۹۴۹ توسط فیزیکدانی به نام متروپلیس شروع شد که در مقاله‌ای مشهور این روش را معرفی کرد.

در استنباط بیزی به چند دلیل استفاده از روش مونت کارلویی بر روش‌های عددی ترجیح داده می‌شود. اول این‌که روش مونت کارلویی مشکل روش‌های عددی در حل انتگرال‌ها را ندارد؛ به‌عبارت دیگر روش‌های مونت کارلویی به نواحی چگال‌تر تابع زیر انتگرال، وزن بیشتری در محاسبه تقریب می‌دهد در حالی‌که در روش‌های عددی همه زیر نواحی به یک شکل وزن داده می‌شوند. دوم، بعد فضای پارامتر در اغلب مسائل بیزی بسیار بالا است که روش‌های عددی پیچیده‌ای برای به‌دست آوردن یک تقریب قابل قبول مورد نیاز است. سوم، دقت روش‌های عددی معمولا با افزایش بعد کاهش می‌یابد، در حالی‌که روش‌های مونت کارلویی از نظر دقت، مستقل از بعد هستند (رابرت و کسلا، ۲۰۰۴ و ۲۰۱۰). ایده اصلی روش‌های مونت کارلویی، تبدیل انتگرال به یک امید ریاضی بر اساس یک تابع چگالی احتمال مشخص، تولید نمونه از این تابع چگالی و استفاده از قانون قوی اعداد بزرگ[۲۹] (گات، ۲۰۰۵) برای تقریب این امید ریاضی است. در واقع مسأله اصلی محاسبه انتگرال زیر است:

$$J = \mathrm{E}_f[h(X)] = \int_{\chi} h(x)f(x)dx$$

[۲۵] Sampling Base Methods
[۲۶] Accept-Reject Methods
[۲۷] Importence Sampling
[۲۸] Markov Chain Monte Carlo Algorithm
[۲۹] Strong Low of Large Number

که در آن χ تکیه‌گاه متغیر تصادفی X، یک یا چند بعدی است. تابع $f(\cdot)$ یک تابع چگالی، دارای شکل بسته یا به‌طور جزئی بسته[۳۰] و $h(\cdot)$ نیز یک تابع حقیقی مقدار است. الگوریتم روش مونت کارلویی را می‌توان به صورت زیر بیان کرد:

- یک نمونه تصادفی mتایی از تابع چگالی $f(\cdot)$ تولید کنید.
- با جایگذاری این مقدار در تابع $h(\cdot)$، مقدار $\frac{1}{m}\sum_{j=1}^{m} h(x_j)$ را محاسبه کنید.

بنابر قانون قوی اعداد بزرگ، مقدار فوق یک برآورد امید ریاضی است، یعنی

$$\bar{h}_m = \frac{1}{m}\sum_{j=1}^{m} h(x_j) \quad \xrightarrow{a.s.} \quad \mathrm{E}_f[h(x)].$$

علاوه بر این، دقت این تقریب مانند دقت یک برآورد آماری است. بنابراین، به محض این‌که نمونه $x_1, \cdots, x_m$ از توزیع $f(\cdot)$ تولید شود، از آن می‌توان برای محاسبه کمیت‌های مورد علاقه توزیع با تابع چگالی احتمال $f(x)$ استفاده کرد.

مسأله اصلی در روش مونت کارلویی، تولید نمونه از یک توزیع دلخواه است که در اغلب موارد تولید مستقیم نمونه امکان‌پذیر نیست و باید از روش‌های غیرمستقیم تولید نمونه استفاده کنیم (چن و همکاران، ۲۰۰۰). الگوریتم متروپلیس ـ هستینگز[۳۱] (MH) و نمونه‌گیری گیبز[۳۲] از جمله روش‌های تولید نمونه غیرمستقیم هستند که در ادامه به معرفی آن‌ها می‌پردازیم.

۵.۱ الگوریتم‌های مونت کارلوی زنجیر مارکوفی

در مواردی که صورت بسته توزیع پسین در دسترس نیست و بعد پارامترها بالا است، برای محاسبه تقریبی آن، از روش‌های تقریبی استفاده می‌شود. یکی از روش‌های تقریب توزیع پسین استفاده از روش‌های شبیه‌سازی مونت کارلویی است. برای رفع این مشکل الگوریتم‌های MCMC (متروپلیس و همکاران، ۱۹۴۹؛ هستینگز، ۱۹۷۰) معرفی شدند که به‌طور غیرمستقیم به تولید نمونه از یک توزیع احتمالی مفروض می‌پردازند. مکانیسم الگوریتم‌های MCMC مبتنی بر تولید نمونه وابسته از یک زنجیر مارکوف است که توزیع مانای آن، همان توزیع پسین مورد نظر است. در این صورت، اگر حجم نمونه مونت کارلویی تولیدشده به اندازه کافی بزرگ باشد، از جایی به بعد، حالت‌های زنجیر حاصل به‌عنوان نمونه‌هایی از توزیع مطلوب، یعنی توزیع پسین، در نظر گرفته می‌شوند.

[۳۰] Partially Closed Form
[۳۱] Metropolis-Hastings Algorithm
[۳۲] Gibbs Sampling

۱.۵.۱ الگوریتم متروپلیس‌ـ هستینگز

الگوریتم متروپلیس‌ـ هستینگز ابتدا توسط متروپلیس و همکاران (۱۹۳۵) پیشنهاد شد و سپس توسط هستینگز (۱۹۷۰) به‌طور کامل تدوین یافت. این الگوریتم ساختاری مشابه با ساختار رد و پذیرش دارد و مشاهداتی را که توسط یک توزیع پیشنهادی به‌عنوان نمونه‌های ممکن از توزیع پسین پیشنهاد می‌شوند، براساس یک قاعده احتمالی مورد ارزیابی قرار داده و با احتمال معینی می‌پذیرد. بر این اساس برای استفاده از این الگوریتم، به یک توزیع پیشنهادی نیاز است که انتخاب آن بر دقت و سرعت همگرایی موثر است. در این الگوریتم فضای حالت زنجیر $X \in R^q$ که $q \geq ۱$ و $\bar{w}$ که یک اندازه احتمال است، را در نظر بگیرید و فرض کنید $X^{(k)} = x$ نشان‌دهنده حالت جاری زنجیر، و $\bar{w}$ دارای تابع مانای $\pi(\cdot)$ باشد و همچنین $g(\cdot|\cdot)$ نشان‌دهنده چگالی پیشنهادی باشد. الگوریتم MH برای به‌روزرسانی $X^{(k+۱)}$ به‌صورت زیر به‌دست می‌آید:

۱) تولید y از تابع چگالی پیشنهادی $g(y|x)$

۲) محاسبه احتمال پذیرش $\alpha(x,y)$ که در آن

$$\alpha(x,y) = \min\left\{۱, \frac{\pi(y)g(x|y)}{\pi(x)g(y|x)}\right\}$$

۳) قرار دادن $X^{(k+۱)} = y$ با احتمال $\alpha(x,y)$ و قرار دادن $X^{(k+۱)} = x$ با احتمال $۱ - \alpha(x,y)$.

ویژگی‌های مطلوب این الگوریتم آن است که شناخت توزیع پسین به‌صورت مناسب کفایت می‌کند. این الگوریتم با شرط داشتن توزیع پسین، مجموعه‌ای از متغیرهای تصادفی $\left\{X^{(k+۱)}\right\}$، برای $k = ۱, ۲, \cdots$، به‌طور پی‌درپی تولید می‌کند. برای تولید این نمونه‌ها، این الگوریتم، ما را ملزم به تعیین چگالی پیشنهادی $g(y|x)$ می‌کند. این تابع، تابع چگالی x، احتمال در زمان $k+۱$، با توجه به مقدار آن در زمان k است.

نسبت موجود در رابطه فوق هستینگز نامیده می‌شود. اگر توزیع پیشنهادی مقادیری نزدیک به x را پیشنهاد کند، تقریبا همیشه پیشنهادها پذیرفته می‌شوند، اما زنجیر خیلی کند حرکت می‌کند و زمان زیادی طول می‌کشد تا الگوریتم همگرا شود. همچنین اگر توزیع پیشنهادی مقادیر دور از x را پیشنهاد کند، پیشنهادها تقریبا همیشه رد می‌شوند و باز هم زنجیر به کندی همگرا می‌شود. با کمی تنظیم می‌توان توزیع پیشنهادی را به نحوی تبدیل کرد که پیشنهادها مکررا رد یا قبول نشوند و الگوریتم از نرخ پذیرش مناسبی برخوردار باشد (رابرتز و همکاران، ۱۹۹۷).

۲.۵.۱ نمونه‌گیر گیبز

روش نمونه‌گیری گیبز ابتدا توسط گمان و گمان (۱۹۸۴) در پردازش مدل‌های تصویری معرفی شد. سپس توسط گلفاند و اسمیت (۱۹۹۰) به‌عنوان روشی بر مبنای نمونه‌گیری برای محاسبه چگالی‌های کناری در چارچوب استنباط بیزی پیشنهاد شد.

الگوریتم گیبز از توزیع پسین شرطی پارامترهای مختلف، به‌شرط سایر پارامترها که اصطلاحا توزیع پسین شرطی کامل[۳۳] نامیده می‌شوند، برای نمونه‌گیری از توزیع پسین هدف استفاده می‌کند. برای معرفی الگوریتم حالت ساده دو مرحله‌ای را در نظر می‌گیریم.

نمونه‌گیری دو مرحله‌ای یک زنجیر مارکوف از توزیع توام دو متغیره به روش زیر به‌دست می‌آید. اگر دو متغیر تصادفی x و y دارای تابع چگالی احتمال توام $f(x,y)$ با چگالی‌های شرطی $f_{y|x}$ و $f_{x|y}$ باشند، نمونه‌گیر گیبز دو مرحله‌ای یک زنجیر مارکوف $(x^{(t)},y^{(t)})$ طبق مراحل زیر تولید می‌کند:

الگوریتم ۱. (نمونه‌گیری گیبز دومرحله‌ای). مقدار مفروض $x^{(۰)}$ را در نظر بگیر. برای $t=۱,۲,\ldots$ نمونه‌های $(x^{(t)},y^{(t)})$ را با اجرای دو مرحله زیر تولید کن:

۱. $y^{(t)} \sim f_{y|x}(\cdot|x^{(t-۱)})$

۲. $x^{(t)} \sim f_{x|y}(\cdot|y^{(t)})$

الگوریتم ۱ در صورتی‌که چگالی‌های شرطی صورت بسته‌ای داشته باشند و تولید نمونه از آن‌ها ساده باشد، به سادگی اجرا می‌شود. اگر $(x^{(t)},y^{(t)})$ از توزیع f باشند، آن‌گاه $(x^{(t+۱)},y^{(t+۱)})$ نیز از توزیع f خواهند بود، به‌عبارت دیگر توزیع مانای این زنجیر، f است. در صورتی‌که بعد متغیرها بیشتر از دو باشد، الگوریتم نمونه‌گیر گیبز چندمرحله‌ای به‌طور مشابه تعمیم داده می‌شود.

۳.۵.۱ الگوریتم متروپلیس هستینگز درون گیبز

نمونه‌گیر گیبز را برای مواقعی که نتوان به سادگی از چگالی‌های شرطی کامل شبیه‌سازی کرد، می‌توان به آسانی تعمیم داد. به‌عنوان مثال، فرض کنید برای $i=۱,۲,\cdots,m$ شبیه‌سازی از

$$f_i(x_i|x_۱^{(t+۱)},\cdots,x_{i-۱}^{(t+۱)},x_{i+۱}^{(t)},\cdots,x_m^{(t)})$$

در مرحله $t+۱$ مشکل باشد. در این حالت می‌توان برای تولید نمونه از آن (در الگوریتم نمونه‌گیری گیبز) از الگوریتم MH استفاده کرد. در نتیجه این گام با یک الگوریتم MH که توزیع مانای آن $f_i(x_i|x_۱^{(t+۱)},\cdots,x_{i-۱}^{(t+۱)},x_{i+۱}^{(t)},\cdots,x_m^{(t)})$ است، جایگزین می‌شود. این نوع الگوریتم از ترکیب دو الگوریتم گیبز و MH تشکیل می‌شود که به آن الگوریتم متروپلیس‌ـ‌هستینگز تحت گیبز[۳۴] می‌گویند. در صورتی‌که شبیه‌سازی بیشتر از یک چگالی شرطی کامل دشوار باشد، می‌توان به همین ترتیب مراحل مختلفی از نمونه‌گیر گیبز را با الگوریتم MH ترکیب کرد (گیلکز و همکاران، ۱۹۹۶).

---

[۳۳] Full Conditional Postrior Distribution

[۳۴] Metropolis-within-Gibbs Algorithm

۶.۱ اسپلاین

در آمار هموار کردن[۳۵] یک نمودار پراکنش، استفاده از الگوهای مهم داده‌ها در ایجاد یک تابع تقریبی است. یک هموارکننده[۳۶] تابعی برای خلاصه‌سازی متغیر پاسخ برحسب تابعی از متغیرهای تبیینی $X_۱, X_۲, \cdots, X_p$ است. یکی از انواع هموارکننده‌های نمودارهای پراکنش، اسپلاین‌ها[۳۷] هستند.

مکانیسم عملکرد روش برازش مبتنی بر اسپلاین‌ها بر اساس تقسیم کل بازه‌ای که مشاهدات در آن قرار دارند به تعدادی زیربازه است. سپس در هر زیربازه یک منحنی از درجه $p \geq ۱$ به مشاهدات برازش داده می‌شود که منحنی‌ها در نقاط مرزی بین زیربازه‌ها (گره‌ها[۳۸]) به هم متصل می‌شوند. فرض کنید تعدادی مشاهده داشته باشیم که در بازه

$$[t_۰ = \min(x_i), t_{m+۱} = \max(x_i)]$$

قرار دارند. اگر طول بازه را به فواصل $[t_۰, t_۱], [t_۱, t_۲], \cdots, [t_m, t_{m+۱}]$ تقسیم کنیم، آن‌گاه چنان‌چه منحنی برازش‌شده f، در هر بازه به‌صورت یک چندجمله‌ای از مرتبه p باشد، یک تابع اسپلاین از مرتبه p با شرایط زیر خواهیم داشت:

- تابع f در هر گره t_j، $j = ۱, ۲, \cdots, m$، پیوسته است.
- مشتق‌های تابع f تا مرتبه $p - ۱$ وجود دارند و در گره‌ها پیوسته هستند.

با این شرایط، ضابطه کلی توابع اسپلاین به صورت زیر است:

$$f(x) = \sum_{k=۰}^{p} \beta_k x^k + \sum_{j=۱}^{m} \beta_{p+j}(x - t_j)_+^p$$

$$(x - t_j)^p = \begin{cases} x - t_j & x > t_j \\ ۰ & x < tj \end{cases}$$

که در آن $\boldsymbol{\beta} = (\beta_۰, \cdots, \beta_{p+m})$ بردار ضرایب است.

توابع پایه

در جبر خطی منظور از یک مجموعه پایه[۳۹] در یک فضای برداری، مجموعه‌ای از بردارهای موجود در آن فضا است به‌طوری‌که مستقل خطی باشند و هر بردار دیگر در فضای برداری را

[۳۵] Smoothing
[۳۶] Smoother
[۳۷] Splines
[۳۸] Nodes
[۳۹] Bases Set

بتوان از ترکیب خطی آن بردارها به‌دست آورد. برای مثال چندجمله‌ای درجه دو با ضرایب حقیقی را می‌توان به‌صورت

$$y = ۱a + bt + ct^۲$$

نوشت که در واقع از ترکیب خطی توابع پایه چندجمله‌ای $\{۱, t, t^۲, \cdots\}$ با ضرایب a، b و c برای سه مؤلفه اول و صفر برای سایر مؤلفه‌ها حاصل می‌شود.

۱.۶.۱ B ـ اسپلاین‌ها

توابع پایه B ـ اسپلاین، چندجمله‌ای‌هایی از مرتبه $k = p - ۱$ هستند که در گره‌ها پیوستگی دارند و هر کدام از این توابع پایه، نواحی کوچکی از کل مشاهدات را پوشش می‌دهند. مکانیزم کار در این نوع اسپلاین، مشابه اسپلاین‌های معمولی است، تنها تفاوت در نحوه ساختن توابع پایه است. اگر $t = (t_{(k-۱)}, \cdots, t_{(m+k)})$ یک دنباله نانزولی از گره‌ها باشد، توابع پایه B ـ اسپلاین با

$$B_j^{k-۱}, \quad j = -(k-۱), \cdots, m$$

نمادگذاری می‌شوند. تعداد توابع پایه B ـ اسپلاینی $p + m + ۱$ است.

تابع پایه B ـ اسپلاینی از مرتبه صفر $(k = ۱)$ به‌صورت زیر تعریف می‌شود:

$$B_j^{\circ} = \begin{cases} ۱ & t_j \le x \le t_{j+۱} \\ ۰ & \text{در غیر این صورت} \end{cases}$$

دی بور (۱۹۷۸) برای ساخت توابع پایه هر مرتبه از توابع پایه مرتبه‌های پایین‌تر استفاده کرد. بر اساس یک رابطه بازگشتی، می‌توان B ـ اسپلاین‌ها را از هر درجه دلخواه به‌صورت زیر محاسبه کرد:

$$B_j^p(x) = B_j^{k-۱}(x) = \frac{x - t_j}{t_{j+k-۱} - t_j} B_j^{k-۲}(x) + \frac{t_{j+k} - x}{t_{j+k} - t_{j+۱}} B_{j+۱}^{k-۲}, j = -(k-۱), \cdots, m. \tag{۹.۱}$$

با توجه به معادله (۹.۱)، می‌توان B ـ اسپلاین‌های مرتبه یک $(k = ۲)$ را به‌دست آورد که بر اساس توابع پایه درجات پایین‌تر تعریف می‌شوند:

$$\begin{aligned} B_j^{۱}(x) &= \frac{x - t_j}{t_{j+۱} - t_j} B_j^{\circ}(x) + \frac{t_{j+۲} - x}{t_{j+۲} - t_{j+۱}} B_{j+۱}^{\circ}(x) \qquad (۱۰.۱) \\ &= \begin{cases} \dfrac{x - t_j}{t_{j+۱} - t_j} & t_j \le x < t_{j+۱} \\ \dfrac{t_{j+۲} - x}{t_{j+۲} - t_{j+۱}} & t_{j+۱} \le x < t_{j+۲} \end{cases} \end{aligned}$$

با جایگذاری گره‌ها در رابطه (۱۰.۱) تعداد معادلات درجه یک به‌دست می‌آیند.

توابع پایه مرتبه‌های دو و سه بین محققان محبوب‌تر واقع شده‌اند. دلیل آن هم ویژگی‌ها و انعطاف‌پذیری خوبی است که این توابع دارند. برای اطلاعات بیشتر در زمینه B ـ اسپلاین‌ها به دایرکس (۱۹۹۳) و دی‌بور (۱۹۷۷) مراجعه کنید.

برخی ویژگی‌های مهم B ـ اسپلاین مرتبه $p = k - ۱$ به شرح زیر هستند:

- هر تابع پایه با درجه p شامل $p + ۱$ تکه از درجه p است.
- تکه‌های چندجمله‌ای مجاور هم در گره‌ها پیوسته هستند.
- در گره‌ها از مرتبه ۱ تا مرتبه $p - ۱$ پیوسته و مشتق‌پذیر است.
- توابع پایه B ـ اسپلاین هم‌پوشانی دارند.
- توابع پایه روی تکیه‌گاه‌شان مثبت هستند، یعنی

$$B_j^{k-۱}(x) > ۰ \quad x \in [t_j, t_{j+m}].$$

- جمع مقادیر این توابع برابر یک است. یعنی

$$\sum_{j=-(k-۱)}^{m} B_j^{k-۱}(x) = ۱.$$

۲.۶.۱ اسپلاین‌های جریمه‌ای (P ـ اسپلاین‌ها)

فرض کنید تعداد n زوج داده داشته باشیم و بخواهیم روند برازش این مشاهدات را با استفاده از B ـ اسپلاین ها برآورد کنیم. تابع B ـ اسپلاینی بهتر است که مقدار کمیت زیر را کمینه کند:

$$S = \sum_{i=۱}^{n} \left\{ y_i - \sum_{j=-(k-۱)}^{m} \alpha_j B_j^k(x) \right\}^۲ .$$

اگر تعداد گره‌ها را نسبتا زیاد در نظر بگیریم، منحنی برازش‌شده بیشترین تغییرات تعدیل‌شده توسط داده‌ها را نشان خواهد داد، به عبارتی از تعداد مشاهدات بیشتری عبور خواهد کرد و در نتیجه مقدار مجموع توان‌های دوم خطا کمتر خواهد شد. بنابراین تابعی که کمینه می‌شود با تعداد رابطه‌ها رابطه معکوس دارد. اما اگر تعدا گره‌ها افزایش یابد، منجر به زیر بازه‌های بیشتری می‌شود و در نتیجه توابع پایه بیشتری خواهیم داشت و پیرو آن باید پارامترهای بیشتری را برآورد کنیم. از آن‌جایی‌که تعداد گره‌ها نامعلوم است، معمولا تعداد گره‌ها را برابر با تعداد مشاهدات در نظر می‌گیرند که در این صورت ممکن است با مسأله بیش‌برازش[۴۰] مواجه شویم. به عبارتی نوسانات منحنی زیاد می‌شود و منحنی حالت هموار بودن خود را از دست می‌دهد. بنابراین باید بین برازش مناسب و کاهش پیچیدگی مدل (برحسب تعداد

[۴۰] Over Prediction

گره‌های کافی) یک تعادل ایجاد شود. روپرت (۲۰۰۲) برای انتخاب تعداد و موقعیت گره‌ها روشی معرفی کرد که در آن تعداد گره‌ها برابر است با $min(\frac{N}{۴},۳۵)$. برای کنترل و جلوگیری از بیش برازش، اسیلوان (۱۹۸۶ و ۱۹۸۸) پیشنهاد داد که جریمه‌ای برای منحنی برازش اضافه شود. او انتگرال توان دوم مشتق دوم منحنی برازش‌شده را به‌عنوان جریمه هموارسازی[۴۱] به تابع توان دوم خطای برازش به صورت زیر افزود:

$$S=\sum_{i=۱}^{n}\left\{y_i-\sum_{j=-(k-۱)}^{m}\alpha_j B_j^k(x_i)\right\}^۲+\lambda\int_{x_{min}}^{x_{max}}\left\{\sum_{j=-(k-۱)}^{m}\alpha_j B_j^n(x)\right\}^۲.$$

بعد از او انتگرال مشتق دوم به‌عنوان جریمه هموارساز رایج شد. هیچ تأکید خاصی برای مشتق دوم وجود ندارد در حقیقت می‌توان از مشتقات مراتب بالاتر یا پایین‌تر نیز استفاده کرد درحالی که مشتق اول منجر به معادلات ساده و برازش خطی می‌شود و مشتق مراتب بالاتر منجر به محسابات پیچیده و طولانی خواهد شد. اسپلاین‌هایی که در برازش آن‌ها از جریمه هموارساز استفاده می‌شود، به‌عنوان اسپلاین‌های جریمه‌ای[۴۲] (P_اسپلاین) معروف شدند که اجزای اصلی آن B_اسپلاین‌ها و جریمه ناهمواری است. در واقع اسپلاین‌های جریمه‌ای ترکیبی از یک پایه B_اسپلاینی و یک جمله جریمه روی اختلاف ضرایب جملات مجموعی از توابع اسپلاین است. جمله جریمه بخش اساسی و نقطه قوت در اسپلاین‌های جریمه‌ای است.

۷.۱ داده‌های فضایی

برخی از داده‌ها به گونه‌ای هستند که در یک ناحیه فضایی گردآوری شده‌اند و دارای مشخصه‌های مکانی هستند و به نوعی وابستگی آن‌ها ناشی از موقعیت و مکان قرار گرفتن آن‌ها در فضای مورد بررسی است. به‌عنوان مثال، غلظت یک نوع ماده معدنی در خاک یک ناحیه، مکان تصادفات جاده‌ای در یک شهر بزرگ و مکان درختان در یک جنگل، داده‌هایی هستند که دارای مشخصه های مکانی هستند. برای تحلیل این نوع داده‌ها از روش های کلاسیک آماری که بر پذیره استقلال و هم توزیعی مشاهدات استوارند، نمی‌توان استفاده کرد. در واقع این نوع داده ها دارای همبستگی‌های فضایی هستند. آمار فضایی شاخه‌ای از آمار است که به تحلیل داده‌های با مشخصه های مکانی که وابستگی آن‌ها تابعی از فاصله بین موقعیت‌های مشاهدات باشد، می‌پردازد. در این حالت مشاهدات نزدیک به هم وابسته‌تر و مشاهدات دورتر از هم، وابستگی کم‌تری دارند. این نوع داده‌ها، داده‌های فضایی نامیده می‌شوند.

در آمار فضایی متغیر مورد اندازه‌گیری ممکن است گسسته یا پیوسته باشد. فضای موقعیت یا مکان مشاهدات نیز ممکن است پیوسته یا گسسته، نقطه‌ای یا ناحیه‌ای، منظم یا نامنظم باشد. وقتی مقدار متغیر در یک ناحیه ثبت می‌شود، موقعیت ناحیه‌ای و اگر مقدار متغیر در

[۴۱] Smoothing Penalty
[۴۲] Penalizaton Spline

یک نقطه ثابت، طول و عرض جغرافیایی معین اندازه‌گیری شود، موقعیت نقطه‌ای است. در موقعیت نقطه‌ای اگر نقاط در فواصل مساوی از هم قرار داشته باشند، موقعیت‌ها منظم و در غیر این صورت نامنظم‌اند. در مورد مکان‌های ناحیه‌ای اگر ناحیه‌ها هم شکل و هم اندازه باشند، ناحیه‌ها منظم و در غیر این صورت نامنظم‌اند. با توجه به انواع موقعیت‌ها، مشاهدات فضایی به سه گروه عمده، داده‌های زمین‌آماری[۴۳]، داده‌های مشبکه‌ای[۴۴] و الگو نقطه‌ای[۴۵] تقسیم می‌شوند.

انواع داده‌های فضایی

- **دادهای زمین‌آماری:** این نوع داده‌ها در موقعیت‌های ثابت و مشخص در ناحیه‌ای پیوسته مشاهده می‌شوند. متغیر مورد بررسی ممکن است گسسته یا پیوسته باشد. به‌عنوان مثال برای حالت پیوسته می‌توان غلظت مواد معدنی در داخل یک معدن، مقدار ریزش باران ثبت‌شده در ایستگاه‌های هواشناسی و برای حالت گسسته می‌توان تعداد نوعی جانور دریایی در تعدادی از مکان‌های نمونه‌گیری‌شده در طول یک ساحل را نام برد.

- **داده‌های مشبکه‌ای:** این نوع داده‌ها مربوط به مکان‌های ناحیه‌ای هستند، که این مکان‌ها ممکن است منظم یا نامنظم باشند. تصاویر ماهواره‌ای از سطح زمین که در آن سطح زمین به تعدادی عنصر تصویر کوچک به‌صورت مشبکه منظم در R^2 تقسیم شده‌است، مثالی برای داده‌های مشبکه‌ای منظم است. تعداد افراد مبتلا به سرطان در تمام مناطق خدمات درمانی کشور مثالی برای داده‌های مشبکه‌ای نامنظم است. به‌طور معمول هدف از تحلیل داده‌های فضایی مشبکه‌ای مدل‌بندی احتمالی مشاهدات است، درصورتی‌که در زمین‌آمار معمولا پیش‌گویی مقدار متغیر در یک موقعیت جدید مد نظر است.

- **الگو نقطه‌ای:** در این حالت مکان یا موقعیت مشاهده‌شده خود متغیر تصادفی است. الگو نقطه‌ای شامل تعدادی متناهی از مکان‌ها در یک ناحیه‌اند که در آن‌ها یک صفت خاص اندازه‌گیری می‌شود. به‌طور معمول الگوهای نقطه‌ای به سه دسته به‌طور کامل تصادفی فضایی، منظم و خوشه‌ای تقسیم و به مدل‌بندی آن‌ها اقدام می‌شود. یک مثال برای این نوع مشاهدات موقعیت گونه‌ای از درختان در یک ناحیه جنگلی یا موقعیت مراکز زلزله است.

---

[۴۳] Geostatistical Data

[۴۴] Lattice Data

[۴۵] Point Pattern

۱.۷.۱ فرآیند تصادفی مارکوفی گاوسی

فرآیند تصادفی مجموعه‌ای از متغیرهای تصادفی با اندیس مرحله یا زمان است که وضعیت یک پدیده آزمایش تصادفی را در طول یک دوره نمایش می‌دهد. متغیرهای تصادفی و اندیس آن‌ها می‌توانند از نوع گسسته و پیوسته و نیز چند بعدی باشند. فرض کنید X_t متغیر تصادفی متناسب با t باشد که در فضای پارامتر T و فضای وضعیت S تعریف شده باشد. در این صورت، مجموعه متغیرهای تصادفی $\{X_t; t \in T\}$ را یک فرآیند تصادفی گویند که در آن X_t به ازای هر $t \in T$ یک متغیر تصادفی است.

حال فرآیند تصادفی $\{X_t; t \in T\}$ را در نظر بگیرید. هرگاه هر ترکیب خطی متناهی از متغیرهای تصادفی X_t دارای توزیع نرمال باشد، X_t را یک فرآیند گاوسی (نرمال) می‌نامند. این فرآیندها کاربردهای مختلفی در علوم مختلف دارند. اگر مجموعه اندیس‌گذار $T \subset \mathbb{R}^d$، $d > ۱$ باشد، آن‌گاه تعریف فرآیند تصادفی به میدان تصادفی تعمیم داده می‌شود. به‌عنوان مثال در داده‌های فضایی $T \subset \mathbb{R}^۲$ است یا در داده‌های فضایی‌ـ‌زمانی $T \subset \mathbb{R}^۳$. در حالتی که فرآیند گاوسی باشد، به میدان حاصل یک میدان تصادفی گاوسی[۴۶] (GRF) می‌گویند.

یک زیررده بسیار پرکاربرد از GRF ها، میدان‌های تصادفی مارکوفی گاوسی[۴۷] (GMRFs) هستند که به‌ویژه برای مدل‌بندی داده‌های مشبکه‌ای به‌طور گسترده استفاده می‌شوند. بهترین راه برای تفهیم ویژگی مارکوفی یک میدان تصادفی گاوسی، استفاده از تعریف گراف است. گراف g مجموعه‌ای از رأس‌ها است که توسط خانواده‌ای از یال‌ها به‌هم وصل شده‌اند و به‌صورت زوج مرتب (ν, ε) نشان داده می‌شود، که در آن ν مجموعه‌ای متناهی و غیرتهی از رئوس و ε یال‌های آن است. اگر رأس‌ها به‌صورت $\nu = ۱, \cdots, n$ باشند، گراف را نشاندار می‌نامند.

فرض کنید بردار تصادفی $\boldsymbol{x} = (x_۱, \cdots, x_n)'$ دارای توزیع نرمال چندمتغیره با میانگین μ و ماتریس کوواریانس Σ باشد. همچنین فرض کنید $g = (\nu, \varepsilon)$ گرافی نشاندار باشد که در آن ε شامل همه زوج‌های i, j است، به‌طوری‌که رأس‌های i و j هیچ یال مشترکی ندارند، اگر و تنها اگر $x_i \perp x_j | \boldsymbol{x}_{-ij}$، در این صورت x یک (GMRF) نسبت به گراف g نامیده می‌شود. منظور از نماد $\cdot \perp \cdot | \cdot$ استقلال شرطی است. قبل از بیان یک تعریف رسمی برای GMRF، ارتباط بین گراف g و پارامترهای توزیع نرمال را بیان می‌کنیم. از آن‌جا که $\boldsymbol{\mu}$ تأثیری بر استقلال شرطی دوبه‌دوی درایه‌های x ندارد، می‌توان نتیجه گرفت اطلاعات مربوط به استقلال شرطی فقط در ماتریس کوواریانس Σ یا معکوس آن یعنی ماتریس دقت، Q، پنهان شده‌اند. پس ماتریس دقت نقش مهمی در تحلیل GMRF بازی می‌کند.

قضیه ۱.۷.۱. (رو و هلد، ۲۰۰۵). اگر x دارای توزیع نرمال چندمتغیره با بردار میانگین $\boldsymbol{\mu}$ و ماتریس دقت معین مثبت $Q > ۰$ باشد، آن‌گاه درایه ijام ماتریس Q، یعنی Q_{ij} برابر صفر است، اگر و تنها اگر $x_i \perp x_j | \boldsymbol{x}_{-ij}$.

---

[۴۶] Gaussian Random Field

[۴۷] Gaussian Markov Random Fields

با توجه به قضیه ۱.۷.۱، درایه‌های ناصفر ماتریس Q، گراف g را مشخص می‌کنند و بر اساس آن‌ها می‌توان استقلال شرطی x_i و x_j را بررسی کرد.

تعریف ۱.۷.۱. بردار تصادفی $\boldsymbol{x} = (x_1, \cdots, x_n)' \in \mathbb{R}^n$ یک میدان تصادفی مارکوفی گاوسی تحت گراف نشاندار $G = (\nu, \varepsilon)$ با میانگین $\boldsymbol{\mu}$ و ماتریس دقت $Q > 0$ است، اگر و تنها اگر تابع چگالی آن به‌صورت

$$f(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} |Q|^{\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' Q (\boldsymbol{x} - \boldsymbol{\mu})\}$$

باشد و برای هر $i, j \in \nu$ که $i \neq j$

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in \varepsilon.$$

مزیت اساسی میدان‌های مارکوفی گاوسی در تنک بودن ماتریس دقت آن‌هاست که این ویژگی دلیل اصلی سرعت بالای روش‌های عددی در تجزیه ماتریس است. این مزیت محاسباتی به‌ویژه در مواردی که تجزیه ماتریس باید بارها و بارها انجام شود، به شدت حائز اهمیت است. یکی از این موارد، اجرای روش‌های نمونه‌گیری MCMC است که در هر تکرار الگوریتم برای تولید نمونه از توزیع پسین یک مدل فضایی باید ماتریس دقت (کوواریانس) تجزیه شود.

فصل ۲

رگرسیون چندکی و گشتاوری

۱.۲ مقدمه

امروزه مدل‌های رگرسیونی متنوعی غیر از رگرسیون میانگین مورد توجه قرار گرفته‌اند که می‌توان به مدل‌های رگرسیون میانه، نما و چندکی اشاره کرد. به‌عنوان مثال در حوزه مطالعات سلامت کودکان، پزشکان علاقه‌مند هستند تا عوامل موثر بر سوءتغذیه کودکان را شناسایی کنند؛ یا در مطالعات زلزله‌شناسی، پژوهشگران به دنبال شناخت شرایطی هستند که در آن‌ها سهمگین‌ترین طوفان‌ها رخ می‌دهد. در این‌گونه مسائل، هدف مدل رگرسیونی بررسی دم توزیع شرطی پاسخ به شرط متغیرهای تبیینی است. مدل‌های معمول برای رسیدن به این هدف، مدل‌های رگرسیونی چندکی هستند.

چالش اصلی استنباط آماری در مدل‌های رگرسیون چندکی، برای برازش داده‌ها، وجود تابع قدرمطلق در تابع هدف مدل است که نیازمند می‌نیمم‌سازی آن هستیم. دلیل این چالش نیز مشتق‌پذیر نبودن تابع هدف در برخی از نقاط دامنه خود است. مدل رگرسیون گشتاوری که اخیرا مورد توجه قرار گرفته است، فاقد این مشکل است؛ دلیل آن هم مشتق‌پذیر بودن تابع هدف در کل دامنه تابع است که این ویژگی به دلیل جایگزین شدن تابع توان دوم با قدرمطلق است.

در این فصل ابتدا دو مدل رگرسیون چندکی و گشتاوری را معرفی می‌کنیم و سپس به بررسی ویژگی‌ها و مقایسه آن‌ها می‌پردازیم.

۲.۲ رگرسیون چندکی

۱.۲.۲ چندک ها

می‌دانیم که تحلیل رگرسیون کلاسیک روی میانگین تمرکز می‌کند. یعنی رابطه بین پاسخ و متغیرهای تبیینی توسط تابع میانگین شرطی پاسخ توصیف می‌شود. ایده مدل‌بندی و برازش تابع میانگین شرطی، هسته خانواده وسیعی از روش‌های مدل‌بندی رگرسیون از جمله مدل رگرسیون خطی ساده، رگرسیون چندمتغیره، مدل‌های با خطای ناهم‌واریانس[۱]، مدل‌های رگرسیون ناپارامتری (هاردل، ۱۹۹۰) و مدل‌های رگرسیون غیرخطی است.

مدل‌های میانگین شرطی در محاسبه و تفسیر، راحت و سرراست هستند اما محدودیت‌های زیر را دارند:

- هنگام خلاصه کردن پاسخ برای مقادیر ثابت متغیرهای تبیینی، مدل میانگین شرطی نمی‌تواند به موقعیت‌های غیرمرکزی تعمیم یابد.
- در جهان واقعی همیشه با پذیره‌های مدل روبه‌رو نمی‌شویم.
- پدیده‌های اجتماعی معمولا دارای توزیع‌های دم سنگین هستند که منجر به افزایش داده‌های پرت می‌شوند و مدل‌های میانگین شرطی نمی‌توانند این موضوع را در نظر بگیرند.

یکی از روش‌هایی که می‌توان برخی از این محدودیت‌ها را برطرف کرد، استفاده از رگرسیون میانه است. در مقایسه با رگرسیون میانگین، مواقعی که توزیع خطاها چوله است، میانه به‌عنوان یک شاخص مرکزی داده‌ها، حاوی اطلاعات مفیدتری است. از نظر محاسباتی، رگرسیون میانگین به راحتی قابل اجرا است ولی برای رگرسیون میانه باید توابعی را می‌نیمم کرد که نیاز به می‌نیمم‌سازی عددی و توان محاسباتی بیشتری است.

مفهوم چندک به عبارت‌های خاص چارک، دهک و صدک تعمیم داده می‌شود. چندک pام، مقداری از پاسخ را مشخص می‌کند که قبل از آن سهم جامعه p است. بنابراین چندک‌ها می‌توانند هر موقعیتی از توزیع را مشخص کنند. میانه یک چندک خاص است که مکان تمرکز توزیع را توصیف می‌کند. رگرسیون میانه شرطی یک حالت خاص از رگرسیون چندکی است که در آن چندک با مرتبه ۰٫۵ به‌عنوان تابعی از متغیرهای تبیینی مدل‌بندی می‌شود. دیگر چندک‌ها نیز می‌توانند وضعیت‌های غیرمرکزی یک توزیع را توصیف کنند.

تعریف ۱.۲.۲. (تابع محدب). فرض کنید $-\infty \leq a < b \leq +\infty$، تابع $f : (a,b) \longrightarrow \mathbb{R}$ را محدب گویند، هرگاه به ازای هر دو عدد $x_1, x_2 \in (a,b)$ و هر t که $0 \leq t \leq 1$ بتوان نوشت

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$

[۱] Heteroscedastic Error

اگر X متغیر تصادفی با تابع توزیع تجمعی

$$\mathrm{F}(x) = \mathrm{P}(X \leq x)$$

باشد، چندک مرتبه τ به‌صورت زیر تعریف می‌شود:

$$\eta(\tau) = \mathrm{Q}(\tau) = \mathrm{F}^{-1}(\tau) = \inf\{x : \mathrm{F}(x) \geq \tau\}$$

که در آن $\tau \in [0, 1]$. توجه کنید که میانه معادل $\mathrm{F}^{-1}(\frac{1}{2})$ است.

چندک‌ها را می‌توان با استفاده از یک مسأله بهینه‌سازی ساده به‌دست آورد. در واقع در این مسأله ساده، هدف به‌دست آوردن یک برآورد نقطه‌ای برای چندک‌های یک متغیر تصادفی با تابع توزیع F است. برای این منظور تابع زیان به‌صورت زیر تعریف می‌شود:

$$\rho_\tau(\nu) = |\nu||\tau - I(\nu \leq 0)|. \qquad (1.2)$$

توجه کنید که تابع زیان (۱.۲) را تابع بررسی[۲] نیز می‌نامند. این تابع، یک تابع محدب و نامتقارن است. اگر X یک متغیر تصادفی پیوسته باشد، یک چندک خاص را می‌توان با می‌نیمم کردن امید ریاضی تابع زیان (۱.۲) در $X - u$ نسبت به u به‌دست آورد. در واقع می‌توان نوشت

$$\begin{aligned}
\min_u \mathrm{E}(\rho_\tau(X-u)) &= \min_u \mathrm{E}[|X-u||\tau - I(X \leq u)|] \\
&= \min_u \mathrm{E}[(I(X \leq u) - \tau)|X-u|] \\
&= \min_u [-\int_{-\infty}^{u} (x-u)dF(x) - \tau \mathrm{E}[|X-u|]] \\
&= \min_u [-\int_{-\infty}^{u} (x-u)dF(x) + \tau \int_{-\infty}^{u} (x-u)dF(x) - \tau \int_{u}^{\infty} (x-u)dF(x)] \\
&= \min_u [(\tau - 1)\int_{-\infty}^{u} (x-u)dF(x) - \tau \int_{u}^{\infty} (x-u)dF(x)].
\end{aligned}$$

با قرار دادن مشتق امید ریاضی تابع زیان برابر صفر می‌توان معادله را حل کرد و چندک مرتبه τ یعنی q_τ از حل معادله زیر به‌دست می‌آید:

$$(\tau - 1)\int_{-\infty}^{q_\tau} d\mathrm{F}(x) + \tau \int_{q_\tau}^{\infty} dF(x) = 0.$$

این معادله به‌صورت زیر قابل بازنویسی است:

$$\mathrm{F}(q_\tau) - \tau = 0.$$

پس

$$\mathrm{F}(q_\tau) = \tau$$

و بنابراین q_τ چندک مرتبه τ متغیر تصادفی X در حالت پیوسته است.

[۲] Check function

چندک‌های نمونه

زمانی‌که F توسط تابع توزیع تجربی یعنی:

$$\mathrm{F}_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq q_\tau)$$

جایگزین شود، چندک نمونه‌ای مرتبه τ از حل مسأله می‌نیمم‌سازی زیر محاسبه می‌شود:

$$\begin{aligned}\hat{q}_t &= \arg\min_{q\in\mathbb{R}} \sum_{i=1}^{n} \rho_\tau(x_i - q) \\ &= \arg\min_{q\in\mathbb{R}} [\tau \sum_{x_i \geq q} (x_i - q) + (\tau - 1) \sum_{x_i < q} (x_i - q)].\end{aligned}$$

این روش را کوئنکر و باست (۱۹۹۷) برای پیدا کردن چندک‌های نمونه‌ای ارائه کردند. در واقع هدف آن‌ها از معرفی این روش، تعمیم آن به مفهوم رگرسیون چندکی است.

۲.۲.۲ محاسبه چندک‌ها از طریق بهینه‌سازی

محاسبه چندک‌ها را می‌توان به‌عنوان مسأله می‌نیمم‌سازی مطرح کرد. درست همان‌طور که می‌توان میانگین نمونه را به‌عنوان حل مسأله می‌نیمم‌سازی مجموع توان دوم باقیمانده‌ها تعریف کرد، مسأله می‌نیمم‌سازی مجموع قدرمطلق باقیمانده‌ها را نیز می‌توان تعریف کرد. تقارن تکه‌ای خطی قدرمطلق مقدار تابع بیان می‌کند که در می‌نیمم‌سازی مجموع قدرمطلق باقیمانده‌ها، تعداد باقیمانده‌های مثبت و منفی باید یکسان فرض شوند. به‌عبارت دیگر، فرض می‌شود تعداد مشاهدات بالا و پایین میانه برابر هستند.

در مورد دیگر چندک‌ها، از آن‌جایی که تقارن مقدار قدرمطلق، نتیجه‌اش میانه است، می‌نیمم‌کردن مجموع نامتقارن وزنی قدرمطلق باقیمانده‌ها، به‌طور ساده، با دادن وزن‌های متفاوت به باقیمانده‌های مثبت و منفی، نتیجه‌اش سایر چندک‌ها خواهد بود.

می‌دانیم که میانگین نمونه، جواب مسأله زیر است:

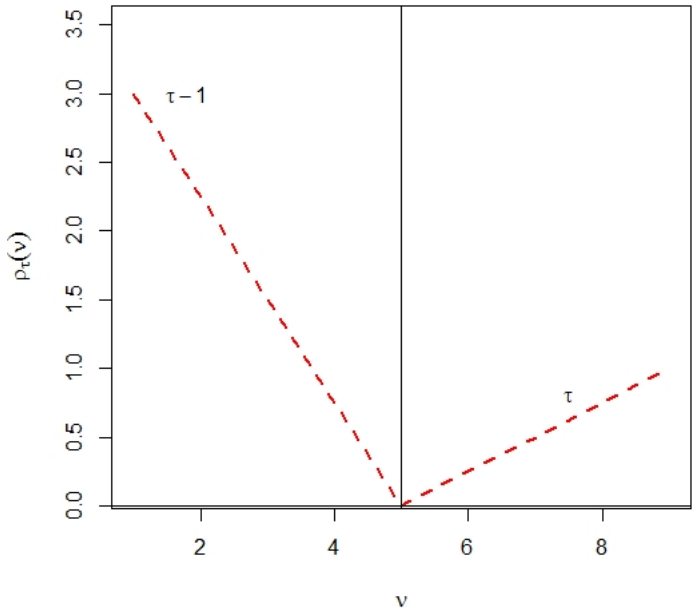
$$\min_{\mu\in\mathbb{R}} \sum_{i=1}^{n} (y_i - \mu)^2.$$

پس اگر خواهان بیان میانگین شرطی متغیر پاسخ y به شرط متغیرهای تبیینی $\boldsymbol{x}$ با بعد p باشیم، به‌طوری‌که $\mu(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}$، بنابراین $\boldsymbol{\beta}$ را می‌توان از طریق حل معادله زیر به‌دست آورد:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n} (y_i - x_i'\boldsymbol{\beta})^2.$$

اگر تابع چندک شرطی را به‌صورت $Q_\tau(y|x) = \boldsymbol{x}'\boldsymbol{\beta}_\tau$ قرار دهیم، $\hat{\boldsymbol{\beta}}_\tau$ جواب معادله زیر است:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\boldsymbol{\beta}_\tau)$$



شکل ۱.۲: تابع زیان

که در آن $\rho_\tau(\cdot)$ تابع زیان معرفی‌شده در رابطه (۱.۲) است. نمودار تابع زیان (۱.۲) در شکل ۱.۲ نمایش داده شده است.

۳.۲.۲ مدل رگرسیون چندکی

رگرسیون چندکی رابطه بین چندک‌های توزیع متغیر پاسخ به شرط متغیرهای تبیینی را ارزیابی می‌کند. مدل رگرسیون چندکی نگرش دقیق‌تری را از ارتباط بین متغیرهای تبیینی و پاسخ فراهم می‌کند. این مدل توسط کوئنکر و باست در سال ۱۹۷۸ میلادی معرفی شد. با این مدل می‌توان شکل توزیع را در سطوح مختلف متغیرهای تبیینی به‌دست آورد.

مدل رگرسیون چندکی مرتبه τ به صورت زیر است:

$$y_i = \eta_{i\tau} + \epsilon_{i\tau} \qquad \mathrm{F}_{\epsilon_{i\tau}}(0) = \mathrm{P}(\epsilon_{i\tau} \leq 0) = \tau \tag{۲.۲}$$

به‌طوری‌که $\mathrm{F}_{\epsilon_\tau}(\cdot)$ تابع توزیع تجمعی $\epsilon_{i\tau}$ است که در صفر مقداردهی شده است. پیش‌گوی η_τ شامل متغیرهای تبیینی است، مثلا پیش‌گوی خطی به‌صورت زیر است:

$$\eta_\tau = \beta_0 + \sum_{k=1}^{p} \beta_k x_k$$

که در آن $\boldsymbol{x} = (x_1, \ldots, x_p)'$ بردار متغیرهای تبیینی با بعد p است. در چنین مدلی $\tau \in (0,1)$ و به‌ازای $\tau = 0.5$ رگرسیون میانه به‌دست می‌آید که مرکز توزیع پاسخ را هدف قرار می‌دهد.

همچنین اگر $\tau \longrightarrow ۰$ یا $\tau \longrightarrow ۱$، دم پایین یا بالای توزیع پاسخ نتیجه می‌شود. پذیرش صفر بودن چندک مرتبه τ توزیع جمله خطا در معادله (۲.۲)، این نتیجه را در بر دارد که η_τ متناظر با چندک مرتبه τ توزیع پاسخ است. هنگامی‌که چندک τ ام متغیر تصادفی ϵ به ازای تمام سطوح متغیر تبیینی برابر صفر باشد، مدل (۲.۲) معادل مدل رگرسیون میانگین کلاسیک است.

رگرسیون چندکی در شرایطی مورد توجه است که در آن نه تنها میانگین به متغیرهای تبیینی وابسته است بلکه سایر ویژگی‌های توزیع پاسخ نیز تابعی از متغیرهای تبیینی هستند. همان‌طور که می‌دانیم در رگرسیون میانگین کلاسیک برای برآورد پارامترهای مدل از روش کمترین توان‌های دوم استفاده می‌شود، اما برآوردگر کلاسیک در رگرسیون چندکی با بهینه‌سازی معیار قدر مطلق باقی‌مانده‌های موزون نامتقارن[۳] به‌دست می‌آید که به‌صورت زیر تعریف می‌شود:

$$\sum_{i=۱}^{n} \omega_\tau(y_i, \eta_{i\tau})|y_i - \eta_{i\tau}| \qquad (۳.۲)$$

که در آن y_i و $\eta_{i\tau}$، $i = ۱, \cdots, n$، پاسخ‌های مشاهده‌شده و مقادیر پیش‌گو هستند و وزن‌ها به‌صورت زیر تعریف می‌شوند (کوئنکر، ۲۰۰۵):

$$\omega_\tau(y_i, \eta_{i\tau}) = \begin{cases} ۱ - \tau & y_i \leq \eta_{i\tau} \\ \tau & y_i > \eta_{i\tau} \end{cases}$$

می‌نیمم‌سازی تابع هدف (۳.۲)، نیاز به روش‌های برنامه‌ریزی خطی[۴] دارد که نسبت به روش کمترین توان‌های دوم، پیچیدگی محاسباتی مضاعف وارد مساله می‌کند.

۳.۲ رگرسیون گشتاوری

۱.۳.۲ تاریخچه رگرسیون گشتاوری

در سال‌های اخیر با توجه به پیشرفت و توسعه مدل‌های رگرسیونی، آماردانان تمرکز ویژه‌ای بر شاخص‌های مختلف توزیع پاسخ در مقایسه با رگرسیون میانگین دارند. در این حالت روش استاندارد، رگرسیون چندکی است که پیش‌تر به معرفی آن پرداختیم. با توجه به مشتق‌ناپذیری تابع هدف مدل رگرسیون چندکی که محاسبات پیچیده‌تری در مقایسه با بهینه‌سازی توان دوم کلاسیک دارا است، می‌نیمم‌سازی این تابع هدف بر حسب پارامترهای رگرسیونی، توسط روش‌های برنامه‌ریزی خطی انجام می‌شود. این چالش محاسباتی رگرسیون چندکی باعث شد تا رگرسیون گشتاوری که متکی بر باقیمانده توان دوم وزنی نامتقارن است، توسط نیوی و پاول (۱۹۸۷) معرفی شود. از آن‌جایی که برآورد رگرسیون گشتاوری را می‌توان با استفاده از روش

[۳] Asymmetrically Weighted Absolute Residual Criterion

[۴] Linear Programming

کمترین توان‌های دوم وزنی ساده به‌دست آورد، این مدل در سال‌های اخیر بسیار مورد توجه قرار گرفته‌است.

امروزه مدل‌های پیچیده‌تری برای تعمیم مدل رگرسیون گشتاوری پیشنهاد شده‌اند. به‌عنوان مثال می‌توان به وارد کردن اثرات غیرخطی متغیرهای تبیینی با اسپلاین‌ها (اشنیبل و ایلزر، ۲۰۰۹)، مدل‌بندی پاسخ‌های فضایی (سوبتکا و نیب، ۲۰۱۲) و مدل‌بندی متغیرهای ابزاری[۵] (سوبتکا و همکاران، ۲۰۱۳) اشاره کرد.

در حالی‌که نتایج اولیه مدل‌های رگرسیون گشتاوری برای برآورد کمترین توان‌های دوم در دسترس هستند (سوبتکا و همکاران، ۲۰۱۳)، روش‌های برآورد جایگزین مانند بوستینگ[۶] برای معرفی گشتاورها توسط سوبتکا و نیب (۲۰۱۲) با تکیه بر روش‌های خودگردان‌سازی[۷] برای انجام استنباط نیز معرفی شده‌اند. یانگ و زو (۲۰۱۵) نوع جدیدی از گشتاور بوستینگ را معرفی کردند و فاروگ و استینوارت (۲۰۱۵) از مدل‌های ماشین بردار پشتیبان[۸] در رویکرد برآورد گشتاورها الگوبرداری کردند. مدل گشتاوری اتورگرسیو برای تحلیل سری‌های زمانی توسط تیلور (۲۰۰۸) معرفی شد. اخیرا ماجومدار و پاول (۲۰۱۶) از یک توزیع نرمال دو متغیره برای معرفی رگرسیون فضایی با گشتاورهای صفر استفاده کردند.

در فصل سوم یک مدل بیزی از رگرسیون گشتاوری را معرفی خواهیم کرد که بر توزیع نرمال نامتقارن[۹] (AND) به‌عنوان توزیع پاسخ متکی است. این روش بسیار شبیه به رگرسیون چندکی بیزی است که در آن از توزیع لاپلاس نامتقارن[۱۰] (ALD) استفاده شده است. البته مدل نرمال نامتقارن را نمی‌توان مانند مدل لاپلاس نامتقارن به‌صورت توزیعی برحسب مکان‌ـ‌مقیاس[۱۱] نوشت، به‌طوری‌که بتوان توزیع پیشنهادی مناسبی برای حالت بیزی بر حسب آن به‌دست آورد. برای جزئیات بیشتر به یو و رو (۲۰۱۱)، کوزومی و کوبایاشی (۲۰۱۱) و رِد و یو (۲۰۰۹) مراجعه کنید.

یکی از مزایای فرمول‌بندی بیزی رگرسیون گشتاوری سر راست بودن آن است، به‌طوری‌که به‌راحتی می‌توان اثرات پیچیده‌تری را در آن گنجاند. به‌عنوان مثال، امکان استفاده از اثرات غیرخطی متغیرهای تبیینی به کمک اسپلاین‌ها، اثرات فضایی به کمک میدان‌های تصادفی مارکوفی، و فرآیندهای انقباضی[۱۲] مانند پیشین‌های میخ و تخته[۱۳] (فهرمیر و همکاران، ۲۰۱۰) به‌راحتی قابل دستیابی است.

---

[۵] Instrumental Variable
[۶] Boosting
[۷] Bootstrap
[۸] Support Vector Machine-type
[۹] Asymmetric Normal Distribution
[۱۰] Asymmetric Laplace Distribution
[۱۱] Location-Scale
[۱۲] Shrinkage Processes
[۱۳] Spike and Slab Priors

۲.۳.۲ مدل رگرسیون گشتاوری

مدل رگرسیون گشتاوری اولین بار توسط نیوی و پاول در سال ۱۹۸۷ پیشنهاد شد و به دلیل مزایای محاسباتی، در سال‌های اخیر، مورد توجه قرار گرفته است. همان مدل (۲.۲) را در نظر بگیرید که به‌صورت

$$y_i = \eta_{i\tau} + \epsilon_{i\tau} \qquad i = ۱, ۲, \cdots, n$$

است، که در آن y_i متغیر پیوسته پاسخ، $\epsilon_{i\tau}$ خطاهای مستقل و $\eta_{i\tau}$ یک پیش‌گوی وابسته به پارامتر نامتقارن $\tau \in (۰, ۱)$ است که محدوده دنباله توزیع پاسخ را تعیین می‌کند.

برای به‌دست آوردن جانشینی مناسب برای رگرسیون چندکی، نیوی و پاول (۱۹۸۷) پیشنهاد کردند تابع بررسی (۱.۲) با تابع زیان توان دوم نامتقارن که به‌صورت زیر تعریف می‌شود، جایگزین شود:

$$\rho_\tau(\nu) = |\tau - I(\nu \leq ۰)|\nu^۲, \quad \tau \in (۰, ۱). \tag{۴.۲}$$

بنابراین برآوردگرهای رگرسیونی $\hat{\boldsymbol{\beta}}_\tau$ از می‌نیمم کردن تابع هدف

$$\sum_{i=۱}^{n} \rho_\tau(y_i - \boldsymbol{x}_i \boldsymbol{\beta}_\tau)$$

روی بردار $\boldsymbol{\beta}$ نتیجه می‌شوند، که در آن تابع بررسی (۴.۲) جانشین تابع بررسی رگرسیون چندکی شده است.

برای شناسایی رده برآوردگرهای پارامترهای رگرسیونی، نیوی و پاول پارامتر عددی $\mu(\tau)$ را در نظر گرفتند که می‌نیمم‌کننده تابع

$$\mathrm{E}[\rho_\tau(Y - m) - \rho_\tau(Y)]$$

روی m است، که در آن امید ریاضی نسبت به توزیع متغیر تصادفی Y با فرض وجود میانگین آن، محاسبه می‌شود. آن‌ها نشان دادند که $\mu(\tau)$ جواب معادله زیر است:

$$\mu(\tau) - \mathrm{E}(Y) = \frac{۲\tau - ۱}{۱ - \tau} \int_{[\mu(\tau), \infty]} (y - \mu(\tau)) \mathrm{dF}(y) \tag{۵.۲}$$

که در آن $\mathrm{F}(y)$ تابع توزیع متغیر تصادفی Y است.

با این تعریف، $\mu(\tau)$ توسط ویژگی‌های امید ریاضی متغیر تصادفی Y به شرط آن‌که Y در دم توزیع قرار گیرد، تعیین می‌شود. با این تفسیر، نیوی و پاول پارامتر $\mu(\tau)$ را به‌عنوان تابعی از τ با نام تابع گشتاور (گشتاور مرتبه τ) خواندند. تابع گشتاور $\mu(\tau)$ به شکلی خیلی مشابه با تابع چندک $Q(\tau)$، ویژگی‌های تابع توزیع را مشخص‌سازی[۱۴] (خلاصه‌سازی) می‌کند. برای اثبات این ادعا نیوی و پاول قضیه زیر را بیان و اثبات کردند. ابتدا فرض کنید I_F بیانگر مجموعه $\{y | ۰ < \mathrm{F}(y) < ۱\}$ است.

[۱۴] Characterize

قضیه ۱.۳.۲. (نیوی و پاول، ۱۹۸۷). فرض کنید $\mathrm{E}(y) = m$ وجود داشته باشد. بنابراین برای هر $۰ < \tau < ۱$، یک پاسخ یکتا مثل $\mu(\tau)$ برای معادله (۵.۲) وجود دارد و دارای ویژگی‌های زیر است:

الف) $\mu(\tau)$ به‌عنوان یک تابع به شکل $\mu(\tau) : (۰, ۱) \to \mathbb{R}$، اکیدا یکنوای صعودی است.

ب) دامنه $\mu(\tau)$، I_F است و $\mu(\tau)$ فاصله $(۰, ۱)$ را به I_{F} تصویر می‌کند.

ج) برای تبدیل خطی $\tilde{Y} = aY + b$، که در آن $a > ۰$، گشتاور مرتبه τ متغیر $\tilde{Y}$، یعنی $\tilde{\mu}(\tau)$، در تساوی

$$\tilde{\mu}(\tau) = a\mu(\tau) + b$$

صدق می‌کند

د) اگر $\mathrm{F}(y)$ مشتق‌پذیر پیوسته باشد، آن‌گاه $\mu(\tau)$ نیز مشتق‌پذیر پیوسته است و برای $y \neq m$ در I_{F} و τ_y به‌طوری که $y = \mu(\tau_y)$ داریم

$$\mathrm{F}(y) = \frac{-[y - m + \tau_y \mu'(\tau_y)(۱ - ۲\tau_y)]}{\mu'(\tau_y)(۱ - ۲\tau_y)^{۲}} \qquad (۶.۲)$$

که در آن $\mu'(\tau)$ مشتق تابع گشتاور نسبت به τ است. در حالت حدی، این تساوی برای $y = m$ و $\tau_y = \frac{۱}{۲}$ برقرار است.

ویژگی ج) قضیه بیان می‌کند، مشابه چندک مرتبه τ، گشتاور مرتبه τ نسبت به مکان و مقیاس هم‌ورداا[۱۵] است. از همه مهمتر، ویژگی‌های ب) و د) بیانگر آن هستند که تابع $\mu(\tau)$ توزیع متغیر تصادفی Y را مشخص‌سازی می‌کند. با توجه به ویژگی ب)، دامنه $\mu(\tau)$ مجموعه I_{F} است و برای هر y در I_{F}، معادله (۶.۲) نمایشی برای $\mathrm{F}(y)$ بر حسب $\mu(\tau)$ و مشتق‌اش ارایه می‌دهد.

یک معادله جانشین برای تابع $\mu(\tau)$، به‌صورت زیر است:

$$\frac{\tau}{۱ - \tau} = \left[\int_{(-\infty, \mu(\tau))} (\mu(\tau) - y)\mathrm{dF}(y)\right]\left[\int_{(\mu(\tau), \infty)} (y - \mu(\tau))\mathrm{dF}(y)\right]^{-۱}.$$

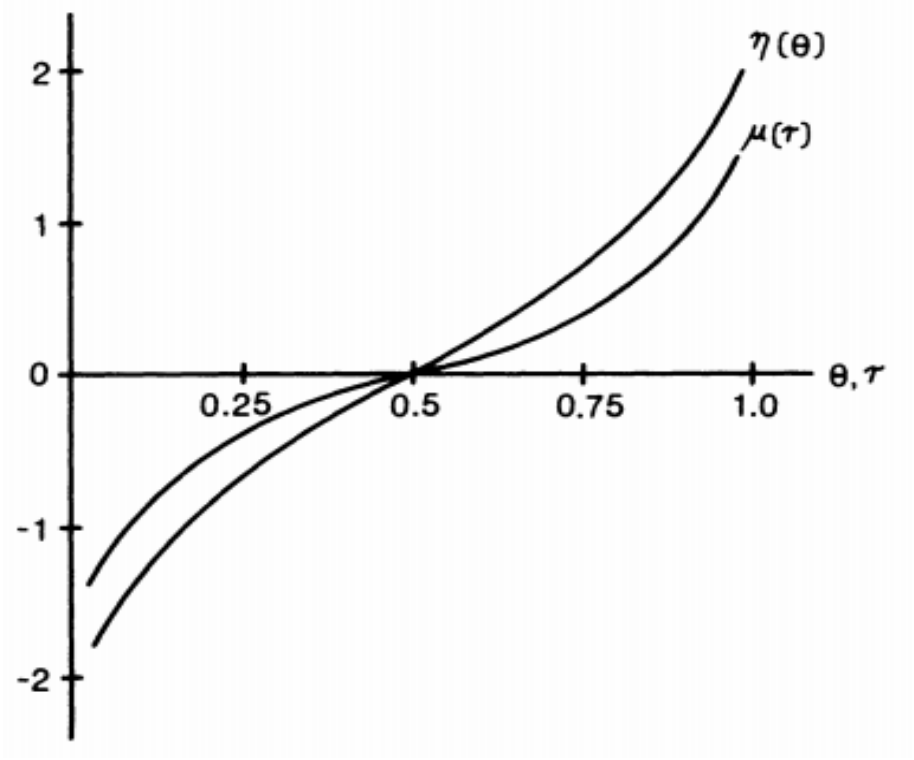
از مقایسه این معادله با رابطه مشابهی برای چندک‌ها، یعنی معادله

$$\frac{\tau}{۱ - \tau} = \frac{\mathrm{F}(\eta(\tau))}{۱ - \mathrm{F}(\eta(\tau))}$$

به‌سادگی می‌توان دید که گشتاورها توسط امید ریاضی‌های دمی، به شکلی مشابه با چندک‌ها، تابع توزیع را شناسایی می‌کنند.

پس در مجموع می‌توان گفت، بنا بر قضیه ۱.۳.۲، گشتاورها ویژگی‌های مشابه چندک‌ها دارند. برای درک شهودی از این شباهت، در شکل ۲.۲ دو تابع چندک و گشتاور برای توزیع نرمال استاندارد رسم شده‌اند. در نزدیکی $\tau = ۰/۵$ تابع گشتاور شیب کمتری نسبت به تابع چندک دارد و در نزدیکی $\tau = ۰$ یا $\tau = ۱$ دارای شیب بزرگ‌تری است.

[۱۵] Equivariant



شکل ۲.۲: تابع چندک $\eta(\theta)$ و تابع گشتاور $\mu(\tau)$ برای توزیع نرمال استاندارد. برگرفته از نیوی و پاول (۱۹۸۷).

در رگرسیون میانگین می‌دانیم $\mathrm{E}(\epsilon_{i\tau}) = 0$ و در رگرسیون چندکی نیز $\mathrm{P}(\epsilon_{i\tau} \leq 0) = \tau$. در مقابل، در رگرسیون گشتاوری عبارت زیر بیان‌گر این است که پیش‌گوی $\eta_{i\tau}$ برابر با گشتاور τام پاسخ y_i است:

$$\arg\min_{\boldsymbol{\beta}} \mathrm{E}(\omega_\tau(\epsilon_{i\tau})|\epsilon_{i\tau} - e|^2) = 0.$$

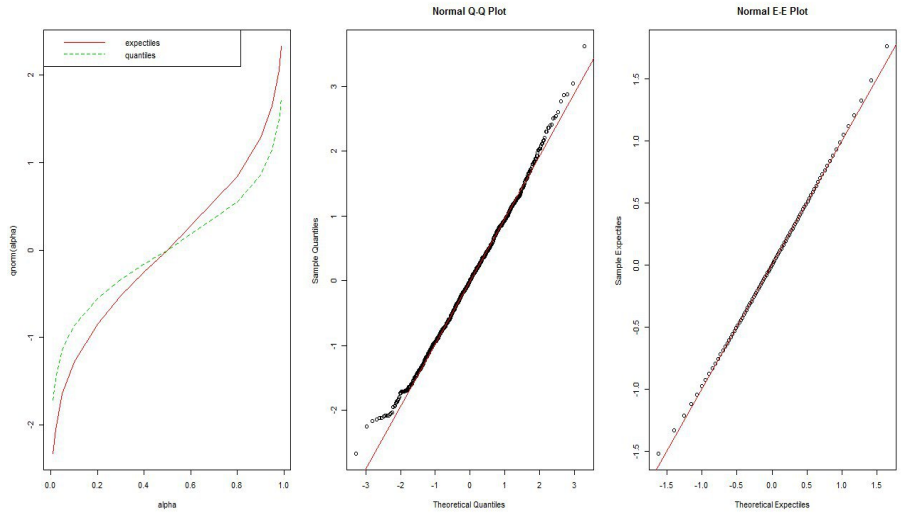
در مدل رگرسیون گشتاوری، از معیار توان دوم باقی‌مانده‌های موزون نامتقارن که به‌صورت زیر تعریف می‌شود، برای برازش مدل استفاده می‌کنند:

$$\sum_{i=1}^{n} \omega_\tau(y_i, \eta_{i\tau})(y_i - \eta_{i\tau})^2. \qquad (۷.۲)$$

استفاده از این معیار این مزیت بزرگ را دارد که برآورد پارامترها (به دلیل مشتق‌پذیر بودن تابع هدف نسبت به پارامترهای مدل) با روش کمترین توان‌های دوم وزنی تکراری قابل محاسبه هستند و نیاز به استفاده از روش‌های برنامه‌ریزی خطی نیست. در واقع، در مدل رگرسیون گشتاوری فرض می‌شود که پیشگوی η_τ گشتاور مرتبه τ پاسخ y است. در این مدل به ازای $\tau = 0.5$، مدل رگرسیون میانگین نتیجه می‌شود که از این منظر که تفسیرپذیری راحتی دارد، نسبت به رگرسیون میانه می‌تواند برتری داشته باشد.

از آن‌جایی که گشتاورها جانشین مناسبی برای چندک‌ها در مشخص‌سازی توزیع پاسخ هستند، این مدل می‌تواند جانشینی برای رگرسیون چندکی در مطالعه ویژگی‌های مختلف

توزیع شرطی پاسخ باشد (نمودارهای سمت چپ شکل ۳.۲ را ببینید). حتی با توجه به شکل ۳.۲ می‌توان دریافت که ممکن است گشتاورها بتوانند توزیع را بهتر مشخص کنند. با میل کردن مرتبه τ به سمت صفر یا یک، بخش‌های فرین توزیع مورد توجه خواهند بود. عمده ویژگی‌های رگرسیون گشتاوری مشابه رگرسیون چندکی است، زیرا به‌جز چندک صفر و گشتاور صفر بودن مرتبه τ خطا، هیچ پذیره دیگری در مورد شرایط خطا در نظر گرفته نشده است. در واقع، خطاها می‌توانند ناهم‌پراش باشند یا ممکن است انواع مختلفی از توزیع‌ها را دنبال کنند.



شکل ۳.۲: چندک‌ها و گشتاورهای نظری توزیع نرمال استاندارد (شکل سمت چپ)، نمودار چندک‌ـ چندک (شکل وسط) و گشتاورـ گشتاور (شکل سمت راست) برای نمونه‌ای هزار تایی از توزیع نرمال استاندارد

برای تعریف رگرسیون گشتاوری در یک چارچوب بیزی، به یک توزیع پاسخ نیاز داریم که معیار بهینه‌سازی (۷.۲) از تابع درستنمایی آن حاصل شود. در رگرسیون چندکی بیزی از توزیع لاپلاس نامتقارن و در رگرسیون گشتاوری از توزیع نرمال نامتقارن استفاده می‌شود. یعنی فرض می‌شود:

$$y \sim AN(\mu, \sigma^2, \tau)$$

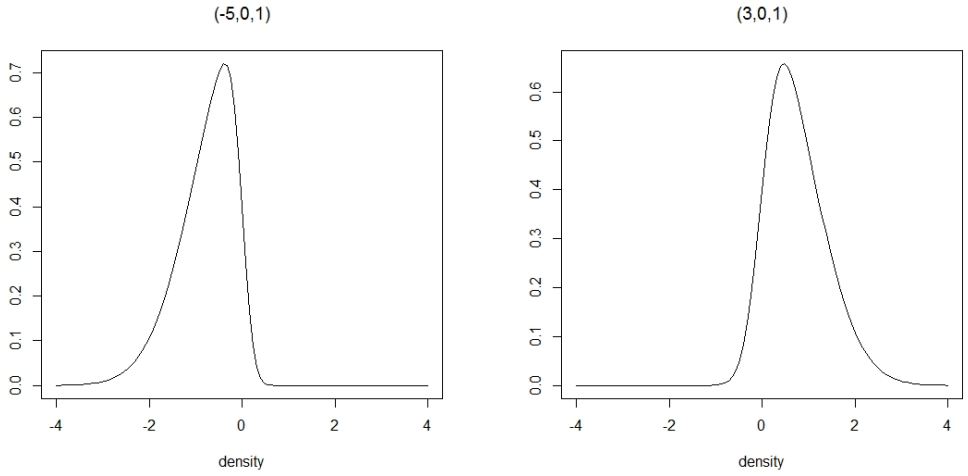
که تابع چگالی، میانگین و واریانس آن به‌صورت زیر است:

$$f(y) = \frac{2}{\sqrt{\sigma^2 \pi}} \left(\sqrt{\frac{1}{1-\tau}} + \sqrt{\frac{1}{\tau}} \right)^{-1} \exp\left(-\frac{1}{\sigma^2} \omega_\tau(y, \mu)(y-\mu)^2 \right)$$

$$\mathrm{E}(y) = \mu + \frac{\sigma}{(\sqrt{\tau}) + \sqrt{1-\tau}} \left(\frac{1-2\tau}{\sqrt{\pi\tau(1-\tau)}} \right)$$

$$\mathrm{Var}(y) = \frac{\sigma^2}{\sqrt{\tau} + \sqrt{1-\tau}} \left[\frac{1}{2} \left(\frac{\sqrt{1-\tau}}{\tau} + \frac{\sqrt{\tau}}{1-\tau} \right) - \frac{1}{\sqrt{\tau} + \sqrt{1-\tau}} \left(\frac{(1-2\tau)^2}{\pi\tau(1-\tau)} \right) \right].$$

در مدل رگرسیونی گشتاوری قرار می‌دهیم $\mu_i = \eta_i = \boldsymbol{x}'\boldsymbol{\beta}$. ماکسیمم کردن درستنمایی این توزیع، معادل با کمینه کردن کمیت (۷.۲) است، چون با هسته لگاریتمی توزیع یکسان است. شکل ۴.۲ منحنی تابع چگالی توزیع نرمال نامتقارن را برای دو مجموعه متفاوت از پارامترهای آن نشان می‌دهد.



(ب) (آ)

شکل ۴.۲: تابع چگالی نرمال نامتقارن به ازای مقادیر متفاوتی از پارامترهای توزیع

۳.۳.۲ تفسیرپذیری رگرسیون گشتاوری

گشتاورها تفسیرپذیری راحتی مانند چندک‌ها ندارند، اما با دیدگاه‌های زیر می‌توان آن‌ها را تفسیر کرد:

- برای نمونه‌های i.i.d مانند $y_1, \cdots, y_n$ برآورد گشتاوری نتیجه‌شده از روی داده‌ها، $\hat{e}_\tau$، یک میانگین وزنی از مشاهدات است. به‌عبارت دقیق‌تر

$$\hat{e}_\tau = \sum_{i=1}^{n} \omega_i y_i$$

که در آن وزن‌های ω_i به گشتاور برآوردشده وابسته است. در نتیجه گشتاورهای رگرسیونی می‌توانند به‌عنوان یک میانگین وزنی مشابه که روی بردار متغیرهای تبیینی شرطی شده‌اند، در نظر گرفته شوند.

- گشتاورها امید ریاضی دم هستند، یعنی مرتبه τ چنین گشتاورهایی در رابطه زیر صدق می‌کند (والدمن و همکاران، ۲۰۱۶):

$$\tau = \frac{\int_{-\infty}^{e_\tau} |y - e_\tau| f(y) dy}{\int_{-\infty}^{\infty} |y - e_\tau| f(y) dy}. \qquad (۸.۲)$$

چنین رابطه‌ای نشان می‌دهد که e_τ توسط یک گشتاور جزئی[۱۶] تعیین می‌شود. برآورد τ در رابطه (۸.۲) بر اساس یک نمونه از مشاهدات به‌صورت زیر قابل محاسبه است:

$$\tau = \frac{\sum_{i:y_i<e_\tau} |y_i - e_\tau|}{\sum_{i=۱}^{n} |y_i - e_\tau|}.$$

- بنابراین می‌توان نتیجه گرفت e_τ یک مرکز ثقل وزنی است (یو و تانگ، ۱۹۹۶). به این معنا که هر چه τ به مرکز نزدیک باشد، وزن‌هایی که به مرکز داده می‌شود، بیشتر است و هر چه τ به دم‌ها نزدیک باشد، وزن دم‌ها بیشتر می‌شود.
- معمولا محققین از یک گشتاور با یک مرتبه خاص استفاده نمی‌کنند، بلکه مجموعه‌ای از گشتاورها را با مرتبه‌های مختلف در نظر می‌گیرند، زیرا مجموعه‌ای از گشتاورها درباره شکل توزیع شرطی پاسخ به شرط متغیرهای تبیینی اطلاعات کاملی در اختیار محقق قرار می‌دهد. به‌ویژه، می‌توان تشخیص داد آیا توزیع شرطی y به شرط متغیرهای تبیینی ناهمگن، چوله یا کشیده است یا خیر. علاوه‌بر این، چندک‌ها به‌وسیله گشتاورها قابل محاسبه هستند. بنابراین اگر کسی علاقه‌مند به برآورد چندک‌ها باشد، با برآورد گشتاورها می‌تواند این کار را انجام دهد. افرون (۱۹۹۱) و والتراپ و همکاران (۲۰۱۵) با جزئیات بیشتر این موضوع را بیان کرده‌اند.

۴.۲ مزایا و معایب رگرسیون گشتاوری و چندکی

در این قسمت، رگرسیون گشتاوری و چندکی را مقایسه کرده و مزایا و معایب آن‌ها را بیان می‌کنیم:

- مزیت اصلی رگرسیون گشتاوری نسبت به رگرسیون چندکی در تابع هدف آن است. تابع هدف مدل گشتاوری نسبت به ضرایب رگرسیونی مشتق‌پذیر است. همین ویژگی امکان استفاده راحت از روش کمترین توان‌های دوم وزنی تکراری را برای برآورد پارامترها و در نتیجه برازش مدل رگرسیون گشتاوری فراهم می‌کند. این موضوع زمانی اهمیت بیشتری پیدا می‌کند که با مدل‌های رگرسیونی پیچیده‌تری نسبت به مدل خطی سروکار داریم. مثلا برازش مدل با اثرات تصادفی، یا مدل‌های با ساختارهای وابستگی زمانی و فضایی از نظر محاسباتی پیچیده است. این مزیت سبب می‌شود که مشکل برنامه‌ریزی خطی رگرسیون چندکی برطرف شود.
- گشتاورها امید ریاضی را به‌عنوان یک حالت خاص در نظر می‌گیرند: گشتاور با مرتبه $\tau = ۰/۵$ برابر با میانگین توزیع است. بنابراین تفسیر رگرسیون گشتاوری با مرتبه ۰/۵ همان تفسیر رگرسیون میانگین است. البته برای سایر مرتبه‌ها مساله برعکس است و

[۱۶] Partial Moment

تفسیرپذیری مدل‌های چندکی راحت‌تر از گشتاوری است. در واقع تفسیر مستقیمی از رگرسیون گشتاوری مشابه آن‌چه که برای رگرسیون چندکی می‌توان مطرح کرد، قابل ارایه نیست. برای رگرسیون چندکی مرتبه τ می‌توان گفت که ۱۰۰τ درصد داده‌ها زیر خط رگرسیون هستند و بقیه بالای خط و به سادگی قابل درک است ولی برای رگرسیون گشتاوری این تفسیر مستقیم وجود ندارد. تفسیری غیرمستقیم توسط والتراپ و همکاران (۲۰۱۴) ارایه شده است.

- نیوی و پاول (۱۹۸۷) نشان دادند که استفاده از گشتاورها در تحلیل داده‌ها، کاراتر از چندک‌ها است. دلیل این کارایی آن است که رگرسیون گشتاوری بر فاصله مشاهدات از پیشگوی رگرسیونی متکی است، در حالی‌که در رگرسیون چندکی داده‌ها از یک مقدار آستانه کمتر یا بیشتر در نظر گرفته می‌شوند. البته این مزیت این هزینه را هم دارد که حساسیت آن نسبت به نقاط پرت افزایش می‌یابد.

- خط رگرسیون چندکی باید از p نقطه (p ضریب رگرسیونی) بگذرد. بنابراین انتظار داریم منحنی برازش‌شده به داده‌ها هموار نباشد. ولی رگرسیون گشتاوری این محدودیت را ندارد و منحنی برازش‌شده هموارتر است (سبوتکا و نیب، ۲۰۱۲).

- می‌توان با برآورد گشتاورها، چندک‌ها را برآورد کرد (والتراپ و همکاران، ۲۰۱۴). در عمل موقعیت‌هایی هستند که چندک‌های مختلف متقاطع[۱۷] هستند، که در این صورت رگرسیون چندکی تفسیر درستی از داده‌ها ارایه نخواهد کرد. والتراپ و همکاران (۲۰۱۴) نشان دادند اگر از طریق گشتاورها، چندک‌ها را بازیابی کنیم این مشکل کمتر رخ خواهد داد.

با توجه به ویژگی‌های بیان‌شده، توصیه می‌شود که از رگرسیون گشتاوری استفاده شود، به دلیل این‌که گشتاورها، چندک‌ها را نتیجه می‌دهند و بعضا ویژگی‌های بهتری از چندک‌ها دارند.

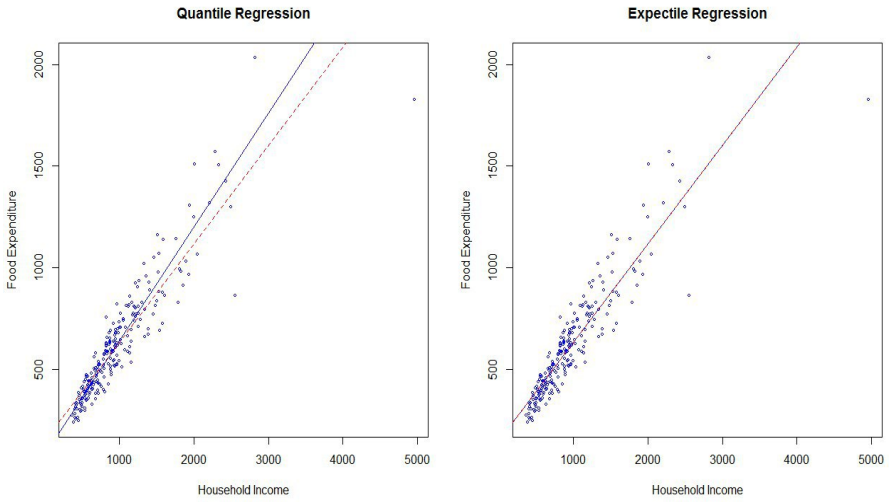
۵.۲ مقایسه عملکرد مدل‌ها: مثال واقعی

برای نمایش عملکرد دو مدل پیشنهادی، از یک مجموعه داده مربوط به هزینه سبد غذایی خانوار و درآمد آن‌ها شامل ۲۳۵ خانوار استفاده کردیم. پیشگوی خطی برای هر دو مدل به صورت

$$\eta_i = \beta_۰ + \beta_۱ \text{Income}_i, \quad i = ۱, \ldots, ۲۳۵$$

در نظر گرفته شد. با استفاده از دو بسته `quantreg` و `expectreg` در محیط `R`، دو رگرسیون خطی چندکی و گشتاوری را به داده‌ها برازش دادیم. برای مقایسه رگرسیون میانگین را نیز برازش دادیم و نتیجه را در شکل ۵.۲ مشاهده می‌کنید. به‌وضوح معلوم است که خط رگرسیون

[۱۷] Crisscross

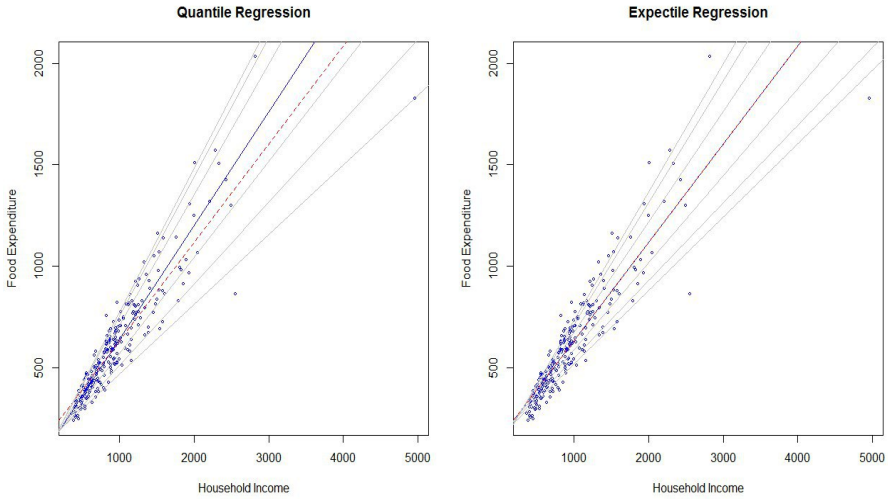


شکل ۵.۲: خط‌های رگرسیون برازش‌شده با مدل‌های رگرسیون گشتاوری (سمت راست) و چندکی (سمت چپ) مرتبه ۵٪ به همراه رگرسیون میانگین

میانگین به سمت دو داده پرت در قسمت پایین پراکنش داده‌ها متمایل شده است، در حالی‌که رگرسیون میانه نسبت به حضور آن‌ها حساس نیست. از طرفی خط رگرسیون گشتاوری مرتبه ۵٪ با خط رگرسیون میانگین یکی شده است. با هدف مطالعه سایر ویژگی‌های توزیع پاسخ که هزینه سبد غذایی خانوار است، خط‌های رگرسیونی با مرتبه‌های مختلف

$$\tau \in \{۰/۰۵, ۰/۱, ۰/۲۵, ۰/۷۵, ۰/۹, ۰/۹۵\}$$

را نیز برای هر دو مدل برازش دادیم. نتیجه در شکل ۶.۲ مشاهده می‌شود. عملکرد هر دو مدل تقریبا یکسان است با این تفاوت که مدل چندکی پراکندگی بیشتری از توزیع شرطی پاسخ را در نظر گرفته است.



شکل ۶.۲: خط‌های رگرسیون برازش‌شده با مدل‌های رگرسیون گشتاوری (سمت راست) و چندکی (سمت چپ) با مرتبه‌های مختلف τ

فصل ۳

مدل بیزی رگرسیون گشتاوری فضایی

در این فصل حالت بیزی مدل رگرسیون گشتاوری که در فصل دوم بیان شد را با تمرکز بر اثرات فضایی معرفی می‌کنیم. از طریق اعمال پیشین‌های مناسب به بحث منظم‌سازی[۱] (جریمه کردن[۲]) پیچیدگی مدل می‌پردازیم. در پایان، استنباط یا تقریب توزیع پسین مدل را بیان می‌کنیم.

قبل از معرفی مدل بیزی، ابتدا درباره تنظیم‌سازی صحبت کرده و روش‌های مختلف آن را معرفی می‌کنیم.

۱.۳ تنظیم‌سازی

روش کمترین توان‌های دوم[۳] (OLS)، یکی از متداول‌ترین روش‌های برآورد ضرایب رگرسیونی در مدل‌های رگرسیونی است. از مهم‌ترین ویژگی‌های این روش، محاسبات نسبتا ساده آن است. علاوه بر آن قضیه معروف گاوس‌ـ مارکوف[۴] نیز در این راستا وجود دارد که ثابت می‌کند بهترین برآوردگر خطی نااریب پارامترهای رگرسیونی، برآوردگر OLS است. اما وضعیت‌های متفاوتی وجود دارند که در برخی از آن‌ها، این برآوردگر به خوبی رفتار نمی‌کند. هنگامی‌که

[۱] Regularisation
[۲] Penalized
[۳] Least Squares
[۴] Gauss-Markov Theorem

خطاها توزیع دم سنگین دارند، داده‌های پرت می‌توانند بر برآوردگر OLS تاثیر زیادی بگذارند. وقتی‌که تعداد پارامترها زیاد است یا وقتی ستون‌های ماتریس طرح رگرسیونی X به شدت همبسته هستند یا به‌عبارت دیگر، در داده‌ها هم‌خطی[۵] وجود دارد، واریانس برآوردگر OLS افزایش می‌یابد. این بدان معناست که قدرمطلق برآوردهای کمترین توان‌های دوم خیلی بزرگ می‌شود و در نتیجه برآوردهای ناپایداری[۶] تولید می‌کند. یعنی با ارائه نمونه‌های متفاوت اندازه این برآوردها، اغلب دارای اریبی کم ولی واریانس زیاد هستند.

یکی از روش‌ها برای مواجهه با این مسائل استفاده از تکنیک جریمه است. برای توضیح، مدل رگرسیون خطی

$$y = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

را در نظر بگیرید، که در آن $\boldsymbol{y} = (y_1, \cdots, y_n)'$ بردار n بعدی پاسخ، $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ بردار p بعدی ضرایب رگرسیونی، Xماتریس $n \times p$ طرح شامل متغیرهای تبیینی و $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$ بردار n بعدی جمله‌های خطا است. در روش کمترین توان‌های دوم جریمه‌ای[۷] به جای کمینه کردن تابع مجموع توان‌های دوم خطا که در رابطه

$$\mathrm{SSE}(\boldsymbol{\beta}) = \boldsymbol{y}'\boldsymbol{y} - ۲\boldsymbol{y}'X\boldsymbol{\beta} + \boldsymbol{\beta}'X'X\boldsymbol{\beta}$$

تعریف شده است، تابع SSE جریمه‌شده زیر کمینه می‌شود:

$$\mathrm{SSE}(\boldsymbol{\beta}) + P_\lambda(\boldsymbol{\beta}) \qquad (۱.۳)$$

که در آن $P_\lambda(\boldsymbol{\beta})$ یک تابع جریمه است که مقادیر بزرگتر پارامترهای نامعلوم را جریمه می‌کند. روش‌های جریمه‌شده هموار با حداقل یک پارامتر تنظیم‌کننده[۸] (λ) ظاهر می‌شود که موازنه بین تابع SSE و تابع جریمه را کنترل می‌کند.

فرض کنید تعدادی از متغیرهای تبیینی مهم وجود دارند اما برخی از آن‌ها با متغیر پاسخ نامرتبط هستند. مدل‌هایی با تعداد متغیرهای کمتر همواره دلخواه تحلیل‌گران آماری هستند چرا که این‌گونه مدل‌ها روابط بین متغیرهای علمی را به صورت ساده بیان می‌کنند و قابلیت تفسیرپذیری بیشتری دارند. هزینه‌های محاسباتی و صرف زمان زیاد، روش‌های انتخاب متغیر سنتی و کلاسیک را با چالش‌های بسیار جدی در زمینه استفاده از داده‌ها با بعد بالا روبه‌رو کرده‌اند. یک روش خوب انتخاب مدل آماری، مدلی را انتخاب می‌کند که استفاده از آن آسان، قابل فهم و قابل توضیح برای دیگران باشد. از آن‌جایی‌که کمینه کردن تابع SSE مدل کامل، برآوردگری با واریانس بزرگ را تولید می‌کند، پیشنهاد شده است یک زیرمجموعه از متغیرها بر اساس معنی‌داری آن‌ها در مدل یا برخی از معیارهای انتخاب مدل، انتخاب شود. مشکل

---

[۵] Colinearity
[۶] Unstable Estimates
[۷] Least Quadratic Penalty
[۸] Tuning Parameter

اساسی که در این رهیافت وجود دارد این است که معیارها زمانی قابل استفاده هستند که همه مدل‌های ممکن شناخته شده باشند. اما تعداد همه زیر مجموعه‌های ممکن به صورت نمایی با افزایش p، زیاد می‌شود و با رایانه‌های امروزی، عملا برای pهای بزرگ‌تر از ۴۰ یا ۵۰ محاسبات ممکن نیست. بدین دلیل روش‌های اصلاح‌شده‌ای مانند انتخاب پیش‌رو[۹]، پس‌رو[۱۰] و گام به گام[۱۱]، به‌عنوان ترکیبی از انتخاب‌های پیش‌رو و پس‌رو، پیشنهاد شده‌اند. یکی از مشکلاتی که می‌توان به آن اشاره کرد، گسسته بودن این روش‌ها است. به این معنی که با تغییر کوچکی در داده‌ها، برآوردهای کاملا متفاوتی به‌دست می‌آیند. در نتیجه رهیافت انتخاب زیرمجموعه[۱۲] اغلب ناپایدار بوده و با تغییرپذیری زیادی همراه است.

ایده جریمه کردن در انتخاب متغیر و کاهش تعداد متغیرها در مدل‌سازی‌های آماری، می‌تواند بسیار مفید باشد. روش جریمه کردن را می‌توان به دو گروه کلی تقسیم کرد: گروه اول، روش‌های مبتنی بر جریمه کردن اندازه مدل هستند که منجر به معرفی معیارهای انتخاب مدل شده‌اند. تعداد زیادی از معیارهای انتخاب مدل بر اساس نظریه اطلاع پدید آمده‌اند. یکی از پرکاربردترین آن‌ها، معیار اطلاع آکاییک[۱۳] (AIC) است که توسط آکاییک (۱۹۷۳) معرفی شد و به‌صورت زیر تعریف می‌شود:

$$\mathrm{AIC} = -۲\ell + ۲p$$

که در آن، ℓ لگاریتم تابع درستنمایی ماکسیمم‌شده و p تعداد پارامترهای موجود در مدل است. هر مدل زیرمجموعه با AIC کمتر، از سایر مدل‌ها برتری دارد. بعد از معرفی این معیار، معیارهای دیگری پدید آمده‌اند که معروف‌ترین آن‌ها معیار اطلاع بیزی[۱۴] (BIC؛ شوارتز، ۱۹۷۸) است و به‌صورت زیر تعریف می‌شود:

$$BIC = -۲\ell + p\ln(n)$$

که در آن n، حجم نمونه است.

گروه دوم، روش‌های مبتنی بر جریمه کردن تک تک متغیرهای تبیینی هستند که به روش‌های کمترین توان‌های دوم جریمه‌شده معروف هستند و به‌ویژه زمانی که تعداد متغیرها از تعداد مشاهدات خیلی بیشتر است (مثل مطالعات بیولوژیکی و پزشکی)، مورد توجه هستند. این روش‌ها در انتخاب متغیر پایدارتر، پیوسته و با روش‌های محاسباتی کارآمدتری قابل دستیابی هستند. فن و لی (۲۰۰۱) سه ویژگی را ارائه کردند که یک تابع جریمه مناسب باید آن‌ها را داشته باشد:

---

[۹] Forward
[۱۰] Backward
[۱۱] Stepwise
[۱۲] Subset selection
[۱۳] Akaike Information Criterion
[۱۴] Bayesian Information Criterion

نااریبی: به‌منظور جلوگیری از اریبی برآوردگر، باید پارامتر برآوردشده به مقادیر بزرگ پارامتر نامعلوم نزدیک باشد.
تنکی[۱۵]**:** برآوردگر باید یک حد آستانه داشته باشد که به‌صورت خودکار پارامترهای برآوردشده کوچک را صفر در نظر بگیرد، تا بدین وسیله از پیچیدگی مدل کاسته شود.
پیوستگی: برآوردگر به‌دست آمده باید پیوسته باشد تا پیش‌گویی‌های مدل پایدار باشد.
برای به‌کار بردن تکنیک جریمه، بهتر است ابتدا متغیرهای تبیینی استاندارد شوند به‌طوری‌که میانگین آن‌ها صفر و واریانس آن‌ها، یک باشد. متغیر پاسخ نیز مرکزی می‌شود، به‌گونه‌ای که میانگین آن برابر صفر شود. به‌عبارت دیگر

$$\bar{Y} = ۰ \qquad \bar{X}_j = ۰ \qquad {X'}_j X_j = n \qquad \forall j.$$

لازم به توضیح است تغییرات در مقیاس را بعد از برازش مدل، می‌توان به‌صورت زیر بازیابی کرد:

$$X_{ij}\beta_j = \frac{X_{ij}}{a} a\beta_j = \tilde{X}_{ij}\tilde{\beta}_j \qquad \tilde{X}_{ij} = \frac{X_{ij}}{a} \qquad \tilde{\beta}_j = a\beta_j.$$

به‌عبارت دیگر اگر X_j بر a تقسیم شود، به سادگی می‌توان پاسخ $\tilde{\beta}_j$ را بر a تقسیم کرد تا β_j اولیه را به‌دست آورد.

برآورد ریج

روش‌های متعددی برای غلبه بر مشکل هم‌خطی در مدل‌های خطی ارائه شده‌اند. یکی از موثرترین روش‌ها برای حل مشکل هم‌خطی، رگرسیون ریج[۱۶] است که توسط هورل و کنارد (۱۹۷۰) معرفی شد. برآورد ریج به‌صورت زیر است:

$$\hat{\boldsymbol{\beta}}_{ridge} = \arg\min_{\boldsymbol{\beta}} \left\{ (\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta}) + \lambda \sum_{j=۱}^{n} {\beta_j}^{۲} \right\} = (X'X + \lambda I_p)^{-۱} X'\boldsymbol{y}$$

که در آن λ پارامتر تنظیم‌کننده است. اگر $\lambda = ۰$ باشد، برآورد ریج به برآورد OLS تبدیل می‌شود و اگر λ بزرگ شود، برآورد ریج به صفر میل می‌کند. هم‌چنین برای هر λ، یک مجموعه متفاوت از برآورد ضرایب به‌دست می‌آید. در حالتی‌که $X'X = I_p$ برآوردگر ریج به‌صورت $\frac{۱}{۱+\lambda}\hat{\boldsymbol{\beta}}$ است که ویژگی اساسی رگرسیون ریج یعنی انقباض را نشان می‌دهد. به‌کارگیری جریمه ریج، برآوردها را به صفر منقبض می‌کند که باعث به‌وجود آمدن اریبی می‌شود اما واریانس برآوردها را کاهش می‌دهد. از طرف دیگر این تابع جریمه شرط تنکی را فراهم می‌سازد.
از طرفی، برآوردگرهای OLS همیشه وجود ندارند زیرا اگر ماتریس $X'X$ رتبه کامل نباشد، $X'X$ معکوس‌پذیر نیست و پاسخ یکتا برای OLS وجود ندارد. اما برآوردگر ریج همیشه وجود

[۱۵] Sparse
[۱۶] Ridge

دارد. زیرا برای هر ماتریس X، ماتریس $X'X + \lambda I_p$ همیشه معکوس‌پذیر است. همین‌طور برای هر $\boldsymbol{\beta}$، λای وجود دارد که MSE برآوردگر ریج کمتر از MSE برآوردگر OLS است که این نتیجه بسیار ارزشمند است، زیرا بیان می‌کند که حتی اگر مدل برازش‌شده کاملا درست باشد و از توزیع دقیق مشخص‌شده پیروی کند، همیشه می‌توان برآوردگر بهتری را با منقبض کردن به سمت صفر به‌دست آورد. اگر چه روش ریج کمک می‌کند تا برآوردهایی با تغییرپذیری کمتر به‌دست آیند، اما دارای نقص‌های زیر نیز هست:

- ضرایب با مقادیر واقعی بزرگتر را نیز به سمت صفر منقبض می‌کند.
- مشکل تفسیرپذیری مدل وجود دارد، به این معنی که متغیرهای تبیینی بی‌اهمیت مدل ممکن است به صفر منقبض شوند ولی هنوز در مدل حضور دارند.

برآورد لاسو

تیبشیرانی (۱۹۹۶) پیشنهاد کرد از تابع جریمه $\lambda \sum_{j=1}^{p} |\beta_j|$ در رابطه (۱.۳) استفاده شود. در نتیجه، برآوردگر $\boldsymbol{\beta}$ به‌صورت زیر به‌دست می‌آید:

$$\hat{\boldsymbol{\beta}}_{lasso} = \arg\min_{\boldsymbol{\beta}} \left\{ (\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta}) + \lambda \sum_{j=0}^{p} |\beta_j| \right\}.$$

تنها تفاوت برآوردگر لاسو با برآوردگر ریج، استفاده از تابع قدرمطلق به جای توان دوم به عنوان جریمه است. اگرچه این تغییر بسیار جزئی است، اما تاثیر چشم‌گیری بر برآورد نتیجه‌شده دارد.

مشابه برآوردگر ریج، جریمه کردن مقادیر قدرمطلق ضرایب، آن‌ها را به سمت صفر منقبض می‌کند اما با این وجود، برخلاف برآوردگر ریج، برخی از ضرایب را دقیقا برابر صفر قرار می‌دهد. لذا این تابع جریمه، انتخاب متغیر نیز انجام می‌دهد. برخلاف روش ریج، لاسو صورت بسته‌ای ندارد و برای هر مقدار λ بردار ضرایب به‌صورت عددی برآورد می‌شوند. روش لاسو به لحاظ پایداری و دقت پیش‌گویی برآوردگرها عملکرد قابل قبولی از خود نشان داده است و نسبت به رگرسیون ریج اکثر مواقع دقت پیش‌گویی بالایی دارد و هم‌چنین، منجر به تولید جواب‌های تنک می‌شود. اما این روش محاسبات بسیار پیچیده‌ای دارد که افرون (۲۰۰۴) رگرسیون کمترین زاویه[۱۷] را برای محاسبه برآوردگر لاسو مطرح کرد. در روش لاسو، ضرایب بزرگ نیز جریمه و به سمت صفر منقبض می‌شوند و ضرایب کوچک‌تر فرصت حضور در مدل را پیدا می‌کنند که باعث بیش برآوردی می‌شود و در نتیجه برآوردگر لاسو، دارای ویژگی نااریبی نیست. در این راستا، فن و لی (۲۰۰۱) تابع جریمه اسکد[۱۸] را معرفی کردند. اگر تعداد واقعی متغیرها بیش از حجم نمونه باشد، برآوردگر لاسو، اجازه حضور حداکثر n متغیر را در مدل نهایی می‌دهد که این مشکل به‌وسیله ژو و هستی (۲۰۰۵) رفع شده است. مشکل دیگر

[۱۷] Least Angle Regression

[۱۸] Smoothly Clipped Absolute Deviation Skad

برآوردگر لاسو، ناسازگاری آن است. ژو (۲۰۰۶) شرط لازم برای سازگاری این برآوردگر و صورت اصلاح‌شده‌ای از آن را معرفی کرد.

انتخاب پارامتر تنظیم‌کننده در رگرسیون جریمه‌شده

بدیهی است که روش‌های جریمه‌شده، شدیدا به پارامتر تنظیم‌کننده λ وابسته هستند. این پارامتر، موازنه بین برازش و جریمه را کنترل می‌کنند. وقتی λ بسیار کوچک باشد، داده‌ها بیش‌برازش می‌شوند و در نتیجه مدلی تولید می‌شود که واریانس بزرگی دارد. وقتی λ خیلی بزرگ باشد، داده‌ها کم‌برازش می‌شوند و منجر به مدل‌هایی می‌شود که اریبی بزرگی دارند. بنابراین پارامتر λ یک پارامتر تنظیم‌کننده است که تغییر آن اجازه می‌دهد تا موازنه بین اریبی و واریانس به تعادل برسد. بدیهی است که انتخاب این پارامتر، اولین قدم در استفاده از هر تابع جریمه‌ای است.

برخلاف معیارهای اطلاع، که در آن‌ها جریمه‌های بعد پارامترها یک مقدار ثابت و از پیش تعیین‌شده است، مقدار پارامتر تنظیم‌کننده در رگرسیون جریمه‌شده بر اساس مشاهدات به‌دست می‌آید. روش‌های مختلفی برای یافتن این پارامترها وجود دارند. اما روشی که بیشترین مورد استفاده را دارد، بر مبنای این است که پیش‌گویی‌ها بر اساس $\hat{\beta}_\lambda$ (برآوردگر به‌ازای λ) تا چه اندازه می‌تواند مقدار واقعی پاسخ را بازیابی کند. اگر از مجموعه داده‌ها هم برای برازش مدل و هم برای برآورد دقت پیش‌گویی استفاده شود، آن‌گاه منجر به بیش‌برازشی خواهد شد.

یک ایده به این صورت مطرح می‌شود که مجموعه داده‌ها به دو قسمت تقسیم شود، یک قسمت از آن برای برازش مدل و دیگری برای ارزیابی این که مدل برازش‌شده تا چه اندازه می‌تواند مشاهدات را در قسمت دوم پیش‌گویی کند. مساله اساسی زمانی است که تعداد مشاهدات آنقدر بزرگ نباشد که بتوان به‌راحتی آن‌ها را به دو قسمت تقسیم کرد و از بخشی از آن‌ها به تنهایی برای برآورد λ استفاده کرد.

روش پیشنهادی دیگر، اعتبارسنجی متقابل[۱۹] است که داده‌ها را به K قسمت تقسیم می‌کند؛ مدل به $K-1$ زیرمجموعه برازش داده می‌شود و سپس مخاطره روی قسمتی که کنار گذاشته شده است، محاسبه می‌شود. این رویه برای هر قسمت تکرار می‌شود و در نهایت میانگین مخاطره‌ها روی همه این نتایج محاسبه می‌شود. معمولا فرض می‌شود K برابر ۵ یا ۱۰ است. اگر $K=n$، بدین معنی است که تنها یکی از مشاهدات کنار گذاشته می‌شود. یک روش برای اندازه‌گیری قدرت پیش‌گویی یک مدل، آزمودن آن بر روی مجموعه‌ای از داده‌هاست که در برازش مدل مورد استفاده قرار نگرفته باشند. متخصصان یادگیری ماشین و داده‌کاوی به چنین مجموعه‌ای، مجموعه آزمون و به مجموعه داده‌ای که برای برازش مدل استفاده شده است، مجموعه آموزشی می‌گویند. در این حالت، این معیار به‌صورت زیر تعریف می‌شود:

$$\mathrm{CV} = \frac{1}{n}\sum (y_i - \hat{y}_\lambda^{-i})^2$$

[۱۹] Cross Validation

که در آن، $\hat{y}_\lambda^{-i}$ بیان‌گر مقدار پیش‌گویی مشاهده iام توسط مدلی است که iامین مشاهده در برازش آن نقش نداشته است. در این‌صورت، λ مورد نظر متناظر با λای خواهد بود که این معیار را کمینه کند.

محاسبه $\mathrm{CV}(\lambda)$ بسیار وقت‌گیر است. کروان و واهبا (۱۹۷۹) معیار اعتبارسنجی متقابل تعمیم‌یافته[۲۰] (GCV) را به‌عنوان جایگزینی برای $\mathrm{CV}(\lambda)$ به‌صورت زیر معرفی کردند:

$$\mathrm{GCV}(\lambda) = \frac{(\boldsymbol{y} - \hat{\boldsymbol{y}}_\lambda)'(\boldsymbol{y} - \hat{\boldsymbol{y}}_\lambda)}{n\left(1 - \frac{tr(A(\lambda))}{n}\right)^2}$$

که در آن، $\hat{\boldsymbol{y}}_\lambda$ بردار مقادیر برازش‌شده تحت ماتریس تصویر[۲۱] $A(\lambda)$ است و $tr(A)$ اثر ماتریس A است.

۱.۱.۳ تنظیم‌سازی بیزی

از آن‌جا که در دیدگاه بیزی همه استنباط‌ها مبتنی بر توزیع پسین انجام می‌شود، پس بدیهی است که جریمه کردن پیچیدگی بیش از حد مدل (و در نتیجه خارج کردن متغیرهای تبیینی با اثرات ناچیز) باید در توزیع پسین ظاهر شود. برای رسیدن به این هدف، در دیدگاه بیزی، جریمه کردن بردار ضرایب متغیرهای تبیینی در یک مدل رگرسیونی، که به تنظیم‌سازی بیزی[۲۲] معروف است، با تعریف توزیع پیشین روی بردار ضرایب و تاثیر آن بر توزیع پسین مدل، وارد می‌شود.

به‌عنوان نمونه، اگر در مدل رگرسیونی خطی برای بردار ضرایب رگرسیونی از توزیع پیشین نرمال استفاده کنیم، ماکسیمم کردن توزیع پسین حاصل معادل با می‌نیمم کردن تابع هدف رگرسیون ریج خواهد شد. یا اگر از توزیع پیشین لاپلاس (نمایی دوگانه) برای پارامترهای رگرسیونی استفاده کنیم، ماکسیمم کردن تابع چگالی توزیع پسین حاصل، معادل با می‌نیمم کردن تابع هدف رگرسیون لاسو خواهد شد. لازم به ذکر است که پارامترهای تنظیم‌ساز نیز در نقش ابرپارامترهای توزیع پیشین ظاهر می‌شوند که برای انتخاب مناسب آن‌ها می‌توان از روش‌های مشابه دیدگاه کلاسیک استفاده کرد.

بنابراین انتخاب روش تنظیم‌سازی یا جریمه کردن بیزی، به انتخاب نوع پیشین ضرایب رگرسیونی منتهی می‌شود. در این فصل، تنظیم‌سازی را برای هر دو نوع اثرات خطی و غیرخطی در نظر می‌گیریم. برای تنظیم‌سازی اثرات خطی از پیشین‌های میخ و تخته استفاده می‌کنیم. پیشین میخ و تخته ابتدا توسط میچل و بیوچامپ (۱۹۸۸) برای انتخاب متغیر بیزی با مدل‌های رگرسیونی خطی نرمال معرفی شد. پیشین‌های میخ و تخته، هر ضریب رگرسیونی β_i را با مخلوطی[۲۳] از یک جرم نقطه‌ای در $\beta_i = 0$ (میخ) و یک توزیع پیوسته، مانند نرمال یا

[۲۰] Generalized Cross Validation
[۲۱] Projection Matrix
[۲۲] Bayesian Regularisation
[۲۳] Mixture

یکنواخت (تخته)، مدل می‌کند. در کار اصلی که توسط میچل و بیوچامپ معرفی شد، توزیع میخ و تخته ترکیبی از اندازه دیراک متمرکز بر نقطه صفر و توزیع یکنواخت است.

۲.۳ معرفی مدل رگرسیون گشتاوری بیزی

با توجه به مدل رگرسیون گشتاوری (۲.۲) برای گشتاور مرتبه τ، از یک نسخه ناپارامتری برای پیش‌گوی η به‌صورت زیر استفاده می‌کنیم:

$$\eta_i = \gamma_\circ + \sum_{j=1}^{p} f_j(z_i)$$

که در آن $\gamma_\circ$ سطح کلی پیش‌گو و توابع $f_j(z_j)$ انواع متفاوتی از اثرات خطی و غیرخطی را برای متغیرهای تبیینی z_i نشان می‌دهند. این تعریف پیش‌گوی جمعی[۲۴] انعطاف لازم در مدل‌بندی رگرسیونی را برای وارد کردن انواع اثرات غیرخطی، زمانی و فضایی ایجاد می‌کند.

برای وارد کردن اثر غیرخطی متغیرهای تبیینی در این توابع از اسپلاین‌ها استفاده می‌کنیم. اسپلاین‌ها توابع پایه‌ساز هستند، و هر کدام از توابع f_j را می‌توان به‌صورت زیر نوشت:

$$f_j(z) = \sum_{k=1}^{K} \beta_{jk} B_k(z),$$

که در آن $B_k(z)$ توابع پایه اسپلاین و β_{jk} ضرایب اسپلاین[۲۵] هستند.

پیشینی که برای ضرایب اسپلاین در مدل بیزی در نظر گرفته‌ایم، نرمال است که می‌توان آن را به‌صورت زیر تعریف کرد:

$$\pi(\boldsymbol{\beta_j}|\boldsymbol{\theta_j}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta_j}' K_j(\boldsymbol{\theta_j})\boldsymbol{\beta_j}\right)$$

که در آن $K_j(\boldsymbol{\theta_j})$ ماتریس دقت است که نشان‌دهنده انواع مختلف پذیره‌های وابستگی ساختاری در مورد تابع f_j مانند همواری و تنظیم‌سازی برای مجموعه‌ای از متغیرهای تبیینی است. بردار $\boldsymbol{\theta_j}$ شامل ابرپارامترهایی است که هر کدام از آن‌ها برای کنترل انقباض یا درجه آزادی هموارسازی استفاده می‌شود.

قبل از معرفی پیشین‌های ضرایب پارامتری و تعیین ماتریس دقت K برای پیشین‌های ضرایب اسپلاین، ابتدا بحث تنظیم‌سازی بیزی اثرات این مدل را مطرح می‌کنیم.

۳.۳ تنظیم‌سازی در مدل رگرسیون گشتاوری بیزی

تنظیم‌سازی در مدل رگرسیون گشتاوری بیزی با اعمال پیشین‌ها به‌دست می‌آید که برای اثرات خطی و غیرخطی متفاوت است.

---

[۲۴] Additive Pridector

[۲۵] Coefficients Spline

۱.۳.۳ اثرات خطی

یک رویکرد مهم برای تنظیم‌سازی اثرات خطی، اعمال روش میخ و تخته بر واریانس توزیع پیشین ضرایب رگرسیونی است (فرمیر و همکاران، ۲۰۱۰). البته روش‌های مختلف دیگری برای این هدف وجود دارند که استفاده از آن‌ها سرراست است؛ ولی در این بخش ما از ابرپیشین میخ و تخته برای ابرپارامتر واریانس توزیع نرمال به‌عنوان توزیع پیشین ضرایب استفاده می‌کنیم. برای تنظیم اثرات خطی، فرض می‌کنیم $K_j(\theta_j)$ یک ماتریسی قطری با مولفه‌های قطری $\frac{۱}{\delta_j^۲}$ است، به‌طوری که یک توزیع میخ و تخته را برای $\delta_j^۲$ها در نظر می‌گیریم.

پارامتر (در واقع ابرپارامتر) $\delta_j^۲$ دارای یک توزیع آمیخته است که قسمتی از آن شامل تخته و قسمت دیگر شامل میخ است که پارامتر رگرسیونی را به سمت صفر میل می‌دهد. به‌طور دقیق، با انتخاب توزیع گامای معکوس (IG) توزیع مورد نظر را به‌صورت

$$\delta_j^۲|I_j \sim (۱-I_j)\mathrm{IG}(a_{\delta_j^۲}, \nu_۰ b_{\delta_j^۲}) + I_j \mathrm{IG}(a_{\delta_j^۲}, b_{\delta_j^۲})$$

تعریف می‌کنیم. با افزودن پارامتر اضافی $\nu_۰$، منقبض کردن ضریب رگرسیونی به سمت صفر در توزیع گامای معکوس رخ می‌دهد. در واقع، اگر این پارامتر به اندازه کافی کوچک انتخاب شود، شکل توزیع گامای معکوس متناظر با شکل میخ می‌شود و در نتیجه ضریب رگرسیونی به سمت صفر میل می‌کند. در این پایان‌نامه، مقدار $\nu_۰$ را برابر ۰/۰۰۰۰۱ و پارامترهای توزیع گامای معکوس را $a_{\delta_j^۲} = ۰/۰۰۰۱$ و $b_{\delta_j^۲} = ۰/۰۰۰۵$ در نظر می‌گیریم. مؤلفه آمیختگی توزیعی که $\delta_j^۲$ از آن نمونه‌گیری می‌شود، با پارامتر I_j کنترل می‌شود که فرض می‌کنیم دارای توزیع پیشین برنولی $I_j \sim \mathrm{Bin}(۱, \omega)$ با ابرپیشین $\omega \sim \mathrm{Beta}(\mathrm{a}_\omega, \mathrm{b}_\omega)$ است.

می‌توان یک توزیع پیشین ناآگاهی‌بخش را جایگزین توزیع بتا برای ω کرد. یک جانشین دیگر تعریف توزیع پیشینی است که تعداد کم یا زیادی از پارامترها را در مدل ایجاد کند (والدمن و همکاران، ۲۰۱۶). در این پایان‌نامه برای هر دو مثال‌های شبیه‌سازی و واقعی با انتخاب $a_\omega = b_\omega = ۱$، از توزیع پیشین ناآگاهی‌بخش یکنواخت استاندارد برای ω استفاده شده است. برای هر دوی این توزیع‌های پیشین، چگالی شرطی کامل $I_j|\cdot$ مزدوج $\mathrm{Bin}(۱, \omega*)$ است که در آن

$$\omega^* = \frac{\omega \mathrm{IG}(\delta_j^۲, a_\omega, b_\omega)}{\omega \mathrm{IG}(\delta_j^۲, a_\omega, b_\omega) + (۱-\omega)\mathrm{IG}(\delta_j^۲, a_\omega, \nu_۰ b_\omega)}$$

$$\omega|\cdot \sim \mathrm{Beta}\left(a_\omega + \sum_{j=۱}^{k} I_j, b_\omega + k - \sum_{j=۱}^{k} I_j\right).$$

۲.۳.۳ اثرات غیرخطی

اثرات غیرخطی متغیرهای تبیینی پیوسته را با اسپلاین‌های جریمه‌ای[۲۶] (ایلرز و مارکس، ۱۹۹۶) مدل‌بندی می‌کنیم. فرمول‌بندی بیزی اسپلاین‌های جریمه‌ای (برزگر و لانگ، ۲۰۰۶) نیاز به ماتریس طرح B_j دارد که شامل توابع پایه B ـ اسپلاین و ماتریس دقت است.

برای اثرات غیرخطی ماتریس دقت $K_j(\theta_j)$ به‌صورت زیر تعریف می‌شود:

$$K_j(\theta_j) = \frac{1}{\delta_j^2} D_k' D_k$$

که در آن D_k ماتریس تفاضلی اسپلاین[۲۷] مرتبه kام (پایان‌نامه اریه، ۱۳۹۴) و δ_j^2 پارامتر هموارساز است. برای پارامتر δ_j^2 از توزیع گامای معکوس به‌عنوان پیشین استفاده می‌کنیم، به‌طوری که

$$\delta_j^2 \sim \mathrm{IG}(a_{\delta_j^2}, b_{\delta_j^2}).$$

در این حالت نیز مقادیر $a_{\delta_j^2}$ و $b_{\delta_j^2}$ را مشابه قبل اختیار می‌کنیم.

۳.۳.۳ اثرات فضایی

در این پایان‌نامه، برای مدل‌بندی اثرات فضایی، از GMRFها (رو و هولد، ۲۰۰۵) استفاده می‌کنیم. در این مدل، ماتریس طرح B_j شامل تابع نشان‌گر برای نواحی و ماتریس دقت $K_j(\theta_j) = \frac{1}{\delta_j^2} K_j$ است، که در آن K_j ماتریس مجاورت[۲۸] یا همسایگی نواحی و δ_j^2 پارامتر هموارساز است.

پارامتر هموارساز δ_j^2 تنها برای اثرات پیوسته فضایی استفاده می‌شود و برای آن از توزیع پیشین گامای معکوس به‌صورت $\delta_j^2 \sim \mathrm{IG}(a_{\delta_j^2}, b_{\delta_j^2})$ استفاده می‌کنیم. این انتخاب به توزیع‌های شرطی کامل از همان خانواده گامای معکوس منتهی می‌شود.

۴.۳ استنباط مدل

برای برازش مدل و استنباط آن باید توزیع پسین مدل محاسبه شود، که به‌صورت زیر نمادگذاری می‌شود:

$$\pi(\{\boldsymbol{\beta_j}\}, \{\delta_j\}, \sigma^2 | \boldsymbol{y}, X)$$

که در آن σ^2 واریانس جمله خطا است. توزیع پیشین برای σ^2 نیز یک توزیع گامای معکوس با پارامترهای $a_\circ$ و $b_\circ$، یعنی $\sigma^2 \sim \mathrm{IG}(a_\circ, b_\circ)$، است. در این پایان‌نامه مقادیر این پارامترها را برابر $a_\circ = ۰/۰۰۰۱$ و $b_\circ = ۱$ انتخاب کردیم.

---

[۲۶] Penalized Splines

[۲۷] Spline Difference Matrix

[۲۸] Adjacency Matrix

پسین مدل با توجه به پیشین‌های تعریف‌شده، صورت بسته‌ای ندارد. در نتیجه برای استنباط بر اساس آن باید از روش‌های تقریبی استفاده کرد. یکی از این روش‌ها نمونه‌گیری گیبز است که حالتی از الگوریتم MCMC است.

برای تولید نمونه از چگالی‌های شرطی کامل زیر استفاده می‌کنیم:

$$\sigma^{۲}|\cdot \sim \mathrm{IG}(a_{۰}+\frac{n}{۲}, b_{۰}+\sum_{i=۱}^{n}\epsilon_i^{۲}\omega_i)$$

$$\boldsymbol{\beta_j}|\boldsymbol{\theta_j} \sim \exp\left(-\frac{۱}{\sigma^{۲}}\left(B_j\boldsymbol{\beta_j}-\boldsymbol{\eta_{-j}}-\boldsymbol{y}\right)' W \left(B_j\boldsymbol{\beta_j}-\boldsymbol{\eta_{-j}}-\boldsymbol{y}\right) - \frac{۱}{۲\delta_j^{۲}}\boldsymbol{\beta_j}' K_j(\boldsymbol{\theta_j})\boldsymbol{\beta_j}\right)$$

$$\delta_j^{۲} \sim \begin{cases} \text{اثرات خطی} & (۱-I_j)\mathrm{IG}\left(a_{\delta_j^{۲}}+۰/۵, \nu_{۰}b_{\delta_j^{۲}}+۰/۵\beta_{j۲}\right) \\ & +I_j\mathrm{IG}\left(a_{\delta_j^{۲}}+۰/۵, b_{\delta_j^{۲}}+۰/۵\beta_{j۲}\right) \\ \text{اثرات غیرخطی} & \mathrm{IG}\left(a_{\delta_j^{۲}}+rk(K_j)/۲, b_{\delta_j^{۲}}+۰/۵\gamma^T K_j\gamma\right) \end{cases}$$

که در آن B_j ماتریس طرح مربوط به ضریب رگرسیونی jام مدل است، $\boldsymbol{y}$ بردار مشاهدات پاسخ، $\boldsymbol{\eta_{-j}} = \boldsymbol{\eta} - B_j\boldsymbol{\beta}_j$ پیش‌گوی کامل بدون j امین مؤلفه و $W = \mathrm{diag}(\omega(y_{۱},\eta_{۱}),\cdots,\omega(y_n,\eta_n))$ ماتریس قطری شامل وزن‌ها است.

چگالی‌های $\sigma^{۲}$ و $\delta_j^{۲}$ شناخته شده‌اند، اما چگالی $\boldsymbol{\beta_j}|\boldsymbol{\theta}_j$ شکل شناخته‌شده‌ای ندارد. بنابراین نمی‌توان به‌طور مستقیم از نمونه‌گیری گیبز استفاده کرد. نمونه‌گیری گیبز در این مرحله توسط یک الگوریتم متروپلیس‌ـهستینگز انجام می‌شود. برای اجرای الگوریتم متروپلیس‌ـهستینگز درون گیبز به توزیع پیشنهادی نیاز داریم. توزیع پیشنهادی معرفی‌شده در این بخش نرمال چندمتغیره است که میانگین و ماتریس کوواریانس آن به‌صورت زیر تعیین می‌شود:

$$\boldsymbol{\mu_j} = \Sigma_j B_j' W(\boldsymbol{y}-\boldsymbol{\eta_{-j}})$$
$$\Sigma_j = \left(B_j' W B_j + \sigma^{۲}K_j(\boldsymbol{\theta_j})\right)^{-۱}.$$

با توجه به این توزیع پیشنهادی به‌روزرسانی $\boldsymbol{\beta_j}$ ها به قاعده زیر منتهی می‌شود که این قاعده از روش کمترین توان‌های دوم وزنی تکراری (IWLS) به‌دست آمده است. به‌همین دلیل آن را پیشنهادی IWLS می‌نامند:

$$\hat{\boldsymbol{\beta}}_j^{[t+۱]} = (B_j' W B_j + \sigma^{۲}K_j(\theta_j))^{-۱} B_j' W(\boldsymbol{y}-\boldsymbol{\eta_{-j}})$$

از نمونه‌های به‌دست آمده برای استنباط درباره پارامترهای مدل استفاده می‌شود.

فصل ۴

ارزیابی مدل با مثال‌های شبیه‌سازی و واقعی

۱.۴ مقدمه

در این فصل با دو مثال شبیه‌سازی عملکرد مدل بیزی رگرسیون گشتاوری پیشنهادی را در انتخاب متغیر و مدل‌بندی ناپارامتری متغیرهای تبینی رگرسیونی، به‌طور هم‌زمان، مورد بررسی قرار داده‌ایم. در مثال اول، یک مطالعه جامع با انتخاب اثرات غیرخطی متفاوت انجام شده است و عملکرد برآورد ناپارامتری اثرات متغیرهای تبیینی در دیدگاه بیزی با دو روش برآورد بوستینگ[۱] و LAWS مقایسه شده است. در مثال دوم به بررسی عملکرد پیشین‌های میخ و تخته در انتخاب متغیرهای تبیینی مهم پرداخته‌ایم.

برای نمایش کاربست مدل پیشنهادی، دو مجموعه داده واقعی شامل سوءتغذیه کودکان در کشورهای تانزانیا و هند، نیز تحلیل شده‌اند که هر دو مجموعه از نوع داده‌های فضایی مشبکه‌ای هستند. در هر دو مثال، هدف بررسی تأثیر متغیرهای تبیینی شامل سن، قد، اطلاعات مادر و کودک و اطلاعات مکان سکونت کودکان بر سوءتغذیه آنان است. با توجه به هدف تحلیل این داده‌ها، تمرکز ما بر دم توزیع شرطی پاسخ به شرط متغیرهای تبیینی است که برای دستیابی به آن از یک مدل رگرسیون گشتاوری استفاده کرده‌ایم.

[۱] Boosting

با توجه به استفاده از روش برآورد بوستینگ برای مقایسه، ابتدا به‌طور خلاصه این روش را معرفی می‌کنیم. مطالب مربوط به این روش برگرفته از مقاله بوهلمن و هوتورن (۲۰۰۷) است.

۲.۴ روش برآورد بوستینگ

بوستینگ یک رهیافت قدرتمند برای برآورد پارامترهای یک مدل آماری است. این روش ابتدا در الگوریتم رده‌بندی آدابوست[۲] که توسط فروند و شاپیر (۱۹۹۶) معرفی شد، به‌کار گرفته شد. اما بعد از آن به برخی از مسائل مهم مانند یادگیری ماشین تعمیم یافت (بوهمن و هوتورن، ۲۰۰۷). روش بوستینگ در ابتدا به‌عنوان روش گروهی پیشنهاد شده است که به تولید پیش‌گوهای چندگانه[۳] متکی است. در اصطلاح آماری بوستینگ به‌عنوان مدل‌سازی جمعی مرحله‌ای[۴] معرفی می‌شود. دقت کنید کلمه ''جمعی'' به برازش مدل که دارای متغیرهای تبیینی جمعی است، دلالت نمی‌کند، بلکه به یک ترکیب جمعی برآوردگرهای ساده اشاره دارد. جنبه کاربردی فرآیند برآورد بوستینگ برای برازش مدل‌های آماری با استفاده از بسته نرم‌افزاری mboost (در نرم‌افزار R) تشریح شده است. این بسته توابعی که برای برازش مدل، در پیش‌گویی و انتخاب متغیر استفاده می‌شود را فراهم کرده است.

در این بخش، الگوریتم بوستینگ را در حالتی که با داده‌های فضایی مواجه‌ایم، بیان می‌کنیم. برای تشریح جزییات الگوریتم، ابتدا بردار پاسخ در n موقعیت فضایی $\{s_1, \ldots, s_n\}$ را با $\boldsymbol{Y} = (Y(s_1), \ldots, Y(s_n))'$ نشان می‌دهیم. علاوه‌بر این، فرض کنید می‌خواهیم برآورد ماکسیمم درستنمایی پارامتر d بعدی $\boldsymbol{\theta}$ را به‌دست آوریم و لگاریتم تابع درستنمایی را با $\ell(\boldsymbol{\theta}, \boldsymbol{y})$ نشان می‌دهیم که در آن $\boldsymbol{y}$ مقدار مشاهده‌شده بردار پاسخ است. الگوریتم بوستینگ برای برآورد بردار پارامتر $\boldsymbol{\theta}$، به‌صورت یک روش مؤلفه به مؤلفه عمل می‌کند. به این معنی که تابع درستنمایی برحسب هر مؤلفه به‌طور جدا، ماکسیمم می‌شود. در پایان تمام مراحل برآوردهای نتیجه‌شده به‌عنوان برآوردهای درستنمایی ماکسیمم پارامترهای مدل نتیجه می‌شود. از این‌رو در هر تکرار، یک مؤلفه انتخاب می‌شود و مقدار جاری $\boldsymbol{\theta}$ با استفاده از تنها مؤلفه انتخاب‌شده به‌روزرسانی می‌شود، به‌گونه‌ای که مقادیر سایر مؤلفه‌های بردار $\boldsymbol{\theta}$ ثابت فرض می‌شوند. به‌طور دقیق‌تر، هر الگوریتم بوستینگ را می‌توان به سه مرحله زیر تقسیم کرد:

۱. نموهای ذاتی $\delta_j^{(t)}$، $j = 1, \cdots, d$، برای هر مؤلفه محاسبه می‌شوند، به‌طوری‌که t شماره تکرار را نشان می‌دهد. هر نمو ذاتی $\delta_j^{(t)}$ بر مبنای در نظر گرفتن $\ell(\boldsymbol{\theta}_j, \boldsymbol{y})$ به‌عنوان تابعی فقط از θ_j با ثابت در نظر گرفتن تمام مؤلفه‌های دیگر در مقادیر قبلی خود، تعیین می‌شود؛ یعنی به ازای تمام $l \neq j$، $\theta_l^{(t)} = \theta_l^{(t-1)}$. برای به‌دست آوردن این برآورد

[۲] Adaboost Algorithm
[۳] Multivariate Predictors
[۴] Stagewise Additive Modeling

درستنمایی ماکسیمم، می‌توان از روش‌های صعودی‌سازی گرادیان یا روش نیوتون–رافسون استفاده کرد.

۲. فرض کنید $\theta^{(t:j)}$ مقداری از $\boldsymbol{\theta}$ باشد که $\theta_j^{(t)} = \theta_j^{(t-1)} + \delta_j^{(t)}$ و به‌ازای $l \neq j$، $\theta_l^{(t)} = \theta_l^{(t-1)}$. در این صورت با در نظر گرفتن هر مؤلفه به‌عنوان پارامتر و ثابت نگه داشتن بقیه مؤلفه‌ها در مقادیر قبلی خود، رده‌ای از بردارهای پارامتری به‌دست می‌آید. حال از بین بردارهای $\theta^{(t:1)}, \ldots, \theta^{(t:d)}$، برداری انتخاب می‌شود که به‌ازای آن تابع درستنمایی بیشترین مقدار را داشته باشد. به‌عبارت دیگر مؤلفه s_t را به‌گونه‌ای اختیار کند که

$$s_t = \arg\max_j \ell(\theta^{(t:j)}, \boldsymbol{y}) \quad j = 1, \cdots, d.$$

۳. بردار $\boldsymbol{\theta}$ با قرار دادن $\theta^{(t)} = \theta^{(t:s_t)}$ به‌روزرسانی می‌شود.

اگر این سه مرحله به‌طور متوالی تکرار شوند، آن‌گاه $\boldsymbol{\theta}^{(t)}$ (تحت شرایط نظم) به برآوردگر درستنمایی ماکسیمم همگرا می‌شود. تکرار الگوریتم تا رسیدن به زمان همگرایی ادامه می‌یابد. ویژگی‌های مناسب روش بوستینگ در برازش مدل‌های به‌ویژه پیچیده، برآورد مؤلفه به مؤلفه پارامترها است، زیرا به‌روزرسانی مدل با استفاده از یک مؤلفه مجزا از لحاظ عددی ساده‌تر از تلاش برای برازش همزمان تمامی مؤلفه‌ها است.

۳.۴ مثال اول: مدل‌بندی اثرات غیرخطی

در این مطالعه شبیه‌سازی، چهار سناریوی مختلف را برای تولید متغیر پاسخ در نظر گرفته‌ایم. این سناریوها شامل توابع غیرخطی با ضوابط متفاوت برای تولید داده‌ها، هستند که به‌صورت زیر فهرست می‌شوند:

$$\mathrm{M}_1 : y = 5 \exp(-0/5 z^2) + \epsilon$$
$$\mathrm{M}_2 : y = 5 \sin(2z) + \epsilon$$
$$\mathrm{M}_3 : y = 5 \exp(-0/5 z^2) + 5 \sin(2z) + \epsilon$$
$$\mathrm{M}_4 : y = \sin(2(4z - 2)) + 2 \exp(-16^2 (z - 0/5)^2) + \epsilon$$

به‌طوری که برای سناریوهای M_1، M_2 و M_3، متغیر تبیینی z دارای توزیع $\mathrm{U}(0,3)$ و برای M_4، دارای توزیع $\mathrm{U}(0,1)$ است. توابع $f(z)$ در این چهار سناریو، طوری در نظر گرفته شده‌اند که انواع مختلفی از توابع غیرخطی و با ناهمواری‌های مختلف را نتیجه دهند.

برای جمله خطا نیز دو توزیع متفاوت در نظر گرفتیم:

$$\begin{aligned} \mathrm{E}_۱ : \mathrm{M}_{۱,۲,۳} \quad & \mathrm{N}(۰, ۰/۵z^۲) \\ \mathrm{M}_۴ \quad & \mathrm{N}(۰, (۰/۲ + |z - ۰/۵|)^۲) \\ \mathrm{E}_۲ : \mathrm{M}_{۱,۲,۳} \quad & \mathrm{Exp}\left(\frac{۱}{z}\right) \\ \mathrm{M}_۴ \quad & \mathrm{Exp}\left(\frac{۱}{۰/۲ + |z - ۰/۵|}\right). \end{aligned}$$

برای هر یک از ترکیب‌های توزیع خطا و تابع $f(z)$، ۱۰۰ مجموعه داده با دو اندازه نمونه $n = ۱۰۰,۵۰۰$ تولید کردیم. همچنین مدل‌های رگرسیون گشتاوری با مرتبه‌های

$$\tau \in \{۰/۰۲, ۰/۰۵, ۰/۱, ۰/۲, ۰/۵, ۰/۸, ۰/۹, ۰/۹۵, ۰/۹۸\}$$

را برای همه داده‌های تولیدشده، تحت سناریوهای مختلف شبیه‌سازی، برازش دادیم.

برای برآورد بیزی مدل‌ها، حجم نمونه مونت کارلویی را برابر ۳۵۰۰۰ با یک دوره سوزاندن برابر ۵۰۰۰ انتخاب کردیم. همچنین به منظور کاهش وابستگی نمونه‌های تولیدشده، طول گام را ۳۰ در نظر گرفتیم. با این تنظیمات، همه نتایج گزارش‌شده برای روش بیزی بر اساس یک نمونه نهایی به حجم ۱۰۰۰ به‌دست آمده‌اند. در همه مدل‌ها، اثر غیرخطی متغیر تبیینی z با استفاده از B-اسپلاین‌های مکعبی و با ۲۲ گره داخلی مدل‌بندی شده است. دو روش برازش بوستینگ و LAWS در بسته نرم‌افزاری expecterg قابل پیاده‌سازی هستند (سبوتکا و همکاران، ۲۰۱۴).

برای ارزیابی عملکرد رهیافت‌های برازش مدل (شامل بیزی، بوستینگ و LAWS)، دو هدف را مد نظر قرار دادیم:

۱. دقت بازیابی تابع غیرخطی $f(z)$. برای سنجش کیفیت تابع برآوردشده f، از معیار ریشه میانگین توان‌های دوم خطا برای تابع برآوردشده استفاده کردیم که به‌صورت زیر تعریف می‌شود:

$$\mathrm{RMSE}(f_\tau) = \sqrt{(f_\tau - \hat{f}_\tau)'(f_\tau - \hat{f}_\tau)}$$

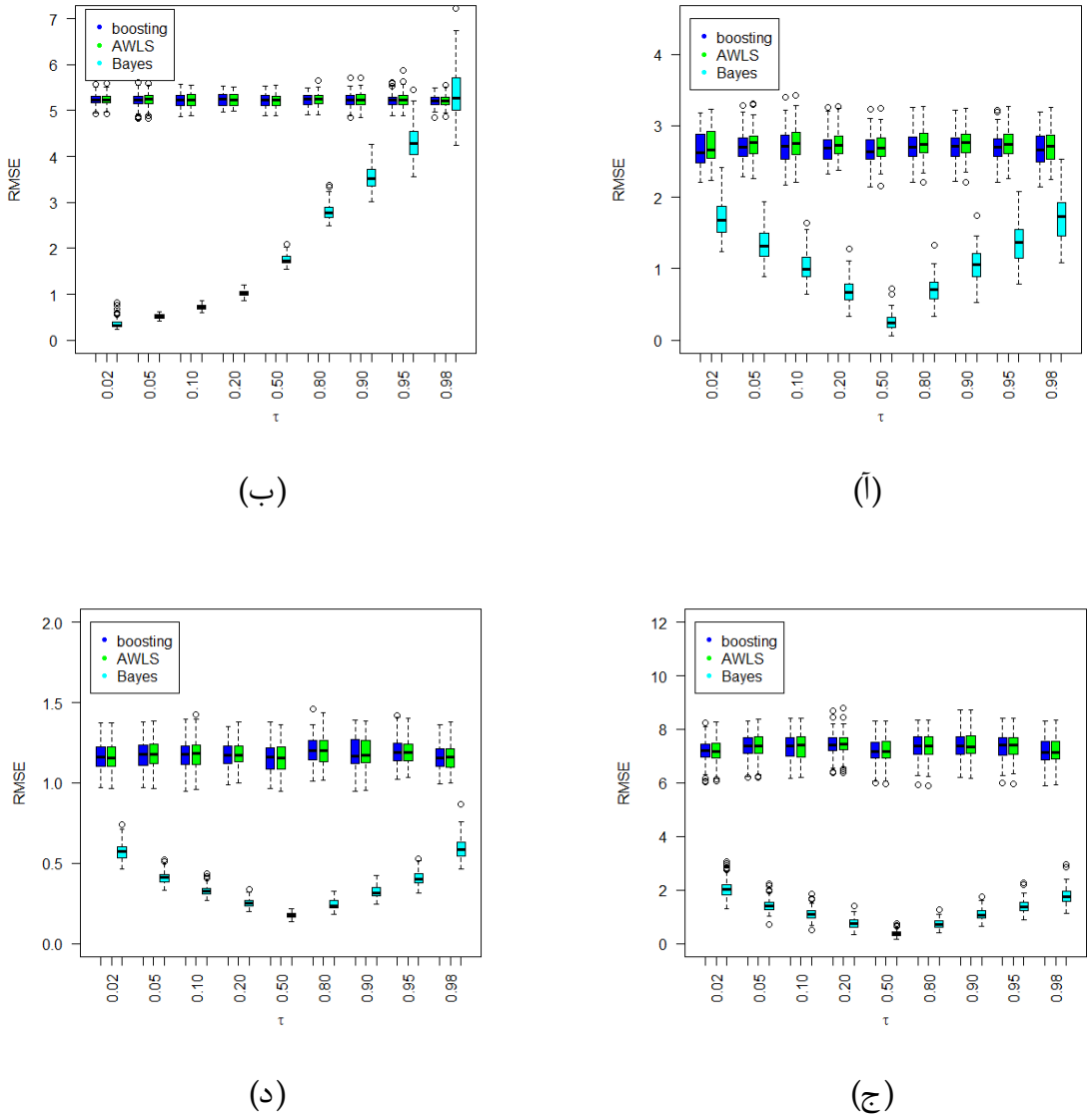
که در آن $\hat{f}_\tau$ تابع برازش‌شده برای رگرسیون گشتاوری با مرتبه τ است.

۲. مقایسه فواصل اطمینان و اعتبار حاصل برای $f(z)$. برای این کار از معیار ماکسیمم طول فواصل اعتبار (برای روش بیزی) و اطمینان (برای روش LAWS) نتیجه‌شده برای $f(z)$ استفاده کردیم که به‌صورت زیر تعریف می‌شود:

$$\max \widehat{\mathrm{Width}}(\mathrm{CI}(f_\tau(z_i))) = \max_k \left(\hat{f}^{[k]}_{\tau,U}(z_i) - \hat{f}^{[k]}_{\tau,L}(z_i)\right)$$

که در آن f_U و f_L به‌ترتیب حد بالایی و پایینی فاصله را نشان می‌دهند، k شمارنده تکرار است و مقادیر z_i شامل ۱۰۰ نقطه در فاصله $(۰,۳)$ برای سه سناریوی اول و فاصله

$(۰,۱)$ برای سناریوی چهارم است که به‌صورت یک دنباله با فاصله‌های یکسان انتخاب شده‌اند. به‌طور مشابه میانگین و می‌نیمم طول فاصله‌ها را بین تمام ۱۰۰ تکرار نیز محاسبه کردیم. لازم به ذکر است که روش بوستینگ در این هدف مورد توجه نبوده است، زیرا محاسبه فاصله اطمینان در این روش از نظر محاسباتی بسیار زمان‌بر است. برای روش LAWS نیز از فاصله‌های اطمینان مجانبی استفاده شده است.



شکل ۱.۴: نمودارهای جعبه‌ای مقادیر RMSE برای سه روش بیزی، بوستینگ و LAWS برای مرتبه‌های مختلف τ. شکل آ تا د به ترتیب مربوط به مدل‌های M_1 تا M_4 هستند.

۱.۳.۴ نتایج شبیه‌سازی و تحلیل آن‌ها

شکل ۱.۴، نمودارهای جعبه‌ای مقادیر RMSE را برای هر سه روش برازش مدل‌ها و مرتبه‌های مختلف τ، زمانی که $n = ۱۰۰$ است، نشان می‌دهد. کاملا واضح است که روش بیزی در بازیابی تابع غیرخطی متغیر تبیینی، به ازای هر مرتبه τ، به‌طور یکنواخت خیلی بهتر از دو روش رقیب خود عمل کرده است. فقط یک استثنا در مدل $M_۲$ برای $\tau = ۰/۹۸$ دیده می‌شود که نسبت به دو رقیب خود کمی ضعیف‌تر ظاهر شده است. از طرفی دو روش LAWS و بوستینگ تقریبا عملکرد یکسانی در تمام مدل‌ها دارند. همچنین در اکثر مدل‌ها، دقت برآورد اثر غیرخطی برای مرتبه‌های مرکزی بالاتر از مرتبه‌های دمی (به‌ویژه دم بالا) است. نتیجه مشابهی برای حالتی که $n = ۵۰۰$ باشد، به‌دست می‌آید که از گزارش آن خودداری کرده‌ایم.

شکل‌های ۲.۴ و ۳.۴ معیارهای مربوط به طول فاصله‌های اعتبار و اطمینان را برای چهار مدل مختلف و چند مرتبه τ منتخب و برای $n = ۱۰۰$ نشان می‌دهند. با توجه به این شکل‌ها کاملا واضح است که از منظر دو معیار ماکسیمم و میانگین پهنای فاصله، روش بیزی کاملا برتر از روش LAWS است؛ ولی بر حسب معیار می‌نیمم پهنا، عملکرد دو روش تقریبا یکسان است و حتی در مواردی روش LAWS بهتر عمل کرده است.

بنابراین، در مجموع، می‌توان روش برازش بیزی مدل رگرسیون گشتاوری و چارچوب مدل‌بندی اثرات غیرخطی با اسپلاین‌های جریمه‌ای را نسبت به روش‌های رقیب موجود، تایید کرد و برتر دانست.

۴.۴ مثال دوم: انتخاب اثرات ثابت با پیشین میخ و تخته

این شبیه‌سازی نزدیک به مدلی است که در مقاله اصلی لاسو (تیبشیرانی، ۱۹۹۴) ارایه شده است. دو سناریوی مختلف را در نظر گرفتیم. در سناریوی اول، داده‌ها را از مدل رگرسیون خطی ساده زیر تولید کردیم:

$$M_۵ : y = ۳ + ۳x_۱ + ۱/۵x_۲ + ۰x_۳ + ۰x_۴ + ۲x_۵ + ۰x_۷ + ۰x_۸ + \epsilon$$

که در آن تنها سه متغیر تبیینی در مدل اثر معنی‌دار دارند و بقیه صفر هستند. در این مدل، بردار متغیرهای تبیینی $\boldsymbol{x}$ با بعد ۸ را از یک توزیع نرمال چندمتغیره با بردار میانگین صفر، و ماتریس کوواریانس Σ تولید کردیم، که در آن درایه‌های ماتریس کوواریانس طوری تعریف شدند که انحراف معیار همه متغیرها برابر ۱ و همبستگی زوجی هر دو متغیر برابر

$$\text{Corr}(x_k, x_l) = ۰/۵^{|k-l|}$$

باشد. همچنین جمله خطا را از توزیع $\epsilon \sim \mathrm{N}(۰, \sqrt{۳})$ تولید کردیم. حجم نمونه داده‌ها را نیز سه اندازه مختلف $n = ۲۰, ۵۰, ۱۰۰$ در نظر گرفتیم و شبیه‌سازی را ۱۰۰ بار تکرار کردیم.

برای ارزیابی توانایی روش پیشنهادی در انتخاب متغیرهای تبیینی با اثرات ثابت، در سناریوی دوم تابع $f(z)$ مدل $\mathrm{M}_۱$ را به مدل $\mathrm{M}_۵$ اضافه کردیم تا ترکیبی از اثرات خطی و غیرخطی داشته باشیم. برای مدل رگرسیونی نیز مرتبه‌های گشتاوری را برابر

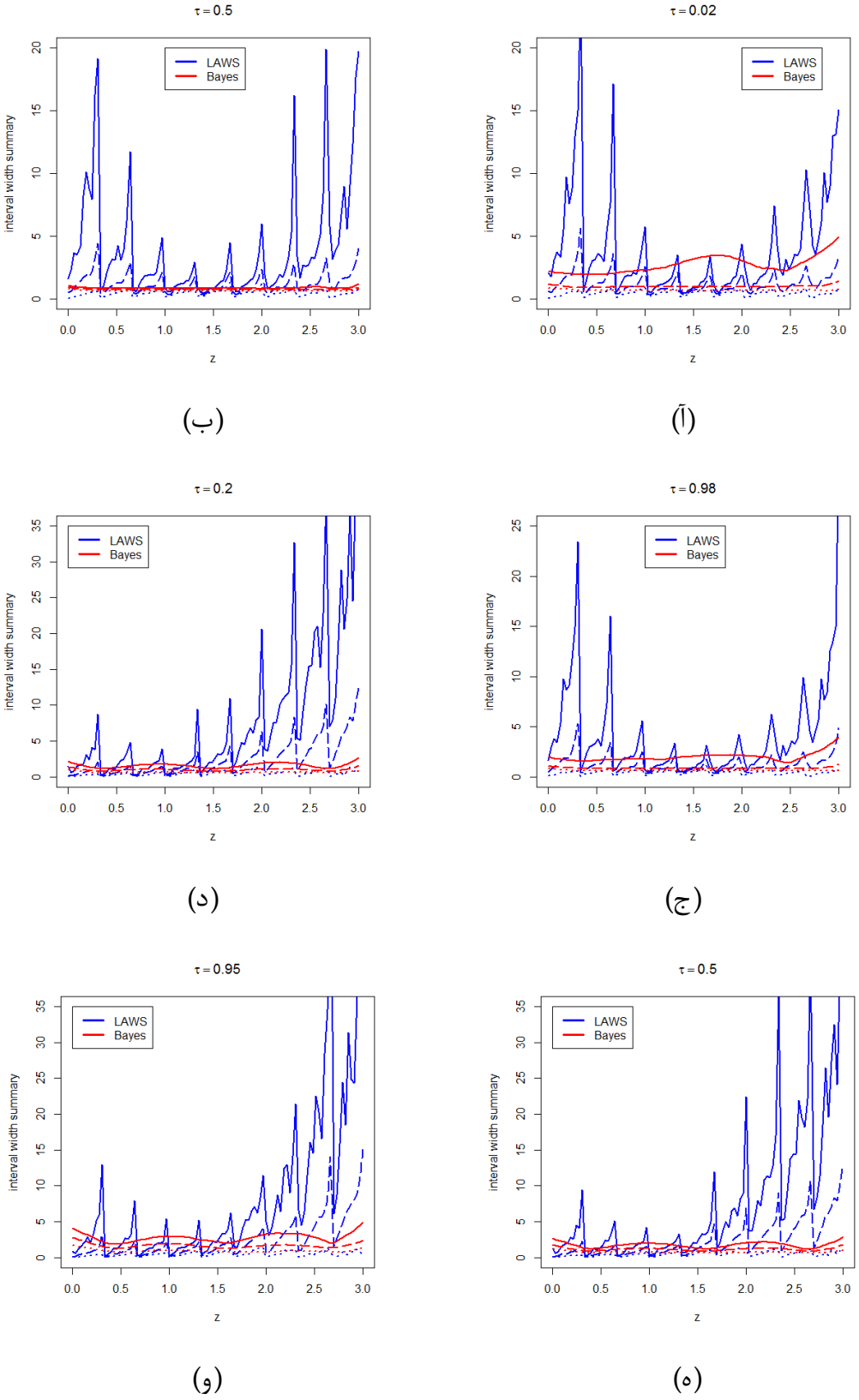
$$\tau \in \{۰/۰۵, ۰/۲, ۰/۵, ۰/۸, ۰/۹۵\}$$

در نظر گرفتیم. حجم نمونه مونت کارلویی و سایر تنظیمات برای روش بیزی، مشابه مثال اول انتخاب شدند.

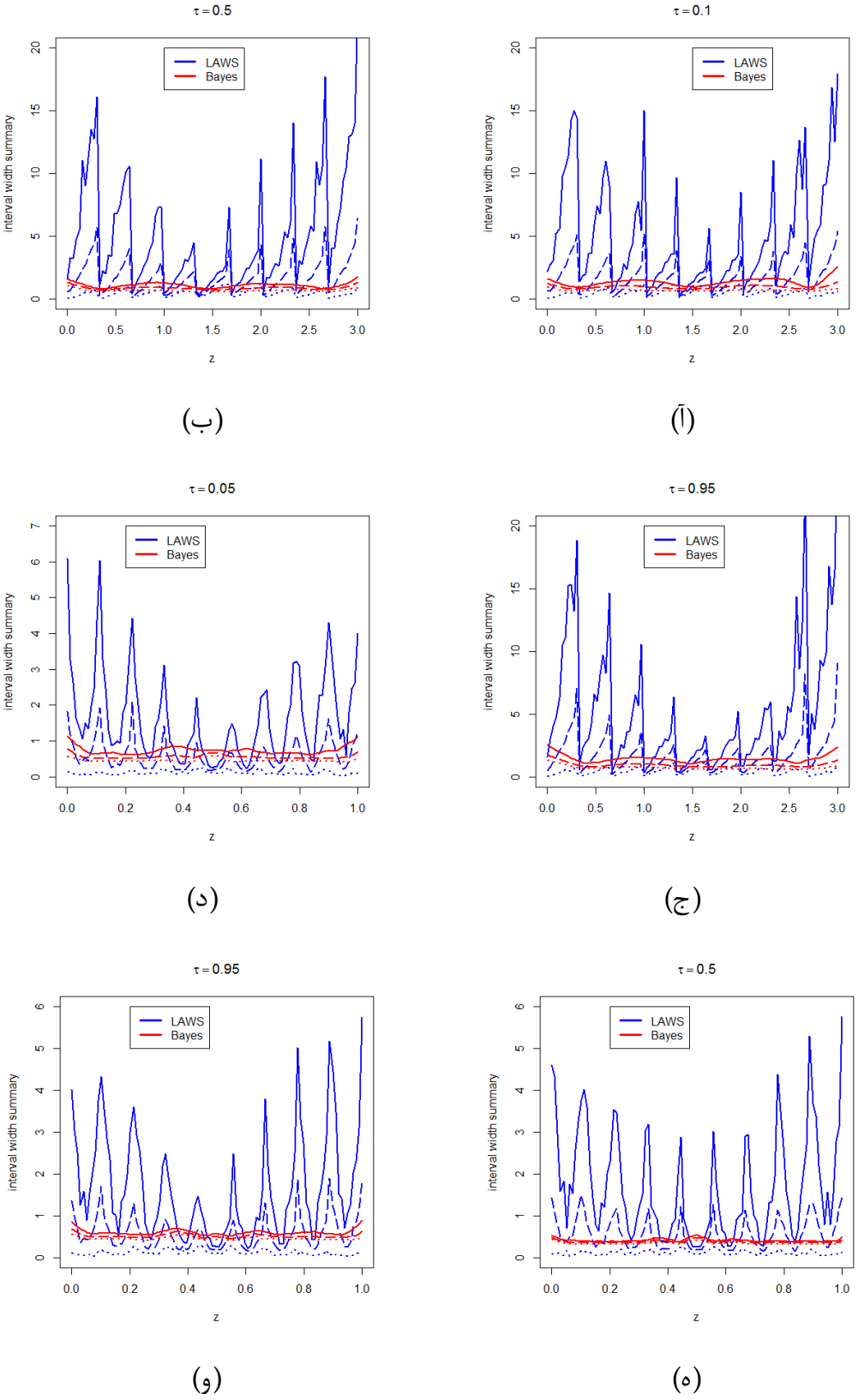
۱.۴.۴ نتایج شبیه‌سازی و تحلیل آن‌ها

برای نمایش نتایج، دو معیار برآورد پارامترهای رگرسیونی و احتمال ورود آن‌ها در مدل (انتخاب متغیر) را مد نظر قرار دادیم. شکل‌های ۴.۴، ۵.۴، و ۶.۴ نمودارهای جعبه‌ای احتمال انتخاب شدن متغیرها را برای هر دو سناریو در ۱۰۰ تکرار نشان می‌دهند و شکل‌های ۷.۴ تا ۹.۴ برآوردهای پارامترهای ضرایب ثابت را در ۱۰۰ تکرار برای هر دو سناریوی مطرح‌شده نمایش می‌دهند.

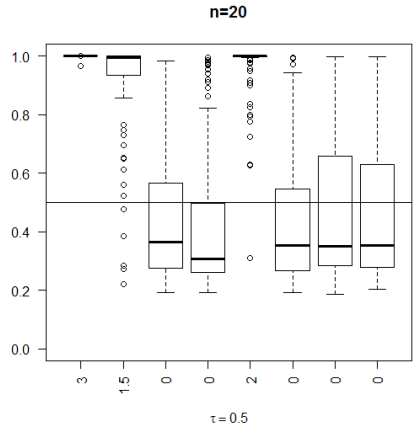
در همه موارد، نتایج نشان می‌دهند که احتمال انتخاب شدن متغیرهای معنی‌دار و مهم $x_۱$، $x_۲$ و $x_۵$ در هر دو سناریوی مطرح‌شده بسیار بالا و حتی نزدیک به ۱ است. در مقابل، برای سایر متغیرهای بی‌اثر، این احتمال کمتر از ۰/۵ است. همچنین نمودارهای برآورد پارامترها حاکی از اریبی کوچک در برآورد مقادیر واقعی ضرایب $\beta_۱$، $\beta_۲$ و $\beta_۵$ است و سایر پارامترهای بی‌اهمیت نزدیک به صفر برآورد شده‌اند. به‌طور کلی، پیشین میخ و تخته عملکرد مناسبی در انتخاب متغیر به ازای مرتبه‌های مختلف گشتاوری در رگرسیون گشتاوری دارد.



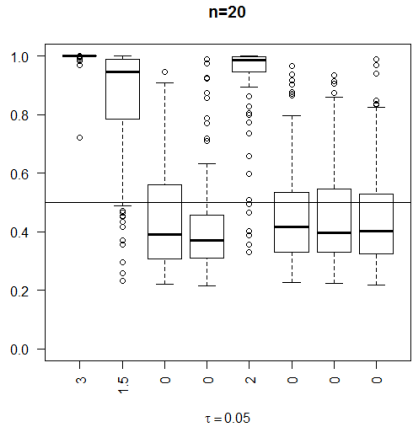
شکل ۲.۴: مقادیر معیارهای پهنای فاصله‌های اعتبار (بیزی) و اطمینان (کلاسیک) با روش‌های برازش پیشنهادی. شکل‌های آ، ب و ج مربوط به مدل M_1 و شکل‌های د، ه و و مربوط به مدل M_2 هستند. هم‌چنین منحنی‌های ممتد مربوط به معیار ماکسیمم، نقطه‌چین مربوط به معیار می‌نیمم و خط‌چین مربوط به معیار میانگین پهنا هستند.



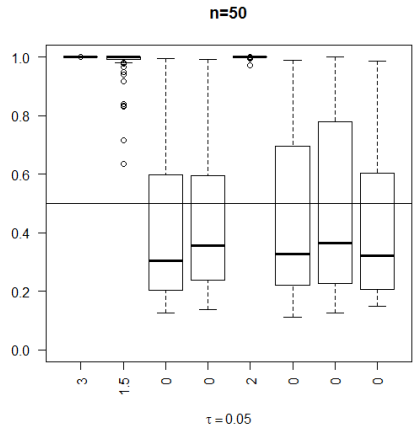
شکل ۳.۴: مقادیر معیارهای پهنای فاصله‌های اعتبار (بیزی) و اطمینان (کلاسیک) با روش‌های برازش پیشنهادی. شکل‌های آ، ب و ج مربوط به مدل M_3 و شکل‌های د، ه و و مربوط به مدل M_4 هستند. هم‌چنین منحنی‌های ممتد مربوط به معیار ماکسیمم، نقطه‌چین مربوط به معیار می‌نیمم و خط‌چین مربوط به معیار میانگین پهنا هستند.



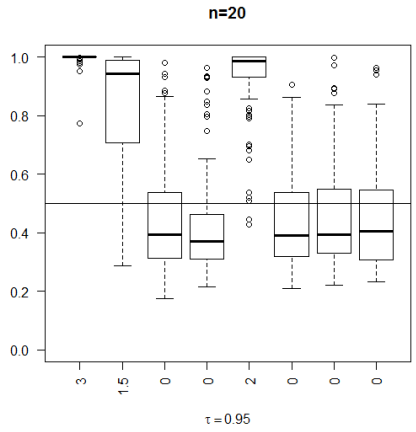
(ب)



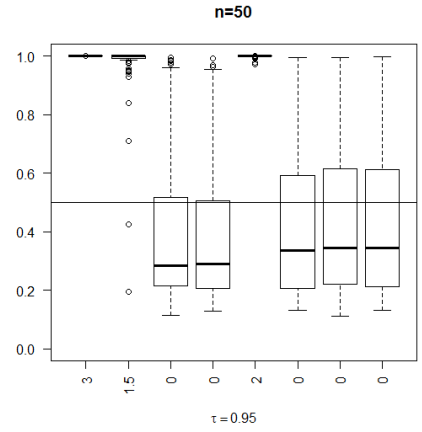
(آ)



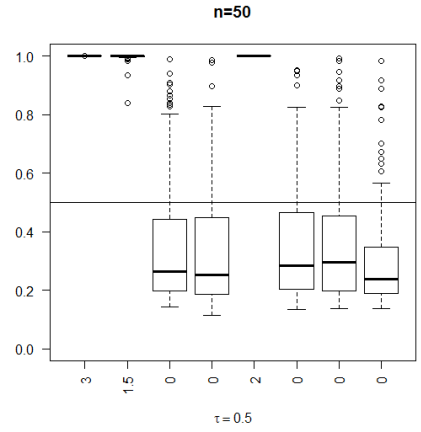
(د)



(ج)

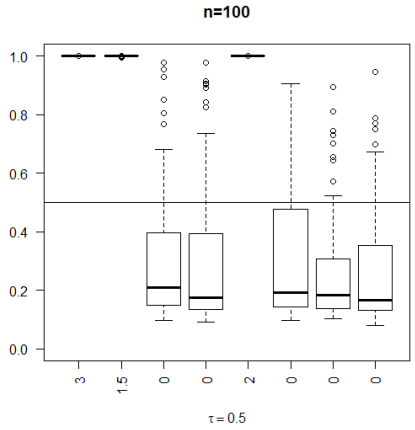


(و)

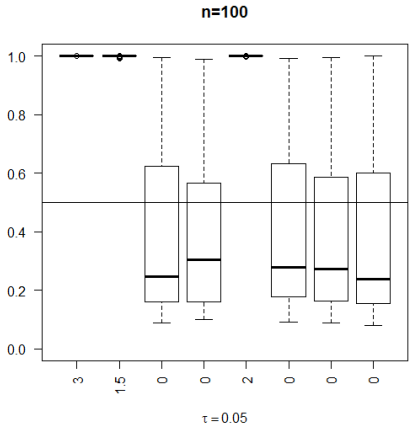


(ه)

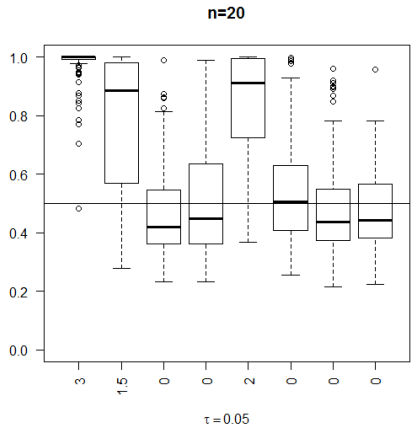
شکل ۴.۴: نمودارهای جعبه‌ای احتمال انتخاب متغیرهای تبیینی با اثرات ثابت برای مرتبه‌های $\tau = ۰/۰۵, ۰/۵, ۰/۹۵$ در مدل $M_۵$. شکل‌های آ، ب و ج مربوط به حجم نمونه $n = ۲۰$ و شکل‌های د، ه و و مربوط به حجم نمونه $n = ۵۰$ هستند.



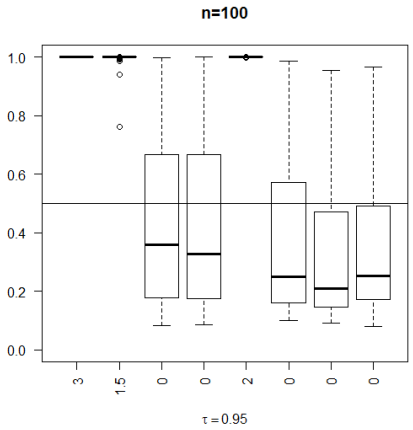
(ب)



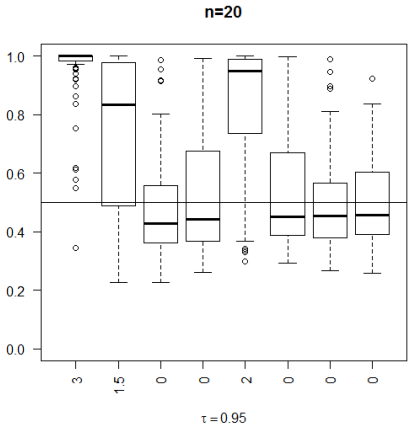
(آ)



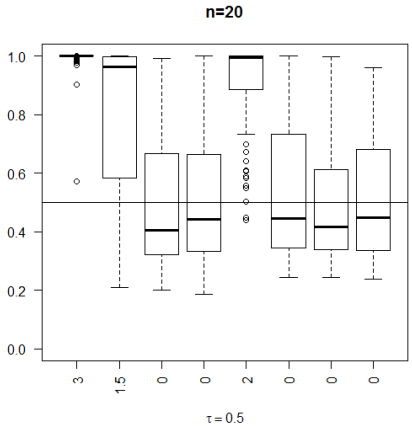
(د)



(ج)

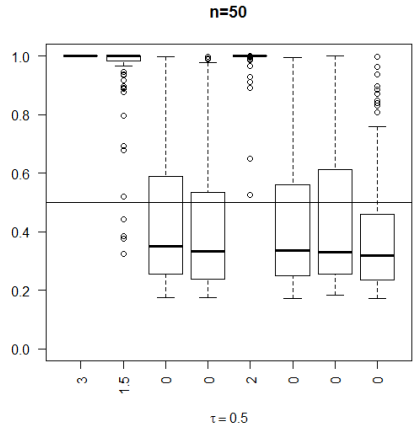


(و)

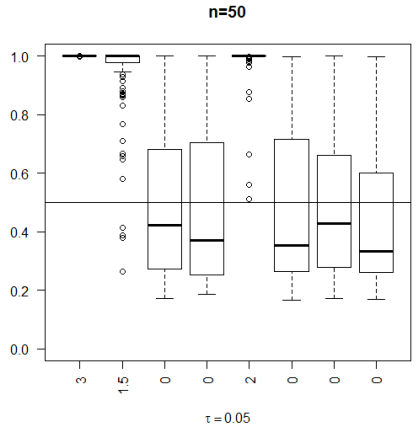


(ه)

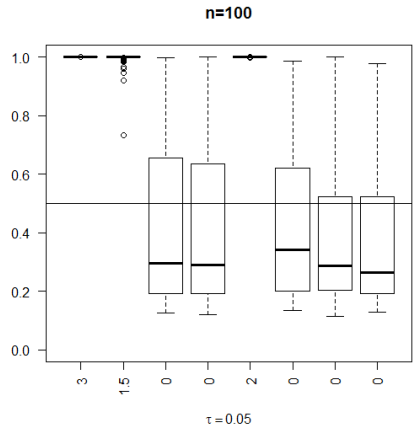
شکل ۵.۴: نمودارهای جعبه‌ای احتمال انتخاب متغیرهای تبیینی با اثرات ثابت برای مرتبه‌های $\tau = ۰/۰۵, ۰/۵, ۰/۹۵$ در مدل ترکیبی $M_۵$ و $M_۱$. شکل‌های آ، ب و ج مربوط به حجم نمونه $n = ۱۰۰$ و شکل‌های د، ه و و مربوط به حجم نمونه $n = ۲۰$ هستند.



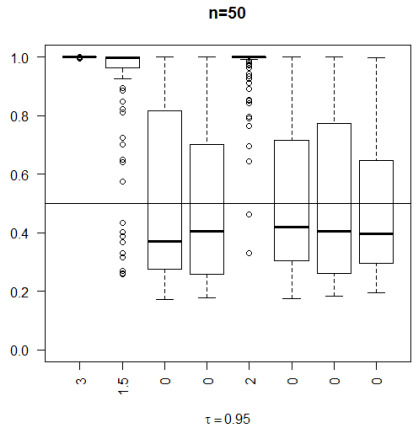
(ب)



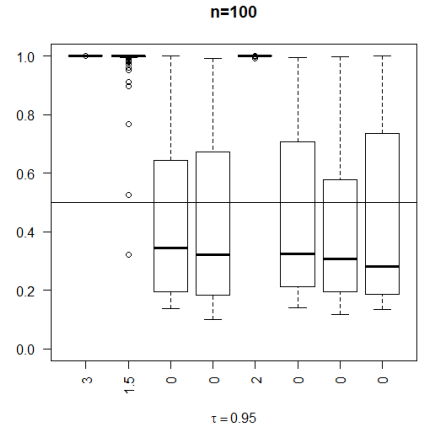
(آ)



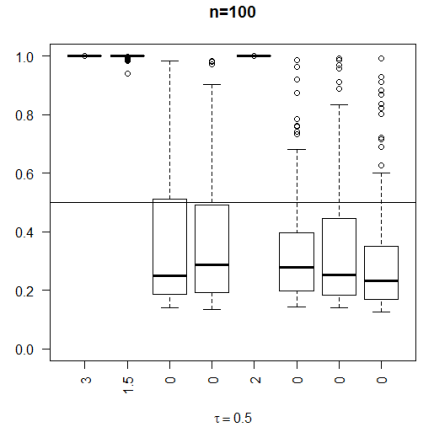
(د)



(ج)

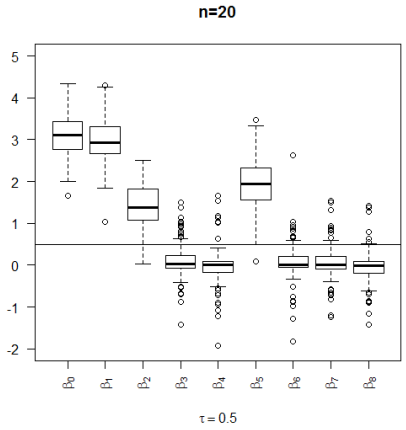


(و)

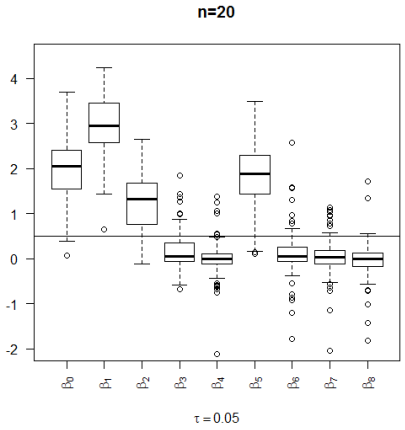


(ه)

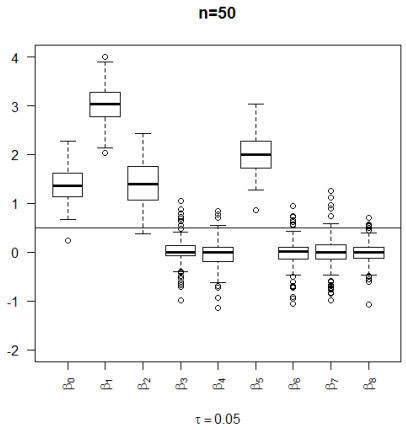
شکل ۶.۴: نمودارهای جعبه‌ای احتمال انتخاب متغیرهای تبیینی با اثرات ثابت برای مرتبه‌های $\tau = ۰/۰۵, ۰/۵, ۰/۹۵$ در مدل ترکیبی $M_۵$ و $M_۱$. شکل‌های آ، ب و ج مربوط به حجم نمونه $n = ۵۰$ و شکل‌های د، ه و و مربوط به حجم نمونه $n = ۱۰۰$ هستند.



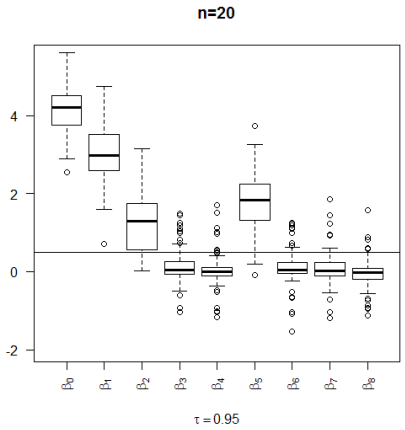
(ب)



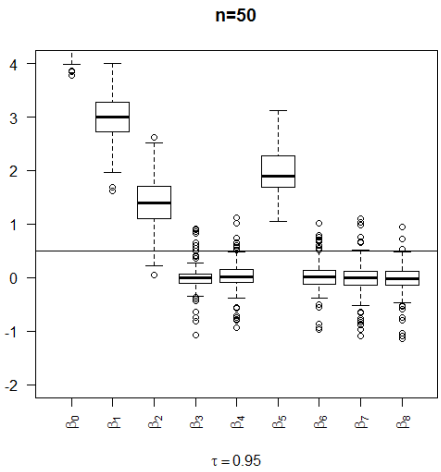
(آ)



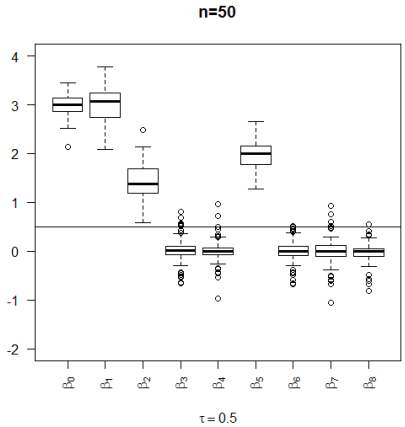
(د)



(ج)

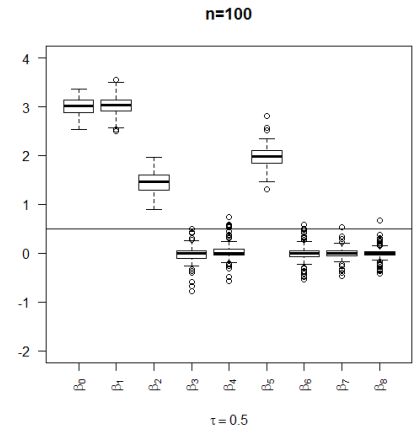


(و)

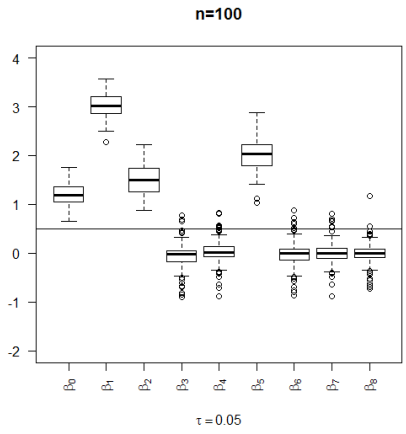


(ه)

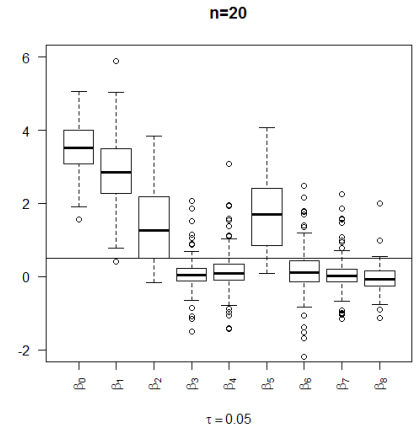
شکل ۷.۴: نمودارهای جعبه‌ای برآورد ضرایب رگرسیونی متغیرهای تبیینی با اثرات ثابت برای مرتبه‌های $\tau = ۰/۰۵, ۰/۵, ۰/۹۵$ در مدل $M_۵$. شکل‌های آ، ب و ج مربوط به حجم نمونه $n = ۲۰$ و شکل‌های د، ه و و مربوط به حجم نمونه $n = ۵۰$ هستند.



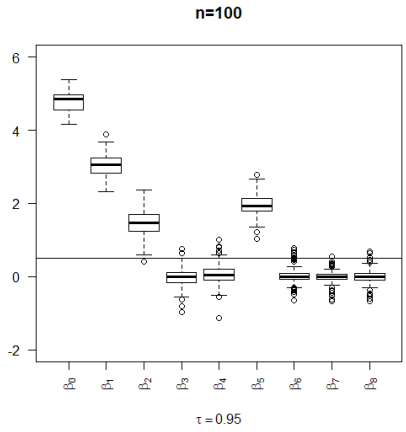
(ب)



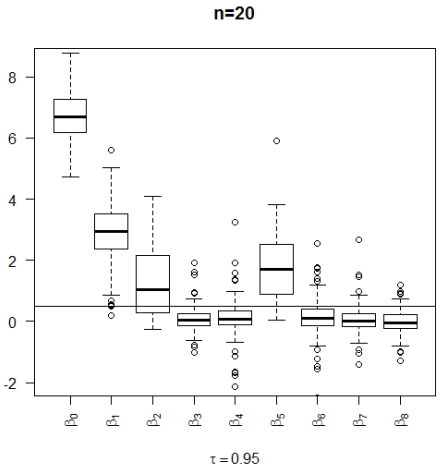
(آ)



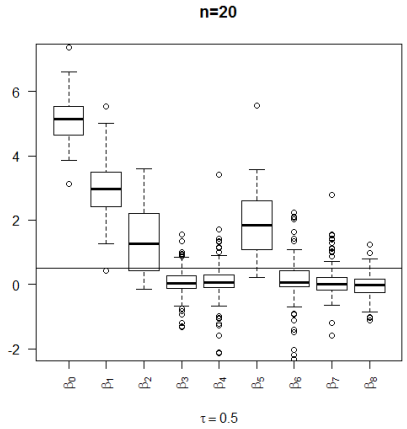
(د)



(ج)

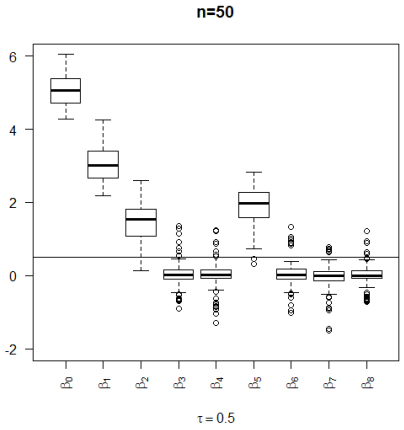


(و)

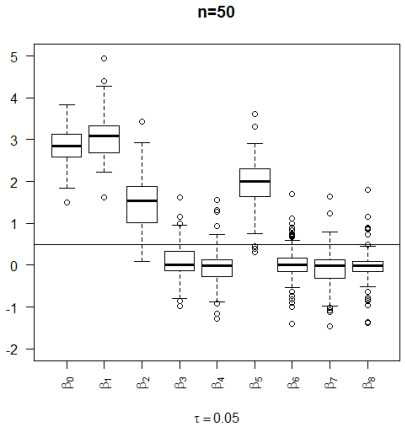


(ه)

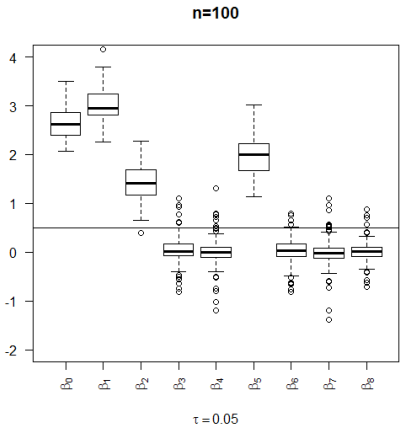
شکل ۸.۴: نمودارهای جعبه‌ای برآورد ضرایب رگرسیونی متغیرهای تبیینی با اثرات ثابت برای مرتبه‌های $\tau = ۰/۰۵, ۰/۵, ۰/۹۵$ در مدل ترکیبی $M_۵$ و $M_۱$. شکل‌های آ، ب و ج مربوط به حجم نمونه $n = ۱۰۰$ و شکل‌های د، ه و و مربوط به حجم نمونه $n = ۲۰$ هستند.



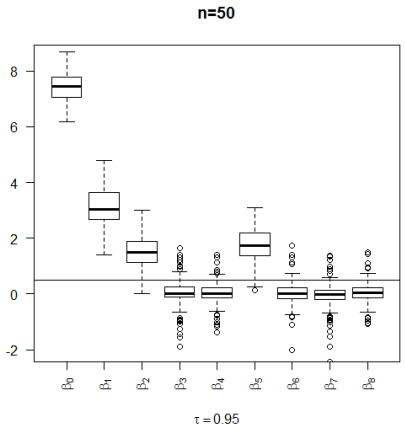
(ب)



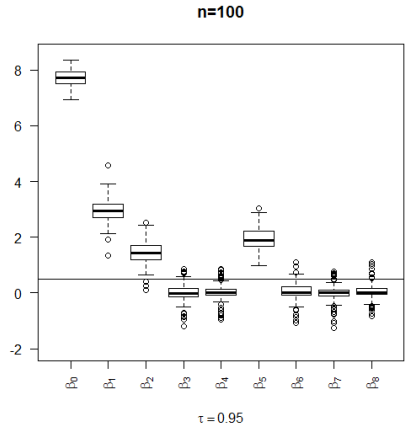
(آ)



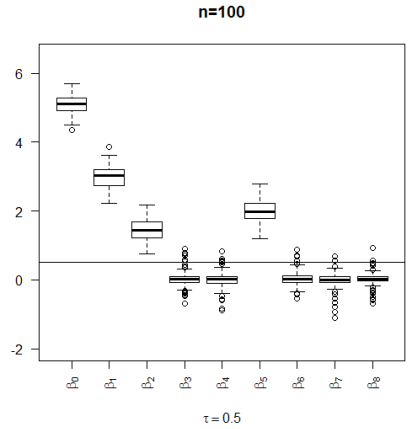
(د)



(ج)

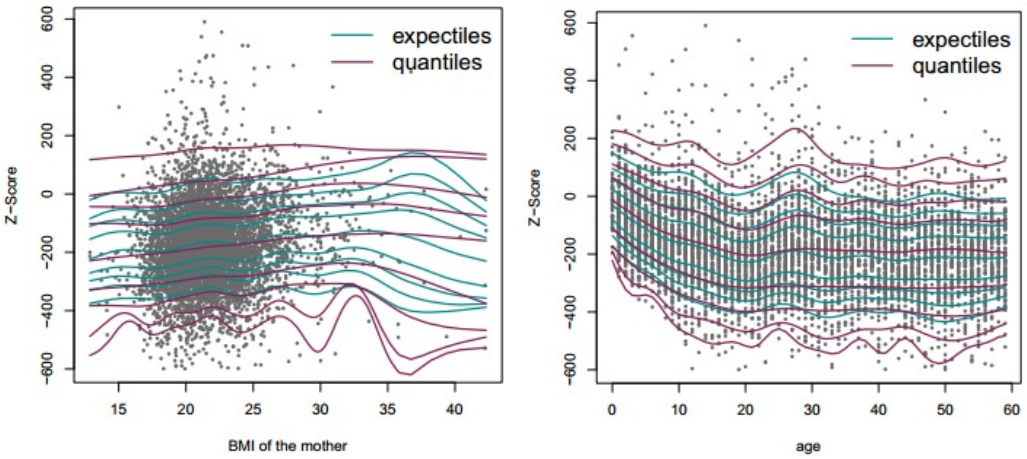


(و)



(ه)

شکل ۹.۴: نمودارهای جعبه‌ای برآورد ضرایب رگرسیونی متغیرهای تبیینی با اثرات ثابت برای مرتبه‌های $\tau = ۰/۰۵, ۰/۵, ۰/۹۵$ در مدل ترکیبی M_5 و M_1. شکل‌های آ، ب و ج مربوط به حجم نمونه $n = ۵۰$ و شکل‌های د، ه و و مربوط به حجم نمونه $n = ۱۰۰$ هستند.



(آ) اثر غیرخطی سن مادر (ب) اثر غیر خطی BMI مادر

شکل ۴.۱۰: اثرات غیرخطی متغیرهای تبیینی پیوسته مجموعه داده تانزانیا

۵.۴ مثال واقعی: سوءتغذیه کودکان تانزانیا

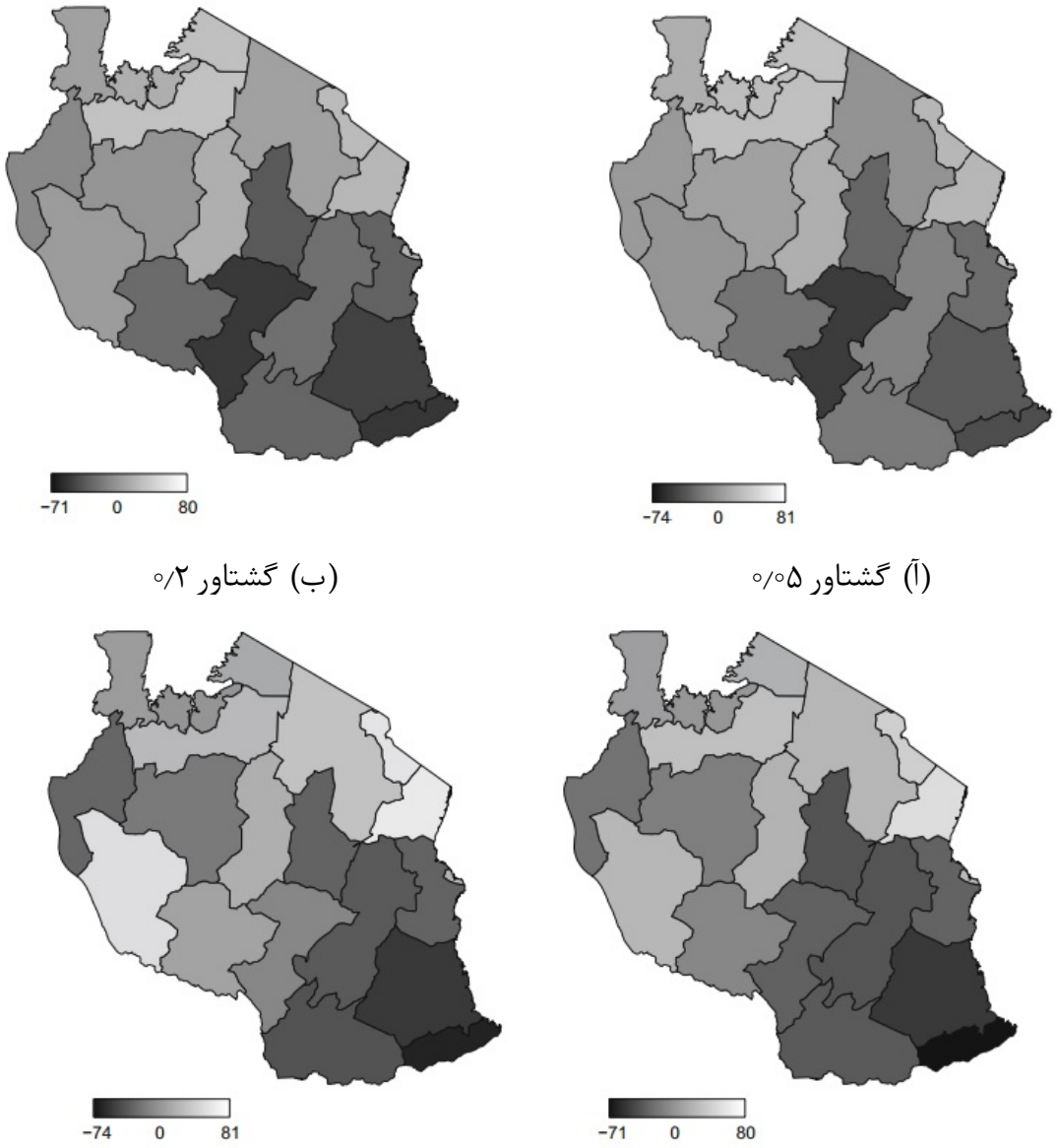
سوءتغذیه کودکان یکی از مهم‌ترین مشکلات در کشورهای در حال توسعه است. سازمان نظارت بر سلامت و جمعیت، آمارگیری نفوس و بهداشت، برنامه‌ریزی خانواده، سلامت مادران و کودکان، مالاریا و تغذیه و موارد دیگر را انجام می‌دهد. داده‌های این برنامه برای ۷۵ کشور به منظور اهداف تحقیقاتی در سایت `www.measuredhs.com` در دسترس هستند.

سوءتغذیه کودکان تانزانیا در سال ۱۹۹۲ مورد بررسی قرار گرفته است. مجموعه داده این مثال شامل ۵۳۸۹ مشاهده است. متغیر پاسخ در این مساله، برای کودک iام، معیاری به نام z-score است که به‌صورت

$$\text{z} - \text{score} = \frac{(\text{AL}_\text{i} - \text{MAI})}{\sigma}$$

محاسبه می‌شود، که در آن AI_i شاخصی است که بر اساس قد و سن کودک محاسبه می‌شود، MAI و σ نیز به‌ترتیب میانه و انحراف استاندارد شاخص مورد نظر در جامعه کودکان سالم قابل مقایسه، هستند. برای این مجموعه داده، میانه جامعه سالم برابر با ۱۷۷/۹− و انحراف استاندارد آن برابر ۱۴۲/۲۴ در نظر گرفته شدند. با این دو مقدار، ٪۹۰ کودکان شاخص تغذیه‌ای کمتر از صفر و ٪۹۵ آن‌ها شاخصی بین ۴۵۵/۰۰− و ۱۰۸/۰۵ دارند.

متغیرهای تبیینی در این مجموعه نیز شامل شاخص توده بدن (BMI)، جنسیت (sex)، و سن کودکان (age)، و اطلاعاتی در مورد مادر مانند وزن مادر در هنگام تولد کودک، سابقه آموزشی او (Edu) در چهار مقطع و اشتغال فعلی او (work: استخدام یا بی‌کار) و اطلاعات مربوط به محل سکونت است. اطلاعات مربوط به سکونت شامل دو قسمت می‌شوند: وضعیت



(آ) گشتاور ۰٫۰۵ (ب) گشتاور ۰٫۲

(ج) گشتاور ۰٫۸ (د) گشتاور ۰٫۹۵

شکل ۱۱.۴: برآورد اثر فضایی داده‌های تانزانیا

سکونت که یک متغیر تبیینی دودویی با دو سطح شهری و روستایی (RS) است، و شاخص فضایی که نشان می‌دهد مادر و فرزند در کدام استان زندگی می‌کنند.

برای این داده‌ها، با توجه به هدف بررسی عوامل موثر بر سوءتغذیه کودکان، یک مدل رگرسیون گشتاوری فضایی در نظر گرفته شد. اثرات غیرخطی با اسپلاین‌های جریمه‌ای بیزی و اثر فضایی با یک میدان تصادفی مارکوفی وارد مدل شدند. برای متغیرهای تبیینی پیوسته اثر غیرخطی در نظر گرفته شدند و برای متغیرهای رسته‌ای یا دودویی اثر ثابت. مدل رگرسیون

مورد نظر به صورت زیر است:

$$\begin{aligned} \text{z} - \text{score}_i = \beta_0 + \beta_1 \text{RS}_{i\tau} + \beta_2 \text{Edu}_{i\tau} + \beta_3 \text{work}_{i\tau} \\ + \beta_4 \text{sex}_{i\tau} + f_1(\text{age}_{i\tau}) + f_2(\text{BMI}_{i\tau}) + \epsilon_{i\tau}. \end{aligned}$$

برای تنظیم‌سازی و تعیین اهمیت متغیرهای تبیینی رسته‌ای در گشتاورهای مرتبه‌های مختلف، مشابه مثال‌های شبیه‌سازی از پیشین‌های میخ و تخته استفاده شد. مدل مورد نظر با در نظر گرفتن مرتبه‌های

$$\tau = \{۰/۰۵, ۰/۱, ۰/۲, ۰/۵, ۰/۸, ۰/۹, ۰/۹۵\}$$

به داده‌ها برازش داده شد. همچنین برای مقایسه، مدل رگرسیون چندکی نیز به این داده‌ها برازش شد.

نتایج گزارش‌شده این مثال برگرفته از مقاله والدمن و همکاران (۲۰۱۷) است. جدول ۱.۴ نتایج برآورد ضرایب رگرسیونی متغیرهای تبیینی با اثرات خطی را نشان می‌دهد. برای هر متغیر، دو سطر از اعداد (برای τهای مختلف) نمایش داده شده‌اند. اعداد بالا برآورد اثر و اعداد زیر احتمال ورود متغیر به مدل را نشان می‌دهند. با توجه به نتایج جدول می‌توان گفت

- استخدام مادر در دم پایین توزیع شرطی اثر مثبت دارد، در حالی‌که در قسمت‌های دیگر اثر منفی دارد. البته به‌جز دم بالا به نظر این متغیر بر سپءتغذیه کودکان چندان معنی‌دار نیست.
- سوءتغذیه در کودکانی که مادران‌شان دارای تحصیلات متوسطه به بالا هستند، بیشتر است. این متغیر به‌ویژه برای مادران با سطح تحصیلات عالی نسبت به بی‌سوادها، کاملا معنی‌دار است.
- تأثیر مثبت سکونت در روستا را نیز می‌توان این‌گونه تفسیر کرد که نزدیکی به مزارع و پیوند نزدیک خانواده‌ها از اهمیت بالایی برخوردار است.
- با توجه به تأثیر منفی جنسیت، می‌توان اظهار کرد که به‌طور کلی پسران از دختران کوتاه‌تر هستند.

اثرات غیرخطی برآوردشده توسط دو مدل گشتاوری و چندکی در شکل ۱۰.۴ نشان داده شده‌اند. این نتایج بسیار نزدیک به نتایج به‌دست آمده توسط کاندالا و همکاران (۲۰۰۱) است، که در آن مجموعه داده‌ها در یک مدل رگرسیون میانگین مورد بررسی قرار گرفته‌اند. همان‌طور که مشاهده می‌کنید با افزایش سن مادر تا ۳۵ سالگی، سوءتغذیه تقریبا در حال افزایش است و پس از آن به تدریج کاهش می‌یابد. نقشه پهنه‌بندی برآورد اثر فضایی، به ازای مرتبه‌های مختلف در رگرسیون گشتاوری، در شکل ۱۱.۴ نمایش داده شده است. با توجه به شکل می‌توان به روشنی دید که در بخش‌های شمالی و شمال غربی تانزانیا، سوءتغذیه کودکان یک مساله جدی است.

جدول ۱.۴: برآورد پارامترهای اثرات ثابت برای داده‌های سوءتغذیه تانزانیا

متغیر/τ	۰/۰۵	۰/۲	۰/۸	۰/۹۵
شغل مادر	۴/۹۵۳	۰/۴۹۵	−۶/۱۵۳	−۱۴/۸۰۳
بی‌کار	۰/۴۵	۰/۱۲	۰/۴۸۳	۰/۷۰۲
تحصیلات مادر	بدون هیچ تحصیلات			
”دبستان”	−۵/۷۲۷	−۱/۰۱۶	−۹/۳۰۵	−۲۰/۷۱۲
	۰/۴۷۸	۰/۱۹۹	۰/۵۶۰	۰/۷۷۶
”راهنمایی”	۱۸/۹۸۷	۱۴/۶۵۹	۱/۷۴۲	−۱۱/۱۳۰
	۰/۷۵۹	۰/۷۰۱	۰/۲۴۶	۰/۶۴۵
” آموزش عالی ”	۶۰/۹۵۵	۵۷/۴۰۴	۶۱/۲۳۴	۷۷/۸۸۴
	۰/۹۵۵	۰/۹۴۷	۰/۹۵۴	۰/۹۸۳
محل سکونت مادر	۱۳/۱۵۵	۱۴/۰۷۱	۱۱/۰۵۰	۴/۸۷۹
”شهری”	۰/۶۵۹	۰/۷۰۰	۰/۶۳۳	۰/۴۲۴
جنسیت کودک	−۵/۶۶۶	−۵/۸۱۸	−۴/۱۲۸	−۳/۳۶۵
”مونث”	۰/۴۹۹	۰/۴۵۹	۰/۳۵۶	۰/۳۴۳

۶.۴ مثال دوم: مجموعه داده سوءتغذیه کودکان هند

سوءتغذیه کودکان هند در سال ۲۰۰۵ و ۲۰۰۶ مورد بررسی قرار گرفته است. داده‌های این مثال شامل ۴۰۰۰ مشاهده است. متغیر پاسخ مشابه مثال اول شاخص سوءتغذیه است و به همان شکل محاسبه می‌شود. متغیرهای تبیینی نیز شامل سن و BMI برای کودکان و سن و BMI برای مادران است و همین‌طور اطلاعات ناحیه محل سکونت کودکان است. اطلاعات مربوط به متغیرهای موجود در این مجموعه داده در جدول ۲.۴ آمده است. این مجموعه داده در بسته `expectreg` در دسترس است.

جدول ۲.۴: متغیرهای تبیینی مدل

y	z-score
cbmi	BMI کودک
cage	سن کودک
mbmi	BMI مادر
mage	سن مادر
disth	محل زندگی کودک

با توجه به ماهیت پیوسته متغیرهای تبیینی، همه اثرات را غیرخطی در نظر گرفتیم. بنابراین مدل رگرسیون گشتاوری فضایی با متغیر پاسخ z-score و متغیر محل سکونت کودک به‌عنوان اثر فضایی، به صورت زیر است:

$$y_i = \beta_0 + f_1(\text{cage}_{i\tau}) + f_2(\text{cbmi}_{i\tau}) + f_3(\text{mage}_{i\tau}) + f_4(\text{mbmi}_{i\tau}) + \delta_{i\tau} + \epsilon_{i\tau}$$

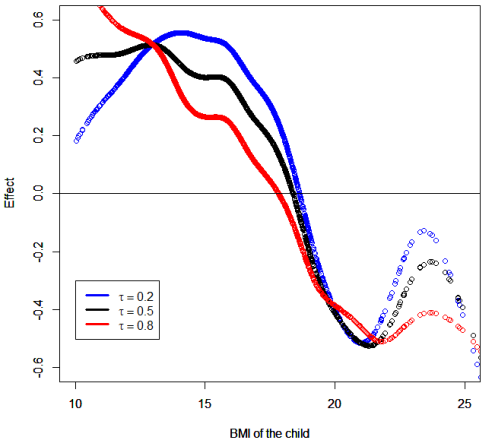
که در آن δ_τ اثر فضایی است که فرض می‌کنیم یک میدان تصادفی مارکوفی با ساختار فضایی مشبکه‌ای است که از طریق ماتریس مجاورت نواحی کشور هند ساخته می‌شود. برای برازش مدل رگرسیون گشتاوری نیز مرتبه‌های مختلف

$$\tau = \{۰/۰۵, ۰/۱, ۰/۲, ۰/۵, ۰/۸, ۰/۹۵, ۰/۹۸\}$$

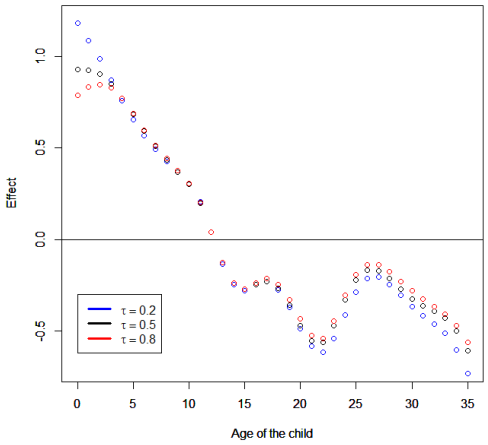
را در نظر گرفتیم. با توجه به تعداد بزرگ نواحی کشور هند (۴۲۲ ناحیه) برازش مدل مورد نظر بسیار زمان‌بر بود.

برآوردهای اثرات غیرخطی متغیرهای تبیینی، به‌ازای $\tau = ۰/۲, ۰/۵, ۰/۸$، در شکل ۱۲.۴ نمایش داده شده‌اند. با توجه به نتایج این شکل، متغیر تبیینی وزن مادر می‌تواند ثابت باشد؛ ولی سایر متغیرها غیرخطی هستند. با توجه به بخش آ این شکل، تا حدود ۱۲ سالگی اثر سن بر سوءتغذیه مثبت ولی نزولی است. پس از آن این اثر معکوس شده و سوءتغذیه را کاهش می‌دهد. این روند برای مرکز و هر دو دم پایین و بالای توزیع شرطی برقرار است. برای متغیر تبیینی BMI کودکان نیز می‌توان روندی مشابه را مشاهده کرد. برای کودکانی که BMI آن‌ها زیر حدود ۱۸ است، اثر سوءتغذیه مثبت و در سایر موارد منفی است. نتیجه جالب در مورد BMIمادران است که نشان می‌دهد سوءتغذیه در مادران چاق‌تر کمتر و در حال نزول است. یعنی هر چه مادر BMI بالاتری دارد، سوءتغذیه در کودکان کمتر دیده می‌شود.

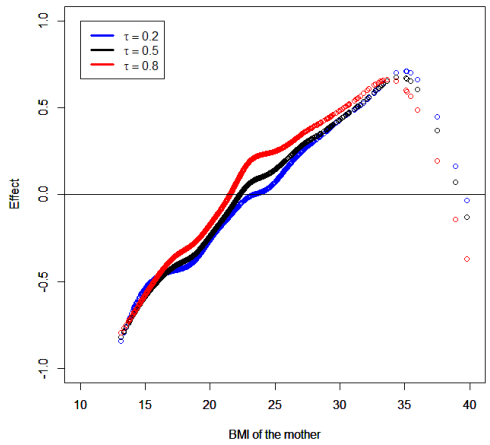
شکل ۱۳.۴ نیز اثر فضایی را برای این مجموعه داده، به ازای مرتبه‌های مختلف، نشان می‌دهد. با توجه به اختلاف بسیار ناچیز در اثر برآوردشده برای مرتبه‌های مختلف، می‌توان گفت اثر فضایی بر ویژگی‌های مختلف توزیع شرطی پاسخ تاثیر چندانی ندارد. در عین حال با توجه به مثبت بودن مقادیر برآوردشده در همه نواحی کشور هند، اثر معنی‌دار فضایی بر افزایش سوءتغذیه کاملا دیده می‌شود. این تاثیر در نواحی شرقی و مرکز کشور هند از سایر نواحی بیشتر است.



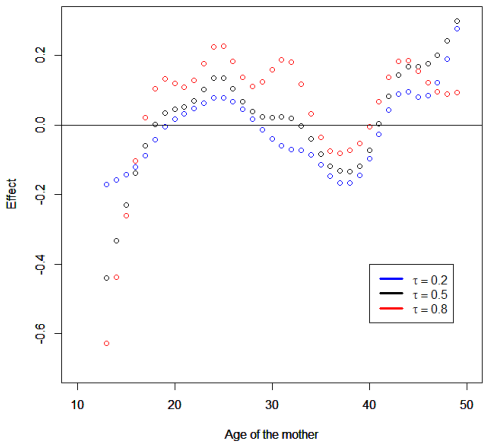
(ب) اثر وزن کودک



(آ) اثر سن کودک

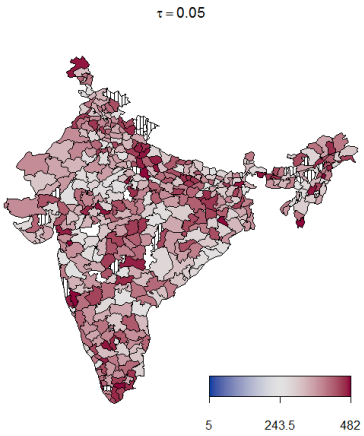


(د) اثر وزن مادر

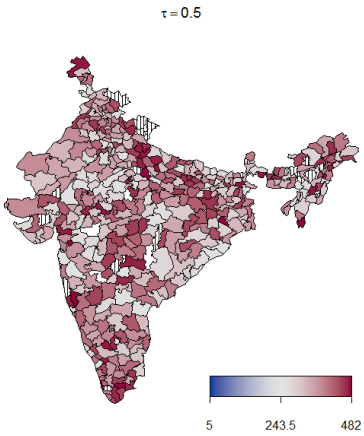


(ج) اثر سن مادر

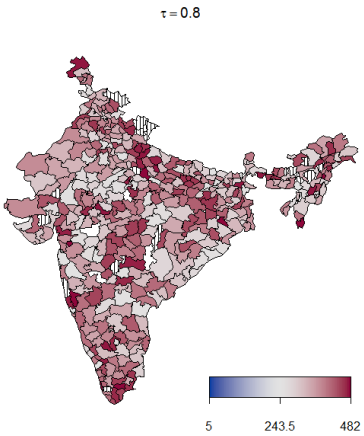
شکل ۱۲.۴: برآوردهای اثرات غیرخطی متغیرهای تبیینی در مجموعه داده هند برای سه مرتبه $\tau = ۰/۲, ۰/۵, ۰/۸$



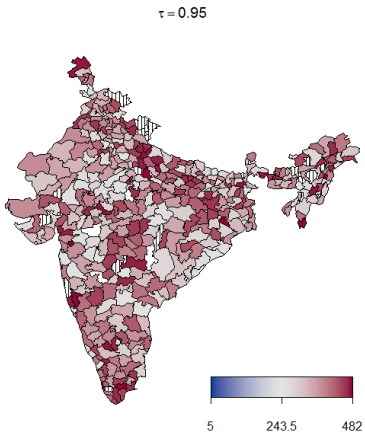
(آ) گشتاور ۰/۰۵



(ب) گشتاور ۰/۵



(ج) گشتاور ۰/۸



(د) گشتاور ۰/۹۵

شکل ۱۳.۴: برآورد اثر فضایی مجموعه داده هند

۷.۴ نتیجه‌گیری و پیشنهادات

در این پایان‌نامه مدل رگرسیون چندکی و گشتاوری را معرفی کردیم و معایب و مزایای هر کدام را مورد بررسی قرار دادیم. مدل بیزی رگرسیون گشتاوری را برای اثرات فضایی در نظر گرفتیم. برای رفع پیچیدگی مدل، با استفاده از پیشین‌های مناسب به تنظیم‌سازی آن پرداختیم و برای تقریب توزیع پسین مدل، الگوریتم MCMC مناسب را بر اساس توزیع‌های پیشنهادی IRWLS بازگو کردیم. در پایان، با مثال‌های شبیه‌سازی و واقعی، عملکرد مدل بیزی گشتاوری پیشنهادی و بحث انتخاب متغیر و مدل‌بندی ناپارامتری اثرات غیرخطی را بررسی کردیم.

معرفی مدل‌های رگرسیون گشتاوری برای پاسخ‌های غیرنرمال، تعمیم مدل گشتاوری به مدل‌های گشتاوری غیرخطی، و ساخت مدل‌های رگرسیون گشتاوری چندمتغیره، از جمله مواردی هستند که به‌عنوان آینده تحقیق قابل بررسی و اقدام می‌باشند.

پیوست آ

آ.۱ کد R مربوط به مثال شبیه‌سازی برای مدل‌بندی اثرات غیرخطی

```
### generating data with Scenario M1-E1
gen.data <- function(n, z){
y1 <- 5*exp(-0.5*z^2)
e <- rnorm(n, 0, sqrt(0.5*z^2))
y <- y1 + e
dt <- data.frame(cbind(z,y1,y))
return(dt)}
z.pred = seq(0, 3, length.out=100)
y1.pred <- 5*exp(-0.5*z.pred^2)
r = 100; n = 100
par1 = array(NA, c(r, 9, n))
par2 = array(NA, c(r, 9, n))
par2.L = array(NA, c(r, 9, length(z.pred)))
par2.U = array(NA, c(r, 9, length(z.pred)))
cov.laws = array(NA, c(r, 9, length(z.pred)))
```

```
width.laws = array(NA, c(r, 9, length(z.pred)))
par3.02 matrix(NA, ncol=r, nrow=n)
#
par3.02.L = matrix(NA, ncol=r,
nrow=length(z.pred))
#
par3.02.U = matrix(NA, ncol=r, nrow=length(z.pred))
cov.bayes.02 = matrix(NA, ncol=r, nrow=length(z.pred))
width.bayes.02 = matrix(NA, ncol=r, nrow=length(z.pred))
#####
for(i in 1:r){
print(i)
#
dt = gen.data(n=n, z=runif(n, 0, 3))
#
z.pred = seq(0, 3, length.out=100)
dt.pred = gen.data(n=length(z.pred), z=z.pred)
## boosting method
fit1 = expectreg.boost(y ~ bbs(z), data=dt, expectiles =
c(0.02, 0.05 ,0.1, 0.2, 0.5, 0.8, 0.9, 0.95, 0.98))$fitted
par1[i, ,] = fit1
## AWLS method
fit22 = expectreg.ls(y ~ rb(z, "pspline"), data=dt,
expectiles = c(0.02, 0.05,0.1, 0.2, 0.5, 0.8, 0.9, 0.95, 0.98))
par2[i, ,] = fit22$fitted
#  pred.laws = predict(fit22, newdata=z)$fitted
fit2 = expectreg.ls(y ~ rb(z, "pspline"), data=dt.pred,
expectiles = c(0.02, 0.05,0.1, 0.2, 0.5, 0.8, 0.9, 0.95, 0.98), ci=TRUE)
par2.L[i, ,] =  matrix(unlist(lapply(confint(fit2),
"[", 1:length(z.pred))), ncol=9, byrow=F)
par2.U[i, ,] = matrix(unlist(lapply(confint(fit2)
,"[", (length(z.pred)+1):(2*length(z.pred)))),
ncol=9, byrow=F)
##
```

```
cov.laws[i,,] = (par2.L[i,,] <= dt.pred$y1 & dt.pred$y1 <= par2.U[i,,])
width.laws[i,,] = par2.U[i,,]- par2.L[i,,]
## Bayesian method
y = dt$y
z = as.matrix(dt$z)
y.pred = dt.pred$y
z.pred = as.matrix(dt.pred$z)
##
fit3 = exp_bayes(y, z=z, tau= 0.02, B=35000, burn = 5000, thin = 30)
par3.02[,i] = fit3$fitted
rmse.bayes.02[i] = sqrt(mean((par3.02[,i]-dt$y1)^2))
# prediction
fit3 = exp_bayes(y.pred, z=z.pred, tau= 0.02,
B=35000, burn = 5000, thin = 30)
par3.02.L[, i] = fit3$lower
par3.02.U[,i] = fit3$upper
##
cov.bayes.02[,i] = (par3.02.L[,i] <= dt.pred$y1
& dt.pred$y1 <= par3.02.U[,i])
width.bayes.02[,i] = par3.02.U[,i] - par3.02.L[,i] }
```

آ.۲ کد R مربوط به مثال شبیه‌سازی بر پایه پیشین‌های میخ و تخته

```
gen.data <- function(n){
x = mvrnorm(n, rep(0,8), H)
e = rnorm(n, 0, sqrt(3))
y = 3+3*x[,1]+1.5*x[,2]+0*x[,3]+0*x[,4]+
2*x[,5]+0*x[,6]+0*x[,7]+0*x[,8] + e # + y1
#
dt = data.frame(cbind(y,x))
return(dt)}
r = 100; n = 100
```

```
beta.hat.05 = matrix(NA, nrow= r, ncol=9)
#
prob.inc.05 = matrix(NA, nrow= r, ncol=8)
#####
for(i in 1:r){
print(i)
dt = gen.data(n=n)
## Bayesian method
y = dt$y
Xspsl <- as.matrix(dt[,-1])
fit3 = exp_bayes(y, xspsl = Xspsl, tau= 0.05, B=3500, burn = 500,
thin = 3)
beta.hat.05[i,] = apply(fit3$b, 2, mean)
prob.inc.05[i,] = apply(fit3$delr, 2, mean)
```

آ.۳ کد R مربوط به سوءتغذیه کودکان هند

```
data(india)
data("india.bnd")
#
z.score = (india$stunting+177.9)/142.24
india = as.data.frame(cbind("y"=z.score, india))
y = india$y
z = as.matrix(india[,3:6])
distH = india$distH
#
fit.02 = exp_bayes(y,
## linear
#x=NULL,
## spike und slab regularisation
#xspsl = NULL,
## nonlinear effect
z=z,
## spatial effect
```

```
geo = distH,
map = india.bnd,
## expectile
tau=0.02,
## length
B=15000,
## burnin
burn = 5000,
## thinning
thin = 10 )
```

مراجع

[۱] پارسیان، ا. (۱۳۹۱). *مبانی آمار ریاضی*، چاپ دهم، مرکز نشر دانشگاه صنعتی اصفهان.

[۲] محمدزاده، م. (۱۳۹۱). *آمار فضایی و کاربردهای آن*، انتشارات دانشگاه تربیت مدرس.

[۳] اریه م، (۱۳۹۴). پایان‌نامه ارشد: ''برآورد مدل نیمه پارامتری به روش B_اسپلاین''، دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود.

[4] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceeding 2nd Inter Symposium on Information Theory*, 267-281, Budapest.

[5] Agresti, A. (2013). *Categorical Data Analysis*. Third Edition, John Wiley and Sons, New York.

[6] Bayes, T. (1763). An essay toward solving problem in a doctrine of chance. *In Philosophical Transactions of the Royal Society of London*, 53, 370-418.

[7] Berger, J.O. and Bernardo, J.M. (1992). On the development of reference priors. *Bayesian Statistics*, 4(4), 35-60.

[8] Brezger, A. and Lang, S. (2006). Generalized additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, 50, 967–991.

[9] Buhlmann, P. and Holthorn, T. (2007). Boosting algorithms: regularization prediction and model fitting. *Statistical Science*, 22, 477-505.

[10] Chen, M.H., Shao, Q.M. and Ebrahim, J. (2000). *Monte Carlo Methods in Bayesian Compution*. Springer, New York.

[11] Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematika*, 31, 377-403.

[12] Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley, New York.

[13] De Boor, C. (1977). Package for calculating with B-splines. *Society for Industrial and Applied Mathematics, Journal on Numerical Analysis*, 14(3), 441-472.

[14] De Rossi, G. and Harvey, A. (2009). Quantiles, expectiles and splines. *Journal of Economics*, 152(2), 179–185.

[15] Dierckx, P. (1993). *A Practical Guide to Splines*. Oxford University Press, London.

[16] Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, 1, 93–125.

[17] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32, 407-451.

[18] Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing using B-splines and penalized likelihood. *Statistical Science*, 11, 89–121.

[19] Fahrmeir, L., Kneib, T. and Konrath, S. (2010). Bayesian regularisation in structured additive regression: a unifying perspective on shrinkage, smoothing and predictor selection. *Statistics and Computing* , 20(2), 203–219.

[20] Fahrmeir, L. and Tutz, G. (1991). *Multivariate Statistical Modelling Based on Generalized Linear Models*. 2nd Edition, Springer-Verlag, Berlin.

[21] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(4), 1348-1360.

[22] Farooq, M. and Steinwart, I. (2015). An SVM-like approach for expectile regression. *arXiv preprint arXiv*, 1507.03887.

[23] Fisher, R.A. (1925). Theory of statistical estimation. *In Mathematical Proceedings of the Cambridge Philosophical Society*, 22, 700-725.

[24] Freund, Y. and Schapire, R. (1996). *Experiments with a New Boosting Algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning*, Morgan Kaufmann, San Francisco, CA.

[25] Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15, pp 246-263.

[26] Gauss, C.F. (1821). *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*. Werke 4, Gottingen, New York.

[27] Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

[28] Gerlach, R. and Chen, C.W.S. (2016). Bayesian expected shortfall forecasting incorporating the intraday range. *Journal of Financial Economics*, 14(1), 128–158.

[29] Gilks, W. R., Richardson, S. and Spiegelhalter, D.J. (1996). *Markov Chain Mont Carlo in Practice*. Chapman and Hall, London.

[30] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distribution, and the Bayesion restoration of image. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721-741.

[31] Gut, A. (2005). *Probability: A Graduate Course*. Springer Texts in Statistics, New York.

[32] Hardle, W. (1990). *Applied Non-parametric Regression*. Economic Society Monographs, vol,19 , Cambridge University Press.

[33] Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57(1), 97-109.

[34] Hoerl, A.E. and Kennard, R. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12, 55-67.

[35] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society*, 186, 453-461.

[36] Kandala, N.B., Lang, S., Klasen, S. and Fahrmeir, L. (2001). *Semiparametric Analysis of the Socio-demographic and Spatial Determinants of Undernutrition in Two African Countries*. University of Munich, Munich.

[37] Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.

[38] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.

[39] Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81, 1565–1578.

[40] Laplace, P.S. (1812). *Thorie Analytique Des Probabilites*. Courcier, Paris.

[41] Majumdar, A. and Paul, D. (2016). Zero expectile processes and Bayesian spatial regression. *Journal of Computational and Graphical Statistics*, 25(3), 727–747.

[42] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Model*. Chapmen and Hall, London.

[43] McCulloch, C. and Searle, S. (2001). *Generalized Linear and Mixed Models*. John Wiley and Sons, New York.

[44] Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*. 44, 335-341.

[45] Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, A, 135, 370–384.

[46] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087-1091.

[47] Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55, 819–847.

[48] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1, 502-518.

[49] O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *Society for Industrial and Applied Mathematics, Journal on Scientific and Statistical Computing*, 9(2), 363-379.

[50] Reed, C. and Yu, K. (2009). *A Partially Collapsed Gibbs Sampler for Bayesian Quantile Regression*. Department of Mathematical Sciences, Brunel University, Technical Report.

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

[52] Robert, C. (2007). *The Bayesian Choice*. Springer, New York.

[53] Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.

[54] Robert, C. and Casella, G. (2010). *Introducing Monte Carlo Methods whit R.* Springer, New York.

[55] Roberts, G.O. and Rosenthal, J. (1997). Geometric ergodicity and hybrid Markov chain. *Electronic Communications in Probability* , 2, 13-25.

[56] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields, Theory and Applications*, Chapman and HallCRC Press, London.

[57] Ruppert, D. (2002). Selecting the number of penalized splines. *Journal of Computational and Graphical Statistics*, 11(4), 735-757.

[58] Schnabel, S.K. and Eilers, P. (2009). Optimal expectile smoothing. *Computational Statistics and Data Analysis* , 53, 4168–4177.

[59] Schwarz, G. (1978). Estimating the dimension of a model. *Institute of Mathematical Statistics*, 6, 461-464.

[60] Sobotka, F. and Kneib, T. (2012). Geoadditive expectile regression. *Computational Statistics and Data Analysis* , 56, 755–767.

[61] Sobotka, F., Kauermann, G., Schulze Waltrup, L. and Kneib, T. (2013). On confidence intervals for semiparametric expectile regression. *Statistics and Computing* , 23(2), 135–148.

[62] Sobotka, F., Marra, G., Radice, R. and Kneib, T. (2013). Estimating the relationship between women's education and fertility in Botswana by using an instrumental variable approach to semiparametric expectile regression. *Journal of the Royal Statistical Society, C*, 62(1), 25–45.

[63] Sobotka, F., Schnabel, S. and Schulze Waltrup, L. (2014). *Expectreg: Expectile and Quantile Regression. R Package Version 0.39*. Georg August University Goettingen, Goettingen.

[64] Taylor, J.W. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Economics*, 6, 231–252.

[65] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, B*, 58, 267–288.

[66] Waldmann, E., Kneib, T., Yue, Y.R., Lang, S. and Flexeder, C. (2013). Bayesian semiparametric additive quantile regression. *Statistical Modeling*, 13(3), 223–252.

[67] Waldmann, E., Sobotka, F. and Kneib, T. (2017). Bayesian regularisation in geoadditive expectile regression. *Statistics and Computing*, 27(6), 1539-1553.

[68] Waltrup, S.L., Sobotka, F., Kneib, T. and Kauermann, G. (2015). Expectile and quantile regression - David and Goliath? *Statistical Modelling*, 15(5), 433-456.

[69] Yang, Y. and Zou, H. (2015). Nonparametric multiple expectile regression via ER-Boost. *Journal of Statistical Computation and Simulation*, 84(1), 84–95.

[70] Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics* , 6(2–3), 273–292.

[71] Yue, Y. and Rue, H. (2011). Bayesian inference for additive mixed quantile regression models. *Computational Statistics and Data Analysis*, 55, 84–96.

[72] Zou, H. (2006). The Adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418-1429.

[73] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, B*, 67(2), 301-320.

Abstract

Today, statisticians are developing a variety of regression models beyond the mean regression where the purpose is to study different aspects of the conditional distribution of the response given some covariates. Primarily, we are interested in examining the effect of covariates on the tail of the distribution of the response variable; e.g., in the area of child health, physicians are interested in identifying the factors affecting child malnutrition; and in earthquake studies, researchers are seeking to know the conditions under which the most devastating earthquakes occur. Conventional models for achieving this goal are quantile regression models. The primary challenge of statistical inference in quantile regression models is their fitting to the data under study, which requires optimization of the objective function based on linear programming. Considering this computational challenge, the class of expectile regression models is an appropriate alternative, particularly for spatial responses. In this thesis, we restate an expectile regression model for spatial responses and use a Bayesian approach to fit the model and implement inference. Also, we consider the modeling of nonlinear functions of covariates by using penalized splines and variable selection of covariates with linear effects by the spike and slab priors.

keywords: Quantile regression, Expectile regression, Gaussian Markov random field, Bayesian regularisation, Penalized splines, Spike and slab prior.

Faculty of Mathematical Sciences

MSc Thesis in Statistics

Bayesian Inference in Spatial Expectile Regression Models

By: Nasim Andakhs

Supervisor:

Dr. Hossein Baghishani

January 2019