
بسم الله الرحمن الرحیم

دانشکده علوم ریاضی

رشته آمار، گرایش آمار ریاضی

پایان‌نامه کارشناسی ارشد

تحلیل پاسخ‌های دو سطحی نامتعادل با مدل رگرسیون مقادیر فرین تعمیم‌یافته

نگارنده: سولماز بلوری کلورزی

استاد راهنما

دکتر حسین باغیشنی

بهمن ۱۳۹۵

تقدیم به

پدرم

مادرم

و آرامم

رضا

سپاس‌گزاری

سپاس خدای را که مرا شوق و توان ادراک آموخت.

مراتب سپاس و تشکر خود را از استاد گرانقدرم جناب آقای دکتر حسین باغیشنی ابراز می‌دارم که در تمام مراحل پژوهش این پایان‌نامه با تلاش‌های بی‌وقفه و راهنمایی‌های ارزنده مرا یاری نمودند. همچنین از اساتید داور، جناب آقای دکتر آرشی و جناب آقای دکتر شاهسونی و از تمامی اساتید فرهیخته گروه آمار دانشگاه صنعتی شاهرود تشکر و قدردانی می‌نمایم.

وظیفه خود می‌دانم که از پدر بزرگوار و مادر مهربانم که پیوسته راهنمای من در زندگی بوده‌اند، خواهر و برادرانم که همواره پشتیبانم بوده‌اند و همسر فداکارم به دلیل صبوری و زحماتش در این دوره، قدردانی به عمل آورم. همچنین از پدر و مادر همسرم و تمامی دوستانی که در طول دوران تحصیل، لحظات تکرارنشدنی را در کنارشان گذراندم، صمیمانه سپاس‌گذاری می‌کنم.

سولماز بلوری کلورزی
بهمن ۱۳۹۵

تعهدنامه

این‌جانب **سولماز بلوری کلورزی** دانشجوی کارشناسی ارشد رشته **آمار** دانشکده **علوم ریاضی** دانشگاه **صنعتی شاهرود**، نویسنده پایان‌نامه با عنوان **تحلیل پاسخ‌های دو سطحی نامتعادل با مدل رگرسیون مقادیر فرین تعمیم‌یافته** 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 شده‌اند.
- 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 شده‌اند.

سولماز بلوری کلورزی
بهمن ۱۳۹۵

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته‌شده) 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

در موقعیت‌های کاربردی متعددی، در یک مدل رگرسیونی، متغیر پاسخ دو سطحی است. معمولاً برای مدل‌بندی این نوع پاسخ‌ها از مدل‌های رگرسیونی لجستیک، پرابیت و Cloglog استفاده می‌شود. در مواردی که متغیر پاسخ دو سطحی نامتعادل (چوله) است، دو مدل (متقارن) لجستیک و پرابیت انتخاب‌های خوبی نیستند. همچنین برای متغیر پاسخ دو سطحی نامتعادل، اگر عدم تعادل پاسخ به نفع صفر باشد، مدل چوله Cloglog که در مدل‌بندی چولگی پاسخ‌ها به راست موفق عمل می‌کند، کارایی خود را از دست خواهد داد. مدل رگرسیونی t-تعمیم‌یافته یک گزینه پیشنهادی در این موارد است. اما گستره مدل‌بندی چولگی موجود در داده‌ها توسط این مدل، به دلیل دامنه تغییرات کوچک پارامتر شکل خود، به شدت محدود است. ونگ و دی (۲۰۱۰) بر اساس توزیع مقادیر فرین تعمیم‌یافته، یک مدل جدید معرفی کردند که از انعطاف‌پذیری بالایی برخوردار است. در این پایان‌نامه، مدل مبتنی بر توزیع مقادیر فرین تعمیم‌یافته را بازگو می‌کنیم و به چگونگی برازش آن از دیدگاه بیزی می‌پردازیم. با توجه به پیچیدگی توزیع پسین مدل، از روش‌های نمونه‌گیری مونت کارلوی زنجیر مارکوفی برای تقریب آن استفاده می‌کنیم. با شبیه‌سازی، کارایی مدل رگرسیونی مقادیر فرین تعمیم‌یافته را بررسی کرده و کاربرد این مدل را در قالب یک مثال واقعی تشریح می‌کنیم.

کلمات کلیدی

الگوریتم‌های مونت کارلوی زنجیر مارکوفی، پیشامدهای نادر، توزیع پسین، توزیع مقادیر فرین تعمیم‌یافته، چولگی، متغیر پنهان.

پیش‌گفتار

نظریه مقادیر فرین[۱] بر دم‌های توزیع تمرکز دارد و رفتار مشاهدات خیلی بزرگ یا خیلی کوچک را در فرآیندی تصادفی تحلیل می‌کند. این نظریه بر اساس توزیع مقادیر فرین تعمیم‌یافته[۲] شکل گرفته است. از مزیت‌های این مدل، انعطاف‌پذیری بالای آن در تشخیص و کنترل چولگی[۳] است. این کنترل چولگی به‌وسیله پارامتر شکل توزیع انجام می‌شود. از کاربردهای گوناگون این توزیع می‌توان به استفاده آن در مهندسی، پزشکی و اقتصاد اشاره کرد. این توزیع یک انتخاب متداول برای مدل‌بندی متغیرهای پاسخ دو سطحی است. مدل مبتنی بر توزیع مقادیر فرین تعمیم‌یافته توسط وانگ و دی (۲۰۱۰) پیشنهاد شد. مدل‌بندی مقادیر فرین در حضور متغیرهای تبیینی با اثرات خطی و غیرخطی، اثرات وابستگی فضایی و زمانی مثل الگوهای دوره‌ای، فصلی و روندها از جمله مسائلی هستند که اخیراً مورد توجه قرار گرفته‌اند. در این پایان‌نامه مدل مقادیر فرین تعمیم‌یافته برای تحلیل پاسخ‌های دو سطحی نامتعادل را بازگو می‌کنیم و از دیدگاه بیزی به استنباط در آن می‌پردازیم. با توجه به این مقدمه، ساختار این پایان‌نامه به‌صورت زیر است:

- در فصل اول، تعاریف و مفاهیم مورد نیاز برای ورود به موضوع اصلی پایان‌نامه را مطرح می‌کنیم.
- در فصل دوم، مدل تابع پیوند t-تعمیم‌یافته و مدل مقادیر فرین تعمیم‌یافته را معرفی می‌کنیم.
- در فصل سوم، به استنباط بیزی مدل مقادیر فرین تعمیم‌یافته و نمونه‌گیری از توزیع پسین آن توسط الگوریتم‌های مونت کارلوی زنجیر مارکوفی می‌پردازیم.
- در فصل چهارم، با مطالعه شبیه‌سازی، عملکرد مدل‌ها و روش پیشنهادی را مورد ارزیابی قرار می‌دهیم. همچنین کاربرد مدل را در یک مثال واقعی نمایش خواهیم داد.
- پیوست آ نیز شامل بخشی از کدهای نوشته‌شده در محیط R برای بازتولید مثال‌های موجود در پایان‌نامه است.

---

[۱] Extreme Value Theory

[۲] Generalized Extreme Value Distribution

[۳] Skewness

فهرست مطالب

فهرست تصاویر

فهرست جداول

فصل ۱

تعاریف و مفاهیم اولیه

۱.۱ مقدمه

آمار علم گردآوری و سازمان‌دهی داده‌ها، تحلیل آن‌ها و استخراج نتایج است. سازمان‌دهی و طبقه‌بندی مجموعه داده‌ها، از اهداف آمار توصیفی و تحلیل و استخراج نتایج از داده‌ها، از اهداف استنباط آماری است. در این پایان‌نامه، در پی استنباط آماری داده‌ها هستیم و عناوین مرتبط دیگر را برای درک مفاهیم اساسی توضیح می‌دهیم.

دامنه تحلیل‌های آماری از تحلیل داده‌های یک تا چند متغیره را در بر می‌گیرد. علم آمار روش‌های مختلفی را برای تعیین رابطه بین متغیرها پیشنهاد می‌کند که شامل انواع مختلف مدل‌های رگرسیونی است. کاربردهای رگرسیون متعدد هستند و تقریباً در هر زمینه‌ای از علوم مانند مهندسی، فیزیک، اقتصاد، مدیریت، علوم زیستی و علوم اجتماعی دیده می‌شوند. در واقع، تحلیل رگرسیونی را می‌توان پرکاربردترین روش تحلیل داده‌ها در آمار دانست. هدف اصلی تحلیل رگرسیون، تعیین بهترین مدل

$$y = f(x) + \epsilon$$

است که بتواند ارتباط بین یک متغیر پاسخ، y، با یک یا چند متغیر تبیینی، $\mathbf{x} = (x_1, \ldots, x_p)'$، را تعیین کند. به طور معمول فرض می‌شود که جمله خطای ϵ دارای توزیع نرمال است که در نتیجه با فرض تصادفی نبودن $\mathbf{x}$، متغیر پاسخ y نیز دارای توزیع نرمال خواهد بود؛ اما در کاربردهای مختلفی ممکن است پذیره نرمال بودن متغیر پاسخ برقرار نباشد. به‌عنوان مثال، متغیرهای پاسخ رسته‌ای[1] یا

[1] Categorical

حضور داده‌های پرت در مشاهدات، نرمال بودن توزیع پاسخ را رد می‌کنند. در این موارد میانگین پاسخ را نمی‌توان به صورت مستقیم، به‌عنوان تابعی از متغیرهای تبیینی، مدل‌سازی نمود، بلکه تبدیلی از آن با استفاده از تابعی به نام تابع پیوند[۲] مورد استفاده قرار می‌گیرد. در این موارد مدلی که بر داده‌ها برازش می‌شود، عضوی از مدل‌های خطی تعمیم‌یافته[۳] (GLM) است (مک‌کلا و نلدر، ۱۹۸۹). مدل رگرسیون خطی حالت خاصی از این مدل‌ها است (جکمن، ۱۹۸۹).

در برخی از موارد، متغیر پاسخ، از نوع رسته‌ای دو سطحی[۴] با سطوح موفقیت (۱) و شکست (۰) است. این نوع از پاسخ‌ها در بسیاری از کاربردها رخ می‌دهند. به‌عنوان مثال، در علوم بانکداری، مدیران به دنبال شناخت نرخ بازپس‌دهی وام‌های واگذارشده به مشتریان، برای برنامه‌ریزی‌های آینده هستند. در این مثال شناخت مشتریان خوش حساب به‌عنوان سطح موفقیت اهمیت دارد؛ یا در پزشکی، محققین به دنبال کشف بیماران مبتلا به سرطان در یک منطقه خاص هستند. مدل‌های رگرسیونی معمول برای تحلیل این نوع پاسخ‌ها، رگرسیون پرابیت[۵]، لجستیک[۶] و Cloglog[۷] هستند.

مثال‌های متنوعی از پاسخ‌های دو سطحی وجود دارند که تعداد مشاهدات در دو رده موفقیت و شکست برابر نیستند (به عبارتی توزیع این داده‌ها به شدت چوله است). این نوع پاسخ را در اصطلاح نامتعادل می‌گویند. در این موارد استفاده از مدل‌های پرابیت و لجستیک برای مدل‌بندی پاسخ‌ها مناسب نیستند؛ زیرا توزیع‌های پشتیبان مدل‌های رگرسیون لجستیک و پرابیت به ترتیب توزیع‌های لجستیک و نرمال هستند که متقارن می‌باشند و برای داده‌های چوله انتخاب خوبی نیستند. رگرسیون Cloglog نیز تنها زمانی که عدم تعادل به سود ۱ها است، می‌تواند گزینه مناسبی باشد. به‌عنوان مثال، نسبت مشتریان خوش حساب در بانکداری، به‌طور معمول کمتر از ۷۰ درصد نیست. اگر هدف، تشخیص مشتریان خوش حساب $(y = ۱)$ از مشتریان بد حساب $(y = ۰)$ باشد، آن‌گاه رگرسیون Cloglog می‌تواند مدل مناسبی باشد؛ اما اگر برعکس، تعداد ۰ها بیشتر باشد، آن‌گاه این مدل نیز به شدت ناکارا خواهد بود. در این‌گونه موارد، مدل رگرسیونی مبتنی بر توزیع مقادیر فرین تعمیم‌یافته پیشنهاد می‌شود که نسبت به داده‌های چوله انعطاف‌پذیری زیادی دارد (وانگ و دی، ۲۰۱۰).

برازش مدل رگرسیونی مقادیر فرین تعمیم‌یافته[۸]، بر پایه هر دو دیدگاه کلاسیک و بیزی قابل انجام است. روش‌های استنباط کلاسیک، معمولاً، روش‌های مبتنی بر درستنمایی هستند که برای محاسبه تابع درستنمایی در رده GLM به کار می‌روند. ماکسیمم‌سازی تابع درستنمایی، حل عددی توابعی را شامل می‌شود که می‌توانند چندمدی و ناهموار و در نتیجه پیچیده و زمان‌بر باشند. پرسکات و والدن (۱۹۸۰) برآوردگرهای درستنمایی ماکسیمم[۹] و نحوه محاسبه ماتریس اطلاع فیشر[۱۰] و هاسکینگ

---

[۲] Link Function
[۳] Generalized Linear Models
[۴] Binary
[۵] Probit Regression
[۶] Logistic
[۷] Complementary Log-Log
[۸] Generalized Extreme Value Regression
[۹] Maximum Likelihood Estimators
[۱۰] Fisher Information Matrix

(۱۹۸۵) الگوریتمی برای محاسبه برآوردگرهای درستنمایی ماکسیمم، برای پارامترهای توزیع مشاهدات خیلی بزرگ یا کوچک یعنی همان مقادیر فرین ماکسیمم یا می‌نیمم مشاهدات را پیشنهاد کردند. اگرچه روش‌های مختلفی برای بهینه‌سازی توابع پیچیده معرفی شده‌اند، اما هر کدام نیز دارای معایب و کاستی‌هایی هستند. بنابراین برای کناره گرفتن از چنین مشکلاتی، در این پایان‌نامه، دیدگاه بیزی را برای استنباط در مدل‌های مورد نظر انتخاب می‌کنیم.

استنباط در دیدگاه بیزی، مبتنی بر توزیع پسین[۱۱] است و می‌تواند انتخابی مناسب برای استنباط مدل‌های پیشنهادی باشد. توزیع پسین در یک مدل بیزی، به‌روز شده باور و دیدگاه اولیه در مورد پارامتر با توجه به مشاهدات است. اولین مرحله در این روش، مشخص ساختن توزیع پیشین[۱۲] است. ورود اطلاعات جدید، توزیع پیشین را به توزیع پسین تبدیل می‌کند. ابزار اساسی برای این تبدیل، قضیه بیز[۱۳] است (گلمن و همکاران، ۲۰۰۳). تحلیل بیزی نیاز به حل انتگرالی دارد که به راحتی نمی‌توان آن را به‌دست آورد. در این صورت می‌توان از روش‌های نمونه‌گیری مبتنی بر شبیه‌سازی مونت کارلویی[۱۴] برای تقریب آن استفاده نمود. از آن‌جایی که نمونه‌گیری مستقیم از توزیع پسین نیز به سادگی صورت نمی‌پذیرد، می‌توان از روش‌های نمونه‌گیری مونت کارلوی زنجیر مارکوفی[۱۵] (MCMC) بهره برد. روش‌های MCMC شامل روش‌هایی است که برای انتگرال‌گیری به روش مونت کارلویی، اغلب در مسائل مربوط به استنباط بیزی، مورد استفاده قرار می‌گیرند. از این روش می‌توان برای تولید نمونه از توزیع پسین استفاده کرد (برانسکام و همکاران، ۲۰۰۷). از جمله روش‌های MCMC می‌توان الگوریتم متروپلیس-هستینگز[۱۶] (متروپلیس و همکاران، ۱۹۵۳ و هستینگز، ۱۹۷۰) و نمونه‌گیر گیبز[۱۷] (گمان و گمان، ۱۹۸۴ و گلفاند و اسمیت، ۱۹۹۰) را نام برد که به‌عنوان ابزار بسیار مفیدی برای تحلیل مدل‌های آماری پیچیده پدید آمده‌اند.

با نظر در متون آماری به این نتیجه می‌رسیم که در دنیای واقعی، رخدادهایی در ناحیه دم توزیع[۱۸] با فراوانی‌های مختلف اتفاق می‌افتند. برخلاف روش‌های کلاسیک آماری در بررسی رفتار توزیع در اطراف میانگین، نظریه مقادیر فرین به تحلیل رفتار دم توزیع‌ها می‌پردازد. در نتیجه با وجود مجموعه داده‌های بزرگ برای مطالعه این نظریه، معمولاً اطلاعات کمی در مورد دم توزیع در اختیار است. هنگام بررسی تغییرات در میانگین توزیع، قضیه حد مرکزی[۱۹] (کسلا و برگر، ۲۰۰۲) نشان می‌دهد که توزیع میانگین‌ها به‌طور مجانبی نرمال بوده و تقریب نرمال برای رفتار میانگین‌ها مناسب می‌باشد. مشابه با قضیه حد مرکزی، قضیه انواع فرین[۲۰] (فیشر و تیپت، ۱۹۲۸) رفتار مجانبی توزیع مقادیر فرین (ماکسیمم‌ها یا

---

[۱۱] Posterior Distribution
[۱۲] Prior Distribution
[۱۳] Bayes Theorem
[۱۴] Monte Carlo Simulation
[۱۵] Markov Chain Monte Carlo
[۱۶] Metropolis-Hastings Algoritm
[۱۷] Gibbs Sampler
[۱۸] Tail of Distribution
[۱۹] Central Limit Theorem
[۲۰] Extremal Types Theorem

می‌نیمم‌ها) را مشخص می‌کند. به‌طور متداول، در روش‌های مختلف آماری به تحلیل مشاهداتی که بخش قابل توجه توزیع را توصیف می‌کنند، پرداخته می‌شود و مقادیر خیلی بزرگ و کوچک کنار گذاشته یا مدل را در مقابل اثر نقاط دورافتاده[۲۱] استوار می‌سازند. در حالی‌که در تحلیل مقادیر فرین تنها این مقادیر حفظ شده و از آن برای توصیف دم توزیع استفاده می‌شود. مدل مقادیر فرین تعمیم‌یافته در رده GLM قرار می‌گیرد. برای معرفی و درک بهتر مدل مورد بحث (در فصل دوم)، ابتدا در این فصل مدل‌های رگرسیونی و مدل‌های خطی تعمیم‌یافته را معرفی می‌کنیم.

۲.۱ مدل‌های رگرسیونی

اولین بار موضوع رگرسیون در سال ۱۸۸۵ توسط گالتن مطرح شد. وی قصد داشت رابطه‌ای بین قد والدین و قد فرزندان پیدا کند. او معتقد بود والدین بلندقد تقریباً فرزندانی بلندقد و والدین کوتاه‌قد معمولاً فرزندانی با قد کوتاه دارند. وی، این مطلب را به صورت یک مدل ریاضی مطرح کرد. برای نشان دادن رابطه بین دو متغیر، می‌توان از ضریب همبستگی نیز استفاده کرد؛ اما ضریب همبستگی اثر یک متغیر بر متغیر دیگر را نشان نمی‌دهد. از طرفی، ضریب همبستگی نمی‌تواند تخمینی از تغییر در یک متغیر را با تغییر در متغیر دیگر ارائه دهد. بنابراین، برای پاسخ دادن به موارد فوق می‌توان از مدل‌های رگرسیونی استفاده کرد (گلمن و همکاران، ۲۰۰۶). تحلیل رگرسیون کاربردهای زیادی دارد. امروزه، می‌توان از آن در زمینه‌هایی چون اقتصاد، بازرگانی، هواشناسی، مهندسی و علوم اجتماعی استفاده نمود.

یک مدل رگرسیونی را در حالت کلی می‌توان به صورت زیر نشان داد:

$$y = f(x_1, x_2, \ldots, x_p) + \epsilon$$

که در آن y متغیر پاسخ[۲۲] نامیده می‌شود که از سایر متغیرها تأثیر می‌پذیرد. متغیر یا متغیرهایی که بر متغیر پاسخ اثر می‌گذارند، یعنی $\mathbf{x} = (x_1, x_2, \ldots, x_p)'$، متغیرهای تبیینی[۲۳] (توضیحی یا پیش‌بین) نامیده می‌شوند.

ساده‌ترین نوع رگرسیون، رگرسیون خطی[۲۴] است. در رگرسیون خطی به دنبال رابطه خطی y با متغیرهای $x_1, \ldots, x_p$ هستیم. الگوی رگرسیون خطی به صورت زیر نشان داده می‌شود:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \qquad (۱.۱)$$

که در آن $\beta_0, \ldots, \beta_p$ ضرایب مدل رگرسیونی نامیده می‌شوند. این ضرایب نامعلوم بوده و مستقیماً برآورد می‌شوند. متغیر تصادفی ε جمله خطای این معادله است که نوسانات تصادفی، خطاهای اندازه‌گیری، یا اثر عوامل خارج از کنترل را نشان می‌دهد. معادله (۱.۱) را به این دلیل خطی گویند که در آن میانگین پاسخ، یک تابع خطی از پارامترهای نامعلوم $\beta_0, \ldots, \beta_p$ است.

مدل‌های رگرسیونی برای اهداف زیر به کار می‌روند:

---

[۲۱] Outliers

[۲۲] Response Variable

[۲۳] Exploratory Variables

[۲۴] Linear Regression

- پیش‌بینی
- توصیف یا تفسیر داده‌ها
- برآورد پارامترها
- انتخاب متغیر
- کنترل پاسخ

در مدل‌های رگرسیونی پذیره‌های بنیادی وجود دارند که از طریق خطاها قابل آزمون هستند. این پذیره‌ها عبارت‌اند از

- مؤلفه تصادفی خطا دارای میانگین صفر است؛ یعنی برای تمام مشاهدات $i = ۱, \ldots, n$، $E(\varepsilon_i) = ۰$ است.
- واریانس مؤلفه خطا، ثابت است؛ یعنی برای $i = ۱, \ldots, n$، $Var(\varepsilon_i) = \sigma^۲$ است.
- خطاها ناهمبسته هستند.
- مؤلفه تصادفی خطا از متغیرهای تبیینی $\mathbf{x}$ مستقل است.

هیچ‌گاه نباید بدون بررسی این پذیره‌ها، از مدل رگرسیونی استفاده نمود. در غیر این صورت اطلاعات به‌دست آمده از رگرسیون می‌تواند گمراه‌کننده باشد. برآورد پارامترها در این مدل به روش‌های کمترین توان‌های دوم و درستنمایی ماکسیمم صورت می‌پذیرد (نیرومند، ۱۳۸۷). حالت‌هایی هم وجود دارند که امکان استفاده از رگرسیون خطی فراهم نیست. مدل‌های خطی تعمیم‌یافته رده بزرگتری از مدل‌های خطی را برای این‌گونه حالت‌ها ارایه می‌دهد.

۳.۱ مدل‌های خطی تعمیم‌یافته

در روش‌های استنباطی مربوط به الگوهای رگرسیون خطی و غیرخطی[۲۵]، فرض بر این است که متغیر پاسخ از توزیع نرمال تبعیت می‌کند؛ اما حالت‌هایی هم وجود دارند که این پذیره برقرار نیست. به‌عنوان مثال، زمانی که متغیر پاسخ دو سطحی یا شمارشی[۲۶] است، پذیره توزیع نرمال برای پاسخ کاملاً غیرواقعی به نظر می‌رسد. در چنین مواقعی، استفاده از مدل‌های خطی امکان‌پذیر نیست. همچنین، مدل‌های خطی محدودیتی را که در میانگین پاسخ برای این‌گونه داده‌ها وجود دارد، در نظر نمی‌گیرند. به‌عنوان مثال، در توزیع پواسن که معمولاً برای داده‌های شمارشی به کار می‌رود، میانگین توزیع پاسخ، μ، باید نامنفی باشد. چنان‌چه از مدل‌های خطی استفاده شود، ممکن است برآورد میانگین پاسخ منفی شود.

[۲۵] Nonlinear

[۲۶] Count

در نتیجه، برای این‌گونه از داده‌ها معمولاً از تبدیلی با نام تابع پیوند لگاریتمی[۲۷] استفاده می‌شود. رده مدل‌های خطی تعمیم‌یافته، به‌عنوان تعمیمی از مدل‌های خطی، توسط نلدر و ودربرن (۱۹۷۲) معرفی شد. ذات این تعمیم مبتنی بر دو جنبه است به‌طوری که

- متغیرهای پاسخ می‌توانند توزیعی غیر از توزیع نرمال داشته باشند. البته توزیع در نظر گرفته‌شده باید متعلق به خانواده توزیع‌های نمایی[۲۸] باشد.
- میانگین پاسخ به صورت مستقیم مدل‌سازی نمی‌شود، بلکه تبدیلی از آن با استفاده از تابع پیوند مدل‌سازی می‌شود.

قبل از بیان مدل‌های خطی تعمیم‌یافته به معرفی خانواده توزیع‌های نمایی می‌پردازیم.

۱.۳.۱ خانواده توزیع‌های نمایی

خانواده نمایی توزیع‌ها یک مؤلفه مهم در مدل‌های خطی تعمیم‌یافته است. این خانواده شامل توزیع‌های مختلفی مانند نرمال، پواسن، هندسی، نمایی، گاما، نرمال وارون[۲۹] و دوجمله‌ای منفی است. تابع چگالی خانواده توزیع‌های نمایی را می‌توان به صورت

$$f(y;\theta,\phi)=\exp\{\frac{y\theta-b(\theta)}{a(\phi)}+c(y,\phi)\}$$

بیان کرد، که در آن $a(\cdot)$، $b(\cdot)$ و $c(\cdot)$ توابعی مشخص می‌باشند. پارامتر θ یک پارامتر مکانی[۳۰] بوده و ϕ را اغلب پارامتر پراکندگی[۳۱] می‌نامند. در این خانواده، میانگین و واریانس به صورت

$$E(Y)=\mu=\frac{db(\theta)}{d\theta}=b'(\theta) \qquad (۲.۱)$$

و

$$Var(Y)=\frac{d^{۲}b(\theta)}{d\theta^{۲}}a(\phi)=b''(\theta)a(\phi) \qquad (۳.۱)$$

محاسبه می‌شوند (بیکل و داکسوم، ۲۰۰۱). به‌عنوان مثال، توزیع پواسن را در نظر بگیرید که تابع احتمال آن به صورت

$$f(y;\mu)=\frac{e^{-\mu}\mu^{y}}{y!}$$

است. می‌توان این تابع احتمال را به صورت خانواده توزیع‌های نمایی نوشت:

$$f(y;\mu)=\exp\{y\ln\mu-\mu-\ln y!\}$$

که در آن $\theta=\ln\mu$، $b(\theta)=e^{\theta}$، $a(\phi)=۱$ و $c(y,\phi)=-\ln(y!)$ است. میانگین و واریانس آن نیز با استفاده از روابط (۲.۱) و (۳.۱) به دست می‌آیند.

---

[۲۷] Log Link Function

[۲۸] Exponential Family Distributions

[۲۹] Inverse Normal

[۳۰] Location Parameter

[۳۱] Dispersion Parameter

۲.۳.۱ مؤلفه‌های مدل خطی تعمیم‌یافته

مدل خطی تعمیم‌یافته دارای سه مؤلفه است که به صورت زیر معرفی می‌شوند:

- مؤلفه تصادفی: توزیع متغیر پاسخ را مشخص می‌کند که باید از خانواده توزیع‌های نمایی باشد.
- مؤلفه اصلی: این مؤلفه دارای یک پیشگوی خطی[۳۲] است که متغیرهای تبیینی را شامل می‌شود و می‌توان آن را به صورت زیر تعریف کرد:
$$\eta_i = \mathbf{x}_i'\beta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i.$$
- تابع پیوند: این مؤلفه پیشگوی خطی را به میانگین متغیر پاسخ مربوط می‌کند. تابع پیوند یک تابع یکنوای مشتق‌پذیر است، به‌طوری که
$$\eta_i = g(E(Y_i|\mathbf{x}_i)) = \mathbf{x}_i'\beta$$

که در آن $g(\cdot)$ تابع پیوند است. از آن‌جایی که تابع پیوند معکوس‌پذیر نیز هست، می‌توان آن را به صورت زیر بیان کرد:
$$E(Y_i|\mathbf{x}_i) = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i'\beta).$$

تابع پیوند انواع گوناگونی دارد که می‌توان به تابع پیوند همانی[۳۳]، $g(p) = p$ (الگوی رگرسیون خطی استاندارد که در معادله (۱.۱) بیان شده یک GLM است)، تابع پیوند لگاریتمی، $g(p) = \log(p)$، تابع پیوند لجیت[۳۴]، $g(p) = \log(\frac{p}{1-p})$، تابع پیوند پرابیت[۳۵]، $g(p) = \Phi^{-1}(p)$، و تابع پیوند Cloglog[۳۶]، $g(p) = \log[-\log[1-p]]$، اشاره کرد.

بسته به انتخاب تابع پیوند g، یک GLM می‌تواند یک الگوی غیرخطی را شامل شود. بنابراین، الگوهای خطی تعمیم‌یافته را می‌توان به‌عنوان یکسان‌سازی الگوهای خطی و غیرخطی تلقی کرد که خانواده غنی توزیع‌های پاسخ نرمال و غیرنرمال را با هم متحد می‌کند (نیرومند، ۱۳۸۴). اینک به معرفی الگوهایی می‌پردازیم که برای متغیر پاسخ دو سطحی کاربرد بسیاری دارند.

۳.۳.۱ رگرسیون لجستیک

رگرسیون لجستیک یکی از اعضای رده مدل‌های خطی تعمیم‌یافته است که از تابع پیوند لجیت برای مدل‌بندی متغیرهای پاسخ دو سطحی با فرض استقلال خطاها استفاده می‌کند. این تابع پیوند از توزیع لجستیک نتیجه می‌شود، به‌طوری که اگر $F(\cdot)$ تابع توزیع لجستیک با پارامتر مکان μ و پارامتر مقیاس[۳۷]

[۳۲] Linear Predictor
[۳۳] Identity Link Function
[۳۴] Logit Link Function
[۳۵] Probit Link Function
[۳۶] Complementary Log-Log Link Function
[۳۷] Scale Parameter

σ باشد، آن‌گاه

$$F(x) = \frac{۱}{۱ + \exp\{(x-\mu)/\sigma\}}.$$

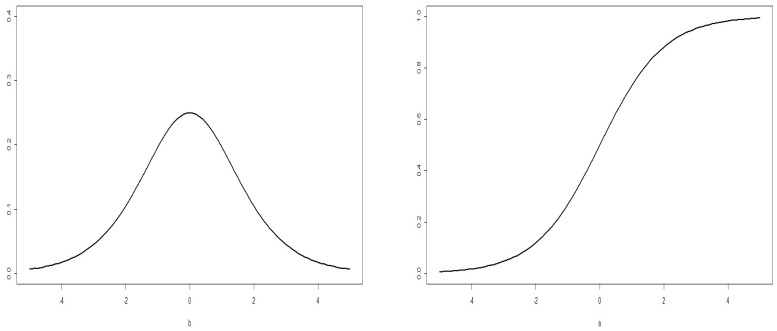
بنابراین، زمانی که متغیر پاسخ دارای توزیع برنولی با احتمال موفقیت p باشد و تابع پیوند لجیت، $\eta = \log \frac{p}{۱-p}$، اختیار شود، مدل رگرسیون لجستیک حاصل می‌شود. بنابراین

$$E(Y|\mathbf{X}=\mathbf{x}) = P(Y = ۱|\mathbf{X}=\mathbf{x}) = p(\mathbf{x}) = \frac{\exp(\beta_۰ + \sum_{i=۱}^{k} \beta_i x_i)}{۱ + \exp(\beta_۰ + \sum_{i=۱}^{k} \beta_i x_i)}$$

و

$$P(Y = ۰|\mathbf{X}=\mathbf{x}) = ۱ - p(\mathbf{x}) = \frac{۱}{۱ + \exp(\beta_۰ + \sum_{i=۱}^{k} \beta_i x_i)}.$$

مدل رگرسیون لجستیک دارای محدودیت $۰ \leq E(Y|\mathbf{X}=\mathbf{x}) \leq ۱$ است. شکل ۱.۱ نمودار تابع توزیع و تابع چگالی لجستیک را نشان می‌دهد. در رگرسیون لجستیک ارتباط بین متغیر پاسخ با متغیرهای تبیینی، غیرخطی و شکل تابع آن به صورت S است که به‌طور یکنواخت افزایش (کاهش) می‌یابد (شکل ۱.۱ قسمت (آ) را ببینید). رگرسیون لجستیک کاربردهای فراوانی دارد. به‌عنوان مثال، در مطالعات کلینیکی در پزشکی وقتی پاسخ دو سطحی است، مقایسه تیمارها توسط این مدل انجام می‌شود. معمولاً برای برآورد پارامترها از روش درستنمایی ماکسیمم استفاده می‌شود (مونت گومری، ۲۰۱۵).



(ب) (آ)

شکل ۱.۱: (آ) نمودار تابع توزیع لجستیک، (ب) نمودار تابع چگالی لجستیک

۴.۳.۱ رگرسیون پرابیت

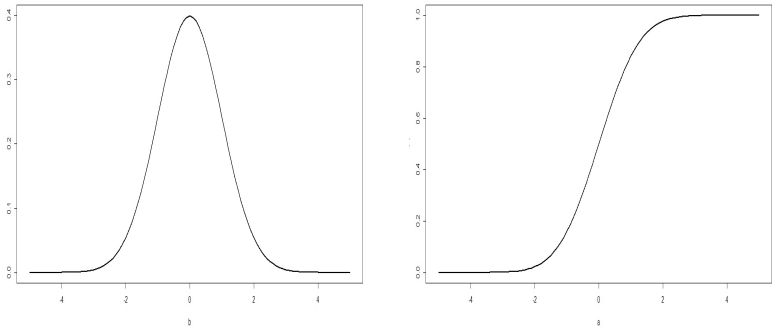
رگرسیون پرابیت نیز برای مدل‌بندی متغیرهای پاسخ دو سطحی با فرض این‌که تابع پیوند پرابیت است، به کار می‌رود. این مدل نسبت به مدل لجستیک از انعطاف‌پذیری کمتری برخوردار است و به اندازه آن طرفدار ندارد.

زمانی که متغیر پاسخ دارای توزیع برنولی با احتمال موفقیت p باشد و تابع پیوند، معکوس تابع توزیع تجمعی نرمال استاندارد، $\eta = \Phi^{-۱}(p)$، اختیار شود، مدل خطی تعمیم‌یافته را مدل پرابیت می‌نامند.

بنابراین در این مدل می‌توان نوشت

$$E(Y|\mathbf{X}=\mathbf{x}) = p(\mathbf{x}) = \Phi(\mathbf{x}_i'\beta) = \Phi(\beta_\circ + \sum_{i=1}^{k}\beta_i x_i)$$

که در آن $\Phi(\cdot)$ تابع توزیع نرمال استاندارد است (آلبرت و چیب، ۱۹۹۳). برآورد پارامترهای این مدل نیز معمولاً به روش درستنمایی ماکسیمم صورت می‌پذیرد. شکل ۲.۱ نمودار تابع توزیع و تابع چگالی پشتیبان مدل پرابیت را نشان می‌دهد.



(ب) (آ)

شکل ۲.۱: (آ) نمودار تابع توزیع نرمال استاندارد، (ب) نمودار تابع چگالی نرمال استاندارد

۵.۳.۱ رگرسیون Cloglog

تابع پیوند دیگر برای مدل‌بندی پاسخ‌های دو سطحی، تابع پیوند Cloglog با ضابطه

$$\eta = \log[-\log[1-p]]$$

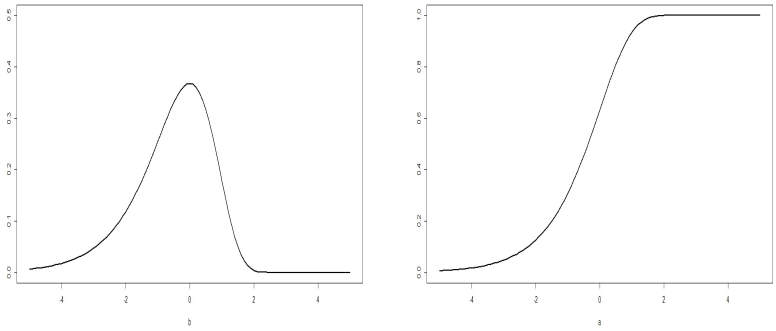
است. مشابه این تعریف، تابع پیوند loglog نیز به صورت $\eta = -\log[-\log[p]]$ معرفی می‌شود که در این پایان‌نامه از هر دو نسخه آن استفاده می‌کنیم. این تابع پیوند از توزیع گامبل[۳۸] نتیجه می‌شود که در فصل دوم آن را معرفی خواهیم کرد. از رگرسیون Cloglog به اندازه دو رگرسیون بالا استفاده نمی‌شود و بیشترین کاربرد آن در تحلیل بقا[۳۹] است. منحنی پاسخ آن S شکل است که در نزدیکی صفر به آرامی کاهش و در نزدیکی ۱ به شدت افزایش می‌یابد. مدل را می‌توان به صورت

$$E(Y|\mathbf{X}=\mathbf{x}) = p(\mathbf{x}) = 1 - \exp(-\exp(\eta))$$

نشان داد. برخلاف پرابیت و لجستیک، مدل Cloglog نامتقارن است و اغلب زمانی به کار برده می‌شود که احتمال یک رویداد بسیار کم یا بسیار زیاد باشد. شکل ۳.۱ نمودار تابع توزیع و تابع چگالی توزیع پشتیبان مدل Cloglog را نشان می‌دهد.

[۳۸] Gumbel Distribution

[۳۹] Survival Analysis



(ب) (آ)

شکل ۳.۱: (آ) نمودار تابع توزیع Cloglog، (ب) نمودار تابع چگالی Cloglog

۴.۱ رهیافت‌های استنباط آماری

مجموعه روش‌های آماری که برای به‌دست آوردن نتایج در مورد ویژگی‌های مورد نظر داده‌ها، به کار می‌روند را استنباط آماری گویند. مهم‌ترین هدف استنباط آماری کسب اطلاعات، بر اساس داده‌های مشاهده‌شده، در مورد موضوع یا مسأله مورد بررسی است. دو رهیافت جامع برای انجام استنباط آماری وجود دارند:

- رهیافت کلاسیک[۴۰]، که به‌طور عمده مبتنی بر روش درستنمایی[۴۱] پیاده‌سازی می‌شوند.
- رهیافت بیزی[۴۲].

۱.۴.۱ رهیافت کلاسیک استنباط آماری

در آمار کلاسیک، پارامترهای مدل کمیت‌هایی ثابت ولی معمولاً 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. درصورتی‌که عدم حتمیت وقوع پیشامد به دلیل عدم آگاهی در مورد آن پیشامد باشد، نمی‌توان خواص احتمالی پارامترها را به دست آورد. بنابراین اصل تکرارپذیری[۴۳] در این رهیافت یکی از اصول پایه‌ای محسوب می‌شود. به‌عنوان مثال، در رهیافت کلاسیک، اگر یک فاصله اطمینان ۹۵٪ برای میانگین نامعلوم جامعه $[-۱, ۱]$ باشد، به این معنی نیست که احتمال این که میانگین در این فاصله قرار گیرد، برابر ۹۵٪ است؛ بلکه بدین معنی است که اگر به دفعات زیادی نمونه‌گیری کنید و

[۴۰] Classical Approach

[۴۱] Likelihood-Based

[۴۲] Bayesian

[۴۳] Repeated Measure Principel

فاصله اطمینان ۹۵٪ بسازید، انتظار می‌رود در ۹۵٪ موارد این فاصله‌ها شامل میانگین باشند. در ادامه به معرفی رهیافت استنباط آماری مبتنی بر روش درستنمایی می‌پردازیم.

روش درستنمایی ماکسیمم

یکی از روش‌های معمول برای برآورد پارامترهای نامعلوم مدل‌های آماری، روش برآورد درستنمایی ماکسیمم نام دارد که به اختصار به‌عنوان MLE یاد می‌شود. این روش شاید متداول‌ترین روش در بین استفاده‌کنندگان از آمار باشد که نخستین بار توسط گاوس (۱۸۲۱) به کار گرفته شد و پس از آن به صورت گسترده‌تری در سال ۱۹۲۵ توسط فیشر در انتقاد از روش گشتاوری[۴۴] مورد استفاده قرار گرفت. برای معرفی این روش، ابتدا تابع درستنمایی را تعریف می‌کنیم.

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

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

ذکر چند نکته در ادامه ضروری است:

- تابع درستنمایی $L(\boldsymbol{\theta})$ لزوماً نسبت به $\boldsymbol{\theta}$ مشتق‌پذیر نیست.
- تابع درستنمایی $L(\boldsymbol{\theta})$ یک تابع چگالی احتمال (تابع احتمال) نیست، بلکه متناسب با تابع چگالی احتمال است.
- اگر $\mathbf{X}_۱, \mathbf{X}_۲, \ldots, \mathbf{X}_n$ یک نمونه تصادفی nتایی از توزیعی از خانواده چگالی‌های $\{f(\mathbf{x}|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$

باشد، آن‌گاه

$$L(\boldsymbol{\theta}) = \prod_{i=۱}^{n} f(\mathbf{x}_i|\boldsymbol{\theta}).$$

تعریف ۲.۴.۱ (برآوردگر درستنمایی ماکسیمم)**.** اگر $\delta(\mathbf{X})$ برآوردگری برای $\boldsymbol{\theta}$ باشد، به‌طوری‌که

$$i) \quad P_{\boldsymbol{\theta}}(\delta(\mathbf{X}) \in \Theta) = ۱ \qquad \forall \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$$

$$ii) \quad L(\delta(\mathbf{X})) \geq L(\boldsymbol{\theta}) \qquad \forall \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$$

آن‌گاه $\delta(\mathbf{X})$ به‌عنوان یک برآوردگر درستنمایی ماکسیمم $\boldsymbol{\theta}$ تعریف می‌شود. معمولاً برآوردگر درستنمایی ماکسیمم $\boldsymbol{\theta}$ را با $\hat{\boldsymbol{\theta}}$ نمایش می‌دهند و لزومی ندارد که منحصر به فرد باشد. همچنین بر اساس تعریف $\hat{\boldsymbol{\theta}}$ (پارسیان، ۱۳۹۱) داریم

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

[۴۴] Method of Moment

روش‌های معمول رهیافت کلاسیک، روش‌های مبتنی بر درستنمایی هستند که برای محاسبه تابع درستنمایی در رده مدل‌های پیچیده، مانند مدل‌های آمیخته خطی[۴۵] و آمیخته خطی تعمیم‌یافته[۴۶]، حل انتگرال با بعد بالا را شامل می‌شوند که دارای محاسبات زمان‌بر و دشواری است. بنابراین، این روش که به صورت شهودی مورد توجه است، بهترین روش نیست و نمی‌تواند همیشه مورد استفاده قرار گیرد. به این دلیل، رهیافت بیزی که استنباط در آن مبتنی بر توزیع پسین است، می‌تواند یک جایگزین مناسب باشد.

۲.۴.۱ رهیافت بیزی استنباط آماری

در رهیافت بیزی، پارامتر واقعی و نامعلوم $\boldsymbol{\theta}$ به‌عنوان تحققی از یک توزیع احتمالی به نام توزیع پیشین در نظر گرفته می‌شود. این در حالی است که آماردانان کلاسیک بر این عقیده‌اند که پارامتر نامعلوم را به سختی می‌توان به صورت یک متغیر تصادفی در نظر گرفت. در ضمن تعیین توزیع احتمالی برای چنین متغیری نیز به سلیقه و عقیده آماردانان بیزی بستگی پیدا می‌کند. آماردانان بیزی بر این عقیده استوارند که با ترکیب اطلاعات شخصی توسط توزیع پیشین و اطلاعات نهفته در داده‌ها توسط تابع درستنمایی، می‌توان به واقعیت نزدیک گردید. در استنباط بیزی اگر میزان اطلاعات حاصل از داده‌ها افزایش یابد، مثلاً حجم نمونه بزرگ انتخاب شود، اثر توزیع پیشین ناچیز می‌شود. پس انتظار می‌رود که عملکرد برآوردگرهای بیزی و کلاسیک یکسان شوند. اما اگر نمونه کوچک باشد، ممکن است اختلاف چشمگیری به وجود آید که در این صورت استنباط بیزی می‌تواند قوی‌تر از کلاسیک عمل کند. همچنین در این دیدگاه می‌توان با استفاده از داده‌ها، توزیع پیشین را به‌روز کرد و توزیع احتمال جدید را که توزیع پسین نام دارد، به‌دست آورد. مکانیسم این به‌روز شدن بر اساس قضیه بیز صورت می‌پذیرد.

قضیه ۳.۴.۱ (قضیه بیز). *فرض کنید $A_1, A_2, \ldots, A_n$ یک افراز از فضای نمونه و E یک پیشامد دلخواه باشد. در این صورت برای هر $i = 1, \ldots, n$ می‌توان نوشت*

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_{i=1}^{n} P(E|A_i)P(A_i)}.$$

این قضیه یک روش رسمی برای به‌روز رساندن شانس A_i از $P(A_i)$ به $P(A_i|E)$، وقتی E مشاهده شده است، می‌باشد. $P(A_i)$ را احتمال پیشین و $P(A_i|E)$ را احتمال پسین می‌نامند (جفریز، ۱۹۳۶).

این قضیه برای تابع چگالی (مدل آماری) به صورت زیر تبدیل می‌شود:

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

که در آن $\pi(\boldsymbol{\theta})$ تابع (چگالی) احتمال توزیع پیشین و $\pi(\boldsymbol{\theta}|\mathbf{x})$ تابع (چگالی) احتمال توزیع پسین θ می‌باشند. از آن‌جا که مقدار $\int_{\Theta} f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ مستقل از θ بوده و حاصل آن یک عدد ثابت است، در محاسبه $\pi(\boldsymbol{\theta}|\mathbf{x})$ معمولاً از محاسبه انتگرال خودداری کرده و توزیع پسین را به صورت زیر بیان می‌کنیم:

$$\pi(\boldsymbol{\theta}|\mathbf{x}) \propto f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

[۴۵] Linear Mixed Models

[۴۶] Generalized Linear Mixed Models

توزیع پیشین

توزیع پیشین $\boldsymbol{\theta}$ را با $\pi(\boldsymbol{\theta})$ نشان می‌دهند که بیان‌کننده اعتقادات و اطلاعات پیشین آماردانان درباره مقدار نامعلوم $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$ 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. اگر واریانس توزیع پیشین خیلی بزرگ باشد، آن‌گاه توزیع پیشین در برابر درستنمایی مغلوب[۴۷] خواهد شد، زیرا هیچ اطلاعاتی در مورد پارامتر در اختیار آماردان نمی‌گذارد.

تعریف ۴.۴.۱ (توزیع‌های پیشین سره[۴۸]). توزیع پیشین با تابع چگالی احتمال $\pi(\boldsymbol{\theta})$ را سره گوییم، هرگاه

$$\int_{\Theta} \pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} < \infty.$$

تعریف ۵.۴.۱ (توزیع‌های پیشین ناسره[۴۹]). توزیع پیشین با تابع چگالی احتمال $\pi(\boldsymbol{\theta})$ را ناسره گوییم، هرگاه

$$\int_{\Theta} \pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = +\infty.$$

اگر توزیع پیشین سره باشد، توزیع پسین حتماً سره خواهد بود؛ اما اگر ناسره باشد، ممکن است توزیع پسین سره شود. توجه داشته باشید زمانی می‌توان از توزیع پیشین ناسره برای $\boldsymbol{\theta}$ استفاده کرد که توزیع پسین به‌دست آمده سره باشد. در غیر این صورت استنباط آماری غیرممکن خواهد بود (غلامی، ۱۳۹۰). 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 ۶.۴.۱ (توزیع پیشین مزدوج[۵۰]). $\pi(\boldsymbol{\theta})$ را برای $f(\mathbf{x}|\boldsymbol{\theta})$ مزدوج گوییم، هرگاه توزیع پسین آن متعلق به خانواده توزیع پیشین $\pi(\boldsymbol{\theta})$ باشد (رابرت، ۲۰۰۷).

تعریف ۷.۴.۱ (توزیع پیشین ناآگاهی‌بخش[۵۱]). یکی از مهم‌ترین پیشین‌ها، پیشین ناآگاهی‌بخش می‌باشد 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. چنین توزیعی را توزیع پیشین ناآگاهی‌بخش می‌نامند. پیشین ناآگاهی‌بخش را به صورت $\pi(\boldsymbol{\theta}) = k$، $(k > 0)$ بیان می‌کنند که انتخاب k اختیاری است. با انتخاب پیشین ناآگاهی‌بخش توزیع پسین به صورت

$$\pi(\boldsymbol{\theta}|\mathbf{x}) \propto kf(\mathbf{x}|\boldsymbol{\theta}) = kL(\boldsymbol{\theta}) \qquad (۴.۱)$$

[۴۷] Dominated
[۴۸] Proper
[۴۹] Improper
[۵۰] Conjugate Prior
[۵۱] Noninformative Prior

بیان می‌شود. اگر متغیر مورد بررسی روی دامنه فضای پارامتر نامحدود باشد، در آن صورت تخصیص یک مقدار ثابت یکنواخت به‌عنوان توزیع پیشین، توزیع پسین ناسره ایجاد می‌کند. نحوه انتخاب k در رابطه (۴.۱) به روش‌های مختلفی صورت می‌پذیرد (برگر و برنارد، ۱۹۹۲). به‌عنوان مثال، می‌توان به پیشین ناآگاهی‌بخش یکنواخت[۵۲] و پیشین ناآگاهی‌بخش جفریز[۵۳] اشاره کرد.

پیشین ناآگاهی‌بخش یکنواخت

برای نخستین بار لاپلاس (۱۸۱۲) با انتخاب $k = ۱$ پیشین ناآگاهی‌بخش را به صورت $\pi(\boldsymbol{\theta}) = ۱$ به کار برد که به پیشین ناآگاهی‌بخش یکنواخت معروف است.

پیشین ناآگاهی‌بخش جفریز

جفریز (۱۹۶۱) توزیع پیشین ناآگاهی‌بخش را که بر اساس ماتریس اطلاع فیشر[۵۴] پایه‌ریزی شده بود ارائه داد. اگر

$$I(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))$$

ماتریس اطلاع فیشر باشد، که در آن

$$I_{ij}(\boldsymbol{\theta}) = -E_{\boldsymbol{\theta}}\left[\frac{\partial^۲}{\partial\boldsymbol{\theta}_i\,\partial\boldsymbol{\theta}_j}\log f(\mathbf{x}|\boldsymbol{\theta})\right]$$

آن‌گاه توزیع پیشین جفریز به صورت

$$\pi(\boldsymbol{\theta}) \propto \sqrt{\det I(\boldsymbol{\theta})}$$

تعریف می‌شود.

توزیع پسین

در ساده‌ترین مسائل استنباط آماری، محاسبه توزیع پسین مستلزم محاسبه انتگرال‌های چندگانه است؛ اما انجام بسیاری از انتگرال‌های چندگانه مشکل است و روش‌های متعارف برای حل آن‌ها وجود ندارند و تنها راه، حل تقریبی آن‌ها است. یک رده کلی از روش‌های تقریب توزیع‌های پسین، روش‌های مبتنی بر نمونه‌گیری مانند روش‌های رد و پذیرش[۵۵] (رابرت و کسلا، ۲۰۱۰)، نمونه‌گیری نقاط مهم[۵۶] و نمونه‌گیری MCMC می‌باشد. 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 روش‌های نمونه‌گیری را به اختصار معرفی می‌کنیم.

---

[۵۲] Noninformative Uniform Prior

[۵۳] Noninformative Jeffreys Prior

[۵۴] Fisher Information

[۵۵] Accept-Reject Methods

[۵۶] Importance Sampling

۵.۱ 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 آن‌ها از اعداد تصادفی برای شبیه‌سازی استفاده می‌شود، روش‌های شبیه‌سازی مونت کارلویی نامیده می‌شوند. کلمه مونت کارلویی در حقیقت برگرفته از مراکز تفریحی مشهور در بندر مونت‌کارلو در کشور فرانسه است که در آن‌ها برای تصادفی کردن بازی‌ها از اعداد تصادفی استفاده می‌شده است (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 را بر روش‌های عددی ترجیح می‌دهیم. اول این‌که روش مونت کارلویی مشکل روش‌های عددی در حل انتگرال‌ها را ندارد؛ به عبارت دیگر روش‌های مونت کارلویی به نواحی چگال‌تر تابع زیر انتگرال، وزن بیشتری در محاسبه تقریب می‌دهد در حالی که در روش‌های عددی همه زیرنواحی 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 معمولاً با افزایش بعد کاهش می‌یابد، درحالی‌که روش‌های مونت کارلویی از نظر دقت، مستقل از بعد می‌باشند (رابرت و کسلا، ۲۰۰۴؛ ۲۰۱۰).

ایده اصلی روش‌های مونت کارلویی، تبدیل انتگرال به یک امید ریاضی بر اساس یک تابع چگالی احتمال مشخص، تولید نمونه تصادفی از این تابع چگالی و استفاده از قانون قوی اعداد بزرگ[۵۷] (گات، ۲۰۰۵) برای تقریب این امید ریاضی است. در واقع مسأله اصلی محاسبه انتگرال زیر است:

$$J = E_f[h(X)] = \int_{\chi} h(x) f(x) \mathrm{d}x$$

که در آن χ تکیه‌گاه متغیر تصادفی X، یک یا چندبعدی است. تابع $f(\cdot)$ یک تابع چگالی، دارای شکل بسته یا به‌طور جزئی بسته[۵۸] است 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 مقادیر 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

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

علاوه بر این، دقت این تقریب مانند دقت یک برآورد آماری است. بنابراین، به محض این که نمونه $(x_1, \ldots, x_m)$ از توزیع $f(\cdot)$ تولید شود، از آن می‌توان برای محاسبه کمیت‌های مورد علاقه توزیع با

[۵۷] Strong Law of Large Number

[۵۸] Partially Closed Form

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-هستینگز (MH) و نمونه‌گیر گیبز از جمله روش‌های تولید نمونه غیرمستقیم هستند که در ادامه به معرفی آن‌ها می‌پردازیم.

۱.۵.۱ روش‌های زنجیر مارکوف مونت کارلویی

روش‌های زنجیر مارکوف مونت کارلو، یک چارچوب کلی از مجموعه روش‌هایی است که برای انتگرال‌گیری به روش مونت کارلویی، توسط متروپلیس و همکاران (۱۹۵۳) معرفی شد و پس از آن توسط هستینگز (۱۹۷۰) به منظور تمرکز روی مسائل آماری تطبیق و تعمیم یافت. این روش‌ها در کنار افزایش قدرت پردازش و سرعت محاسبات کامپیوترها، همواره در حال تکمیل و گسترش بوده‌اند. روش‌های نمونه‌گیری از زنجیرهای مارکوف با توزیع مانای[۵۹] $f(\cdot)$، برای تولید نمونه غیرمستقیم از $f(\cdot)$ طراحی شده‌اند که در ابعاد بالا کارایی خود را حفظ می‌کنند. در این روش‌ها توزیعی مانند $f(\cdot)$، که اغلب آن را توزیع هدف[۶۰] می‌نامند، در نظر گرفته می‌شود. اگر $f(\cdot)$ به اندازه کافی پیچیده باشد که نتوان مستقیماً از آن نمونه‌گیری کرد، یک روش نمونه‌گیری غیرمستقیم، ساختن یک زنجیر مارکوف تحویل‌ناپذیر[۶۱] و بازگشتی[۶۲] (ایبلدریم، ۲۰۱۲) با توزیع مانای $f(\cdot)$ است که با نمونه‌گیری از این زنجیر و استفاده از قضیه ارگودیک[۶۳] (بیرخوف، ۱۹۴۲) می‌توان به برآورد تابع دلخواه از توزیع مورد نظر دست یافت.

تعریف ۱.۵.۱ (زنجیر مارکوف)**.** یک زنجیر مارکوف $\{X^{(t)}\}$ دنباله‌ای از متغیرهای تصادفی وابسته $(X^{(0)}, X^{(1)}, \ldots, X^{(t)})$ است، به‌طوری‌که توزیع احتمال $X^{(t)}$ به شرط مفروض بودن متغیرهای گذشته، فقط به $X^{(t-1)}$ وابسته است. یعنی

$$X^{(t-1)}|X^{(0)}, X^{(1)}, \ldots, X^{(t)} \sim K(X^{(t)}|X^{(t-1)})$$

که در آن توزیع احتمال شرطی $K(\cdot|\cdot)$، هسته مارکوف[۶۴] نامیده می‌شود.

از آن‌جایی که نمونه تولیدشده بر اساس زنجیر مارکوف وابسته است، فاقد کیفیت یک نمونه مستقل با حجم مشابه است. کیفیت چنین نمونه‌ای را می‌توان از طریق کمیتی به نام حجم نمونه مؤثر[۶۵] (ESS) سنجید. هرچه مقدار ESS محاسبه‌شده برای هر پارامتر به تعداد نمونه‌های تولیدشده نزدیک‌تر باشد، نمونه‌های تولیدشده از کیفیت بالاتری برخوردار هستند.

---

[۵۹] Stationary Distribution

[۶۰] Target Distribution

[۶۱] Ergodic Markov Chain

[۶۲] Recurrent

[۶۳] Ergodic Theorem

[۶۴] Markov Kernel

[۶۵] Effective Sample Size

تعریف ۲.۵.۱. (سالین، ۲۰۱۱) اگر n تعداد نمونه‌های تولیدشده از زنجیر باشد، آن‌گاه ESS برای پارامتر مورد نظر به صورت

$$ESS = \frac{n}{۱ + ۲\sum_{k=۱}^{\infty} \rho_k}$$

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

اولین قدم در الگوریتم MCMC، با فرض داشتن چگالی هدف $f(\cdot)$، به‌دست آوردن هسته مارکوف K با توزیع مانای f است. دومین قدم تولید زنجیر مارکوف $X^{(t)}$ با استفاده از این هسته است که توزیع حدی زنجیر $X^{(t)}$ چگالی هدف $f(\cdot)$ است که با افزایش t زنجیر به توزیع هدف همگرا می‌شود. در این حالت، زنجیر تولیدشده نسبت به توزیع هدف باید:

- برگشت‌پذیر باشد، به این معنی که زنجیر به هر مجموعه دلخواه غیرقابل اغماض، به دفعات نامتناهی بازخواهد گشت.

- تحویل‌ناپذیر باشد، به این معنی که هسته K اجازه حرکت بر روی تمام وضعیت‌های فضای حالت زنجیر را دارد. به عبارت ساده‌تر، صرف‌نظر از مقدار اولیه $X^{(۰)}$، دنباله $\{X^{(t)}\}$ با احتمال مثبت به هر ناحیه از فضای حالت می‌تواند سر بزند. یکی از مهم‌ترین ویژگی‌های زنجیرهای مارکوف وجود توزیع احتمال مانا است، یعنی توزیع احتمال f وجود دارد که اگر

$$X^{(t)} \sim f$$

آن‌گاه

$$X^{(t+۱)} \sim f.$$

این ویژگی شرط تحویل‌ناپذیری را بر هسته مارکوف K تحمیل می‌کند.

- ارگودیک باشد، به این معنی که توزیع حدی $X^{(t)}$ به ازای هر مقدار اولیه $X^{(۰)}$، همان توزیع مانای زنجیر یعنی f است.

اگر زنجیر دارای این شرایط نباشد، زنجیر مفیدی نخواهد بود؛ زیرا تقریب‌های به‌دست آمده از این زنجیر در شرایطی معتبر هستند و به مقادیر نظری خود همگرا می‌شوند که زنجیر ارگودیک، تحویل‌ناپذیر و بازگشتی باشد. در این صورت، در مراحل انتهایی زنجیر به یک توزیع پایا می‌رسیم. بنابراین، با توجه به شرط ارگودیک بودن زنجیر مارکوف، اگر $t \to \infty$، آن‌گاه

$$\frac{۱}{T}\sum_{t=۱}^{T} h(X^{(t)}) \longrightarrow E_f(h(X))$$

که به آن قضیه ارگودیک نیز می‌گویند و طبق قانون قوی اعداد بزرگ برای رهیافت MCMC نیز قابل کاربرد است (رابرت و کسلا، ۲۰۰۴). در ادامه دو روش مهم که به کمک زنجیرهای مارکوف از توزیع‌های پیچیده نمونه‌گیری می‌کنند را به اختصار توضیح می‌دهیم.

الگوریتم متروپلیس-هستینگز

اساس کار با روش‌های زنجیر مارکوف مونت‌کارلویی به این صورت است که با فرض داشتن چگالی هدف، $f(\cdot)$، یک هسته مارکوف K را با توزیع مانای f می‌سازیم. پس از آن یک زنجیر مارکوف $\{X^{(t)}\}$ را با استفاده از این هسته به‌گونه‌ای تولید می‌کنیم که تابع چگالی توزیع حدی $X^{(t)}$، f باشد و سپس کمیت‌های مورد نظر را که در قالب انتگرال بیان می‌شوند، مانند میانگین و واریانس، توسط قضیه ارگودیک به‌دست می‌آوریم. ساختن هسته مارکوف K که به چگالی دلخواه $f(\cdot)$ نسبت داده می‌شود، ممکن است کار دشواری باشد. روش‌هایی وجود دارند که می‌توان به کمک آن‌ها هسته‌هایی که عمومیت بیشتری دارند را برای چگالی هدف $f(\cdot)$، به‌دست آورد. الگوریتم متروپلیس-هستینگز نمونه‌ای از این روش‌ها است. در این روش برای تولید نمونه از چگالی هدف، چگالی شرطی $q(y|x)$ را که چگالی پیشنهادی[۶۶] نامیده می‌شود، انتخاب می‌کنیم. توزیع پیشنهادی باید شکل واضح و صریحی داشته باشد تا شبیه‌سازی از آن به روش‌های معمول ساده باشد و تکیه‌گاه آن باید به قدری وسیع باشد که کل تکیه‌گاه f را پوشش دهد. الگوریتم MH را می‌توان به صورت زیر ارایه داد:

با داشتن $x^{(t)}$، برای $t = ۱, ۲, \ldots$

- برای مقدار جاری[۶۷] $x^{(t)}$، یک مقدار پیشنهادی را از توزیع پیشنهادی با تابع چگالی $q(y|x^{(t)})$ تولید کنید. یعنی
$$y_t \sim q(y|x^{(t)}).$$
- نسبت زیر را محاسبه کنید:
$$\frac{f(y_t)}{f(x^{(t)})}\frac{q(x^{(t)}|y_t)}{q(y_t|x^{(t)})}.$$
این الگوریتم فقط به نسبت‌های $\frac{f(y_t)}{f(x^{(t)})}$ و $\frac{q(x^{(t)}|y_t)}{q(y_t|x^{(t)})}$ وابسته است.
- مقدار y_t را به‌عنوان مقدار بعدی زنجیر، با احتمال $\rho(x^{(t)}, y_t)$، بپذیرید و قرار دهید
$$X^{(t+۱)} = y_t.$$
در غیر این صورت قرار دهید
$$X^{(t+۱)} = X^{(t)}$$
که در آن
$$\rho(x^{(t)}, y_t) = \min\{۱, \frac{f(y_t)}{f(x^{(t)})}\frac{q(x^{(t)}|y_t)}{q(y_t|x^{(t)})}\}.$$

به $\rho(x^{(t)}, y_t)$ احتمال پذیرش متروپلیس-هستینگز گویند. اگر توزیع‌های پیشنهادی متقارن باشند، یعنی $q(y|x) = q(x|y)$، آن‌گاه $\rho(x^{(t)}, y_t) = \min\{۱, \frac{f(y_t)}{f(x^{(t)})}\}$ و بنابراین، احتمال پذیرش مستقل

[۶۶] Proposal Density

[۶۷] Current Value

از $q(\cdot|\cdot)$ خواهد بود. عملکرد الگوریتم معمولاً برحسب نرخ پذیرش آن ارزیابی می‌شود که نرخ پذیرش الگوریتم، میانگین احتمال پذیرش بر روی همه تکرارها است. یعنی

$$\bar{\rho} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \rho(X^{(t)}, Y_t) = \int \rho(x,y) f(x) q(y|x)\, \mathrm{d}y\, \mathrm{d}x.$$

کارایی روش MH به تولید نمونه از تابع چگالی $q(y|x)$ وابسته است. در واقع یک نمونه خوب وقتی تولید می‌شود که تابع چگالی پیشنهادی یک تقریب مناسب برای تابع چگالی هدف باشد (رابرت و کسلا، ۲۰۱۰). دو حالت خاص الگوریتم متروپلیس-هستینگز مستقل[۶۸] و الگوریتم متروپلیس-هستینگز قدم زدن تصادفی[۶۹] را در ادامه ذکر می‌کنیم.

◄ الگوریتم متروپلیس-هستینگز مستقل

در الگوریتم MH در حالتی که تابع چگالی پیشنهادی $q(\cdot|\cdot)$ مستقل از $X^{(t)}$ باشد، آن را الگوریتم MH مستقل گویند. در این حالت $q(x|y) = g(y)$ است. با شرط داشتن $x^{(t)}$،

- $y_t \sim g(y)$ را تولید کنید.
- مقدار y_t را به‌عنوان مقدار جدید $X^{(t+1)}$ با احتمال $\min\{1, \frac{f(y_t)g(x^{(t)})}{f(x^{(t)})g(y_t)}\}$ بپذیرید و قرار دهید $X^{(t+1)} = y_t$.

در غیر این صورت

$$X^{(t+1)} = X^{(t)}.$$

بنابراین، مقدار پیشنهادی از یک توزیع مورد علاقه که مستقل از مقدار جاری است، نتیجه می‌شود. برای اطلاع از ویژگی‌های این نوع الگوریتم به رابرت و کسلا (۲۰۱۰) مراجعه کنید.

◄ الگوریتم متروپلیس-هستینگز قدم زدن تصادفی

در این الگوریتم مقدار جدید پیشنهادی بر طبق مقدار شبیه‌سازی‌شده قبلی تولید می‌شود، که در آن یک جستجوی محلی[۷۰] حول یک همسایگی از مقدار جاری زنجیر مارکوف صورت می‌گیرد. در این ایده، برای تولید بعدی زنجیر، الگوریتم نمونه تولیدشده قبلی را به حساب می‌آورد. به این معنی که در همسایگی موقعیت جاری زنجیر، کاوش موضعی بیشتری را انجام می‌دهد. اجرای این ایده، شبیه‌سازی y_t از رابطه $y_t = X^{(t)} + \epsilon_t$ است، که در آن ϵ_t یک خطای تصادفی مستقل از $X^{(t)}$ با تابع چگالی $g(\cdot)$ است. در این حالت تابع چگالی توزیع پیشنهادی عبارت است از

$$q(y_t|x^{(t)}) = g(y_t - x^{(t)}).$$

در صورتی‌که تابع چگالی $g(\cdot)$ حول صفر متقارن باشد، که معمولاً این‌طور است، زنجیر مارکوف متناظر با آن یک قدم زدن تصادفی است. با داشتن $x^{(t)}$ الگوریتم به صورت زیر است:

---

[۶۸] Independent Metropolis-Hastings

[۶۹] Random Walk Metropolis-Hastings

[۷۰] Local Exploration

- برای مقدار جاری زنجیر یک مقدار پیشنهادی را از توزیع $g(\cdot)$ تولید کنید. یعنی
$$y_t \sim g(y_t - x^{(t)}).$$
- نسبت زیر را محاسبه کنید:
$$\frac{f(y_t)}{f(x^{(t)})}.$$
- مقدار y_t را به‌عنوان مقدار جدید $X^{(t+1)}$ با احتمال $\min\{1, \frac{f(y_t)}{f(x^{(t)})}\}$ بپذیرید و قرار دهید
$$X^{(t+1)} = y_t.$$

در غیر این صورت

$$X^{(t+1)} = X^{(t)}.$$

متذکر می‌شویم که احتمال پذیرش به $g(\cdot)$ وابسته نیست. به این معنی که، برای یک جفت مفروض $(x^{(t)}, y_t)$، احتمال پذیرش چه y_t از توزیع نرمال تولید شود چه از توزیع کوشی، یکسان خواهد بود. اما باید توجه داشت که انتخاب $g(\cdot)$های مختلف منجر به دامنه‌های مختلف مقادیر برای y_tها خواهد شد، در نتیجه نرخ‌های پذیرش متفاوت می‌شوند. بنابراین، می‌توان گفت انتخاب $g(\cdot)$ بر رفتار الگوریتم تأثیر دارد.

در این الگوریتم توزیع پیشنهادی ϵ_t یک پارامتر مقیاس دارد که انتخاب آن در عملکرد الگوریتم بسیار مؤثر است. انتخاب بهینه برای این پارامتر طوری است که نرخ پذیرش الگوریتم در فاصله $[۰/۱۵, ۰/۵]$ قرار گیرد (گلمن و همکاران، ۲۰۰۳).

الگوریتم نمونه‌گیر گیبز

یکی دیگر از روش‌های نمونه‌گیری MCMC، نمونه‌گیر گیبز است. نام نمونه‌گیر گیبز از مقاله معروف گمان و گمان (۱۹۸۴) گرفته شده است؛ اما استفاده شایع از این الگوریتم با کار گلفاند و اسمیت (۱۹۹۰) شروع شد که از حل این الگوریتم برای حل مسائلی، در استنباط بیزی که قبلاً بدون حل مانده بودند، استفاده کردند. برای اجرای این روش باید بتوانیم توابع چگالی شرطی کامل[۷۱] هر مؤلفه $\mathbf{X}$ را به شرط بقیه مؤلفه‌ها به‌دست آوریم. این الگوریتم زمانی می‌تواند کارا باشد که شبیه‌سازی از چگالی‌های شرطی کامل ساده و وابستگی مؤلفه‌ها کم باشد. دو نسخه دو مرحله‌ای و چند مرحله‌ای این الگوریتم را در ادامه معرفی می‌کنیم.

◄ نمونه‌گیر گیبز دومرحله‌ای

نمونه‌گیر گیبز دومرحله‌ای[۷۲]، با استفاده از توزیع توأم، یک زنجیر مارکوف را تشکیل می‌دهد. اگر دو متغیر X و Y وجود داشته باشند که دارای توزیع توأم $f(x, y)$ باشند، به‌طوری‌که چگالی‌های شرطی متناظرشان $f(x|y)$ و $f(y|x)$ باشند، آن‌گاه الگوریتم یک زنجیر مارکوف (X_t, Y_t) را با شرط مفروض بودن $X_۰ = x_۰$، برای $t = ۱, ۲, \ldots$ بر اساس مراحل زیر تولید می‌کند:

[۷۱] Full Conditional Densities

[۷۲] Two-Stage Gibbs Samples

- $Y_t \sim f_{Y|X}(\cdot|x_{t-1})$
- $X_t \sim f_{X|Y}(\cdot|y_t)$

◄ **نمونه‌گیر گیبز چند مرحله‌ای**

نمونه‌گیر گیبز دو مرحله‌ای را به سادگی می‌توان به یک نسخه چند مرحله‌ای[۷۳] تعمیم داد. فرض کنید متغیر تصادفی $\mathbf{X}$ به صورت $\mathbf{X} = (X_1, \ldots, X_m)$ باشد. مؤلفه‌های X_i ممکن است یک بعدی یا چند بعدی باشند.

الگوریتم نمونه‌گیر گیبز تحت انتقال از $X^{(t)}$ به $X^{(t+1)}$ با تکرار $t = 1, 2, 3, \ldots$ و با فرض $\mathrm{x}^{(t)} = (x_1^{(t)}, x_2^{(t)}, \ldots, x_m^{(t)})$ طی گام‌های زیر اجرا می‌شود:

۱. تولید $X_1^{(t+1)} \sim f_1(x_1|x_2^{(t)}, \ldots, x_m^{(t)})$

۲. تولید $X_2^{(t+1)} \sim f_2(x_2|x_1^{(t+1)}, x_3^{(t)}, \ldots, x_m^{(t)})$

⋮

m. تولید $X_m^{(t+1)} \sim f_m(x_m|x_1^{(t+1)}, \ldots, x_{m-1}^{(t+1)})$

یک ویژگی نمونه‌گیر گیبز این است که چگالی‌های با بعد کوچکتر از بعد چگالی توأم برای شبیه‌سازی مورد استفاده قرار می‌گیرند. بنابراین، حتی مسأله‌ای با ابعاد بالا را می‌توان به چند مسأله با بعدهای کوچک‌تر تقسیم کرد که این مزیت روش نمونه‌گیر گیبز به حساب می‌آید.

۶.۱ معیارهای سنجش همگرایی الگوریتم‌های MCMC

در کنار استفاده‌های بسیاری که از روش‌های نمونه‌گیری MCMC در استنباط بیزی می‌شود، دو موضوع مهم تشخیص همگرایی و ویژگی‌های آمیختگی[۷۴] زنجیر و زمان طولانی محاسبات مطرح هستند.

شبیه‌سازی از یک زنجیر MCMC به دو بخش تقسیم می‌شود: بخش اول، قبل از همگرایی زنجیر و بخش دوم بعد از همگرایی زنجیر است. بخش قبل از همگرایی به دوره سوزاندن[۷۵] مشهور است و به‌عنوان بخشی از زنجیر که در آن نمونه‌های تولیدشده را نمی‌توان به‌عنوان یک معرف خوب از توزیع پسین در نظر گرفت، معرفی می‌شود. پیدا کردن دوره داغیدن و حذف نمونه‌ها باعث صرفه‌جویی در زمان خواهد شد. در بخش بعد از همگرایی، نمونه‌های تولیدشده از $f(\cdot)$ برای استنباط در مورد کمیت‌های مورد علاقه استفاده می‌شوند. این موضوع، نظارت بر همگرایی را با مقایسه استنباط‌های به‌دست آمده از چند دنباله تولیدشده مستقل با نقاط شروع متفاوت، پیشنهاد می‌کند (گلمن و روبین، ۱۹۹۲).

---

[۷۳] Multi-Stage Gibbs Samples

[۷۴] Mixing Properties

[۷۵] Burn-In

بنابراین تشخیص همگرا شدن زنجیر و زمان آن، برای استفاده از نمونه‌های تولیدشده بعد از همگرایی در استنباط بیزی کلیدی است. روش‌ها و معیارهای متعددی برای بررسی همگرایی زنجیرهای MCMC وجود دارند. از جمله آن‌ها می‌توان به روش‌های معرفی‌شده توسط گلمن و روبین (۱۹۹۲)، جی‌وک (۱۹۹۲) و هیدلبرگ و ولچ (۱۹۸۳) اشاره کرد. برای یک زنجیر مشخص، این معیارها توسط توابعی از بسته‌های موجود در نرم‌افزار R، از جمله بسته coda (پلامر و همکاران، ۲۰۰۶) قابل محاسبه هستند. در این پایان‌نامه از دو معیار همگرایی گلمن-روبین[۷۶] و جی‌وک[۷۷] استفاده می‌کنیم.

۱.۶.۱ معیار همگرایی جی‌وک

جهت تشخیص همگرایی زنجیر MCMC، جی‌وک در سال ۱۹۹۲ روشی را بر پایه آزمون برابری میانگین‌ها برای دو قسمت مختلف از زنجیر پیشنهاد کرد. این معیار بر پایه سه مرحله تشکیل می‌شود:

- مرحله اول، انتخاب دو پنجره از زنجیر که هیچ هم‌پوشانی[۷۸] با یکدیگر نداشته باشند. انتخاب این دو پنجره بنا بر پیشنهاد جی‌وک به‌طور معمول ۰/۱ ابتدای زنجیر و ۰/۵ انتهای زنجیر است. از آن‌جایی که در زنجیر مارکوف، فرض بر این است که زنجیر از ابتدا همگرا نشده است، بنابراین طول زنجیر اول را کوتاه در نظر می‌گیرند که فقط شامل مرحله داغیدن باشد. در نتیجه، انتخاب پنجره در روش جی‌وک از اهمیت بالایی برخوردار است. برای زنجیری به طول n، پنجره‌های A و B به صورت زیر تعریف می‌شوند:

$$A = \{t : 1 \leq t \leq n_A\}$$
$$B = \{t : n^* \leq t \leq n\}$$

که به ازای $1 < n_A < n^* < n$ و $n_B = n - n^* + 1$، $\frac{n_A + n_B}{n} < 1$ است.

- مرحله دوم، محاسبه آماره آزمون Z_n است که در ادامه معرفی می‌شود. برای اطمینان از همگرایی زنجیر، آزمون فرضیه به صورت

$$\begin{cases} H_0 : & \text{A و B هم‌توزیع هستند} \\ H_1 : & \text{A و B هم‌توزیع نیستند} \end{cases}$$

در نظر گرفته می‌شود. با فرض این‌که $\theta(X)$ یک تابع باشد، θ^t به صورت زیر تعریف می‌شود:

$$\theta^t = \theta(x^{(t+n_0)})$$

که در آن n_0 برابر با اولین تکراری است که آزمون انجام می‌شود. با توجه به میانگین و واریانس پنجره‌های A و B به صورت

$$\bar{\theta}_A = \frac{1}{n_A} \sum_{t \in A} \theta^t, \qquad Var(\theta_A) = \frac{\hat{S}_\theta^A(0)}{n_A}$$

[۷۶] Gelman-Rubin Convergence Criterion

[۷۷] Geweke Convergence Criterion

[۷۸] Overlap

و

$$\bar{\theta}_B = \frac{1}{n_B}\sum_{t\in B}\theta^t, \qquad Var(\theta_B) = \frac{\hat{S}_\theta^B(0)}{n_B}$$

که در آن $\hat{S}_\theta(0)$ چگالی طیفی[۷۹] در نقطه صفر است، آماره آزمون به صورت زیر محاسبه می‌شود:

$$Z_n = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\frac{\hat{S}_\theta^A(0)}{n_A} + \frac{\hat{S}_\theta^B(0)}{n_B}}}.$$

این آماره بنابر قضیه حد مرکزی، زمانی‌که $n \to \infty$، به‌طور مجانبی به توزیع نرمال استاندارد همگرا می‌شود. بنابراین، فرضیه صفر زمانی رد می‌شود که مقدار Z_n بسیار بزرگ باشد.

- مرحله سوم، پس از رد شدن فرضیه صفر در مرحله دوم، ۰/۱ ابتدای زنجیر را قطع کرده و مراحل اول و دوم تکرار می‌شوند. این عمل قطع کردن برای پذیرش فرضیه صفر، تا زمانی ادامه می‌یابد که نیمی از زنجیر اولیه قطع شده باشد. در غیر این صورت، زنجیر همگرا نیست.

۲.۶.۱ معیار همگرایی هیدلبرگ-ولچ

معیار همگرایی هیدلبرگ-ولچ[۸۰]، توسط هیدلبرگ و ولچ (۱۹۸۳) پیشنهاد شد. در این روش، روند همگرایی زنجیر در دو مرحله صورت می‌پذیرد. اگر زنجیر در مرحله اول همگرا شود، وارد مرحله دوم همگرایی می‌شود.

- مرحله اول، محاسبه آماره آزمون کرامر-وان میسس[۸۱] (وان میسس، ۱۹۳۱) است. برای این آماره، آزمون فرضیه

$$\begin{cases} H_0: & \text{زنجیر دارای توزیع مانا است} \\ H_1: & \text{زنجیر دارای توزیع مانا نیست} \end{cases}$$

در نظر گرفته می‌شود. این آماره آزمون به روش زیر محاسبه می‌شود:

$$T = \int_0^1 (B_n(t))^2 \mathrm{d}t.$$

با توجه به تعریف

$$Y_0 = 0 \qquad Y_n = \sum_{j=0}^{n} X^j \qquad \bar{X} = \frac{1}{n}\sum_{j=0}^{n} X^j$$

$B_n(t)$ به صورت

$$B_n(t) = \frac{Y_{[nt]} - [nt]\bar{X}}{\sqrt{n\hat{S}(0)}} \qquad 0 \le t \le 1$$

[۷۹] Spectral Density

[۸۰] Measure Convergence Heidelberg and Welch

[۸۱] Cramer-Von Mises Statistic

محاسبه می‌شود، که در آن $[x]$ بزرگترین عدد صحیح است که از x بیشتر نیست. در این مرحله، اگر فرضیه صفر پذیرفته شود، آن‌گاه زنجیر همگرا است و وارد مرحله دوم همگرایی می‌شود. در صورتی که فرضیه صفر رد شود، ۰/۱ زنجیر قطع شده و آماره آزمون دوباره محاسبه می‌شود. این روند تا زمانی‌که فرضیه صفر پذیرفته شود یا نیمی از زنجیر قطع شود، ادامه می‌یابد. در غیر این صورت زنجیر همگرا نخواهد بود.

- مرحله دوم، برای زنجیر پذیرفته‌شده در مرحله قبل یک فاصله اطمینان ۹۵٪ حول میانگین در نظر گرفته می‌شود. اگر شرط زیر برقرار باشد:

$$\frac{\text{۰/۵ [طول بازه]}}{\text{میانگین}} < \epsilon$$

فرضیه صفر پذیرفته می‌شود. بنابراین، زنجیر همگرا خواهد بود. مقدار ϵ توسط کاربر مشخص می‌شود. این مرحله به آزمون نیم‌فاصله[۸۲] معروف است.

۷.۱ معیارهای انتخاب مدل

در چارچوب مدل‌بندی آماری، مقایسه مدل‌ها با تعریف اندازه برازش، که آماره انحراف[۸۳] می‌باشد و همچنین پیچیدگی[۸۴]، که تعداد پارامترهای آزاد در مدل است، همراه می‌شود (وهتری و اجانن، ۲۰۱۲؛ واتاناب، ۲۰۱۳). هرچه پیچیدگی افزایش یابد، برازش بهتری از مدل خواهیم داشت. برای مقایسه مدل‌ها، تلفات قابل انتظار در داده‌های تکراری اندازه‌گیری می‌شود که می‌توانند به‌عنوان معیار قدرت پیش‌بینی مورد استفاده قرار گیرند (اسمولی و کوپیوس، ۲۰۰۹). هرچه میزان تلفات کمتر باشد، مدل بهتری انتخاب می‌شود. برای مقایسه مدل، سه معیار اطلاع انحراف[۸۵] (DIC)، درستنمایی حاشیه‌ای[۸۶] و معیار اطلاع بیزی[۸۷] (BIC) را معرفی می‌کنیم.

معیار اطلاع انحراف

یکی از معیارهای مورد استفاده برای مقایسه مدل‌های بیزی، معیار اطلاع انحراف است. این معیار، تقریبی از انحراف پیشگوی مورد انتظار است و به‌عنوان شاخصی برای برازش مدل، زمانی که هدف بالا بردن توانایی مدل با بهترین پیشگویی است، به کار می‌رود. لازم به ذکر است که مقادیر کوچک DIC برازش بهتر مدل را نشان می‌دهند.

فرض کنید $\boldsymbol{\theta}$ مجموعه همه پارامترهای موجود در مدل مورد نظر باشد. اگر مدلی را برای داده‌های مشاهده‌شده $\boldsymbol{y}$ با چگالی $f(\mathbf{y}|\boldsymbol{\theta})$ در نظر بگیرید، در این صورت منهای دو برابر لگاریتم درستنمایی را

[۸۲] Halfwide Test
[۸۳] Deviance Statistic
[۸۴] Complexity
[۸۵] Deviance Information Criterion
[۸۶] Marginal Likelihood
[۸۷] Bayesian Information Criterion

که با

$$D(\boldsymbol{\theta}) = -۲ \log\{f(\mathbf{y}|\boldsymbol{\theta})\}$$

نشان می‌دهند، آماره انحراف نامیده می‌شود. بنابراین انحراف، تابعی از پارامتر $\boldsymbol{\theta}$ است. میانگین انحراف پسینی[۸۸]، $\overline{D(\boldsymbol{\theta})} = E_{\boldsymbol{\theta}|\boldsymbol{y}}[D(\boldsymbol{\theta})]$، به‌عنوان اندازه‌ای از برازش مدل، توسط اسپیگل‌هالتر و همکاران (۲۰۰۲) پیشنهاد شد. در صورتی که $D(\boldsymbol{\theta})$ صورت بسته‌ای داشته باشد، $\overline{D(\boldsymbol{\theta})}$ به راحتی می‌تواند با استفاده از روش‌های MCMC تقریب زده شود (اسپیگل‌هالتر و همکاران، ۲۰۰۲).

به‌طور کلی جریمه‌ای[۸۹] که باید برای پیچیدگی در مدل به کار گرفته شود، مشخص نیست. اسپیگل‌هالتر و همکارانش، $\frac{۱}{۲}V_{\boldsymbol{\theta}|\boldsymbol{y}}\{D(\boldsymbol{\theta})\}$ را به‌عنوان جریمه معرفی کردند که در آن $V_{\boldsymbol{\theta}|\boldsymbol{y}}(\cdot)$ واریانس توزیع پسین است. معیار معرفی‌شده با این جریمه، معیار اطلاع انحراف نامیده شد که به صورت زیر تعریف می‌شود:

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

که در آن P_D، تعداد پارامترهای مؤثر یک مدل بیزی است که به صورت $P_D = \overline{D(\boldsymbol{\theta})} - D(\tilde{\boldsymbol{\theta}})$ نوشته می‌شود، $\tilde{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\boldsymbol{y})$ و $D(\cdot)$ همان آماره انحراف است. در معیار DIC مقدار پیچیدگی بر اساس P_D تعریف می‌شود و $D(\bar{\boldsymbol{\theta}})$ نیز سنجش برازش مدل را به عهده دارد.

عامل بیزی و درستنمایی حاشیه‌ای

فرض کنید $\boldsymbol{Y} = (Y_۱, \ldots, Y_n)$ یک نمونه تصادفی nتایی از متغیرهای تصادفی مستقل باشند. همچنین $\{M_i\,,\, i \in I\}$ یک مجموعه متناهی از مدل‌های نامزد مورد نظر برای توزیع این مشاهدات باشد. یک روش ساده برای انتخاب بهترین مدل و تعیین تعداد مؤلفه‌ها، محاسبه عامل بیزی[۹۰] است. عامل بیزی یک معیار مقایسه‌ای بر پایه درستنمایی حاشیه‌ای است. عامل بیزی B برای مدل $M_۰$ و $M_۱$ به صورت $B = \frac{f(\boldsymbol{y}|M_۱)}{f(\boldsymbol{y}|M_۰)}$ تعریف می‌شود که در آن درستنمایی حاشیه‌ای به صورت

$$f(\boldsymbol{y}|M_i) = \int f(\boldsymbol{y}|\boldsymbol{\theta}_i, M_i) f(\boldsymbol{\theta}_i|M_i)\, \mathrm{d}\boldsymbol{\theta}_i \qquad (۵.۱)$$

محاسبه می‌شود، به‌طوری که $\boldsymbol{\theta}_i$ بردار پارامتری مدل M_i و $f(\boldsymbol{\theta}_i|M_i)$ تابع چگالی پیشین پارامترهای مدل M_i است. بر اساس عامل بیزی اگر $B > ۱$ باشد، آن‌گاه مدل $M_۱$ از $M_۰$ بهتر است. درستنمایی حاشیه‌ای را می‌توان از روش‌های نمونه‌گیری MCMC که توسط چیب (۱۹۹۵) و چیب و جلیزکوف (۲۰۰۱) ارائه شده‌اند، به‌دست آورد. محاسبه درستنمایی حاشیه‌ای وابسته به احتمال پیشین است (کاس و رفتری، ۱۹۹۵).

معیار اطلاع بیزی

یک روش کلاسیک برای تقریب (۵.۱) استفاده از معیار اطلاع بیزی است، برای زمانی که از الگوریتم MCMC برای برآورد ماکسیمم درستنمایی پارامترهای مدل استفاده شود.

[۸۸] Posteriori Mean of Deviance

[۸۹] Penalty

[۹۰] Bayes Factor

معیار اطلاع بیزی، معیاری برای مقایسه مدل است (ایچوارز، ۱۹۸۷). ممکن است در مواردی، اطلاعات پیشین نسبت به اطلاعات موجود در داده‌ها کم باشد. تحت این شرایط BIC معمولاً به‌عنوان یک تقریب اعمال می‌شود و از توزیع پیشین ارزیابی نمی‌شود (کاس و رفتری، ۱۹۹۵). معیار اطلاع بیزی با p پارامتر و n مشاهده به صورت زیر تعریف می‌شود:

$$BIC = -۲\log\{f(\boldsymbol{y}|\hat{\boldsymbol{\theta}})\} + p\log(n).$$

با این تعریف مدلی که دارای کمترین مقدار BIC باشد، مدل منتخب خواهد بود.

۸.۱ معیارهای چولگی

چولگی[۹۱] معیاری برای کمی‌سازی تقارن یا عدم تقارن، در یک توزیع آماری است. برای یک توزیع متقارن[۹۲]، چولگی برابر با صفر، برای یک توزیع نامتقارن[۹۳] با کشیدگی دمی به سمت مقادیر مثبت، چولگی به سمت راست و برای یک توزیع نامتقارن با کشیدگی دمی به سمت مقادیر منفی، چولگی به سمت چپ است. تلاش‌های بسیاری در زمینه اندازه‌گیری و تعریف چولگی صورت گرفته است. معیارهایی که برای تعیین چولگی یک توزیع به کار می‌روند، نخستین بار توسط کارل پیرسون (۱۸۹۵) مطرح شدند. پس از آن به‌طور گسترده‌تری توسط افراد بسیاری از جمله اجورث (۱۹۰۴)، چارلیر (۱۹۰۵) و باولی (۱۹۲۰) توسعه یافتند. از جمله این معیارها می‌توان به

$$S_k = \frac{(\mu - M)}{\sigma}$$

$$\gamma_۱ = \frac{E(X-\mu)^۳}{\sigma^۳}$$

$$b = \frac{(Q_۳ + Q_۱ - ۲m)}{(Q_۳ - Q_۱)}$$

و

$$\gamma'_m = \frac{(\mu - m)}{\sigma}$$

اشاره کرد، که در آن σ انحراف معیار، $Q_۱$ و $Q_۳$ اولین و سومین چارک[۹۴]های توزیع، μ میانگین، m میانه و M مد توزیع هستند. وان زوییت (۱۹۶۴) با توجه به معیارهای چولگی، مفهوم مرتب‌سازی[۹۵] دو توزیع را بیان کرد، که این موضوع به‌طور گسترده‌ای مورد توجه قرار گرفته است (اوجا، ۱۹۸۱).

تعریف ۱.۸.۱. برای متغیرهای تصادفی با تابع توزیع تجمعی $F(x)$ و $G(x)$ و توابع چگالی $f(x)$ و $g(x)$ با بازه محدود، $G(x)$ را بیشترین چوله به راست نسبت به $F(x)$ می‌نامند، اگر

$$R(x) = G^{-۱}(F(x))$$

[۹۱] Skewness
[۹۲] Symmetric
[۹۳] Asymmetric
[۹۴] Quartile
[۹۵] Ordering

محدب باشد. می‌نویسیم $F <_c G$ و می‌گوییم F از نظر ترتیب محدب کوچکتر از G است.

بنابراین، با توجه به تعریف مرتب‌سازی وان زوییت، یک معیار چولگی مانند $\gamma(\cdot)$ باید دارای ویژگی‌های زیر باشد:

- برای تابع توزیع متقارن F، $\gamma(F) = ۰$.
- برای مقادیر ثابت c و d، $\gamma(cF + d) = \gamma(F)$.
- $\gamma(-F) = -\gamma(F)$.
- اگر $F <_c G$ باشد، آن‌گاه $\gamma(F) \leq \gamma(G)$.

وان زوییت (۱۹۶۴) نشان داد که همه گشتاورهای مرکزی مفرد استاندارد[۹۶]، از مرتبه ۳ یا بالاتر، خاصیت محدب بودن را حفظ می‌کنند. اما معیار $\gamma_۱$ برای بسیاری از توابع توزیع، نمی‌تواند چولگی راست را به خوبی بیان کند (لی و موریس، ۱۹۹۱). سایر معیارها نیز تحت شرایطی، ویژگی محدب بودن را حفظ می‌کنند. آرنولد و گرینولد (۱۹۹۵) معیاری برای چولگی بر پایه مد معرفی کردند که هر چهار ویژگی چولگی را تحت شرایط زیر، برای بسیاری از خانواده‌های توزیع‌های به راست چوله حفظ می‌کند و می‌تواند به‌عنوان رقیب بسیار خوبی برای سایر معیارها در نظر گرفته شود. این شرایط به صورت زیر بیان می‌شوند:

- متغیر تصادفی X با تابع توزیع پیوسته $F(x)$ و تابع چگالی $f(x) > ۰$ در بازه $I_X = (a, b)$ در نظر گرفته می‌شود، که a و b می‌توانند به ترتیب $-\infty$ و $+\infty$ باشند.
- فرض کنید برای $x < M_x$، $f'(x) > ۰$، برای $x > M_x$، $f'(x) < ۰$ و $f'(M_x) = ۰$ باشند. بنابراین، M_x تنها مد X است.

فرض کنید X یک متغیر تصادفی پیوسته با تابع توزیع $F(x)$ و M_x تنها مد توزیع باشد. بنابراین $\gamma_M(X)$ به صورت

$$\gamma_M(X) = ۱ - ۲F(M_x)$$

تعریف می‌شود، به‌طوری که $-۱ \leq \gamma_M(X) \leq ۱$. گرینولد و میدن (۱۹۷۷) نشان دادند، به ازای تنها یک مقدار x، اگر

$$f(M_x + x) - f(M_x - x) \geq ۰ \qquad x \geq ۰$$

آن‌گاه $M_x < m_x$ است. بنابراین، شرط کافی برای این‌که چولگی، γ_M، مثبت باشد، به صورت

$$\gamma_M(X) = ۱ - ۲F(M_x) > ۱ - ۲F(m_x) = ۰$$

بیان می‌شود. در نتیجه، اگر $M_x < m_x$ باشد، آن‌گاه $F(M_x) < \frac{۱}{۲}$ است. بنابراین، تابع توزیع، چولگی مثبت و تابع پیوند ساخته‌شده بر اساس تابع توزیع برای داده‌های دو سطحی، چولگی منفی دارد. بر عکس، اگر $M_x > m_x$ باشد، آن‌گاه $F(M_x) > \frac{۱}{۲}$ است. بنابراین، تابع توزیع، چولگی منفی و تابع پیوند، چولگی مثبت دارد.

[۹۶] Odd Standardized Central Moments

۹.۱ سایر تعاریف

تعریف ۱.۹.۱ (توزیع تباهیده). یک توزیع تباهیده[۹۷]، توزیع احتمال یک متغیر تصادفی است که فقط یک تک نقطه را شامل می‌شود. یعنی توزیع تباهیده در یک نقطه، روی یک خط حقیقی متمرکز شده است (کسلا، ۲۰۰۴). اگر X دارای توزیع تباهیده باشد، تابع احتمال و تابع توزیع آن به صورت زیر است:

$$f(x;c)=\begin{cases}۰ & x\neq c\\ ۱ & x=c\end{cases}\qquad F(x;c)=\begin{cases}۰ & x<c\\ ۱ & x\geq c\end{cases}$$

قضیه ۲.۹.۱ (قضیه فوبینی[۹۸]). *اگر $f(x,y)$ بر ناحیه مستطیلی $R:a\leq x\leq b, c\leq y\leq d$ پیوسته باشد، آن‌گاه*

$$\int_R f(x,y)\,\mathrm{d}A=\int_c^d\int_a^b f(x,y)\,\mathrm{d}x\mathrm{d}y=\int_a^b\int_c^d f(x,y)\,\mathrm{d}y\mathrm{d}x.$$

حالت قوی‌تر: فرض کنید $f(x,y)$ روی ناحیه‌ای چون R پیوسته باشد.

- *اگر تعریف R به صورت $a\leq x\leq b, f_۱(x)\leq y\leq f_۲(x)$ باشد، با این شرط که $f_۱$ و $f_۲$ بر $[a,b]$ پیوسته هستند، آن‌گاه*

$$\int\int_R f(x,y)\,\mathrm{d}x\mathrm{d}y=\int_a^b\int_{f_۱(x)}^{f_۲(x)} f(x,y)\,\mathrm{d}y\mathrm{d}x.$$

- *اگر تعریف R به صورت $c\leq y\leq d, g_۱(y)\leq x\leq g_۲(y)$ باشد، با این شرط که $g_۱$ و $g_۲$ بر $[c,d]$ پیوسته هستند، آن‌گاه*

$$\int\int_R f(x,y)\,\mathrm{d}x\mathrm{d}y=\int_c^d\int_{g_۱(y)}^{g_۲(y)} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

برای اطلاعات بیشتر به کتاب رودین (۱۹۶۴) مراجعه کنید.

تعریف ۳.۹.۱ (توزیع دم‌کلفت). توزیع دم‌کلفت[۹۹] در آمار به دسته‌ای از توابع توزیع احتمال اطلاق می‌شود که کشیدگی بزرگی دارند (رسنیک، ۲۰۰۷). در چنین تابع توزیعی، مقادیر دمی از مقدار میانگین فاصله زیادی دارند. به‌ویژه دم‌کلفتی یک تابع توزیع در مقایسه با یک توزیع نرمال تعیین می‌گردد. توزیع متغیر تصادفی X دم‌کلفت خوانده می‌شود اگر عددی مانند $a>۰$ وجود داشته باشد، به‌طوری که

$$P(X>x)\sim x^{-a}\qquad x\to\infty$$

که در آن نماد $\sim$ به معنی هم ارز بودن است.

تعریف ۴.۹.۱ (مدل با اثرهای آمیخته). مدل با اثرهای آمیخته[۱۰۰] معمولاً برای اندازه‌های مکرر، داده‌های وابسته و برخی داده‌های طبقه‌بندی‌شده به کار می‌رود. در حالت کلی یک مدل خطی با اثرهای آمیخته به صورت زیر معرفی می‌شود:

$$y=X\beta+Zu+\epsilon$$

[۹۷] Degenerate Distribution

[۹۸] Fubini's Theorem

[۹۹] Heavy-Tailed Distribution

[۱۰۰] Mixed Effects Model

که در آن y بردار n بعدی مقادیر مشاهده‌شده متغیر پاسخ، X ماتریس $n \times p$ پررتبه ستونی شامل متغیرهای تبیینی، β بردار p بعدی پارامترهای نامعلوم اثرهای ثابت، u بردار q بعدی اثرهای تصادفی و ماتریس Z شامل برخی متغیرهای تبیینی و نشانگر است که با توجه به فرض‌های موجود درباره اثرهای تصادفی u در مدل ساخته می‌شوند. جمله ϵ نیز بردار n بعدی مؤلفه‌های خطای تصادفی مدل است (لرد و ویر، ۱۹۸۲؛ مولر و استوارت، ۲۰۰۶).

تعریف ۵.۹.۱ (متغیرهای پنهان)**.** متغیرهای پنهان[۱۰۱] در آمار، در مقابل متغیرهای مشاهده‌شده قرار می‌گیرند. این متغیرها که به صورت مستقیم قابل مشاهده نیستند، می‌توانند از میان متغیرهای دیگر که قابل مشاهده هستند، توسط یک مدل آماری شناخته شوند. در مدل‌بندی آماری از این مفهوم، بسته به زمینه کاربردی مورد نظر، برای بیان متغیرهای تصادفی و پارامترهای مدل به کار می‌رود (بورسبوم و همکاران، ۲۰۰۳).

تعریف ۶.۹.۱ (شناسایی‌پذیری)**.** در شناسایی‌پذیری[۱۰۲] مدل، یکتا بودن پارامترها بررسی می‌شود. به بیانی دیگر، برای بررسی مقدار یگانه هر یک از پارامترهای آزاد در مدل از روی داده‌های مشاهده‌شده، از شناسایی‌پذیری مدل استفاده می‌گردد. در عمل برای بررسی شناسایی‌پذیری مدل از چندین مجموعه مقادیر اولیه برای برآورد پارامترها استفاده می‌شود. چنان‌چه مقادیر متفاوت اولیه، منجر به برآورد مشابهی برای پارامترها شود، مدل شناسایی‌پذیر خواهد بود (دایتون و مک ریدی، ۱۹۸۸).

[۱۰۱] Latent Variables

[۱۰۲] Identifiability

فصل ۲

مدل رگرسیون مقادیر فرین تعمیم‌یافته

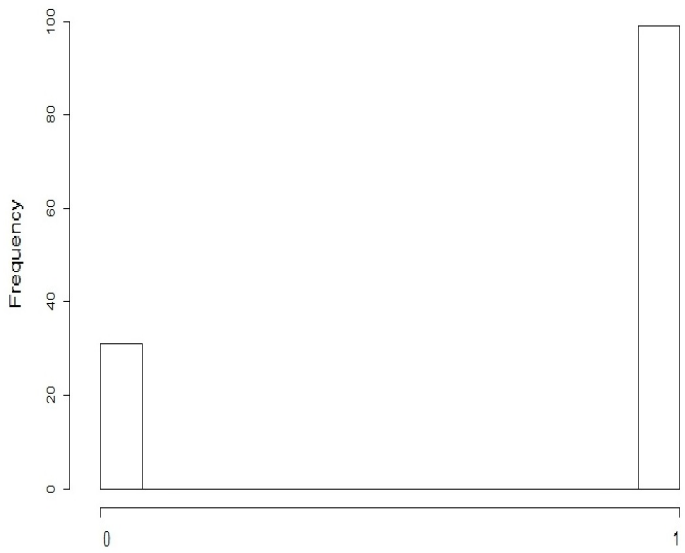
۱.۲ مدل‌بندی پاسخ‌های دو سطحی

در مدل‌های رگرسیونی، در موقعیت‌های کاربردی مختلف، ماهیت متغیر پاسخ دو سطحی است که می‌توان آن‌ها را در قالب اعداد ۱ و ۰ نمایش داد. به‌عنوان مثال، داده‌های متعلق به ۶۵ پسر و ۶۵ دختر دبیرستانی، که توسط شوماخر (۱۹۹۶) گردآوری شده‌اند، را در نظر بگیرید. ساختار این مجموعه داده، شامل سه متغیر است. این متغیرها شامل جنسیت (عاملی با دو سطح مذکر و مؤنث)، درجه حرارت بدن (عاملی با دو سطح طبیعی و غیرطبیعی) و ضربان قلب هستند. هدف از تحلیل این داده‌ها بررسی تأثیر جنسیت و ضربان قلب افراد بر درجه حرارت طبیعی و غیرطبیعی بدن آن‌ها است. معمولاً برای تحلیل این نوع از داده‌ها، از مدل‌های رگرسیونی لجیت، پرابیت و Cloglog استفاده می‌شود. اما همیشه استفاده از این توابع پیوند نتایج خوبی را در بر نخواهند داشت (زادو و سنتنر، ۱۹۹۲).

۱.۱.۲ پاسخ‌های دو سطحی نامتعادل (پیشامدهای نادر)

در بسیاری از موارد ممکن است یک رویداد نادر باشد یعنی با احتمال کم یا بسیار کم اتفاق بیفتد؛ یعنی با داده‌هایی مواجه شویم که منحنی متغیر پاسخ آن‌ها به شدت چوله باشد. به‌عنوان مثال، تفاوت معنی‌داری در چولگی منحنی پاسخ، بین تعداد افرادی که درجه حرارت طبیعی دارند (۰) با افرادی که درجه حرارت غیر طبیعی دارند (۱)، وجود دارد. هیستوگرام متغیر پاسخ با دو سطح ۰ و ۱ در شکل ۱.۲ نشان داده شده است.

همان‌طور که در فصل اول اشاره کردیم، مدل‌های معمول پرابیت و لجیت برای تحلیل داده‌های



شکل ۱.۲: نمودار فراوانی متغیر پاسخ در داده‌های حرارت بدن

چوله، خوب عمل نمی‌کنند؛ زیرا آن‌ها انعطاف لازم را برای برازش مناسب این نوع پاسخ‌ها ندارند. مدل Cloglog نیز تنها زمانی که چولگی به سود ۱ها باشد مناسب است. به عبارتی اگر جای ۰ و ۱ در مدل‌بندی داده‌ها تغییر کند، این مدل نیز ناکارا خواهد بود. بنابراین برای پاسخ‌های دوسطحی نادر که به شدت چوله هستند، استفاده از مدل‌های معمول توصیه نمی‌شود و باید از مدل‌های رگرسیونی منعطف‌تری برای برازش آن‌ها استفاده کرد.

۲.۲ راهکارهای جانشین معرفی‌شده

مسأله تعیین تابع پیوند مناسب، برای داده‌های دو سطحی، توسط محققین مختلفی مورد بررسی قرار گرفته است. به‌عنوان مثال، زادو و سنتنر (۱۹۹۲) نشان دادند که استفاده نادرست از تابع پیوند لجستیک می‌تواند منجر به افزایش شدید اریبی[۱] و میانگین توان دوم خطا[۲] (MSE) برآوردگرهای پارامترها شود. از سوی دیگر، کارهای زیادی برای ورود انعطاف لازم به توابع پیوند مدل‌های خطی تعمیم‌یافته انجام شده‌اند. آراندا اورداز (۱۹۸۱)، دو مدل یک پارامتری جداگانه را برای تقارن و عدم تقارن مدل لجستیک معرفی کرد. گررو و جانسون (۱۹۸۲)، یک تبدیل باکس و کاکس[۳] یک پارامتری بخت‌ها[۴] را پیشنهاد دادند. مورگان (۱۹۸۳)، یک مدل لجستیک مکعبی[۵] یک پارامتری را ارائه داد. استوکل (۱۹۸۸)، یک

[۱] Bias
[۲] Mean Squared Error
[۳] Box-Cox
[۴] Odds
[۵] Cubic Logistic

رده کلی از مدل‌های لجستیک تعمیم‌یافته، برای مدل‌سازی داده‌های دو سطحی پیشنهاد داد. مدل‌های استوکل کلی هستند و می‌توانند توسط بسیاری از خانواده‌ها تقریب زده شوند. این مدل‌ها در حضور متغیرهای تبیینی، برای بسیاری از پیشین‌های ناآگاهی‌بخش دارای پسین ناسره هستند (چن، دی و شائو، ۱۹۹۹).

در سال ۲۰۰۸ کیم و همکارانش تابع پیوند چوله t-تعمیم‌یافته را معرفی کردند که از انعطاف بالایی برای مدل‌بندی داده‌های دو سطحی نامتعادل برخوردار است. در واقع، محققان همواره در جستجوی توزیع‌هایی هستند که رفتار منعطف‌تری از خود بروز دهند یا به بیان دیگر، توزیع‌هایی که علاوه بر تنظیم پارامتر چولگی بتوانند رفتار دم توزیع را نیز کنترل کنند. از دیدگاه آماری توزیع‌هایی با دم‌های کلفت‌تر از توزیع نرمال مد نظر هستند. توزیع t-تعمیم‌یافته به دلیل داشتن قدرت کنترل در رفتار دم توزیع از انعطاف‌پذیری بالایی برخوردار است. در ادامه این مدل را معرفی می‌کنیم.

۱.۲.۲ توزیع t- تعمیم‌یافته

تابع چگالی t-تعمیم‌یافته[۶] که توسط آرلانو-واله و بولفارین (۱۹۹۵) پیشنهاد شده است به صورت

$$f_{gt,v_1,v_2}(w) = \frac{\Gamma(\frac{v_1+1}{2})}{\sqrt{\pi v_2}\Gamma(\frac{v_1}{2})}\left(1+\frac{w^2}{v_2}\right)^{-\frac{(v_1+1)}{2}} \qquad (1.2)$$

است، که در آن $v_1 \in \mathbb{R}^+$ پارامتر شکل[۷] توزیع و $v_2 \in \mathbb{R}^+$ پارامتر مقیاس است. اگر W یک متغیر تصادفی با تابع چگالی t-تعمیم‌یافته باشد که تکیه‌گاه آن مجموعه اعداد حقیقی است، می‌توان نوشت $W \sim P_{gt,v_1,v_2}$. میانگین و واریانس این توزیع به ترتیب عبارت‌اند از

$$E(W) = 0 \qquad v_1 > 1$$

$$Var(W) = \frac{v_2}{(v_1-2)} \qquad v_1 > 2.$$

در حالت خاص $v_1 = v_2 = v$، توزیع t-تعمیم‌یافته به توزیع t-استودنت با v درجه آزادی تبدیل می‌شود. تابع چگالی احتمال توزیع t-تعمیم‌یافته، مشابه t-استودنت، حول محور صفر متقارن است. نمودار آن زنگی-شکل[۸] می‌باشد. به ازای مقادیر کوچک پارامتر شکل، این توزیع به یک توزیع دم‌کلفت تبدیل می‌شود و برای مقادیر بزرگ پارامتر شکل، $v_1 \to \infty$، این توزیع به توزیع نرمال همگرا می‌شود. دم‌کلفتی یک تابع توزیع در مقایسه با یک توزیع نرمال تعیین می‌گردد. همان‌طور که در شکل ۲.۲ مشاهده می‌شود، چگالی توزیع t-تعمیم‌یافته سیر نزولی کندتری نسبت به چگالی نرمال استاندارد دارد.

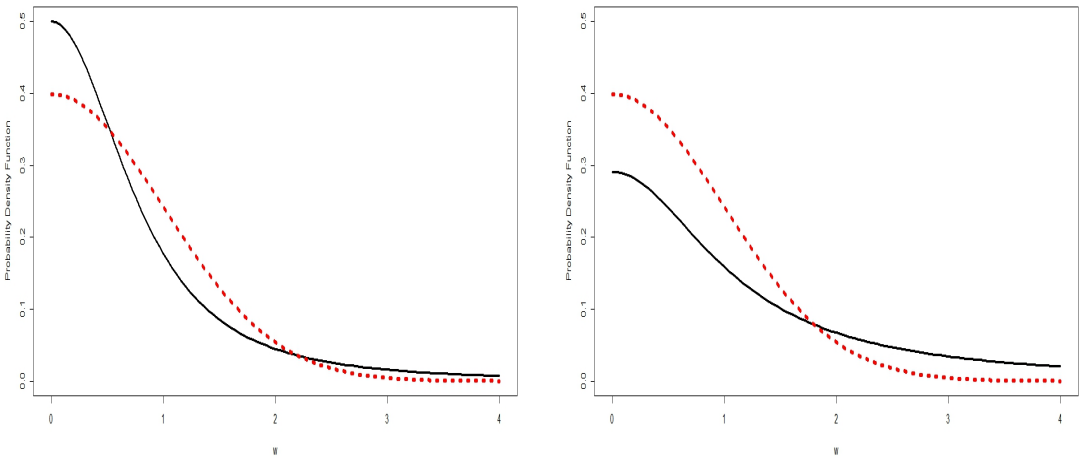
تابع پیوند متقارن t- تعمیم‌یافته

فرض کنید $\mathrm{y} = (y_1, \ldots, y_n)'$ بردار پاسخ‌های مشاهده‌شده مستقل دو سطحی، $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ بردار ضرایب رگرسیونی، $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$، $i = 1, 2, \ldots, n$، بردار متغیرهای تبیینی p

[۶] Generalized T-Distribution

[۷] Shape Parameter

[۸] Bell-Shaped



(ب) (آ)

شکل ۲.۲: نمودار تابع چگالی نرمال استاندارد (خط‌چین) و تابع چگالی t-تعمیم‌یافته (خط پر) با (آ) $v_1 = ۱/۲$ و $v_2 = ۱$ و (ب) $v_1 = ۲$ و $v_2 = ۱$

بعدی و $\mathrm{w} = (w_1, w_2, \ldots, w_n)'$ بردار متغیرهای پنهان مستقل باشند. رهیافت متغیر پنهان که توسط آلبرت و چیب (۱۹۹۳) ارائه شده است، می‌تواند برای نمایش یک مدل رگرسیون دو سطحی به کار رود. فرض کنید w_i یک متغیر پنهان باشد که به صورت زیر تعریف می‌شود:

$$y_i = \begin{cases} ۱ & w_i > ۰ \\ ۰ & w_i \leq ۰ \end{cases}$$

به‌طوری که

$$w_i = \mathrm{x}_i'\boldsymbol{\beta} + \epsilon_i \qquad (۲.۲)$$

که در آن $\epsilon_i \sim F$ و F یک تابع توزیع تجمعی است. فرض کنید تابع توزیع t-استودنت به صورت

$$F_{t_v}(w) = \int_{-\infty}^{w} f_{t_v}(u)\mathrm{d}u$$

باشد، که در آن $f_{t_v}(\cdot)$ تابع چگالی توزیع t-استودنت با v درجه آزادی است. اگر در رابطه (۲.۲) $F = F_{t_v}$ شود، آن‌گاه تابع پیوند به‌دست آمده، تابع پیوند رابیت[۹] نامیده می‌شود (لیو، ۲۰۰۴). این تابع پیوند زمانی‌که $v \to \infty$، به تابع پیوند پرابیت و برای زمانی‌که $v = ۱$ است به تابع پیوند کوشی[۱۰] تبدیل می‌شود. تابع پیوند لجیت نیز می‌تواند به‌وسیله تابع پیوند رابیت تقریب زده شود (آلبرت و چیب، ۱۹۹۳). از آن‌جایی که تحلیل پاسخ‌های چوله به‌وسیله توابع پیوند متقارن کارا نیست، مدل تابع پیوند چوله t-تعمیم‌یافته[۱۱] معرفی شده است.

[۹] Robit Link

[۱۰] Cauchy Link

[۱۱] Skewed Generalized T-link Function

تابع پیوند چوله t- تعمیم‌یافته

مدل تابع پیوند چوله t-تعمیم‌یافته توسط چن، دی و شائو (۱۹۹۹) بر اساس رهیافت متغیر پنهان معرفی شد که در قالب مدلی با ساختار اثرات آمیخته ارایه گردید. با توجه به مفروضات قبل، فرض کنید w_i یک متغیر پنهان باشد که به صورت زیر تعریف می‌شود:

$$y_i = \begin{cases} ۱ & w_i > ۰ \\ ۰ & w_i \leq ۰ \end{cases}$$

به‌طوری که

$$w_i = \mathrm{x}_i'\boldsymbol{\beta} + \sigma z_i + \epsilon_i \qquad (۳.۲)$$

که در آن $z_i \sim G$، $\epsilon_i \sim F$ و ϵ_i و z_i از هم مستقل‌اند. همچنین G تابع توزیع تجمعی یک توزیع چوله و F تابع توزیع تجمعی یک توزیع متقارن است. مدل با این ساختار دارای ویژگی‌های زیر است:

- این مدل یک مدل با اثرهای آمیخته محسوب می‌شود.
- پارامتر مقیاس σ چولگی را کنترل می‌کند. به این معنی که، اگر $\sigma > ۰$ چولگی مثبت و اگر $\sigma < ۰$ چولگی منفی است.
- اگر G یک توزیع تباهیده در صفر باشد یا $\sigma = ۰$ باشد، مدل با تابع پیوندی متقارن است.
- استفاده از الگوریتم نمونه‌گیر گیبز به راحتی قابل اجرا است.

با این‌که مدل دارای ویژگی‌های مناسبی است، اما درهم آمیختن[۱۲] عرض از مبدأ و σ با یکدیگر در رابطه (۳.۲) موجب بروز محدودیت در مدل می‌شود. چن و همکاران (۱۹۹۹) مدلی بدون عرض از مبدأ پیشنهاد دادند که در این مدل نیز ممکن است برازش به خوبی صورت نگیرد. راه حل ممکن دیگر این است که عرض از مبدأ در مدل باشد اما $\sigma = ۱$ در نظر گرفته شود. این مدل نیز موجب از دست رفتن انعطاف‌پذیری در کنترل چولگی می‌شود. برای غلبه بر این مشکلات، کیم و همکاران (۲۰۰۸) مدلی را به صورت زیر ارائه دادند.

فرض کنید w_i^* متغیری پنهان باشد که به صورت زیر تعریف می‌شود:

$$y_i = \begin{cases} ۱ & w_i^* > ۰ \\ ۰ & w_i^* \leq ۰ \end{cases}$$

به‌طوری که

$$w_i^* = \mathrm{x}_i'\boldsymbol{\beta}^* + \{z_i - E(z)\} + \epsilon_i^*$$

که در آن $z_i \sim G$، $\epsilon_i^* \sim F_{gt,v_۱,v_۲}$، $E(z)$ میانگین توزیع G است، و ϵ_i^* و z_i نیز از هم مستقل‌اند. با استفاده از بازپارامتری کردن[۱۳] مدل، قرار می‌دهیم

$$w_i = \frac{w_i^*}{\sqrt{v_۲}} \qquad \beta = \frac{\beta^*}{\sqrt{v_۲}} \qquad \sigma = \frac{۱}{\sqrt{v_۲}} \qquad \epsilon_i = \frac{\epsilon^*}{\sqrt{v_۲}}.$$

[۱۲] Confounding

[۱۳] Reparametrization

بنابراین، مدل تابع پیوند چوله به صورت زیر ارائه می‌شود:

$$y_i = \begin{cases} ۱ & w_i > ۰ \\ ۰ & w_i \leq ۰ \end{cases}$$

به‌طوری که

$$w_i = \mathrm{x}_i'\boldsymbol{\beta} + \sigma\{z_i - E(z)\} + \epsilon_i$$

که در آن $\sigma > ۰$ و $\epsilon_i \sim f_{gt,v_۱,v_۲=۱}$. در این مدل، پارامتر چولگی به صورت $۰ < \sigma \leq ۱$ تعیین می‌گردد که این محدودیت اجازه وجود عرض از مبدأ در مدل را فراهم می‌کند. پارامتر $v_۱$ نیز سنگینی دنباله را کنترل می‌کند. در عمل برای اطمینان از شناسایی‌پذیری مدل، چندین توزیع مختلف برای G تعریف می‌شوند و سپس با استفاده از معیارهای انتخاب مدل مانند DIC بهترین توزیع متناسب با داده‌ها تعیین می‌گردد. توزیع G به صورت‌های زیر معرفی می‌شود:

- در صفر تباهیده است و به صورت $\triangle_{\{۰\}}$ نشان داده می‌شود و منظور از آن

$$P(z_i = ۰) = ۱$$

است. استفاده از این توزیع تابع پیوند متقارن t-تعمیم‌یافته را نتیجه می‌دهد.

- نمایی استاندارد $(\mathcal{E})$ با تابع چگالی احتمال

$$f(z_i) = \begin{cases} e^{-z_i} & z_i > ۰ \\ ۰ & \text{سایر جاها} \end{cases}$$

که یک تابع پیوند چوله مثبت را نتیجه می‌دهد.

- نمایی استاندارد منفی[۱۴] $(\mathcal{NE})$ با تابع چگالی احتمال

$$f(z_i) = \begin{cases} e^{z_i} & z_i < ۰ \\ ۰ & \text{سایر جاها} \end{cases}$$

که یک تابع پیوند چوله منفی را نتیجه می‌دهد.

- توزیع نیم‌نرمال[۱۵] $(\mathcal{HN})$ با تابع چگالی احتمال

$$f(z_i) = \begin{cases} (\frac{۲}{\sqrt{۲\pi}})e^{-\frac{z_i^۲}{۲}} & z_i > ۰ \\ ۰ & \text{سایر جاها} \end{cases}$$

که یک تابع پیوند چوله مثبت را نتیجه می‌دهد. در این تابع پیوند نرخ سرعت نزدیک شدن به یک بیشتر از نرخ سرعت نزدیک شدن به صفر است.

[۱۴] Negative Standard Exponentail Distribution

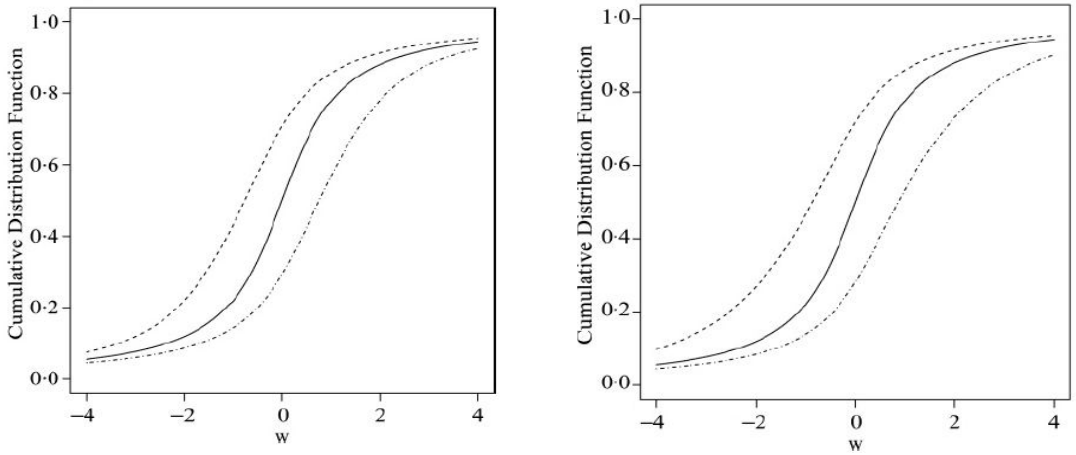
[۱۵] Half Normal Distribution

- نیم‌نرمال منفی[۱۶] ($\mathcal{NHN}$) با تابع چگالی احتمال

$$f(z_i) = \begin{cases} (\frac{۲}{\sqrt{۲\pi}})e^{-\frac{z_i^۲}{۲}} & z_i < ۰ \\ ۰ & \text{سایر جاها} \end{cases}$$

که یک تابع پیوند چوله منفی را نتیجه می‌دهد و نرخ سرعت نزدیک شدن به صفر در آن بیشتر است.

نمودار ۳.۲ تابع توزیع t-تعمیم‌یافته را با توزیع‌های مختلف برای G نشان می‌دهد.



(ب) (آ)

شکل ۳.۲: نمودار تابع توزیع t-تعمیم‌یافته برای $\sigma z_i + \epsilon_i$ با $v_۱ = ۱/۲$، $v_۲ = ۱$ و $\sigma = ۱$ (آ) $G = \mathcal{E}$ (نقطه‌چین)، $G = \mathcal{NE}$ (نقطه‌خط) و $G = \triangle_{\{۰\}}$ (خط پر) و (ب) $G = \mathcal{HN}$ (نقطه‌چین)، $G = \mathcal{NHN}$ (نقطه‌خط) و $G = \triangle_{\{۰\}}$ (خط پر)

به دلیل محدودیت در پارامتر چولگی و همچنین سنگینی محاسبات در این روش، وانگ و دی (۲۰۱۰) تابع پیوند مقادیر فرین تعمیم‌یافته را معرفی کردند که انعطافی مشابه با و حتی بیشتر از تابع پیوند چوله t-تعمیم‌یافته دارد و اجرای آن ساده‌تر است. برای معرفی این مدل، ابتدا توزیع مقادیر فرین تعمیم‌یافته را معرفی می‌کنیم.

۲.۲.۲ توزیع مقادیر فرین تعمیم‌یافته

نظریه مقادیر فرین که بر دم‌های توزیع تمرکز دارد و رفتار مشاهدات خیلی بزرگ یا خیلی کوچک را در فرآیندی تصادفی تحلیل می‌کند، در زمینه‌های علمی مختلف مورد مطالعه قرار گرفته است. از جمله زمینه‌های کاربردی این نظریه می‌توان به علوم زیست‌محیطی (اسمیت، ۱۹۸۹؛ تامسون و همکاران، ۲۰۰۱)، زمین‌شناسی در تحلیل بزرگی زمین‌لرزه‌ها (کایرس و همکاران، ۱۹۹۹)، هواشناسی در تحلیل مقدار بارش (کولی و همکاران، ۲۰۰۷)، مهندسی در تحلیل خوردگی (رایس و تامس، ۲۰۰۷) و تحلیل شکل‌های DNA و بافت ماهیچه‌ای (درایدن و ذمپلنی، ۲۰۰۶) اشاره نمود.

[۱۶] Negative Half Normal Distribution

فرض کنید $\{X_t; t \geq ۱\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع دلخواه $F(\cdot)$ باشد. ماکسیمم n متغیر تصادفی اول این دنباله را با M_n نشان می‌دهیم. یعنی

$$M_n = \max\{X_۱, \ldots, X_n\}.$$

بحث توزیع مقادیر فرین، با تابع توزیع متغیر تصادفی M_n و ویژگی‌های مجانبی آن، وقتی $n \to \infty$ میل می‌کند، در هم آمیخته است. تابع توزیع M_n به صورت

$$\begin{aligned} F_{M_n}(x) &= P(X_۱ \leq x, \ldots, X_n \leq x) \\ &= P(X_۱ \leq x) \times \cdots \times P(X_n \leq x) \\ &= F^n(x) \end{aligned} \qquad (۴.۲)$$

تعیین می‌شود. توجه کنید وقتی n به سمت بی‌نهایت میل کند، اگر x^* کوچکترین مقداری باشد که $F(x^*) = ۱$ آن‌گاه به ازای هر $x < x^*$ توزیع (۴.۲) تباهیده در x^* خواهد شد (کلز، ۲۰۰۱). در عمل تابع توزیع F نامعلوم است و باید برآورد شود. هرگونه نقصان در برآورد F موجب نقصان خیلی بزرگتری در برآورد F^n می‌شود. برای این منظور، با قبول دو فرض هم‌توزیعی و استقلال، دو دنباله از ثابت‌های حقیقی $\{a_n; n \geq ۱\}$ و $\{b_n; n \geq ۱\}$ طوری تعیین می‌شوند که توزیع $P(\frac{M_n - b_n}{a_n} \leq x)$ با افزایش n $(n \to \infty)$ به تابع توزیع غیرتباهیده $G(x)$ همگرا شود. قضیه انواع فرین که در زیر بیان می‌شود، ابتدا توسط فیشر و تیپت (۱۹۲۸) ارائه گردید و سپس ندنکوف (۱۹۴۳) صورت کامل‌تری از آن را اثبات کرد.

قضیه ۱.۲.۲ (انواع فرین). *اگر G یک تابع توزیع غیرتباهیده*[۱۷] *باشد و دنباله‌های ثابت حقیقی $\{a_n; n \geq ۱\}$ و $\{b_n; n \geq ۱\}$ طوری وجود داشته باشند که*

$$\lim_{n\to\infty} P(\frac{M_n - b_n}{a_n} \leq x) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$$

آن‌گاه G یکی از توزیع‌های گامبل، فرشه[۱۸] *و وایبل*[۱۹] *با توابع توزیع به ترتیب*

$$I : G(x) = \exp\{-\exp\{-(\frac{x-b}{a})\}\} \qquad -\infty < x < \infty$$

$$II : G(x) = \begin{cases} ۰ & x \leq b \\ \exp\{-(\frac{x-b}{a})^{-\alpha}\} & x > b \end{cases}$$

$$III : G(x) = \begin{cases} \exp\{-\{-(\frac{x-b}{a})\}^{\alpha}\} & x < b \\ ۱ & x \geq b \end{cases}$$

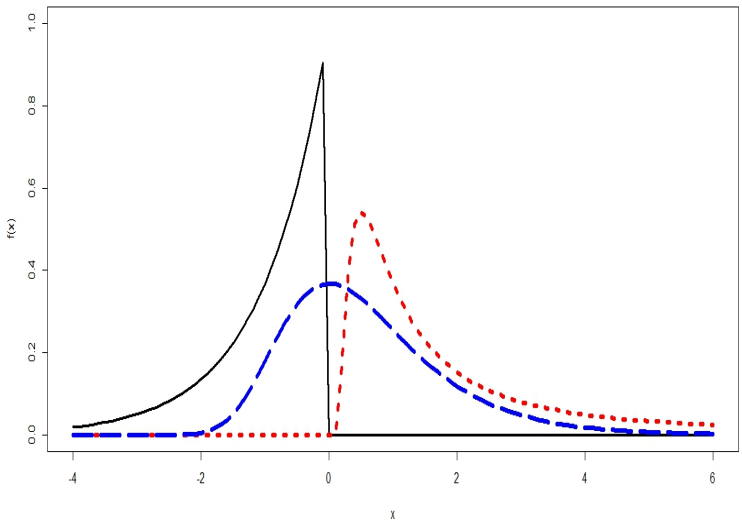
است، که در آن‌ها $a > ۰$ پارامتر مقیاس، $b \in \mathbb{R}$ پارامتر مکان و $\alpha > ۰$ پارامتر شکل توزیع می‌باشند.

در واقع این قضیه نشان می‌دهد که ماکسیمم دنباله‌ای از متغیرهای تصادفی مستقل و هم توزیع، در صورت وجود، از یکی از توزیع‌های گامبل، فرشه و وایبل پیروی می‌کند. شکل ۴.۲ توابع چگالی توزیع‌های وایبل $(a = ۱, b = ۰, \alpha = -۰/۵)$، فرشه $(a = ۱, b = ۰, \alpha = ۰/۵)$ و گامبل

[۱۷] Nondegenerat Distribution

[۱۸] Frechet

[۱۹] Weibull



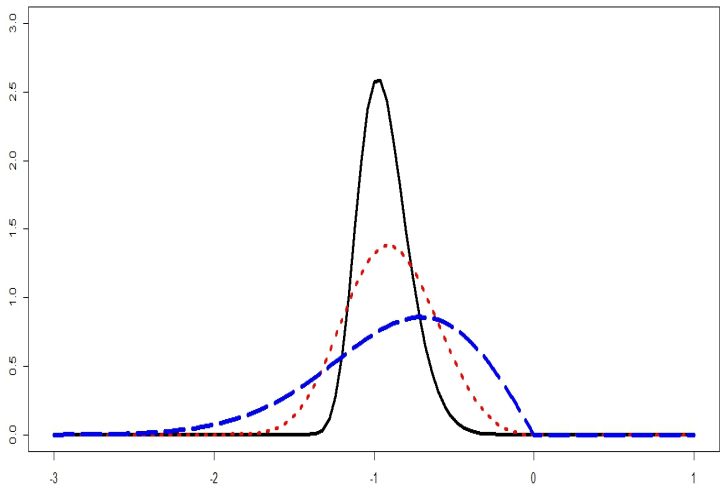
شکل ۴.۲: نمودار تابع چگالی توزیع‌های مقادیر فرین: توزیع وایبل (خط پر)، توزیع فرشه (نقطه‌چین) و توزیع گامبل (خط‌چین)

$(a = ۱, b = ۰)$ را نشان می‌دهد. توزیع فرشه دارای پارامتر شکل $\alpha > ۰$ و توزیع وایبل دارای پارامتر شکل $\alpha < ۰$ هستند. توزیع‌های فرشه و گامبل چوله به راست بوده و هر سه توزیع تک مدی هستند. توزیع وایبل خانواده‌ای غنی از توزیع‌های تک مدی را فراهم می‌کند. شکل ۵.۲ چگالی توزیع وایبل را به ازای مقادیر مختلف α نشان می‌دهد. همان‌طور که ملاحظه می‌شود، این توزیع به ازای $\alpha > -۳/۶$، $\alpha = -۳/۶$ و $\alpha < -۳/۶$ به ترتیب چوله به چپ، متقارن و چوله به راست است.

وان میسس (۱۹۳۶) با بازپارامتری کردن توزیع‌های گامبل، فرشه و وایبل به مدل واحدی دست یافت که توزیع مقادیر فرین تعمیم‌یافته نامیده می‌شود. این توزیع از طریق $\alpha = \frac{۱}{\xi}$ با توزیع‌های وایبل، فرشه و گامبل مرتبط می‌شود، به‌طوری‌که تابع توزیع آن بر روی مجموعه $\{x : ۱ + \xi\frac{(x-\mu)}{\sigma} > ۰\}$، به صورت

$$G(x) = \exp\left[-\left\{۱ + \xi\frac{(x-\mu)}{\sigma}\right\}_+^{\frac{-۱}{\xi}}\right]$$

تعریف می‌شود که در آن $x_+ = \max\{x, ۰\}$ است. این تابع توزیع دارای سه پارامتر مکان $\mu \in \mathbb{R}$، مقیاس $\sigma \in \mathbb{R}^+$ و شکل $\xi \in \mathbb{R}$ است و آن را با نماد $GEV(\mu, \sigma, \xi)$ نشان می‌دهیم. پارامتر شکل ξ رفتار دم توزیع را کنترل می‌کند (وانگ و دی، ۲۰۱۰). پارامتر شکل مثبت، $\xi > ۰$، توزیع فرشه و مقدار منفی آن، $\xi < ۰$، توزیع وایبل را مشخص می‌کند. مقدار $\xi = ۰$ را به‌عنوان $\xi \to ۰$ تفسیر می‌کنیم که توزیع گامبل را نتیجه می‌دهد. مقدار $\xi > ۰$ توزیعی با دم کلفت است که با مرتبه چندجمله‌ای، به سمت صفر کاهش می‌یابد. همچنین مقدار $\xi = ۰$ توزیعی با دم متوسط است که با مرتبه نمایی، به سمت صفر کاهش می‌یابد. سرعت نزول دم توزیع با مرتبه نمایی، بیشتر از دم توزیع با مرتبه چندجمله‌ای است. مقدار $\xi < ۰$ نیز توزیعی با دم باریک است (اسمیت، ۲۰۰۳). میانگین، واریانس و مد این توزیع



شکل ۵.۲: نمودار تابع چگالی توزیع وایبل با $\alpha = -۷$ (خط پر)، $\alpha = -۳/۶$ (نقطه‌چین) و $\alpha = -۲$ (خط‌چین)

عبارت است از:

$$E(X) = \mu - \frac{\sigma}{\xi} + \frac{\sigma}{\xi}\Gamma(۱ - \xi)$$

$$Var(X) = \frac{\sigma^۲}{\xi^۲}(\Gamma(۱ - ۲\xi) - \Gamma^۲(۱ - \xi))$$

$$M(X) = \mu + \sigma\frac{(۱ + \xi)^{-\xi} - ۱}{\xi}$$

که در آن‌ها $\Gamma(\cdot)$ تابع گاما به صورت زیر است:

$$\Gamma(\alpha) = \int_۰^\infty x^{\alpha - ۱}e^{-x}\mathrm{d}x.$$

تابع توزیع مقادیر فرین تعمیم‌یافته برای می‌نیمم‌ها با تعریف

$$m_n = \min\{X_۱, \ldots, X_n\}$$

و رابطه آن با تابع توزیع ماکسیمم‌ها، یعنی $P(\frac{m_n - d_n}{c_n} \leq x) \longrightarrow ۱ - G(-x)$ بر مجموعه

$$\{x : ۱ - \xi\frac{(x - \mu)}{\sigma} > ۰\}$$

قابل تعریف است. در این حالت

$$G(x) = ۱ - \exp\left[-\left\{۱ - \xi\frac{(x - \mu)}{\sigma}\right\}_+^{\frac{-۱}{\xi}}\right].$$

اکنون می‌توانیم مدل رگرسیون مقادیر فرین تعمیم‌یافته را معرفی و ویژگی‌های آن را بیان کنیم.

۳.۲.۲ مدل رگرسیون مقادیر فرین تعمیم‌یافته

فرض کنید $\mathrm{y} = (y_۱, \ldots, y_n)'$ بردار پاسخ‌های مشاهده‌شده دو سطحی، $\boldsymbol{\beta} = (\beta_۱, \ldots, \beta_p)'$ بردار ضرایب رگرسیونی و $\mathbf{x}_i = (x_{i۱}, x_{i۲}, \ldots, x_{ip})'$، $i = ۱, ۲, \ldots, n$، بردار متغیرهای تبیینی p بعدی باشند. همچنین، فرض کنید پاسخ‌های y_i، $i = ۱, ۲, \ldots, n$، از توزیع برنولی پیروی کنند. اگر p_i احتمال موفقیت $y_i = ۱$ و $۱ - p_i$ احتمال شکست $y_i = ۰$ باشد، برای یک نمونه nتایی $y_۱, \ldots, y_n$ مستقل از هم، تابع احتمال توأم آن‌ها برابر است با

$$\prod_{i=۱}^{n} p_i^{y_i}(۱ - p_i)^{۱-y_i} = \exp\left[\sum_{i=۱}^{n} y_i \ln\left(\frac{p_i}{۱ - p_i}\right) + \sum_{i=۱}^{n} \ln(۱ - p_i)\right]$$

که عضوی از خانواده توزیع‌های نمایی است. بنابراین، برای مدل‌سازی احتمال موفقیت بر اساس متغیرهای تبیینی مورد نظر، باید از مدل‌های خطی تعمیم‌یافته استفاده کرد. برای مدل‌سازی احتمال‌های p_i با پیروی از الگوی مدل‌های خطی تعمیم‌یافته، باید تابعی مانند $g(\cdot)$ وجود داشته باشد که

$$g(p_i) = \mathbf{x}_i'\boldsymbol{\beta}.$$

با توجه به این که p_i یک احتمال است و همواره در بازه $[۰, ۱]$ قرار می‌گیرد، مقدار برازش‌شده $\mathbf{x}_i'\boldsymbol{\beta}$ ممکن است با تغییر مقادیر $\mathbf{x}_i'$ در خارج از این بازه قرار گیرد. برای اجتناب از چنین وضعیتی، باید از توابعی استفاده نمود که برد تابع معکوس آن‌ها در محدوده $[۰, ۱]$ باشد. بنابراین چارچوب مدل خطی تعمیم‌یافته در این حالت به صورت

$$p_i = P(y_i = ۱) = g^{-۱}(\mathbf{x}_i'\boldsymbol{\beta}) = F(\mathbf{x}_i'\boldsymbol{\beta}) \qquad (۵.۲)$$

نشان داده می‌شود، که در آن $F(\cdot)$ تابع توزیع تجمعی و $g = F^{-۱}(\cdot)$ تابع پیوند یکنوا[۲۰] و مشتق‌پذیر می‌باشند.

زمانی‌که F یک تابع توزیع متقارن است، تابع پیوند آن نیز متقارن خواهد بود. رایج‌ترین توابع پیوند متقارن پرابیت و لجیت به ترتیب با شکل‌های $F^{-۱}(p_i) = \Phi^{-۱}(p_i)$ و $F^{-۱}(p_i) = \log(\frac{p_i}{۱-p_i})$ هستند (آراندا اورداز، ۱۹۸۱). همچنین تابع پیوند نامتقارن از تابع توزیع نامتقارن تعریف می‌شود. به‌عنوان مثال، می‌توان به تابع پیوند Cloglog با شکل $F^{-۱}(p_i) = \log\{-\log(۱ - p_i)\}$ اشاره کرد (چن، ۱۹۹۹). با توجه به صورت خانواده توزیع‌های مقادیر فرین که می‌توانند متقارن، نامتقارن چوله به راست یا نامتقارن چوله به چپ باشند، چارچوب مدل خطی برای مدل مقادیر فرین تعمیم‌یافته با فرض $\epsilon_i \sim GEV(\mu = ۰, \sigma = ۱, \xi)$ در مدل (۲.۲) به صورت

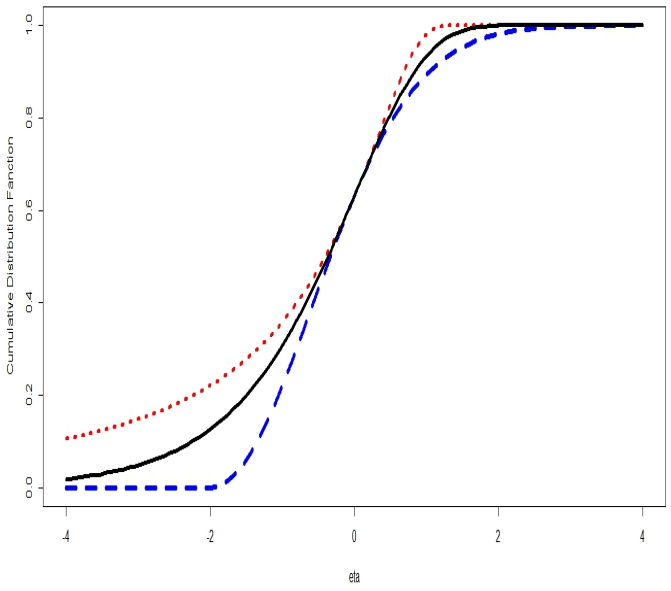
$$p_i = P(y_i = ۱) = ۱ - \exp\{-(۱ - \xi \mathrm{x_i}'\boldsymbol{\beta})_+^{-\frac{۱}{\xi}}\} = ۱ - GEV(-\mathbf{x}_i'\boldsymbol{\beta}; \xi) \qquad (۶.۲)$$

تعریف می‌شود، که در آن $GEV(x; \xi)$ تابع احتمال تجمعی توزیع مقادیر فرین تعمیم‌یافته با پارامترهای $\mu = ۰$، $\sigma = ۱$ و پارامتر شکل نامعلوم ξ در نقطه x است. بنابراین، تابع پیوند مقادیر فرین تعمیم‌یافته به صورت زیر است:

$$F^{-۱}(p_i) = \frac{(\log(۱ - p_i))^{\xi} + ۱}{\xi(\log(۱ - p_i))^{\xi}}.$$

[۲۰] Monotone Link Function

شکل ۶.۲، تابع توزیع مقادیر فرین تعمیم‌یافته را به ازای مقادیر متفاوت پارامتر شکل و مقادیر صفر و یک برای به ترتیب پارامترهای مکان و مقیاس نشان می‌دهد. همان‌طور که در شکل ۶.۲ مشاهده می‌شود، محدوده چولگی به‌دست آمده در مدل GEV با $\xi \in [-۰/۵, ۰/۵]$، نسبت به مدل t-تعمیم‌یافته با توزیع‌های مشخص، بسیار عریض‌تر است.



شکل ۶.۲: نمودار مدل رگرسیونی مقادیر فرین تعمیم‌یافته به ازای $\xi = ۰$ (خط پر)، $\xi = ۰/۵$ (نقطه‌چین) و $\xi = -۰/۵$ (خط‌چین)

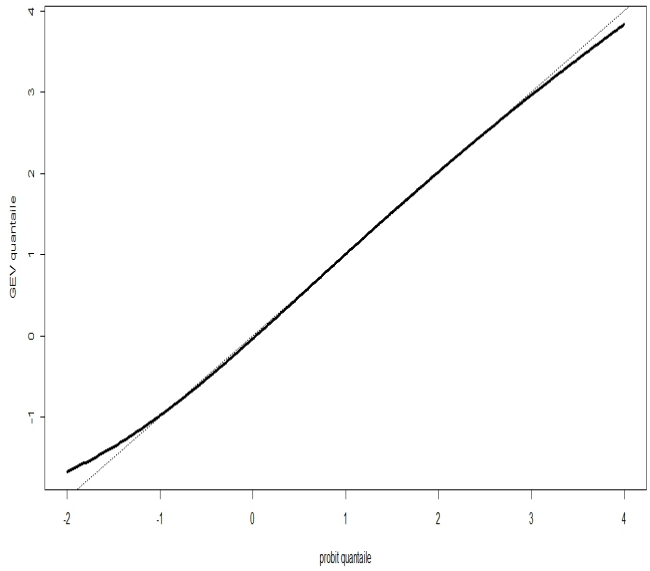
در حالت خاص، اگر $\mu \approx -۰/۳۵۵۷۹$، $\sigma \approx ۰/۹۹۰۳$ و $\xi \approx -۰/۲۷۷۶۰$ در نظر گرفته شوند، آن‌گاه توزیع نرمال استاندارد توسط توزیع مقادیر فرین تقریب زده می‌شود. یعنی تابع پیوند مقادیر فرین به تابع پیوند پرابیت تبدیل می‌شود. در شکل ۷.۲ چندک‌های توزیع مقادیر فرین در مقابل مدل پرابیت ترسیم شده‌اند. با توجه به این نمودار می‌توان نرمال مجانبی بودن مدل مقادیر فرین را نتیجه گرفت. اختلاف عمده در ناحیه دم‌ها به چشم می‌خورد.

ارزیابی چولگی مدل مقادیر فرین تعمیم‌یافته

معمولاً از معیار $\mu_۳ = \{E(x-\mu)^۳\}\{E(x-\mu)^۲\}^{\frac{-۳}{۲}}$ برای سنجش میزان چولگی توزیع‌ها استفاده می‌شود. اما این معیار در مدل GEV، به ازای مقادیر بزرگ و مثبت پارامتر شکل، ξ، وجود ندارد. بنابراین، از معیار $\gamma(\cdot)$، که در فصل اول معرفی شد، برای تعیین چولگی استفاده می‌شود.

منحنی پاسخ برای تابع پیوند مدل مقادیر فرین تعمیم‌یافته به صورت زیر تعریف می‌شود:

$$F(\mathbf{x}_i'\boldsymbol{\beta}) = h(\mathbf{x}_i) = ۱ - e^{\{-(۱-\xi\mathbf{x}_i'\boldsymbol{\beta})\}^{-\frac{۱}{\xi}}}.$$



شکل ۷.۲: نمودار چندک‌های مقادیر فرین با مقادیر $\mu \approx -۰/۳۵۵۷۹$، $\sigma \approx ۰/۹۹۹۰۳$ و $\xi \approx -۰/۲۷۷۶۰$ در مقابل چندک‌های پرابیت، خط توپر نمودار چندک و خط چین، خط مرجع است

بدون از دست دادن کلیت مسأله، فرض کنید تنها یک متغیر کمکی x وجود دارد. برای محاسبه نرخ شیب منحنی پاسخ که اندازه‌گیری آن در نزدیکی ۰ و ۱ اهمیت دارد، داریم

$$\begin{aligned}\frac{\partial h(x)}{\partial x} &= \xi\,\beta\, e^{\{-(۱-\xi \mathrm{x}\beta)\}^{-\frac{۱}{\xi}}} \times \frac{۱}{\xi}\,(۱-\xi \mathrm{x}\beta)^{-\frac{۱}{\xi}-۱} \\ &= \beta\, e^{\{-(۱-\xi \mathrm{x}\beta)\}^{-\frac{۱}{\xi}}} \times (۱-\xi \mathrm{x}\beta)^{-\frac{۱}{\xi}-۱}.\end{aligned}$$

تغییرات نرخ $\frac{\partial h(x)}{\partial x}$ نیز به صورت زیر به‌دست می‌آید:

$$\frac{\partial^۲ h(x)}{\partial x^۲} = \beta^۲\, e^{\{-(۱-\xi \mathrm{x}\beta)\}^{-\frac{۱}{\xi}}} \times (۱-\xi \mathrm{x}\beta)^{-\frac{۱}{\xi}-۲} \times [(۱+\xi)-(۱-\xi x\beta)^{-\frac{۱}{\xi}}].$$

هنگامی‌که $\xi \leq -۱$ باشد، $\frac{\partial^۲ h(x)}{\partial x^۲} < ۰$ است. بنابراین، زمانی‌که x افزایش می‌یابد به این معنی که β مثبت است، نرخ $\frac{\partial h(x)}{\partial x}$ کاهشی است. در نتیجه، سرعت زیاد شدن منحنی پاسخ در نزدیکی ۰ بیشتر از سرعت زیاد شدن آن در نزدیکی ۱ است. بنابراین، تابع پیوند GEV برای $\xi \leq -۱$ همیشه چولگی منفی دارد. زمانی‌که x کاهش می‌یابد به این معنی که β منفی است، در نتیجه، سرعت زیاد شدن منحنی پاسخ در نزدیکی ۱ بیشتر از سرعت زیاد شدن آن در نزدیکی ۰ است. بنابراین، تابع پیوند GEV برای $\xi \leq -۱$ چولگی منفی دارد.

هنگامی‌که $\xi > -۱$ باشد، با توجه به شرط $(۱-\xi x\beta) > ۰$، دو حالت اتفاق می‌افتد:

- برای آن‌که $\frac{\partial^۲ h(x)}{\partial x^۲}$ مثبت باشد، باید

$$[(۱+\xi)-(۱-\xi x\beta)^{-\frac{۱}{\xi}}] > ۰.$$

بنابراین

$$[(۱+\xi)-(۱-\xi x\beta)^{-\frac{۱}{\xi}}]^{-\xi}>۰.$$

اکنون می‌توان نتیجه گرفت

$$(۱+\xi)^{-\xi}>(۱-\xi x\beta).$$

پس

$$\xi x\beta>۱-(۱+\xi)^{-\xi}$$

و در نهایت برای

$$x>\frac{۱-(۱+\xi)^{-\xi}}{\xi\beta}$$

نرخ $\frac{\partial h(x)}{\partial x}$ افزایشی است.

- برای منفی بودن $\frac{\partial^۲ h(x)}{\partial x^۲}$، باید

$$[(۱+\xi)-(۱-\xi x\beta)^{-\frac{۱}{\xi}}]<۰.$$

بنابراین

$$[(۱+\xi)-(۱-\xi x\beta)^{-\frac{۱}{\xi}}]^{-\xi}<۰.$$

در نتیجه

$$(۱+\xi)^{-\xi}<(۱-\xi x\beta).$$

که به راحتی می‌توان نوشت

$$\xi x\beta<۱-(۱+\xi)^{-\xi}.$$

بنابراین برای

$$x<\frac{۱-(۱+\xi)^{-\xi}}{\xi\beta}$$

نرخ $\frac{\partial h(x)}{\partial x}$ کاهشی است.

پس مقدار ماکسیمم که نزدیک به نرخ $\frac{\partial h(x)}{\partial x}$ باشد، در نقطه $x^*=\frac{۱-(۱+\xi)^{-\xi}}{\xi\beta}$ اتفاق می‌افتد، که این حداکثر مقدار همان مد است. تابع پیوند توزیع مقادیر فرین تعمیم‌یافته، تک مدی است و مقدار تابع چگالی احتمال آن در مد، برابر است با

$$f(x^*)=۱-e^{-(۱+\xi)}.$$

تحت شرایطی که در بخش (۸.۱) بیان شد، چولگی متغیر تصادفی X به صورت

$$\gamma(X)=۱-۲F(M_x)$$

تعریف می‌شود. چولگی تابع پیوند (۶.۲) با توجه به رابطه‌ای که برای مد به دست آمد، به ازای مقدار $\xi > -۱$ به صورت

$$\begin{aligned}\gamma_M(X) &= ۱ - ۲F(M_x)\\ &= ۱ - ۲(۱ - \exp[-(۱ + \xi\beta M_x)_+^{-\frac{۱}{\xi}}])\\ &= ۱ - ۲(۱ - \exp[-(۱ + \xi\beta\frac{(۱+\xi)^{-\xi} - ۱}{\xi\beta}]_+^{\frac{-۱}{\xi}})\\ &= ۲\exp[-(۱+\xi)^{-\xi}]_+^{\frac{-۱}{\xi}}] - ۱\\ &= ۲\exp[-(۱+\xi)] - ۱\end{aligned}$$

محاسبه می‌شود. بنابراین بر حسب پارامتر شکل مدل، می‌توان مقادیری از ξ را که به ازای آن‌ها چولگی تابع پیوند منفی است، به دست آورد. برای چولگی منفی باید

$$۲e^{[-(۱+\xi)]} - ۱ > ۰.$$

پس خواهیم داشت

$$e^{[-(۱+\xi)]} > \frac{۱}{۲}.$$

در نتیجه

$$\xi < \log ۲ - ۱.$$

پس توزیع GEV هنگامی‌که $\xi < \log ۲ - ۱$ است، چولگی مثبت دارد. بنابراین، تابع پیوند GEV هنگامی‌که $\xi < \log ۲ - ۱$ است، چولگی منفی دارد. در نتیجه، سرعت زیاد شدن منحنی پاسخ در نزدیکی ۰ بیشتر از سرعت زیاد شدن آن در نزدیکی ۱ است. به‌طور مشابه برای بررسی چولگی مثبت تابع پیوند، باید

$$۲e^{[-(۱+\xi)]} - ۱ < ۰.$$

بنابراین

$$e^{[-(۱+\xi)]} < \frac{۱}{۲}$$

که نتیجه می‌شود

$$\xi > \log ۲ - ۱.$$

بنابراین، هنگامی‌که $\xi > \log ۲ - ۱$ است، تابع پیوند چولگی مثبت دارد. در نتیجه، سرعت زیاد شدن منحنی پاسخ در نزدیکی ۱ بیشتر از سرعت زیاد شدن آن در نزدیکی ۰ است. این نتایج در مقاله آرنولد و گرینولد (۱۹۹۵) بیان شده است.

فصل ۳

مدل بیزی مقادیر فرین تعمیم‌یافته

پارامترهای مدل مقادیر فرین را می‌توان با روش‌های گشتاوری، درستنمایی و بیزی برآورد کرد. از میان روش‌های متفاوت برآوردیابی، روش درستنمایی ماکسیمم در عین سادگی دارای مزیت‌هایی است. این روش علاوه بر دارا بودن ویژگی‌های مجانبی بهینه، از لحاظ کاربردی دارای انعطاف‌پذیری مناسبی در مدل‌بندی اطلاعات متغیرهای تبیینی می‌باشد. مشکلی که استفاده از برآوردگرهای درستنمایی ماکسیمم را، در صورت وجود، محدود می‌کند، کوچک بودن حجم نمونه است. در عمل با وجود مهیا بودن مجموعه داده‌های بزرگ بعد از تعریف مقادیر فرین، مشاهدات کمی برای برازش باقی می‌مانند. از طرف دیگر ممکن است حجم داده‌ها مناسب باشد، اما پراکندگی زیاد داده‌ها مشکلاتی را در برآورد پارامترها با روش درستنمایی ماکسیمم ایجاد کند. این محدودیت‌ها موجب می‌شود که برآوردهای درستنمایی ماکسیمم از دقت کافی برخوردار نباشند.

در مقابل، استنباط بیزی در این مدل می‌تواند جانشین بهتری برای استنباط مبتنی بر درستنمایی باشد و از مشکلات نظری موجود برای برآوردگرهای درستنمایی ماکسیمم مبرا باشد. در این فصل، یک نسخه بیزی را برای مدل رگرسیون مقادیر فرین تعمیم‌یافته ارایه داده و آن را با استفاده از الگوریتم‌های MCMC برازش می‌دهیم.

۱.۳ استنباط درستنمایی ماکسیمم

برآوردگرهای درستنمایی ماکسیمم که از ماکسیمم کردن تابع درستنمایی حاصل می‌شوند، دارای ویژگی‌های مجانبی بهینه هستند که در صورت برقرار بودن شرایط نظم برای استنباط‌های بعدی مفید می‌باشند؛

اما به دلیل آن‌که تکیه‌گاه توزیع مقادیر فرین تعمیم‌یافته به پارامتر وابسته است، شرایط نظم ممکن است برقرار نباشند. اسمیت (۱۹۸۵) برآوردگرهای درستنمایی ماکسیمم پارامترهای توزیع مقادیر فرین تعمیم‌یافته را بررسی نمود و نشان داد:

- اگر $\xi > -۰/۵$ باشد، برآوردگرهای درستنمایی ماکسیمم دارای ویژگی‌های مجانبی معمول هستند.
- اگر $-۱ < \xi < -۰/۵$ باشد، برآوردگرهای درستنمایی ماکسیمم موجود بوده اما دارای ویژگی‌های مجانبی استاندارد نیستند.
- اگر $\xi < -۱$ باشد، برآوردگرهای درستنمایی ماکسیمم وجود ندارند.

بنابراین تنها به ازای $\xi > -۰/۵$ ویژگی‌های مجانبی مانند سازگاری، کارایی و نرمال بودن مجانبی توزیع برآوردگرها قابل حصول هستند.

فرض کنید $Y_۱, \ldots, Y_n$، متغیرهای تصادفی مستقل با تابع توزیع مقادیر فرین تعمیم‌یافته باشند. لگاریتم تابع درستنمایی مشروط بر آن‌که برای هر $i = ۱, \ldots, n$، $۱ + \xi(\frac{y_i - \mu}{\sigma}) > ۰$ باشد، به ازای $\xi \neq ۰$ برابر است با

$$\begin{aligned} \ell(\mu, \sigma, \xi) = -n \log \sigma - (۱ + \frac{۱}{\xi}) \sum_{i=۱}^{n} \log \left[۱ + \xi(\frac{y_i - \mu}{\sigma}) \right] \\ - \sum_{i=۱}^{n} \left[۱ + \xi(\frac{y_i - \mu}{\sigma}) \right]^{-\frac{۱}{\xi}} \end{aligned} \tag{۱.۳}$$

و برای $\xi = ۰$ عبارت است از

$$\ell(\mu, \sigma, \xi) = -n \log \sigma - \sum_{i=۱}^{n} \left(\frac{y_i - \mu}{\sigma} \right) - \sum_{i=۱}^{n} \exp \left\{ - \left(\frac{y_i - \mu}{\sigma} \right) \right\}. \tag{۲.۳}$$

ماکسیمم کردن (۱.۳) و (۲.۳) با روش‌های عددی مانند الگوریتم نیوتن-رافسون بر حسب پارامترهای $(\mu.\sigma, \xi)$ امکان‌پذیر است. این الگوریتم‌ها برآوردهای گشتاوری معمولی را به‌عنوان مقدار اولیه به کار می‌گیرند. هرچند که برای نمونه‌های با حجم کوچک گاهی چنین الگوریتم‌هایی همگرا نمی‌شوند، اما به‌طور قابل توجهی استوار هستند.

با توجه به معایب اشاره‌شده برای انجام استنباط مبتنی بر درستنمایی در مدل مقادیر فرین تعمیم‌یافته، از رهیافت بیزی در این پایان‌نامه برای استنباط در این مدل استفاده می‌کنیم.

۲.۳ استنباط بیزی

تحلیل مقادیر فرین با رهیافت بیزی از چند نظر دارای اهمیت است. نخست به دلیل پراکندگی زیاد مقادیر فرین، استفاده از اطلاعات پیشین می‌تواند موجب بهبود استنباط شود. از طرف دیگر وقتی پارامتر $\xi < -۱$ است، برآوردگرهای درستنمایی ماکسیمم موجود نیستند.

فرض کنید $D_{obs} = (n, \mathrm{y}, X)$ داده‌های مشاهده‌شده در مدل مقادیر فرین تعمیم‌یافته باشند، که در آن X ماتریس طرح مدل است که ستون‌های آن مشاهدات متغیرهای تبیینی هستند، y بردار مشاهدات پاسخ است و n حجم نمونه است. با توجه به روابط (۵.۲) و (۶.۲) و با در نظر گرفتن تابع درستنمایی مشاهدات به صورت

$$L(\boldsymbol{\beta}, \xi|\mathrm{y}, X) = \prod_{i=1}^{n} \{1 - GEV(-\mathrm{x}_i'\boldsymbol{\beta}; \xi)\}^{y_i} \{GEV(-\mathrm{x}_i'\boldsymbol{\beta}; \xi)\}^{1-y_i}$$

توزیع پسین پارامترها به شرط مشاهدات به صورت

$$\pi(\boldsymbol{\beta}, \xi|D_{obs}) \propto L(\boldsymbol{\beta}, \xi|\mathrm{y}, X)\pi(\boldsymbol{\beta}|\xi)\pi(\xi) \qquad (۳.۳)$$

حاصل می‌شود.

تحلیل بیزی نیازمند تعیین توزیع پیشین برای پارامترها است. در حالت کلی، هر توزیع پیشینی می‌تواند استفاده شود. وقتی اطلاعات پیشین در مورد پارامترها در اختیار نباشند و استنباط با رهیافت بیزی صرفاً برای برطرف نمودن محدودیت‌های محاسباتی استفاده شود، معمولاً اطلاعات به‌دست آمده از مشاهدات در تابع درستنمایی بر اطلاعات پیشین ارجحیت دارند. در نتیجه از پیشین‌های ناآگاهی‌بخش استفاده می‌شود. در برخی از موارد استفاده از پیشین‌های ناآگاهی‌بخش می‌تواند منجر به پسین ناسره گردد. استنباط با توزیع پسین ناسره امکان‌پذیر نیست. در نتیجه قبل از انجام استنباط، باید سره بودن توزیع پسین حتما بررسی و تایید شود.

در این پایان‌نامه برای پارامتر شکل ξ از توزیع پیشین سره استفاده می‌کنیم. از دیدگاه نظری، دامنه تغییرات ξ مجموعه اعداد حقیقی $\mathbb{R}$ است؛ اما همان‌طور که در کولز (۲۰۰۱) و کاتز و نادارجا (۲۰۰۰) پیشنهاد شده است، تجربه تحلیل داده‌های واقعی نشان داده که در نظر گرفتن $-1/2 < \xi < 1/2$ تقریبا همیشه در کاربردهای عملی راضی‌کننده است. مقادیر مثبت بزرگ ξ، تحت توزیع‌های پیشین ناسره برای $\boldsymbol{\beta}$ مثل توزیع‌های پیشین یکنواخت، به پسین‌های ناسره منجر می‌شوند و مقادیر منفی بزرگ برای ξ به ندرت در عمل دیده می‌شوند (کولز، ۲۰۰۱). بنابراین منطقی است که برای محاسبات بیزی بپذیریم که $-1 \leq \xi < 1$. در این فصل توزیع پیشین یکنواخت در فاصله $[-1, 1)$ را برای ξ در نظر می‌گیریم. البته در فصل بعد، برای اجرای شبیه‌سازی‌ها از توزیع پیشین $N(0, \sigma_\xi^2)$ برای ξ استفاده خواهیم کرد.

تحت تابع پیوند مقادیر فرین، توزیع پسین به ازای بسیاری از پیشین‌های ناآگاهی‌بخش برای پارامترهای رگرسیونی $\boldsymbol{\beta}$، سره است. از جمله این پیشین‌ها، می‌توان به پیشین ناآگاهی‌بخش جفریز و پیشین یکنواخت ناسره اشاره کرد. در ادامه این دو پیشین را در مدل GEV برای پارامتر $\boldsymbol{\beta}$ معرفی می‌کنیم و سره بودن توزیع پسین را نسبت به توزیع پیشین ناسره یکنواخت بررسی می‌کنیم. در فصل بعدی برای اجرای مطالعه شبیه‌سازی از توزیع‌های پیشین نرمال $\boldsymbol{\beta}_j \sim N(0, \sigma_{\beta_j}^2)$، $j = 1, \ldots, p$، استفاده می‌کنیم.

پیشین جفریز

پیشین جفریز برای پارامترهای رگرسیونی $\boldsymbol{\beta}$ در مدل GEV به صورت $\pi(\boldsymbol{\beta}|\xi) \propto |I(\boldsymbol{\beta}|\xi)|^{\frac{۱}{۲}}$ تعریف می‌شود. در قضیه زیر ماتریس اطلاع فیشر برای پارامتر β در مدل GEV را به‌دست می‌آوریم.

قضیه ۱.۲.۳. *ماتریس اطلاع فیشر، $I(\boldsymbol{\beta}|\xi)$، به صورت $X'\Omega X$ نوشته می‌شود، که در آن*

$$\Omega = diag(w_۱, \ldots, w_n)$$

و برای $i = ۱, \ldots, n$

$$w_i = \{(۱ - \xi x_i'\boldsymbol{\beta})^{\frac{-۲}{\xi}-۲}\}[\exp\{(۱ - \xi x_i'\boldsymbol{\beta})^{\frac{-۱}{\xi}}\} - ۱]^{-۱}.$$

برهان. لگاریتم تابع درستنمایی عبارت است از

$$\begin{aligned}
\ln L(\boldsymbol{\beta}, \xi|\mathrm{y}, X) &= \sum_{i=۱}^{n} [\, y_i \log(۱ - e^{-(۱-x_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}) + (۱ - y_i)\log(e^{-(۱-x_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}})] \\
&= \sum_{i=۱}^{n} [\, y_i \log\left(\frac{۱ - e^{-(۱-x_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}}{e^{-(۱-x_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}}\right) - (۱ - x_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}] \qquad (۴.۳) \\
&= \sum_{i=۱}^{n} [\, y_i \log(e^{(۱-x_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}} - ۱) - (۱ - x_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}].
\end{aligned}$$

به ازای $j = ۱, \ldots, p$ و $k = ۱, \ldots, p$، از رابطه (۴.۳) نسبت به $\boldsymbol{\beta}$ مشتق می‌گیریم. با قرار دادن $\ell(\boldsymbol{\beta}, \xi|\mathrm{y}, X) = \ln L(\boldsymbol{\beta}, \xi|\mathrm{y}, X)$، داریم

$$\begin{aligned}
&\frac{\partial \ell(\boldsymbol{\beta}, \xi|\mathrm{y}, X)}{\partial \boldsymbol{\beta}_j} \\
&= \sum_{i=۱}^{n} \left[y_i \frac{e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}\, x_{ij}\,(۱ - \mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۱}}{(e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}} - ۱)} - x_{ij}\,(۱ - \mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۱} \right] \\
&= \sum_{i=۱}^{n} x_{ij}\,(۱ - \mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۱} \left[\frac{y_i\, e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}}{(e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}} - ۱)} - ۱ \right] \\
&= \sum_{i=۱}^{n} \mathrm{x}_i\,(۱ - \mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۱} \left[y_i - ۱ + \frac{y_i}{(e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}} - ۱)} \right].
\end{aligned}$$

با قرار دادن $A = \mathrm{x}_i\,(۱ - \mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۱}$ و $B = \left[y_i - ۱ + \frac{y_i}{(e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}} - ۱)} \right]$، داریم

$$\begin{aligned}
\frac{\partial^۲ \ell(\boldsymbol{\beta}, \xi|\mathrm{y}, X)}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_k} &= \sum_{i=۱}^{n} (B\frac{\partial A}{\partial \boldsymbol{\beta}_k} + \frac{\partial B}{\partial \boldsymbol{\beta}_k} A) \\
\frac{\partial A}{\partial \boldsymbol{\beta}_k} &= x_{ij}\, x_{ik}\,(۱ + \xi)(۱ - \mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۲} \\
\frac{\partial B}{\partial \boldsymbol{\beta}_k} &= -y_i\,(e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}} - ۱)^{-۲}\, e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}\, x_{ik}\,(۱ - \mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۱}.
\end{aligned}$$

در نتیجه می‌توان نوشت

$$\frac{\partial^{۲}\ell(\boldsymbol{\beta},\xi|\mathrm{y},X)}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_k}=\sum_{i=۱}^{n}[(\,y_i-۱+\frac{y_i}{(e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}-۱)})\,x_{ij}\,x_{ik}\,(۱+\xi)(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۲}$$
$$-\frac{y_i\,e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}}{(e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}-۱)^{۲}}\,x_{ik}\,(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۱}\,x_{ij}\,(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}-۱}].$$

از آن‌جایی که $E(y_i)=p_i=۱-e^{-(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}$، داریم

$$E\left(\frac{\partial^{۲}\ell(\boldsymbol{\beta},\xi|\mathrm{y},X)}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_k}\right)=۰-\sum_{i=۱}^{n}\frac{(۱-e^{-(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}})e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}}{(e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}-۱)^{۲}}x_{ij}x_{ik}(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۲}{\xi}-۲}$$
$$=-\sum_{i=۱}^{n}\frac{x_{ij}x_{ik}(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۲}{\xi}-۲}}{e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}-۱}.$$

بنابراین درایه (j,k) ماتریس اطلاع فیشر به صورت

$$I_{jk}=-E\left(\frac{\partial^{۲}\ell(\boldsymbol{\beta},\xi|\mathrm{y},X)}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_k}\right)=\sum_{i=۱}^{n}\frac{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۲}{\xi}-۲}}{e^{(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)^{-\frac{۱}{\xi}}}-۱}x_{ij}x_{ik}$$

است. این نتیجه مشابه نتیجه‌ای است که توسط گوپتا و ابراهیم (۲۰۰۸) بیان شده است. در این حالت، توزیع پسین به صورت

$$\pi(\boldsymbol{\beta},\xi|D_{obs})\propto\prod_{i=۱}^{n}\{۱-GEV(-\mathrm{x}_i'\boldsymbol{\beta};\xi)\}^{y_i}\{GEV(-\mathrm{x}_i'\boldsymbol{\beta};\xi)\}^{۱-y_i}|I(\boldsymbol{\beta}|\xi)|^{\frac{۱}{۲}}\frac{۱}{۲}I_{[-۱,۱)}(\xi)$$

حاصل می‌شود. □

پیشین یکنواخت

با در نظر گرفتن پیشین یکنواخت ناسره برای $\boldsymbol{\beta}$، $\pi(\boldsymbol{\beta})\propto ۱$، توزیع پسین به صورت

$$\pi(\boldsymbol{\beta},\xi|D_{obs})\propto\prod_{i=۱}^{n}\{۱-GEV(-\mathrm{x}_i'\boldsymbol{\beta};\xi)\}^{y_i}\{GEV(-\mathrm{x}_i'\boldsymbol{\beta};\xi)\}^{۱-y_i}\times ۱\times\frac{۱}{۲}I_{[-۱,۱)}(\xi)$$

حاصل می‌شود. با انتخاب این توزیع پیشین، توزیع پسین در رابطه (۳.۳) سره می‌شود، اگر و فقط اگر

$$\int_{-۱}^{۱}\int_{\mathbb{R}^p}L(\boldsymbol{\beta},\xi|\mathrm{y},\mathrm{X})\,\mathrm{d}\boldsymbol{\beta}\,\mathrm{d}\xi<\infty. \qquad (۵.۳)$$

برای اثبات (۵.۳)، قضیه ۴.۲.۳ را در زیر بیان می‌کنیم. قبل از بیان قضیه مورد نظر، باید یک لم و یک قضیه دیگر را مطرح کنیم، که توسط چن و شائو (۲۰۰۱) ارائه گردیده‌اند.

برای هر $i=۱,\ldots,n$، تعریف می‌کنیم $\tau_i=۱$ هر گاه $y_i=۱$ و $\tau_i=-۱$ هر گاه $y_i=۰$ باشد. فرض کنید X ماتریس طرح $n\times p$ با سطرهای x_i'، برای $i=۱,\ldots,n$، باشد. همچنین $X^*_{l,m}=(\tau_i\mathrm{x}_i', l<i\leq m)$ را یک ماتریس $(m-l)\times p$ با سطرهای $\tau_i\mathrm{x}_i'$، $l<i\leq m$، تعریف می‌کنیم، به‌طوری که $۰\leq l<m\leq n$. قضیه ۲.۲.۳، شرایط لازم و کافی برای برقراری رابطه (۵.۳) را مطرح کرده است.

قضیه ۲.۲.۳. *فرض کنید شرایط زیر برقرار باشند:*

C۱) ماتریس طرح X پر رتبه باشد.

C۲) بردار مثبت $\boldsymbol{a} = (a_1, \ldots, a_p)' \in \mathbb{R}^p$ وجود داشته باشد، به این معنی که هر مولفه $a_i > 0$ باشد، به‌طوری که

$$X^{*\prime}\boldsymbol{a}_i = 0.$$

C۳)

$$\int_{-\infty}^{\infty} |u|^k \mathrm{d}F(u) < \infty.$$

در این صورت

$$\int_{-1}^{1} \int_{\mathbb{R}^p} L(\boldsymbol{\beta}, \xi | y, X) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\xi < \infty.$$

لم ۳.۲.۳. *فرض کنید شرایط $C1$) و $C2$) قضیه ۲.۲.۳ برقرار باشند. همچنین $X^*_{m_{l-1},m_l}$ یک ماتریس پر رتبه باشد، به‌طوری که برای هر $l = 1, \ldots, p$، $a_l' X^*_{m_{l-1},m_l} = 0$ شود. زمانی که*

$$X^*\boldsymbol{\beta} \leq \boldsymbol{u}$$

برقرار باشد، ثابتی مثل k وجود دارد به‌طوری که

$$\|\boldsymbol{\beta}\| \leq k\|\boldsymbol{u}\|$$

یا

$$\|\boldsymbol{\beta}\| \leq k \min_{1 \leq l \leq p} \max_{m_{l-1} < i \leq m_l} |\boldsymbol{u_i}|.$$

اکنون به بیان قضیه مورد نظر می‌پردازیم.

قضیه ۴.۲.۳. *فرض کنید $r > p$، $0 = m_0 < \ldots < m_r \leq n$، و بردارهای مثبت $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$ وجود داشته باشند به‌طوری که $X^*_{m_{l-1},m_l}$ یک ماتریس پر رتبه باشد و برای هر $l = 1, \ldots, r$، $\boldsymbol{a}_l' X^*_{m_{l-1},m_l} = 0$. تحت توزیع پیشین یکنواخت ناسره $\pi(\boldsymbol{\beta}) \propto 1$، توزیع پسین در رابطه (۳.۳) سره خواهد شد.*

برهان. فرض کنید $\boldsymbol{u} = (u_1, \ldots, u_n)'$ متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع مقادیر فرین تعمیم‌یافته با $\mu = 0$، $\sigma = 1$ و پارامتر شکل ξ است که با F نمایش می‌دهیم. با توجه به روابط

$$1 - F(-x) = EI(u_i > -x)$$

و

$$F(-x) = EI(-u_i \geq -(-x))$$

می‌توان نوشت

$$[1 - F(-\mathrm{x_i}'\boldsymbol{\beta})]^{y_i} [F(-\mathrm{x_i}'\boldsymbol{\beta})]^{1-y_i} \leq EI[\tau_i u_i \geq \tau_i(-\mathrm{x_i}'\boldsymbol{\beta})]$$

و

$$[1 - F(-\mathrm{x_i}'\boldsymbol{\beta})]^{y_i} [F(-\mathrm{x_i}'\boldsymbol{\beta})]^{1-y_i} \geq EI[\tau_i u_i > \tau_i(-\mathrm{x_i}'\boldsymbol{\beta})].$$

فرض کنید $\boldsymbol{u}^* = (\tau_1 u_1, \ldots, \tau_n u_n)$ باشد. با توجه به لم ۳.۲.۳ و قضیه فوبینی، می‌توان نتیجه گرفت

$$\begin{aligned}
&\int_{-1}^{1}\int_{\mathbb{R}^p} L(\boldsymbol{\beta},\xi|y,\mathrm{X})\,\mathrm{d}\boldsymbol{\beta}\,\mathrm{d}\xi \\
=\;& \int_{-1}^{1}\int_{\mathbb{R}^n}[\int_{\mathbb{R}^p} I(-\tau \mathrm{x_i}'\boldsymbol{\beta} < \tau_i u_i,\ 1 \le i \le n)\,\mathrm{d}\boldsymbol{\beta}]\,\mathrm{d}F(\boldsymbol{u})\,\mathrm{d}\xi \\
=\;& \int_{-1}^{1}\int_{\mathbb{R}^n}[\int_{\mathbb{R}^p} I(\mathrm{X}^*\boldsymbol{\beta} < \boldsymbol{u}^*)\,\mathrm{d}\boldsymbol{\beta}]\,\mathrm{d}F(\boldsymbol{u})\,\mathrm{d}\xi \\
\le\;& \int_{-1}^{1}\int_{\mathbb{R}^n}[\int_{\mathbb{R}^p} I(\|\boldsymbol{\beta}\| \le k \min_{1\le l\le r}\max_{m_{l-1}<i\le m_l}|\boldsymbol{u}^*|)\mathrm{d}\beta]\mathrm{d}F(\boldsymbol{u})\,\mathrm{d}\xi \\
\le\;& k\int_{-1}^{1}\int_{\mathbb{R}^n}\prod_{l=1}^{r} E(\max_{m_{l-1}<i\le m_l}|\boldsymbol{u}^*|^{\frac{r}{p}})\mathrm{d}F(\boldsymbol{u})\,\mathrm{d}\xi. \qquad (۶.۳)
\end{aligned}$$

برای این‌که بتوانیم نشان دهیم رابطه (۶.۳) متناهی است، ابتدا باید نشان دهیم $E_\xi(|u|^a)$ به ازای $0 < a < 1$ یک تابع توزیع پیوسته از ξ، برای $-1 \le \xi < 1$، است. تابع چگالی u برای زمانی‌که $u \sim F(\cdot)$ برابر است با

$$f_\xi(u) = \begin{cases} \exp\left[-\{1+\xi u\}^{\frac{-1}{\xi}}\right]\frac{1}{\{1+\xi u\}^{\frac{1}{\xi}+1}}, & u > \frac{-1}{\xi} \text{ اگر } \xi > 0;\ u < \frac{-1}{\xi} \text{ اگر } \xi < 0 \\ \exp(-\exp(-u))\exp(-u), & -\infty < u < \infty \text{ اگر } \xi = 0. \end{cases}$$

هنگامی که $\xi \ne 0$ اگر تعریف کنیم $t = \{1+\xi u\}^{\frac{-1}{\xi}}$، می‌توان نوشت

$$E_\xi(|u|^a) = \int_0^\infty |\frac{1}{\xi}(t^{-\xi}-1)|^a e^{-t}\mathrm{d}t.$$

به‌طور مشابه هنگامی‌که $\xi = 0$ باشد اگر قرار دهیم $t = e^{-u}$، می‌توان نوشت

$$E_\xi(|u|^a) = \int_0^\infty |-\log t|^a e^{-t}\mathrm{d}t.$$

به‌ازای $t > 0$ داریم

$$h_t(\xi) = \begin{cases} \frac{t^{-\xi}-1}{\xi} & \text{اگر} \quad \xi \ne 0 \\ -\log t & \text{اگر} \quad \xi = 0. \end{cases}$$

بنابراین برای $t > 0$ و $-1 \le \xi < 1$ داریم

$$-(t+1) \le h_t(\xi) \le \frac{1}{t} + 1.$$

پس اگر $|\xi| \le 1$ باشد، آن‌گاه برای $t > 0$

$$\left|\frac{t^{-\xi}-1}{\xi}\right| \le \max(\frac{1}{t}+1, t+1)$$
$$|\log t| \le \max(\frac{1}{t}+1, t+1).$$

حال فرض کنید

$$h(t) = \begin{cases} \left(\frac{1}{t}+1\right)^a & \text{اگر} \quad 0 < t < 1 \\ (t+1)^a & \text{اگر} \quad t \ge 1. \end{cases}$$

بنابراین اگر $|\xi| \leq ۱$ باشد، برای $t > ۰$ می‌توان نتیجه گرفت

$$\left|\frac{۱}{\xi}(t^{-\xi} - ۱)\right|^a \leq h(t)$$
$$|\log t|^a \leq h(t).$$

در نتیجه برای $۰ < a < ۱$

$$\int_۰^\infty h(t)e^{-t}\mathrm{d}t < \infty.$$

از آن‌جایی که $\left|\frac{۱}{\xi}(t^{-\xi} - ۱)\right|^a$ تابعی پیوسته برای ξ است، طبق قضیه مقدار میانگین $E_\xi(|u|^a)$ نیز پیوسته است. در نتیجه برهان قضیه کامل می‌شود. □

۳.۳ برازش مدل با استفاده از نمونه‌گیری MCMC

توزیع پسین تحت پیشین ناآگاهی‌بخش جفریز به صورت

$$\begin{aligned}
&\pi(\boldsymbol{\beta}, \xi|D_{obs}) \propto L(\boldsymbol{\beta}, \xi|\mathrm{y}, X)\pi(\boldsymbol{\beta}|\xi)\pi(\xi) \\
&= \prod_{i=۱}^{n}\{۱ - GEV(-\mathrm{x}_i'\boldsymbol{\beta}; \xi)\}^{y_i}\{GEV(-\mathrm{x}_i'\boldsymbol{\beta}; \xi)\}^{۱-y_i}\pi(\boldsymbol{\beta}|\xi)\pi(\xi) \\
&= \prod_{i=۱}^{n}\left(۱ - e^{-(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-۱}{\xi}}}\right)^{y_i}\left(e^{-(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-۱}{\xi}}}\right)^{۱-y_i}|I(\boldsymbol{\beta}|\xi)|^{\frac{۱}{۲}}\frac{۱}{۲}I_{[-۱,۱)}(\xi)
\end{aligned} \tag{۷.۳}$$

و تحت پیشین ناآگاهی‌بخش یکنواخت به صورت

$$\begin{aligned}
&\pi(\boldsymbol{\beta}, \xi|D_{obs}) \propto L(\boldsymbol{\beta}, \xi|\mathrm{y}, X)\pi(\boldsymbol{\beta}|\xi)\pi(\xi) \\
&= \prod_{i=۱}^{n}\{۱ - GEV(-\mathrm{x}_i'\boldsymbol{\beta}; \xi)\}^{y_i}\{GEV(-\mathrm{x}_i'\boldsymbol{\beta}; \xi)\}^{۱-y_i}\pi(\boldsymbol{\beta}|\xi)\pi(\xi) \\
&= \prod_{i=۱}^{n}\left(۱ - e^{-(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-۱}{\xi}}}\right)^{y_i}\left(e^{-(۱-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-۱}{\xi}}}\right)^{۱-y_i}\frac{۱}{۲}I_{[-۱,۱)}(\xi)
\end{aligned} \tag{۸.۳}$$

است. با توجه به این‌که توزیع مقادیر فرین تعمیم‌یافته دارای توزیع پیشین مزدوج نیست (کلز و پاول، ۱۹۹۶)، توزیع پسین پارامترها با توزیع‌های پیشین در نظر گرفته‌شده، صورت بسته‌ای ندارد. یکی از روش‌های معمول برای تقریب این‌گونه توزیع‌های پیچیده استفاده از الگوریتم‌های MCMC است. چنانچه بخواهیم از نمونه‌گیر گیبز برای تولید نمونه از پسین‌های (۷.۳) و (۸.۳) استفاده کنیم، باید توزیع‌های شرطی کامل $\pi(\boldsymbol{\beta}|\mathrm{y}, X, \xi)$ و $\pi(\xi|\mathrm{y}, X, \boldsymbol{\beta})$ را محاسبه و از آن‌ها تولید نمونه کنیم. اما این کار به راحتی امکان‌پذیر نیست. زیرا توزیع‌های شرطی کامل پارامترها دارای شکل بسته و مشخص نیستند. به‌عنوان مثال، توزیع‌های شرطی کامل پارامترهای $\boldsymbol{\beta}$ و ξ با پیشین ناآگاهی‌بخش جفریز برای

$\boldsymbol{\beta}$، به صورت زیر خواهند بود:

$$\begin{aligned}\pi(\boldsymbol{\beta}|\mathrm{y},X,\xi)&=\frac{\pi(\boldsymbol{\beta},\mathrm{y},X,\xi)}{\pi(\mathrm{y},X,\xi)}=\frac{L(\boldsymbol{\beta},\xi|\mathrm{y},X)\pi(\boldsymbol{\beta}|\xi)\pi(\xi)}{L(\boldsymbol{\beta},\xi|\mathrm{y},X)\pi(\xi)}\\&=\frac{\prod_{i=1}^{n}\left(1-e^{-(1-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-1}{\xi}}}\right)^{y_i}\left(e^{-(1-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-1}{\xi}}}\right)^{1-y_i}\times|I(\boldsymbol{\beta}|\xi)|^{\frac{1}{2}}}{\int_{\mathbb{R}^p}\prod_{i=1}^{n}\left(1-e^{-(1-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-1}{\xi}}}\right)^{y_i}\left(e^{-(1-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-1}{\xi}}}\right)^{1-y_i}\mathrm{d}\boldsymbol{\beta}}\end{aligned}$$

و

$$\begin{aligned}\pi(\xi|\mathrm{y},X,\boldsymbol{\beta})&=\frac{\pi(\xi,\mathrm{y},X,\boldsymbol{\beta})}{\pi(\mathrm{y},X,\boldsymbol{\beta})}=\frac{L(\boldsymbol{\beta},\xi|\mathrm{y},X)\pi(\boldsymbol{\beta}|\xi)\pi(\xi)}{L(\boldsymbol{\beta},\xi|\mathrm{y},X)\pi(\boldsymbol{\beta}|\xi)}\\&=\frac{\prod_{i=1}^{n}\left(1-e^{-(1-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-1}{\xi}}}\right)^{y_i}\left(e^{-(1-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-1}{\xi}}}\right)^{1-y_i}\frac{1}{2}I_{[-1,1)}(\xi)}{\int_{\mathbb{R}^p}\prod_{i=1}^{n}\left(1-e^{-(1-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-1}{\xi}}}\right)^{y_i}\left(e^{-(1-\mathrm{x}_i'\boldsymbol{\beta}\xi)_+^{\frac{-1}{\xi}}}\right)^{1-y_i}\mathrm{d}\xi}.\end{aligned}$$

بنابراین، باتوجه به پیچیده بودن چگالی‌های شرطی کامل نمی‌توان از آن‌ها تولید نمونه کرد و باید از الگوریتم‌های متروپلیس-هستینگز تحت گیبز[1] استفاده نمود. در زیربخش بعدی ساختار کلی این نوع از الگوریتم‌های MCMC را معرفی می‌کنیم.

۱.۳.۳ الگوریتم متروپلیس-هستینگز تحت 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 $i=1,2,\ldots,m$ شبیه‌سازی از

$$f_i(x_i|x_1^{(t+1)},\ldots,x_{i-1}^{(t+1)},x_{i+1}^{(t)},\ldots,x_m^{(t)})$$

در مرحله $t+1$ 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 $f_i(x_i|x_1^{(t+1)},\ldots,x_{i-1}^{(t+1)},x_{i+1}^{(t)},\ldots,x_m^{(t)})$ است، جایگزین می‌شود. این نوع الگوریتم از ترکیب دو الگوریتم گیبز و MH تشکیل می‌شود که به آن الگوریتم متروپلیس-هستینگز تحت گیبز می‌گویند.

در صورتی که شبیه‌سازی بیشتر از یک چگالی شرطی کامل مشکل باشد، می‌توان به همین ترتیب مراحل مختلفی از نمونه‌گیر گیبز را با الگوریتم MH ترکیب کرد (گیلکز و همکاران، ۱۹۹۶).

[1] Metropolis-Hastings Within Gibbs

فصل ۴

ارزیابی عملکرد مدل در مطالعه شبیه‌سازی و تحلیل داده‌های واقعی

در این فصل، ابتدا عملکرد مدل مقادیر فرین تعمیم‌یافته را برای داده‌های دو سطحی نامتعادل، با استفاده از دو مثال شبیه‌سازی بررسی می‌کنیم. سپس به تحلیل یک مجموعه داده واقعی می‌پردازیم. هدف، نشان دادن انعطاف‌پذیری تابع پیوند مقادیر فرین تعمیم‌یافته، در برازش داده‌های تولیدشده از مدل‌های رقیب است. در شبیه‌سازی اول، داده‌ها از یک مدل رگرسیونی Cloglog شبیه‌سازی شده‌اند و در شبیه‌سازی دوم، از یک مدل رگرسیونی پرابیت. همگرایی نمونه‌های تولیدشده از توزیع پسین مدل‌ها با روش‌های تشخیص همگرایی جی‌وک و هیدلبرگ-ولچ ارزیابی شده‌اند. بررسی رفتار آمیختگی زنجیر نیز با رسم نمودارهای اثر[۱] و خودهمبستگی[۲] صورت گرفته‌اند. همچنین، برای مقایسه مدل‌های مطرح‌شده در این پایان‌نامه از معیارهای انتخاب مدل DIC، BIC و درستنمایی حاشیه‌ای استفاده کرده‌ایم.

همگی روش‌های ذکرشده و اجرای الگوریتم‌های نمونه‌گیری MCMC، در محیط نرم‌افزار R (تیم برنامه‌نویسی R، ۲۰۱۶) پیاده‌سازی شده‌اند. برای تولید داده‌ها از مدل‌های مورد نظر از بسته‌های VGAM و evd استفاده کرده‌ایم. همچنین برای ارزیابی همگرایی و رفتار آمیختگی زنجیرها از توابع بسته coda کمک گرفته‌ایم.

[۱] Trace Plots

[۲] Autocorrelation Plots

۱.۴ مثال اول: تولید از مدل Cloglog

در مثال اول سه مدل لجیت، GEV و Cloglog را بر روی مجموعه داده‌هایی که از یک مدل رگرسیون Cloglog شبیه‌سازی شده‌اند، برازش دادیم. به یاد بیاورید که مدل Cloglog حالت خاصی از مدل GEV است که پارامتر شکل واقعی آن برابر $\xi = ۰$ است. برای این مثال چهار متغیر تبیینی را در نظر گرفتیم. متغیر تبیینی اول $x_۱$ را یک بردار nتایی با مقادیر ۱ تعریف کردیم. این متغیر در واقع برای وارد کردن عرض از مبدا در نظر گرفته شد. متغیر تبیینی دوم $x_۲$، که متغیری پیوسته است، را از توزیع نرمال استاندارد تولید کردیم. متغیر تبیینی سوم را یک متغیر اسمی با سه سطح از توزیع چندجمله‌ای با احتمال‌های برابر $\frac{۱}{۳}$ تعریف کردیم. برای وارد کردن مشاهدات این متغیر در مدل رگرسیونی از دو متغیر ظاهری[۳]، یعنی $x_۳$ و $x_۴$، استفاده کردیم. متغیر تبیینی چهارم $x_۵$ را نیز از یک توزیع برنولی با احتمال موفقیت ۰٫۵ در نظر گرفتیم. سه حجم نمونه $n = ۲۰۰, ۱۰۰۰, ۵۰۰۰$ را برای اجرای شبیه‌سازی‌ها در نظر گرفتیم.

پیشگوی خطی در همه مدل‌های رگرسیونی برازش‌شده به داده‌ها به صورت

$$\mathrm{x_i}'\beta = \beta_۱ + \beta_۲ x_{i۲} + \cdots + \beta_۵ x_{i۵}, \quad i = ۱, \ldots, n$$

می‌باشد. ضرایب رگرسیونی واقعی را برابر $\beta = (۰, ۱, ۱, ۰٫۵, -۰٫۵)$ تعریف کردیم. این ضرایب به گونه‌ای تعریف شده‌اند که تعداد پاسخ‌های با مقدار ۱ تقریبا ۷۵٪ کل مشاهدات را شامل شود. دلیل آن هم ایجاد عدم تعادل متغیر پاسخ به سود ۱ ها است که مدل Cloglog می‌تواند در این حالت کارا باشد. هدف از این شبیه‌سازی مقایسه نتایج مدل GEV با نتایج مدل Cloglog است.

برای تکمیل مدل بیزی در شبیه‌سازی‌ها، باید توزیع‌های پیشین را نیز تعریف کنیم. در تمام مدل‌های برازش‌شده در این مثال، از توزیع‌های پیشین نرمال $N(۰, \sigma^۲_{\beta_j})$ برای ضرایب رگرسیونی β_j، $j = ۱, \ldots, p$، استفاده کردیم که در همه آن‌ها $\sigma_{\beta_j} = ۱۰$ انتخاب شد. برای پارامتر شکل ξ نیز توزیع پیشین $N(۰, \sigma^۲_\xi)$ را با $\sigma_\xi = ۱۰$ در نظر گرفتیم. برای برازش مدل‌ها، الگوریتم‌های MCMC را با حجم نمونه مونت کارلویی ۵۰۰۰۰ تایی اجرا کردیم. از این تعداد، ۲۰۰۰۰ نمونه اول زنجیر را به‌عنوان دوره داغیدن سوزاندیم (حذف کردیم) و از مابقی نمونه‌ها، مولفه‌های زنجیر در گام‌های ۱۰ را انتخاب کردیم. توزیع‌های پیشنهادی برای همه پارامترها $N(\theta^{(t-۱)}, ۱۰^۲)$ در نظر گرفته شدند که در آن‌ها $\theta = (\beta, \xi)$ است. بنابراین، همه استنباط‌ها بر اساس یک نمونه نهایی ۳۰۰۰ تایی استخراج شده‌اند.

در دو جدول ۱.۴ و ۲.۴، مقادیر آماره‌های خلاصه‌شده‌ای از توزیع پسین گزارش شده‌اند. این مقادیر خلاصه‌شده شامل برآوردهای بیزی پارامترها (میانگین توزیع پسین)، خطای استاندارد برآوردها، SE، و فاصله‌های اعتبار ۹۵٪ پارامترهای مدل، به ازای حجم‌های نمونه متفاوت، برای سه مدل لجیت، GEV و Cloglog هستند. نتایج نشان می‌دهند که برآوردهای پارامترهای مدل GEV حتی با حجم نمونه کوچک، بسیار نزدیک به برآوردهای متناظر در مدل واقعی Cloglog و نزدیک به مقادیر واقعی هستند. به عبارت دیگر می‌توان گفت هر دو مدل GEV و Cloglog تقریباً برازش مشابهی بر داده‌ها

[۳] Dummy Variable

دارند. همچنین، طول فواصل اعتبار پارامترها برای هر دو مدل تقریباً برابر هستند.

مقدار برآوردشده پارامتر شکل در مدل GEV نیز وجود چولگی مثبت در پاسخ‌ها را تایید می‌کند. از طرف دیگر، مقدار برآوردشده، برای هر سه حجم نمونه، به مقدار واقعی $\xi = ۰$ نزدیک هستند و فاصله‌های اعتبار ۹۵٪ متناظر نیز صفر را شامل می‌شوند. این به معنی سازگاری داده‌های شبیه‌سازی‌شده با مدل GEV برازش‌شده است.

جدول ۱.۴: برآوردهای بیزی پارامترها و خطای استاندارد آن‌ها در مثال شبیه‌سازی اول

n	پارامترها	Cloglog		GEV		Logit	
		میانگین	SE	میانگین	SE	میانگین	SE
۲۰۰	$\beta_۱$	−۰/۱۵۳	۰/۲۴۸	−۰/۱۵۵	۰/۲۶۱	۰/۴۵۵	۰/۳۹۶
	$\beta_۲$	۱/۰۱۰	۰/۱۴۸	۰/۹۸۳	۰/۱۶۲	۱/۵۷۰	۰/۲۳۵
	$\beta_۳$	۰/۶۸۱	۰/۲۷۶	۰/۶۴۰	۰/۲۷۶	۱/۲۲۸	۰/۴۵۴
	$\beta_۴$	۰/۶۴۷	۰/۲۸۴	۰/۶۳۸	۰/۲۸۴	۱/۰۸۹	۰/۴۵۴
	$\beta_۵$	−۰/۴۰۹	۰/۲۲۵	−۰/۳۸۷	۰/۲۲۱	−۰/۶۵۰	۰/۳۸۵
	ξ			۰/۰۱۹	۰/۲۹۷		
۱۰۰۰	$\beta_۱$	۰/۰۵۱	۰/۰۹۶	۰/۰۳۱	۰/۱۰۸	۰/۸۱۹	۰/۱۶۱
	$\beta_۲$	۱/۰۰۴	۰/۰۷۰	۰/۹۹۳	۰/۰۷۲	۱/۶۱۲	۰/۱۱۷
	$\beta_۳$	۰/۸۸۶	۰/۱۲۰	۰/۸۸۱	۰/۱۲۳	۱/۴۹۷	۰/۲۰۷
	$\beta_۴$	۰/۳۷۷	۰/۱۲۰	۰/۳۷۹	۰/۱۲۰	۰/۶۰۴	۰/۱۹۷
	$\beta_۵$	−۰/۵۷۰	۰/۰۹۸	−۰/۵۵۷	۰/۱۰۰	−۰/۹۷۱	۰/۱۷۰
	ξ			۰/۰۴۵	۰/۱۴۲		
۵۰۰۰	$\beta_۱$	۰/۰۱۵	۰/۰۴۳	۰/۰۷۱	۰/۰۴۸	۰/۷۶۶	۰/۰۷۴
	$\beta_۲$	۰/۹۵۵	۰/۰۲۹	۰/۹۵۲	۰/۰۳۰	۱/۵۵۷	۰/۰۵۰
	$\beta_۳$	۱/۰۴۴	۰/۰۵۷	۱/۰۳۸	۰/۰۵۶	۱/۷۱۳	۰/۰۹۷
	$\beta_۴$	۰/۴۸۱	۰/۰۵۴	۰/۴۷۹	۰/۰۵۳	۰/۷۴۹	۰/۰۸۸
	$\beta_۵$	−۰/۵۰۹	۰/۰۴۳	−۰/۵۰۸	۰/۰۴۴	−۰/۸۴۵	۰/۰۷۷
	ξ			۰/۰۳۶	۰/۰۶۶		

همان‌گونه که در نتایج جدول‌ها مشاهده می‌شود، مدل (متقارن) لجیت اصلا به خوبی دو مدل چوله پارامترها را برآورد نکرده است. دقت برآوردها نیز (بر اساس خطای استاندارد و فاصله‌های اعتبار بیزی) در این مدل با دو مدل GEV و Cloglog قابل رقابت نیست.

برای مقایسه مدل‌ها از معیارهای انتخاب مدل استفاده می‌کنیم. این معیارها با حجم‌های نمونه متفاوت در جدول ۳.۴ گزارش شده‌اند. همان‌طور که ملاحظه می‌شود، مقادیر به‌دست آمده از معیارهای

جدول ۲.۴: فاصله‌های اعتبار ۹۵٪ برای پارامترهای مدل‌های برازش‌شده در مثال شبیه‌سازی اول

n	پارامترها	Cloglog		GEV		Logit	
		٪۲٫۵	٪۷٫۵	٪۲٫۵	٪۷٫۵	٪۲٫۵	٪۷٫۵
۲۰۰	β_1	−۰٫۷۰۷	۰٫۲۹۵	−۰٫۷۱۰	۰٫۲۹۹	−۰٫۲۵۶	۱٫۳۰۳
	β_2	۰٫۷۲۶	۱٫۳۰۳	۰٫۶۵۳	۱٫۲۸۸	۱٫۰۹۶	۲٫۰۲۶
	β_3	۰٫۱۳۲	۱٫۲۱۲	۰٫۰۸۷	۱٫۱۶۵	۰٫۳۶۶	۲٫۱۲۷
	β_4	۰٫۰۸۹	۱٫۲۰۸	۰٫۰۶۹	۱٫۱۷۴	۰٫۱۶۸	۰٫۹۴۷
	β_5	−۰٫۸۵۲	۰٫۰۳۵	−۰٫۸۳۵	۰٫۰۲۸	−۱٫۳۹۴	۰٫۰۹۶
	ξ			−۰٫۵۲۳	۰٫۶۱۹		
۱۰۰۰	β_1	−۰٫۱۲۴	۰٫۲۴۸	−۰٫۱۸۹	۰٫۲۲۶	۰٫۵۰۰	۱٫۱۲۸
	β_2	۰٫۸۷۱	۱٫۱۴۸	۰٫۸۵۱	۱٫۱۲۳	۱٫۳۸۷	۱٫۸۵۰
	β_3	۰٫۶۴۷	۱٫۱۲۲	۰٫۶۴۹	۱٫۱۲۹	۱٫۰۸۴	۱٫۹۰۲
	β_4	۰٫۱۴۳	۰٫۶۱۵	۰٫۱۳۶	۰٫۶۰۷	۰٫۲۴۱	۱٫۰۰۵
	β_5	−۰٫۷۵۹	−۰٫۳۷۸	−۰٫۷۶۲	−۰٫۳۶۸	−۱٫۲۷۹	۰٫۶۱۳
	ξ			−۰٫۲۳۰	۰٫۳۲۱		
۵۰۰۰	β_1	−۰٫۰۶۸	−۰٫۰۱۴	−۰٫۰۷۹	۰٫۱۱۳	۰٫۶۲۰	۰٫۹۰۷
	β_2	۰٫۸۹۷	۰٫۹۳۴	۰٫۸۹۶	۱٫۰۱۶	۱٫۴۵۸	۱٫۶۵۱
	β_3	۰٫۹۳۷	۱٫۱۶۰	۰٫۹۲۸	۱٫۱۵۳	۱٫۵۰۹	۱٫۸۹۵
	β_4	۰٫۳۷۸	۰٫۵۹۳	۰٫۳۶۶	۰٫۵۷۸	۰٫۵۷۸	۰٫۹۲۹
	β_5	−۰٫۵۹۰	−۰٫۴۱۷	−۰٫۵۹۴	−۰٫۴۲۳	−۰٫۹۹۰	−۰٫۶۸۷
	ξ			−۰٫۰۹۶	۰٫۱۶۳		

درستنمایی حاشیه‌ای، DIC و BIC برای مدل لجیت در نمونه ۲۰۰ تایی کمتر از مدل GEV است. این نشان می‌دهد که تعداد نمونه‌ها در مدل چوله GEV می‌تواند در کارایی برازش تاثیرگذار باشد و برای آشکار کردن چولگی موجود در داده‌ها، این مدل نیاز به حجم نمونه بزرگ دارد. اما مقادیر به‌دست آمده از این معیارها، در حجم‌های نمونه ۱۰۰۰ و ۵۰۰۰ در مدل GEV کمتر از مدل لجیت هستند. با این حال، مدل Cloglog در تمام شرایط عملکرد بهتری نسبت به سایر مدل‌ها داشته است. این بهتر بودن هم طبیعی است، زیرا شبیه‌سازی از این مدل انجام شده و مدل Cloglog مدل واقعی است.

نتایج استنباطی گزارش‌شده زمانی معتبر هستند که نمونه‌های حاصل از زنجیرهای MCMC از یک زنجیر همگرا استخراج شده باشند. بنابراین سنجش همگرایی زنجیرهای اجراشده مهم است. برای بررسی همگرایی زنجیرها از دو آزمون رسمی بررسی همگرایی زنجیرهای MCMC هیدلبرگ-ولچ و جی‌وک استفاده کردیم. آزمون هیدلبرگ-ولچ برای مدل‌های مورد نظر در این مثال، به ازای حجم‌های نمونه متفاوت، بررسی و نتایج در جدول ۴.۴ گزارش شده‌اند. با توجه به نتایج جدول، زنجیر تولیدشده برای بعضی از پارامترها در مرحله اول آزمون همگرا شده ولی در مرحله دوم همگرا نشده است. برای

جدول ۳.۴: معیارهای مقایسه مدل‌های Logit، GEV و Cloglog در مطالعه شبیه‌سازی اول

n	پارامترها	Cloglog	GEV	Logit
۲۰۰	$D(\tilde{\theta})$	۱۸۶٫۶۰۰	۱۸۶٫۶۳۰	۱۸۷٫۸۵۵
	DIC	۱۹۶٫۵۱۶	۱۹۸٫۲۱۷	۱۹۷٫۹۳۸
	P_D	۴٫۹۵۸	۵٫۷۹۳	۵٫۰۴۱
	Marginal Likelihood	−۹۳٫۳۰۰	−۹۳٫۳۱۵	−۹۳٫۹۲۷
	BIC	۲۱۳٫۰۹۱	۲۱۸٫۴۲۰	۲۱۴٫۳۴۷
۱۰۰۰	$D(\tilde{\theta})$	۸۸۹٫۹۲۵	۸۸۹٫۸۳۲	۹۰۱٫۰۶۸
	DIC	۸۹۹٫۸۵۲	۹۰۱٫۸۹۵	۹۱۱٫۱۲۷
	P_D	۴٫۹۶۳	۶٫۰۳۱	۵٫۰۲۹
	Marginal Likelihood	−۴۴۴٫۹۶۲	−۴۴۴٫۹۱۶	−۴۵۰٫۵۳۴
	BIC	۹۲۴٫۴۶۴	۹۳۱٫۲۷۹	۹۳۵٫۶۰۷
۵۰۰۰	$D(\tilde{\theta})$	۴۳۷۴٫۸۳۸	۴۳۷۴٫۵	۴۴۳۰٫۳۷۴
	DIC	۴۳۸۴٫۸۱۷	۴۳۸۶٫۶۴۱	۴۴۴۰٫۳۸۵
	P_D	۴٫۹۸۹	۶٫۰۷۰	۵٫۰۰۵
	Marginal Likelihood	−۲۱۸۷٫۴۱۹	−۲۱۸۷٫۲۵	−۲۲۱۵٫۱۸۷
	BIC	۴۴۱۷٫۴۲۴	۴۴۲۵٫۶۰۳	۴۴۷۲٫۹۶

بعضی از پارامترها نیز همگرایی در همان مرحله اول رد شده است. در مجموع می‌توان نتیجه گرفت که برای بسیاری از پارامترها، نمونه‌های تولیدشده تقریباً از توزیع پسین مانای متناظرشان آمده‌اند.

آزمون دوم ارزیابی همگرایی زنجیرها برای هر پارامتر، آزمون جی‌وک است. با رسم نمودارهای آماره آزمون جی‌وک، می‌توان همگرایی را برای هر یک از پارامترها بررسی کرد. در این نمودارها آماره آزمون جی‌وک، Z_n، بارها و بارها محاسبه می‌شود. اولین Z_n با تمام تکرارهای زنجیر، دومین Z_n بعد از گذاشتن بخش اول و سومین بعد از کنار گذاشتن دو بخش اول و به همین ترتیب محاسبه می‌شوند. در نهایت، آخرین Z_n تنها با استفاده از نیمه دوم زنجیر محاسبه خواهد شد. نمودارهای جی‌وک برای هر یک از پارامترها در هر سه مدل، به ازای حجم‌های نمونه متفاوت، در شکل‌های ۱.۴، ۲.۴ و ۳.۴ گزارش شده‌اند. با توجه به این نمودارها مشاهده می‌کنید که تعداد زیادی از Z_nهای محاسبه‌شده در فاصله مشخص‌شده قرار دارند. بنابراین می‌توان گفت نمونه‌های تولیدشده از قسمت مانای زنجیر، گرفته شده‌اند.

کیفیت نمونه‌های تولیدشده را نیز می‌توان با محاسبه معیار ESS سنجید. جدول ۵.۴، مقادیر ESS را به ازای حجم‌های نمونه مختلف، برای هر یک از پارامترها نشان می‌دهد. همان‌طور که مشاهده می‌شود، نمونه‌های تولیدشده برای دو مدل لجیت و Cloglog از کیفیت خوبی برخوردارند ولی برای مدل

جدول ۴.۴: نتایج آزمون هیدلبرگ-ولچ در مطالعه شبیه‌سازی اول

n	پارامترها	برآورد آماره			آزمون نیم‌فاصله		
		Cloglog	GEV	Logit	Cloglog	GEV	Logit
۲۰۰	β_1	۵۰۱	-	۱٫۰۰	رد	-	پذیرش
	β_2	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	β_3	۵۰۱	-	۱٫۰۰	پذیرش	-	پذیرش
	β_4	۵۰۱	-	۱٫۰۰	پذیرش	-	پذیرش
	β_5	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	ξ		۱٫۰۰			رد	
۱۰۰۰	β_1	۱٫۰۰	۱٫۰۰	۱٫۰۰	رد	رد	پذیرش
	β_2	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	β_3	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	β_4	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	β_5	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	ξ		۱٫۰۰			رد	
۵۰۰۰	β_1	۱٫۰۰	۱٫۰۰	۱٫۰۰	رد	رد	پذیرش
	β_2	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	β_3	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	β_4	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	β_5	۱٫۰۰	۱٫۰۰	۱٫۰۰	پذیرش	پذیرش	پذیرش
	ξ		۱٫۰۰			رد	

GEV بعضی از پارامترها معیار ESS کوچکی دارند. این نشان‌دهنده آن است که نمونه‌های تولیدشده در مدل GEV، برای برخی از پارامترها، وابستگی زیادی با یکدیگر دارند. برای داشتن نمونه‌های با کیفیت نزدیک به مشاهدات مستقل، باید مقادیر به‌دست آمده برای هر یک از پارامترها، نزدیک به تعداد نمونه‌های تولیدشده از توزیع پسین، یعنی ۳۰۰۰، باشند. البته قضیه ارگودیک، در صورتی که زنجیر همگرا شده باشد، مجوز استفاده از نمونه‌های (وابسته) حاصل از آن را صادر می‌کند.

نمودارهای موجود در شکل ۴.۴، نمودارهای اثر هر یک از پارامترهای مدل GEV را نشان می‌دهند. با رسم این نمودارها می‌توان رفتار آمیختگی زنجیرها را در فضای پارامتر مشاهده کرد. با توجه به این نمودارها و نوسانات منظم زنجیر در هر یک از آن‌ها می‌توان نتیجه گرفت که زنجیر تولیدشده برای پارامترها از آمیختگی خوبی برخوردار است. این آمیختگی به این معنی است که نواحی چگالی پسین به خوبی مورد کاوش قرار گرفته‌اند. برای سایر مدل‌ها نتیجه نمودارهای اثر مشابه مدل GEV است که از گزارش آن‌ها صرف‌نظر کرده‌ایم.

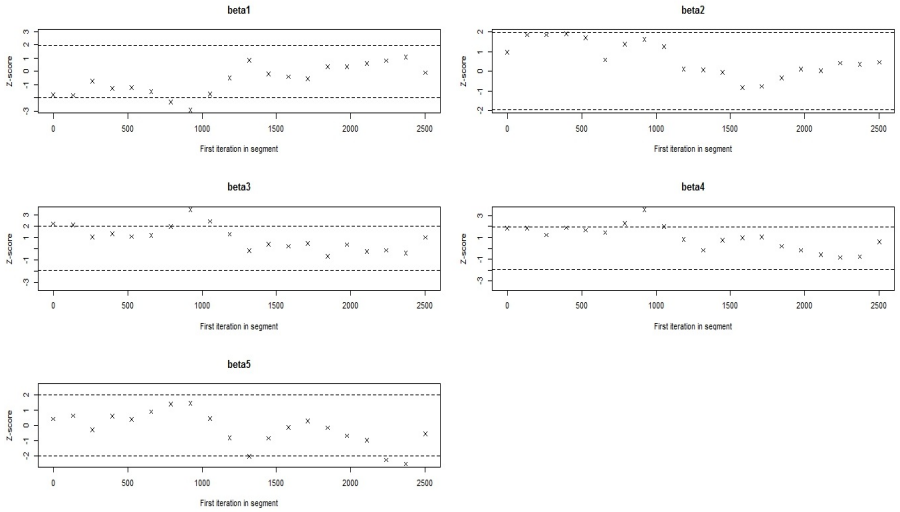
راه جانشین دیگر برای ارزیابی کیفیت زنجیر و همگرایی آن، ارزیابی مقادیر خودهمبستگی بین

جدول ۵.۴: مقادیر ESS برای پارامترهای مدل‌های مورد نظر در مطالعه شبیه‌سازی اول

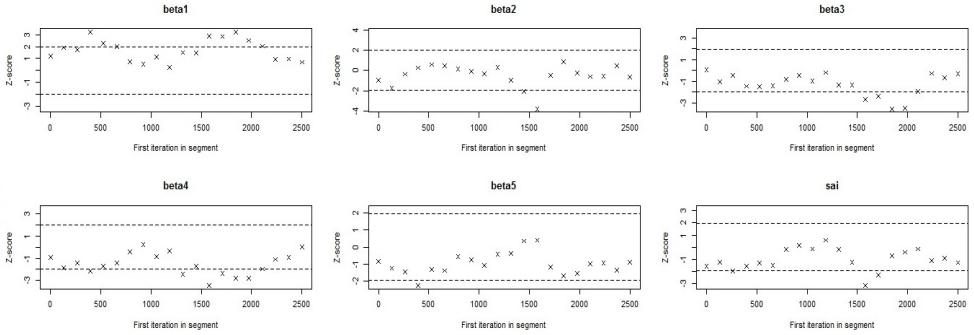
n	پارامترها	ESS		
		Cloglog	GEV	Logit
۲۰۰	β_1	۶۶۷٫۹۱۰	۷۰۹٫۹۰۳	۱۴۵۵٫۱۷۱
	β_2	۴۷۱۸٫۵۲۷	۳۵۷۰٫۲۴۹	۴۲۵۶٫۳۰۱
	β_3	۱۱۰۴٫۴۳۵	۱۲۷۰٫۴۵۱	۲۶۸۱٫۸۳۸
	β_4	۱۱۷۶٫۳۱۷	۱۱۹۳٫۵۷۷	۲۳۳۲٫۹۸۸
	β_5	۱۶۷۴٫۲۵۴	۱۷۵۰٫۸۵۹	۲۰۱۰٫۰۱۳
	ξ		۲۷۱۲٫۸۵۴	
۱۰۰۰	β_1	۵۲۴٫۵۸۷	۳۹۴٫۹۰۸	۱۰۴۱٫۶۰۹
	β_2	۳۷۲۱٫۸۰۵	۴۲۴۵٫۸۳۹	۳۹۳۰٫۹۶۷
	β_3	۱۳۴۸٫۲۷۴	۱۱۴۵٫۱۳۴	۲۵۴۴٫۱۸۳
	β_4	۱۰۰۵٫۵۴۴	۸۶۴٫۰۵۴	۲۳۳۵٫۹۸۸
	β_5	۱۳۴۲٫۶۶۱	۱۱۵۱٫۳۴۴	۲۱۹۷٫۷۹۲
	ξ		۱۳۸۳٫۸۶۳	
۵۰۰۰	β_1	۱۴۱۹٫۵۲۶	۱۷۰٫۴۹۲	۱۴۵۲٫۱۵۶
	β_2	۴۲۸۳٫۸۳۳	۴۲۲۶٫۸۰۷	۳۹۷۷٫۲۹۵
	β_3	۲۰۰۷٫۳۷۲	۶۶۰٫۳۹۰	۲۵۹۰٫۶۰۳
	β_4	۲۱۷۳٫۰۴۰	۵۳۲٫۵۱۴	۲۲۳۰٫۹۳۲
	β_5	۳۰۹۲٫۰۲۰	۶۷۸٫۰۷۱	۲۲۵۵٫۴۰۰
	ξ		۱۱۶۹٫۹۴۷	

زنجیرهای مارکوف است. از آن‌جا که یک مقدار نمونه در زمان t به مقدار قبلی آن در زمان $t-1$ وابسته است، نمونه‌ها همواره همبسته خواهند بود. ارزیابی همبستگی در نمونه‌های موجود در زنجیرها با نمودار خودهمبستگی در مقابل تأخیر صورت می‌پذیرد. اگر زنجیر همگرا شود و نمونه‌ها همبستگی ناچیزی داشته باشند، انتظار داریم با افزایش تأخیر، همبستگی به سرعت کاهش یابد. هدف از رسم این نمودار، مشاهده کاهش سریع همبستگی یک زنجیر به سمت صفر است که این پدیده تاییدکننده این مطلب است که نمونه‌های تولیدشده به توزیع مانای خود همگرا می‌شوند.

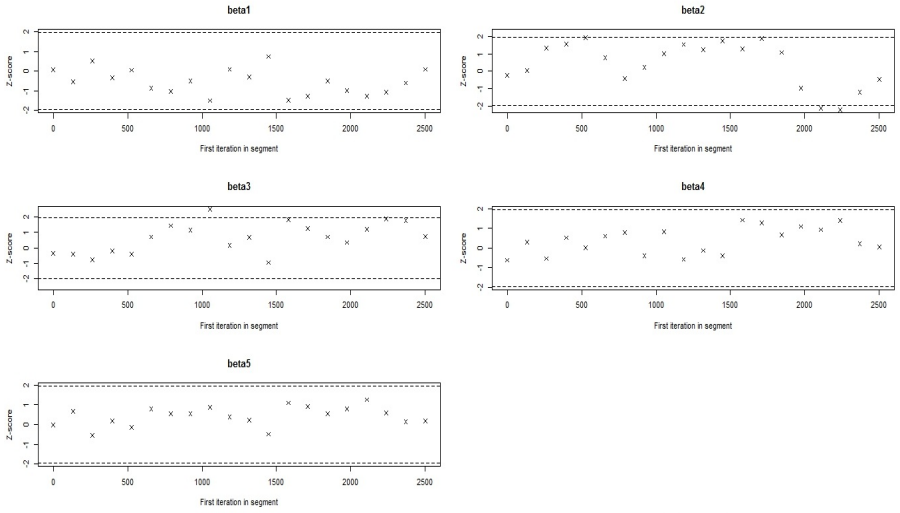
نمودارهای موجود در شکل‌های ۵.۴، ۶.۴ و ۷.۴ نمودارهای خودهمبستگی را به ازای هر یک از پارامترها در سه مدل برازش‌شده نشان می‌دهند. در این نمودارها مشاهده می‌شود، از وابستگی نمونه‌ها کاسته شده و به سرعت به سمت صفر میل می‌کنند. به عبارت دیگر، زنجیر برای هر پارامتر همگرا است و کیفیت نمونه‌ها برای استخراج استنباط‌های معتبر، قابل پذیرش است.



(آ)

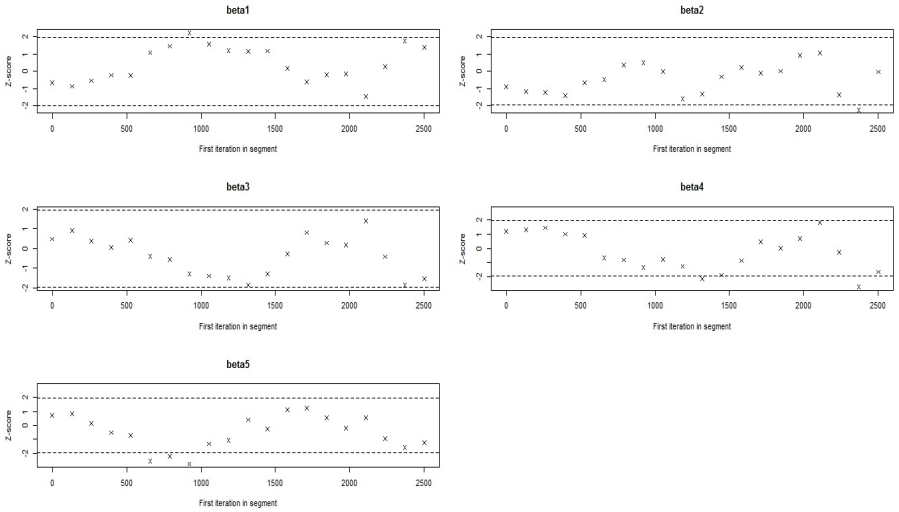


(ب)

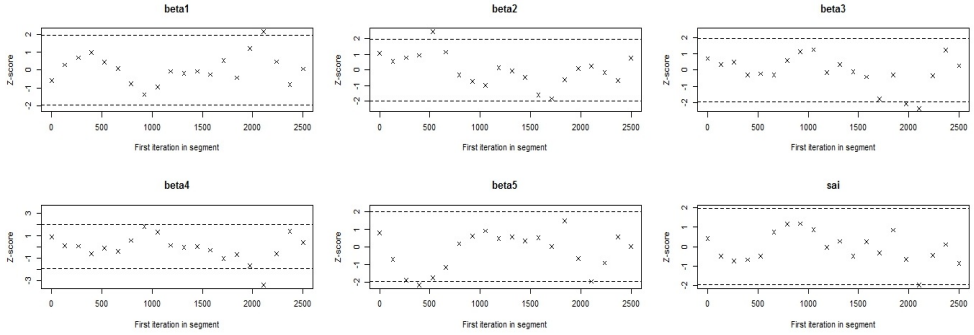


(ج)

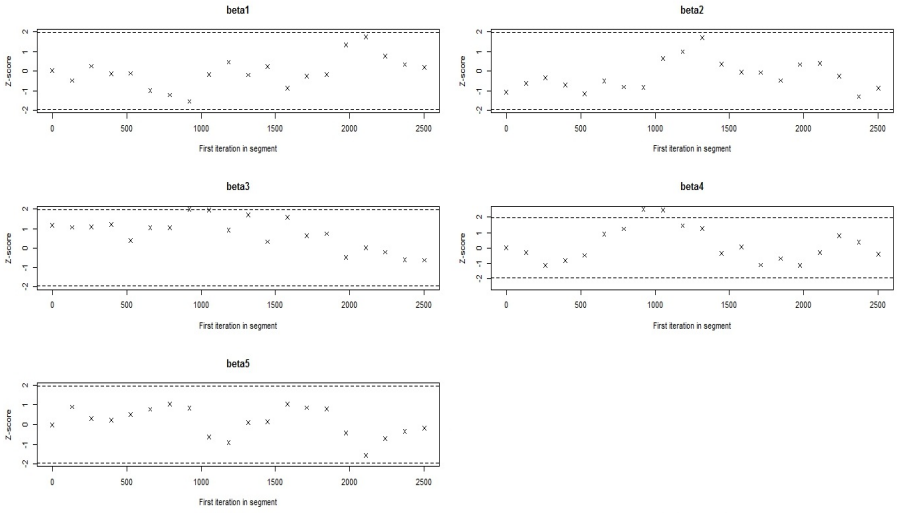
شکل ۱.۴: نمودارهای جی‌وک پارامترها برای (آ) مدل Cloglog، (ب) مدل GEV، (ج) مدل لجیت در مطالعه شبیه‌سازی اول با حجم نمونه ۲۰۰



(آ)

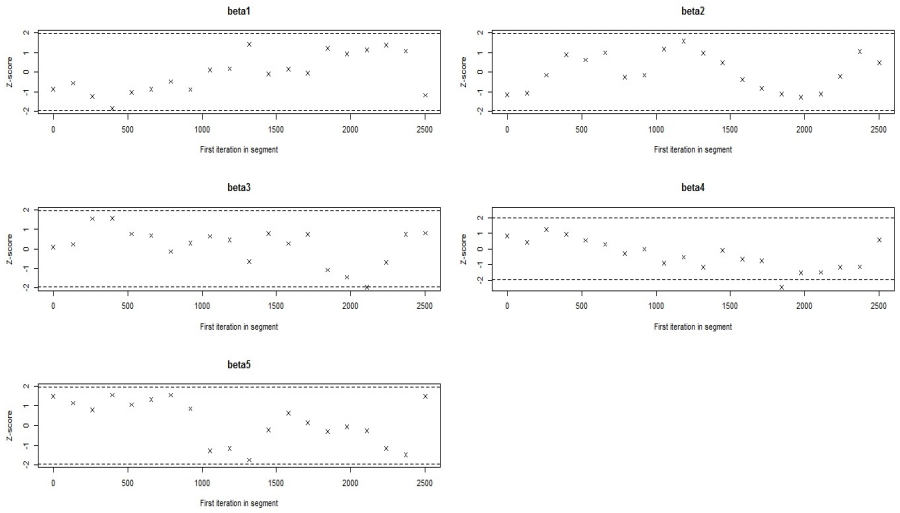


(ب)

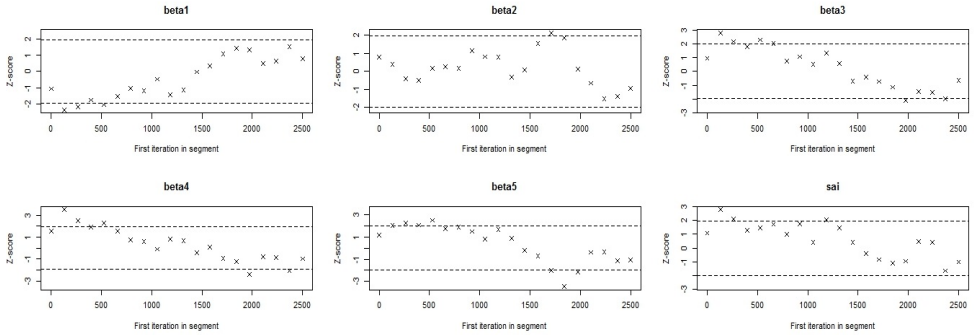


(ج)

شکل ۲.۴: نمودارهای جی‌وک پارامترها برای (آ) مدل Cloglog، (ب) مدل GEV، (ج) مدل لجیت در مطالعه شبیه‌سازی اول با حجم نمونه ۱۰۰۰



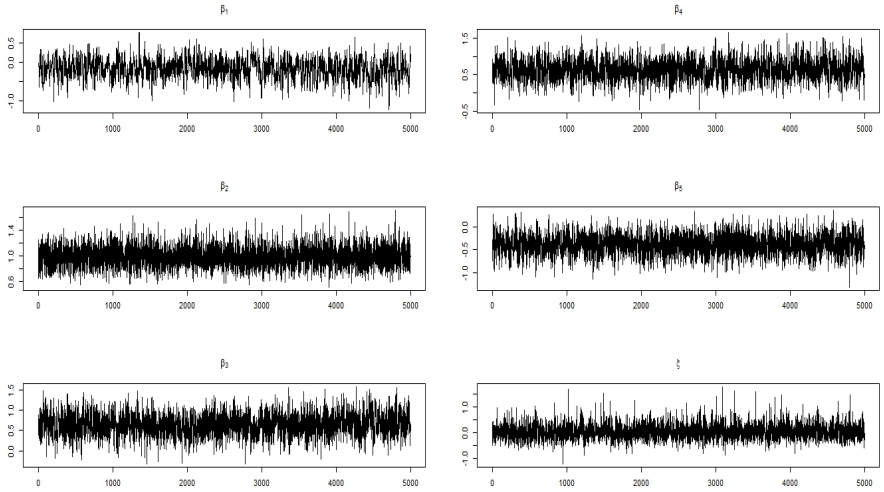
(آ)



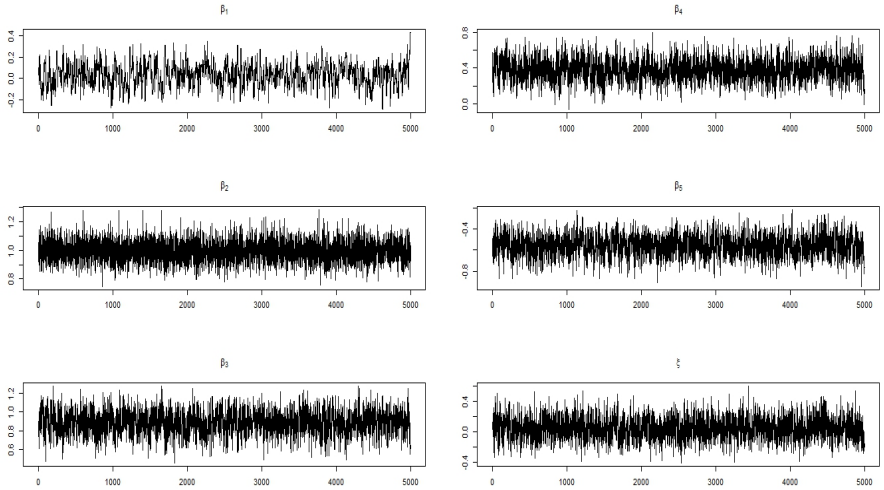
(ب)

(ج)

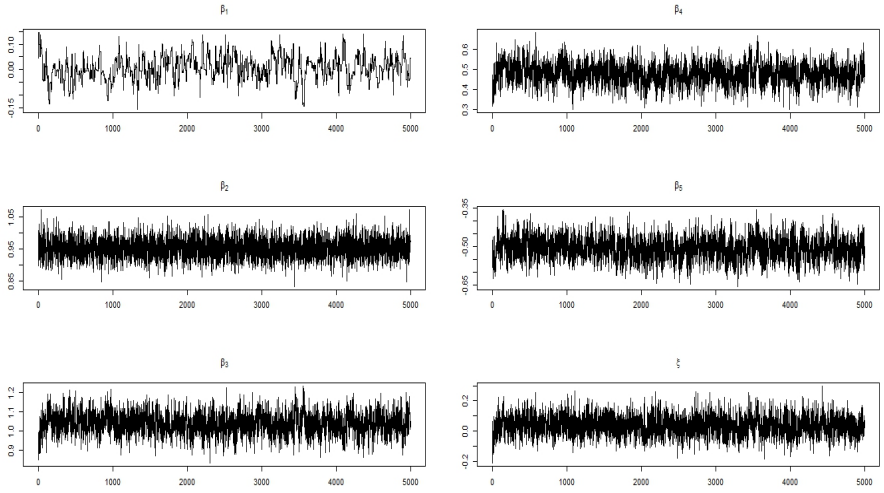
شکل ۳.۴: نمودارهای جی‌وک پارامترها برای (آ) مدل Cloglog، (ب) مدل GEV، (ج) مدل لجیت در مطالعه شبیه‌سازی اول با حجم نمونه ۵۰۰۰



(آ)

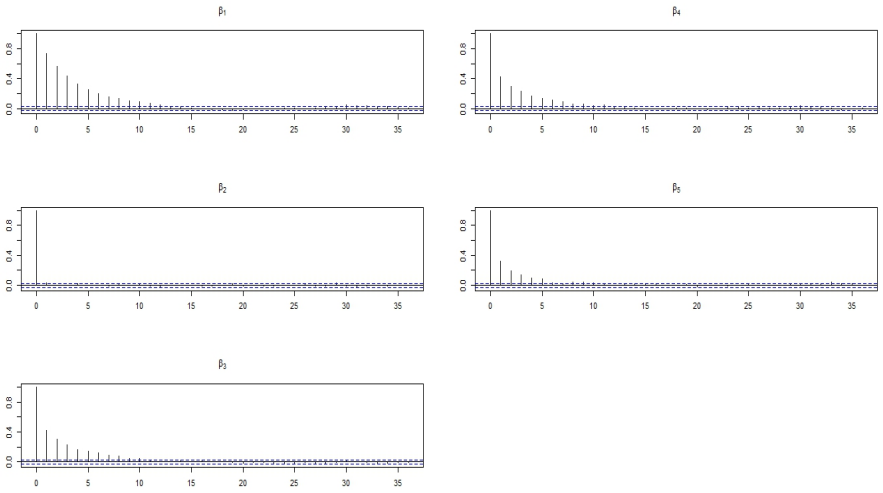


(ب)

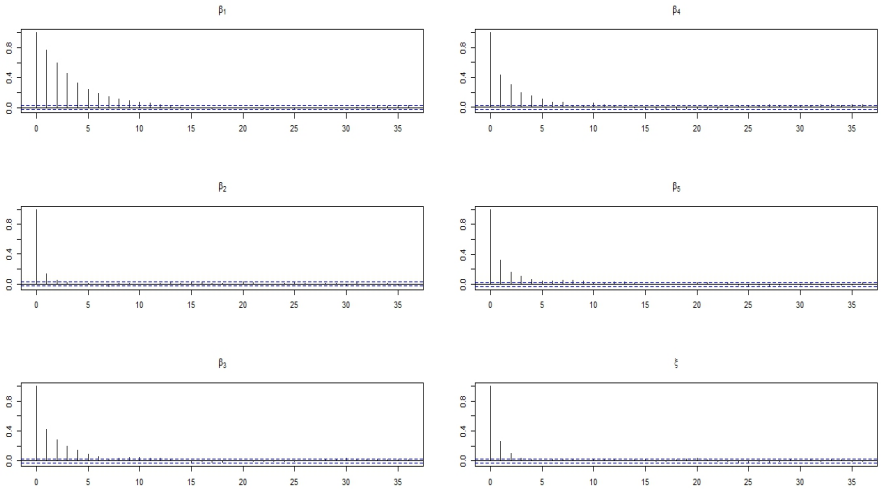


(ج)

شکل ۴.۴: نمودارهای اثر پارامترها برای مدل GEV با (آ) حجم نمونه ۲۰۰، (ب) حجم نمونه ۱۰۰۰ و (ج) حجم نمونه ۵۰۰۰ در مطالعه شبیه‌سازی اول



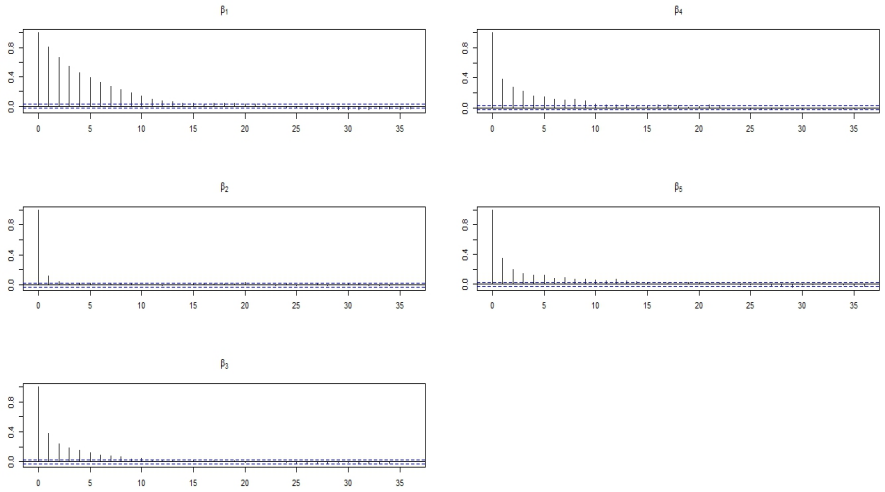
(آ)



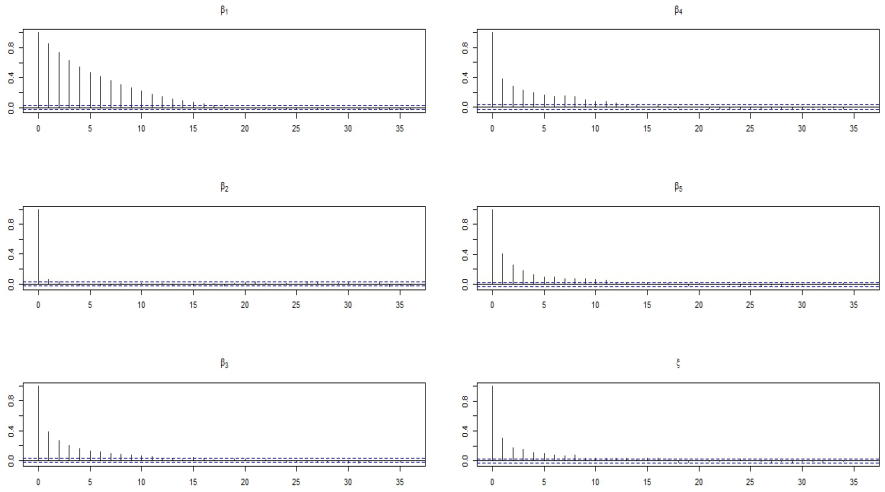
(ب)

(ج)

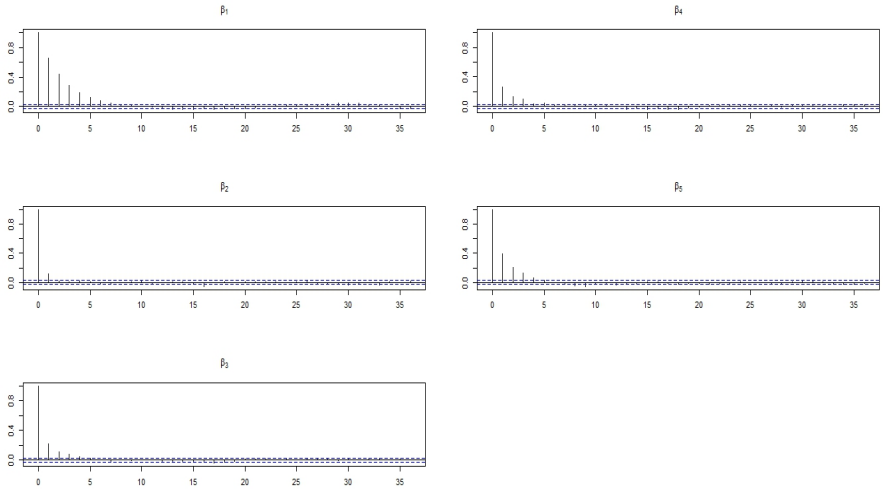
شکل ۵.۴: نمودارهای خودهمبستگی پارامترها برای (آ) مدل Cloglog، (ب) مدل GEV، و (ج) مدل لجیت در مطالعه شبیه‌سازی اول با حجم نمونه ۲۰۰



(آ)

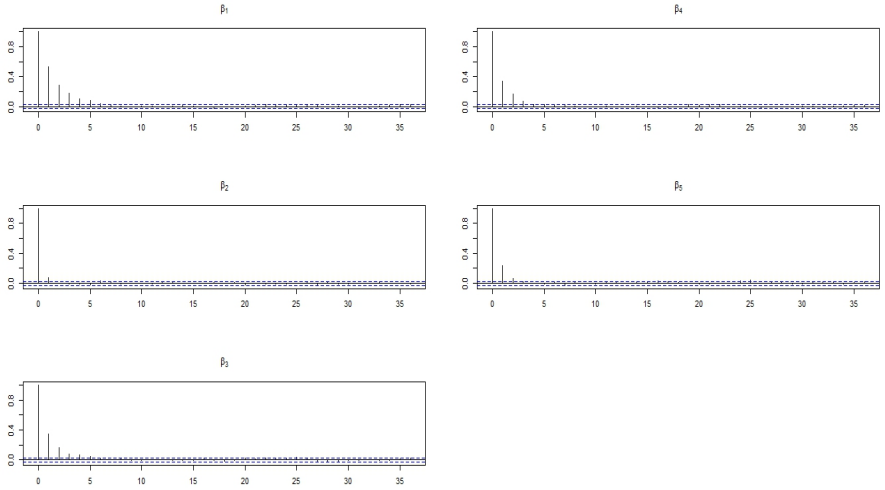


(ب)

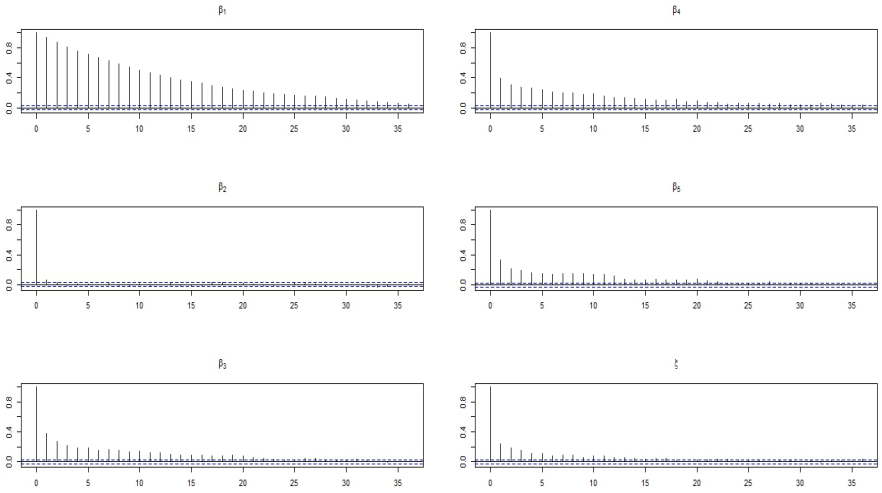


(ج)

شکل ۶.۴: نمودارهای خودهمبستگی پارامترها برای (آ) مدل Cloglog، (ب) مدل GEV، و (ج) مدل لجیت در مطالعه شبیه‌سازی اول با حجم نمونه ۱۰۰۰



(آ)



(ب) (ج)

شکل ۷.۴: نمودارهای خودهمبستگی پارامترها برای (آ) مدل Cloglog، (ب) مدل GEV، و (ج) مدل لجیت در مطالعه شبیه‌سازی اول با حجم نمونه ۵۰۰۰

۲.۴ مثال دوم: تولید از مدل پرابیت

در مثال دوم، مدل‌های پرابیت، لجیت، GEV و Cloglog را برای مجموعه داده‌هایی که از مدل رگرسیونی پرابیت شبیه‌سازی شده‌اند، برازش دادیم. در این مثال، متغیرهای تبیینی را مشابه مثال اول تولید کردیم. بنابراین، پیشگوی خطی برای مدل‌های این مثال، مشابه مثال اول، به صورت

$$\mathrm{x_i}'\beta = \beta_۱ + \beta_۲ x_{i۲} + \cdots + \beta_۵ x_{i۵}, \quad i = ۱, \ldots, n$$

است. در این مثال نیز سه حجم نمونه $n = ۲۰۰, ۱۰۰۰, ۵۰۰۰$ را در نظر گرفتیم. ضرایب رگرسیونی واقعی را برابر $\beta = (۰, ۱, ۱, ۱/۲۵, -۰/۲۵)$ تعریف کردیم. این ضرایب نیز، در این مثال، به شکلی تعیین شده‌اند که تعداد پاسخ‌های با مقدار ۱ تقریبا ۶۵٪ کل مشاهدات را شامل شود. دلیل آن هم ایجاد تعادل نسبی بین مقادیر ۰ و ۱ متغیر پاسخ است که مدل پرابیت می‌تواند کارا باشد. هدف از این شبیه‌سازی مقایسه نتایج مدل GEV با نتایج مدل پرابیت است.

توزیع‌های پیشین پارامترهای چهار مدل، مشابه مثال قبل تعیین شدند. حجم نمونه مونت کارلویی، دوره داغیدن و طول گام انتخاب نیز به‌طور مشابه تعیین شدند. مشابه مثال اول، توزیع‌های پیشنهادی برای همه پارامترها $N(\theta^{(t-۱)}, ۱۰^۲)$ در نظر گرفته شدند که در آن‌ها $\theta = (\beta, \xi)$ است. بنابراین، همه استنباط‌ها بر اساس یک نمونه نهایی ۳۰۰۰ تایی به‌دست آمده‌اند. نتایج برآورد پارامترها و فواصل اعتبار آن‌ها در دو جدول ۶.۴ و ۷.۴ گزارش شده‌اند. با توجه به این نتایج، می‌توان گفت

- برآوردهای پارامترهای مدل GEV حتی با حجم نمونه کوچک، بسیار نزدیک به برآوردهای متناظر در مدل واقعی پرابیت و نزدیک به مقادیر واقعی هستند.
- مدل GEV، در حجم نمونه بالا، نسبت به مدل لجیت برازش بهتری دارد.
- طول فواصل اعتبار پارامترهای سه مدل پرابیت، GEV و Cloglog تقریباً یکسان هستند.
- مدل پرابیت حالت خاصی از مدل GEV، با پارامتر شکل $\xi = -۰/۲۷۸$ ، محسوب می‌شود. برآورد بیزی این پارامتر در مدل GEV، برای هر سه حجم نمونه، به این مقدار واقعی نزدیک است و فاصله اعتبار ۹۵٪ متناظر نیز این مقدار را شامل می‌شود. این به معنی سازگاری داده‌های شبیه‌سازی‌شده با مدل GEV برازش‌شده است.
- مدل لجیت به خوبی سه مدل دیگر پارامترها را برآورد نکرده است. دقت برآوردها نیز با مقادیر متناظر در سایر مدل‌ها اختلاف دارد.

معیارهای انتخاب مدل، برای حجم‌های نمونه متفاوت، در جدول ۸.۴ گزارش شده‌اند. مقادیر به‌دست آمده از معیارهای درستنمایی حاشیه‌ای، DIC و BIC برای مدل لجیت در نمونه ۲۰۰ تایی کمتر از مدل GEV است. این نتیجه، مشابه مثال شبیه‌سازی اول، نشان می‌دهد که حجم نمونه کوچک برازش خوبی برای مدل GEV نتیجه نمی‌دهد. اما مقادیر به‌دست آمده از این معیارها، در حجم‌های نمونه ۱۰۰۰ و ۵۰۰۰ در مدل GEV کمتر از مدل لجیت هستند. مدل پرابیت نیز به‌عنوان مدل واقعی پشتیبان داده‌ها، به‌طور یکنواخت نسبت به سایر مدل‌ها برتر است.

جدول ۶.۴: برآوردهای بیزی پارامترها و خطای استاندارد آن‌ها در مثال شبیه‌سازی دوم

n	پارامترها	Probit		Logit		Cloglog		GEV	
		میانگین	SE	میانگین	SE	میانگین	SE	میانگین	SE
۲۰۰	β_1	۰/۲۶۵	۰/۲۴۴	۰/۴۳۶	۰/۴۲۲	−۰/۱۸۴	۰/۲۴۵	−۰/۰۸۲	۰/۲۴۸
	β_2	۰/۹۹۵	۰/۱۴۴	۱/۷۷۳	۰/۲۶۵	۰/۹۹۵	۰/۱۵۱	۰/۹۶۲	۰/۱۴۹
	β_3	۰/۶۸۲	۰/۲۶۷	۱/۲۷۴	۰/۴۷۳	۰/۷۰۸	۰/۲۷۴	۰/۶۳۶	۰/۲۷۶
	β_4	۱/۵۶۵	۰/۳۰۸	۲/۸۶۳	۰/۵۸۸	۱/۵۱۹	۰/۳۰۰	۱/۵۰۸	۰/۳۰۲
	β_5	−۰/۴۲۴	۰/۲۴۲	−۰/۷۶۳	۰/۴۲۶	−۰/۴۷۴	۰/۲۲۵	−۰/۳۹۱	۰/۲۴۱
	ξ							−۰/۲۸۶	۰/۲۰۰
۱۰۰۰	β_1	۰/۰۱۸	۰/۰۹۲	۰/۰۶۳	۰/۱۶۴	−۰/۴۴۸	۰/۱۱۱	−۰/۳۲۹	۰/۱۰۵
	β_2	۱/۰۹۳	۰/۰۶۹	۱/۸۸۸	۰/۱۳۴	۱/۱۰۰	۰/۰۷۳	۱/۰۹۲	۰/۰۷۲
	β_3	۱/۱۲۷	۰/۱۲۵	۱/۸۹۵	۰/۲۱۶	۱/۱۹۴	۰/۱۳۵	۱/۱۳۳	۰/۱۳۱
	β_4	۱/۱۶۲	۰/۱۲۷	۲/۰۰۵	۰/۲۲۶	۱/۱۰۱	۰/۱۳۲	۱/۱۶۷	۰/۱۲۵
	β_5	−۰/۳۴۷	۳/۱۰۸	−۰/۶۱۴	۰/۱۷۹	−۰/۳۲۵	۰/۱۰۴	−۰/۳۴۲	۰/۱۰۰
	ξ							−۰/۳۰۶	۰/۱۰۵
۵۰۰۰	β_1	۰/۰۱۱	۰/۰۴۳	۰/۲۴	۰/۷۴	۰/۴۵۲	۰/۰۴۶	۰/۳۴۵	۰/۰۴۹
	β_2	۱/۰۶۲	۰/۰۳۰	۱/۸۳۶	۰/۰۵۶	۱/۰۶۲	۰/۰۳۰	۱/۰۶۸	۰/۰۳۱
	β_3	۱/۰۷۵	۰/۰۵۶	۱/۸۵۲	۰/۰۹۸	۱/۰۹۴	۰/۰۵۶	۱/۰۸۲	۰/۰۵۷
	β_4	۱/۳۰۱	۰/۰۵۸	۲/۲۴۲	۰/۱۰۴	۱/۳۱۰	۰/۰۵۸	۱/۳۱۲	۰/۰۵۹
	β_5	−۰/۳۵۱	۰/۰۴۴	−۰/۶۰۹	۰/۰۷۶	−/۰/۳۴۲	۰/۰۴۵	−/۰۳۵۴	۰/۰۴۵
	ξ							−۰/۲۷۱	۰/۰۳۶

تمام این نتایج مبتنی بر تایید همگرایی زنجیر تولیدشده بر اساس معیارهای همگرایی هیدلبرگ-ولچ و جی‌وک و سایر معیارهای اشاره‌شده در مثال اول، استخراج شده‌اند. برای صرف‌نظر از حجیم شدن پایان‌نامه از گزارش آن‌ها در این مثال خودداری کردیم.

جدول ۷.۴: فاصله‌های اعتبار ۹۵٪ برای پارامترهای مدل‌های برازش‌شده در مثال شبیه‌سازی دوم

n	پارامترها	Probit		Logit		Cloglog		GEV	
		٪۲٫۵	٪۷٫۵	٪۲٫۵	٪۷٫۵	٪۲٫۵	٪۷٫۵	٪۲٫۵	٪۷٫۵
۲۰۰	β_1	−۰٫۲۱۷	۰٫۷۱۴	−۰٫۴۰۲	۱٫۲۳۴	−۰٫۶۷۷	۰٫۲۸۳	−۰٫۵۵۴	۰٫۴۰۶
	β_2	۰٫۷۱۸	۱٫۲۸۰	۱٫۲۳۳	۲٫۳۰۳	۰٫۷۱۴	۱٫۳۰۵	۰٫۶۷۱	۱٫۲۵۱
	β_3	۰٫۱۵۸	۱٫۲۰۸	۰٫۳۴۱	۲٫۲۰۸	۰٫۱۵۷	۱٫۲۱۳	۰٫۱۰۴	۱٫۱۸۴
	β_4	۰٫۹۵۲	۲٫۲۶۵	۱٫۶۹۳	۴٫۰۱۸	۰٫۹۴۳	۲٫۱۱۹	۰٫۹۳۶	۲٫۱۰۷
	β_5	−۰٫۸۷۹	۰٫۰۶۹	−۱٫۵۴۹	۰٫۱۱۰	−۰٫۹۰۶	−۰٫۰۲۶	−۰٫۸۴۷	۰٫۰۸۹
	ξ							−۰٫۶۷۷	۰٫۱۲۷
۱۰۰۰	β_1	−۰٫۱۴۷	۰٫۲۱۴	−۰٫۲۴۷	۰٫۳۸۶	−۰٫۶۷۳	−۰٫۲۳۷	−۰٫۵۳۵	−۰٫۱۳۰
	β_2	۰٫۹۵۲	۱٫۲۲۴	۱٫۶۲۵	۲٫۱۵۰	۰٫۹۵۴	۱٫۲۴۱	۰٫۹۵۴	۱٫۲۳۶
	β_3	۰٫۸۷۹	۱٫۳۷۳	۱٫۴۹۹	۲٫۳۴۶	۰٫۹۲۶	۱٫۴۵۷	۰٫۸۶۹	۱٫۳۷۹
	β_4	۰٫۹۱۸	۱٫۴۱۵	۱٫۵۶۳	۲٫۴۵۰	۰٫۸۴۰	۱٫۳۷۰	۰٫۹۳۱	۱٫۴۲۲
	β_5	−۰٫۵۳۶	−۰٫۱۵۶	−۰٫۹۴۵	−۰٫۲۴۶	−۰٫۵۳۹	−۰٫۱۳۲	−۰٫۵۳۸	−۰٫۱۴۵
	ξ							−۰٫۵۰۱	−۰٫۰۸۹
۵۰۰۰	β_1	۰٫۰۷۳	۰٫۰۹۲	۰٫۱۱۳	۰٫۱۷۲	۰٫۵۴۱	۰٫۳۵۹	۰٫۴۳۷	۰٫۲۵۳
	β_2	۱٫۰۰۲	۱٫۱۲۰	۱٫۷۱۸	۱٫۹۴۱	۱٫۰۰۳	۱٫۱۲۴	۱٫۰۰۹	۱٫۱۳۱
	β_3	۰٫۹۶۶	۱٫۱۸۵	۱٫۶۵۹	۲٫۰۴۱	۰٫۹۸۳	۱٫۲۰۱	۰٫۹۷۲	۱٫۱۹۴
	β_4	۱٫۱۸۶	۱٫۴۱۱	۲٫۰۵۰	۲٫۴۵۴	۱٫۲۰۱	۱٫۴۲۸	۱٫۱۹۸	۱٫۴۲۹
	β_5	−۰٫۴۳۷	−۰٫۲۶۴	−۰٫۷۵۹	−۰٫۴۶۵	−۰٫۴۳۱	−۰٫۲۵۳	−۰٫۴۴۵	−۰٫۲۶۶
	ξ							−۰٫۳۴۱	−۰٫۱۹۷

جدول ۸.۴: معیارهای مقایسه مدل‌های Probit، Logit، GEV و Cloglog در مطالعه شبیه‌سازی دوم

n	پارامترها	Probit	Logit	Cloglog	GEV
۲۰۰	$D(\tilde{\theta})$	۱۵۸٫۸۶۷	۱۵۸٫۴۴۷	۱۶۱٫۶۷۹	۱۵۹٫۱۱۰
	DIC	۱۶۹٫۰۵۴	۱۶۸٫۵۶۳	۱۷۱٫۵۶۱	۱۷۰٫۹۷۳
	P_D	۵٫۰۹۳	۵٫۰۵۷	۴٫۹۴۱	۵٫۹۳۱
	Marginal Likelihood	−۷۹٫۴۳۳	−۷۹٫۲۲۳	−۸۰٫۸۳۹	−۷۹٫۵۵۹
	BIC	۱۸۵٫۳۵۹	۱۸۴٫۹۳۹	۱۸۸٫۱۷۱	۱۹۰٫۹۰۰
۱۰۰۰	$D(\tilde{\theta})$	۸۲۹٫۲۵۴	۸۳۲٫۳۸۵	۸۳۶٫۱۴۲	۸۲۸٫۴۹۵
	DIC	۸۳۹٫۱۶۸	۸۴۲٫۵۶۷	۸۴۶٫۵۴۲	۸۴۰٫۳۴۳
	P_D	۴٫۹۵۷	۵٫۰۹۰	۵٫۱۹۹	۵٫۹۲۴
	Marginal Likelihood	−۴۱۴٫۶۲۷	−۴۱۶٫۱۹۲	−۴۱۸٫۰۷۱	−۴۱۴٫۲۴۷
	BIC	۸۶۳٫۷۹۳	۸۶۶٫۹۲۴	۸۷۰٫۶۸۱	۸۶۹٫۹۴۱
۵۰۰۰	$D(\tilde{\theta})$	۴۱۱۲٫۹۳۲	۴۱۲۲٫۳۳۶	۴۱۵۲٫۰۶۳	۴۱۱۲٫۳۸۱
	DIC	۴۱۲۳٫۰۲۸	۴۱۳۲٫۲۸۹	۴۱۶۱٫۸۹۴	۴۱۲۴٫۴۴۵
	P_D	۵٫۰۴۸	۴٫۹۷۶	۴٫۹۱۵	۶٫۰۳۱
	Marginal Likelihood	−۲۰۵۶٫۴۶۶	−۲۰۶۱٫۱۶۸	−۲۰۷۶٫۰۳۲	−۲۰۵۶٫۱۹۱
	BIC	۴۱۵۵٫۵۱۸	۴۱۶۴٫۹۲۲	۴۱۹۴٫۶۴۹	۴۱۶۳٫۴۸۴

۳.۴ کاربرد مدل رگرسیونی مقادیر فرین تعمیم‌یافته

در این بخش، تحلیل بیزی هر چهار مدل رگرسیون پرابیت، لجیت، Cloglog و GEV را بر روی یک مجموعه داده واقعی تشریح می‌کنیم. این مجموعه، شامل داده‌های اداره خدمات مشترکین شرکت توزیع برق استان گیلان است. برای برازش مدل‌های مورد نظر در این بخش، یک نمونه مونت کارلویی به حجم ۴۰۰۰۰ با دوره داغیدن ۱۰۰۰۰ و طول گام انتخاب ۱۰ را در نظر گرفتیم. بنابراین، استنباط‌ها مبتنی بر یک نمونه ۳۰۰۰ تایی استخراج شده‌اند.

۱.۳.۴ مشتریان خوش‌حساب و بدحساب شرکت توزیع برق

این مجموعه داده شامل اطلاعات ثبت‌شده از مشترکین شرکت برق استان گیلان در بین سال‌های ۱۳۵۰ تا ۱۳۹۵ است. متغیر پاسخ در این مجموعه یک متغیر دو سطحی با سطوح مشتری خوش‌حساب (۱) و مشتری بدحساب (۰) است. متغیرهای تبیینی نیز شامل سه متغیر متوسط مصرف برق، مدت زمان اشتراک و نوع مصرف هستند. دو متغیر تبیینی اول پیوسته و متغیر تبیینی سوم رسته‌ای با ۵ دسته مصرف خانگی، کشاورزی، تولید، عمومی و سایر مصارف هستند. متغیر تبیینی رسته‌ای را برای وارد کردن در مدل به چهار متغیر ظاهری تبدیل کردیم. تعداد نمونه‌ها بعد از انجام عملیات پاکسازی[۴]، شامل حذف رکوردهای گم‌شده و پرت، برابر ۵۳۱۲ است. از بین این تعداد نمونه، ۴۱۷۹ پاسخ برابر یک (مشتریان خوش‌حساب) و بقیه صفر (مشتریان بدحساب) هستند. به عبارت دیگر تقریبا ۷۹٪ پاسخ‌ها را یک شامل می‌شود.

پیشگوی خطی برای هر چهار مدل رگرسیونی، عبارت است از

$$\mathbf{x}_i'\boldsymbol{\beta} = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_7 x_{i7}, \quad i = 1, \ldots, n.$$

توزیع‌های پیشین پارامترهای رگرسیونی در مدل‌های لجیت، پرابیت، Cloglog و GEV مشابه مثال‌های شبیه‌سازی در مدل‌های متناظرشان انتخاب شدند. یعنی برای همه پارامترهای رگرسیونی، توزیع پیشین $N(0, 10^2)$ انتصاب شد. برای پارامتر شکل در مدل GEV نیز توزیع پیشین یکنواخت $U[-1, 1)$ در نظر گرفته شد.

معیارهای سنجش همگرایی و آمیختگی زنجیر، شامل آزمون‌های هیدلبرگ-ولچ و جی‌وک، نمودارهای اثر و خودهمبستگی، و معیار ESS برای زنجیر تولیدشده در هر چهار مدل راضی‌کننده بودند و کیفیت خوب نمونه‌های تولیدشده را تایید می‌کردند که به دلیل دوری از حجیم شدن پایان‌نامه از گزارش آن‌ها خودداری کردیم. بنابراین با اطمینان از کیفیت نمونه تولیدی، نتایج استنباط را استخراج کردیم.

معیارهای مقایسه مدل‌ها در جدول ۹.۴ گزارش شده‌اند. با بررسی مقادیر این معیارها برای هر چهار مدل، می‌توان نتیجه گرفت همه مدل‌ها، بر اساس این معیارها، عملکرد یکسانی در برازش داده‌های شرکت توزیع برق دارند. به‌طور جزیی‌تر، به‌جز BIC، سایر معیارها برای هر چهار مدل یکسان و مشابه هستند. تنها معیار BIC برای مدل GEV از سایر مدل‌ها که مقداری مشابه با هم دارند، بیشتر است. البته این اختلاف نیز چندان قابل اعتنا نیست. بنابراین با توجه به بزرگ‌تر و منعطف‌تر بودن رده

[۴] Cleaning

مدل‌های رگرسیونی GEV، که در جدول فاصله‌های اعتبار پارامترها نیز دیده می‌شود، از آن برای توصیف پارامترهای مدل استفاده می‌کنیم.

جدول ۹.۴: معیارهای مقایسه مدل‌های Probit، Logit، GEV و Cloglog برای داده‌های واقعی

پارامترها	Probit	Logit	Cloglog	GEV
$D(\tilde{\theta})$	۵۴۰۳/۵۸۵	۵۴۰۳/۴۵۸	۵۴۰۳/۷۵۳	۵۴۰۳/۷۵۳
DIC	۵۴۱۷/۵۶۵	۵۴۱۷/۴۸۱	۵۴۱۷/۷۷	۵۴۱۷/۴۹۹
P_D	۶/۹۸۹۷	۷/۰۱۱۵	۷/۰۰۸۵	۶/۸۷۳۱
Marginal Likelihood	−۲۷۰۱/۷۹۲	−۲۷۰۱/۷۲۹	−۲۷۰۱/۸۷۷	−۲۷۰۱/۸۷۶
BIC	۵۴۴۶/۴۷۴	۵۴۴۶/۳۴۷	۵۴۴۶/۶۴۲	۵۴۵۵/۲۱۹

در دو جدول ۱۰.۴ و ۱۱.۴ برآوردهای بیزی پارامترها، که همان میانگین توزیع پسین هستند، خطای استاندارد برآوردها و فاصله‌های اعتبار ۹۵٪ برای هر چهار مدل گزارش شده‌اند. مشاهده می‌کنید پارامترهای برآوردشده در مدل GEV نزدیک به دو مدل پرابیت و Cloglog هستند و می‌توان گفت هر سه مدل تقریباً برازش مشابهی دارند. این مسأله در فاصله اعتبار پارامتر ξ نیز قابل درک است. زیرا این فاصله اعتبار هم مقدار $\xi = 0$ را در بر دارد و هم مقدار $\xi = -0/277$ که به ترتیب مقادیر واقعی و تقریبی در مدل‌های Cloglog و پرابیت هستند. طول فواصل اعتبار پارامترها برای هر سه مدل نیز به یکدیگر نزدیک هستند.

جدول ۱۰.۴: برآوردهای بیزی و خطای استاندارد آن‌ها برای مدل‌های Probit، Logit، Cloglog و GEV

پارامترها	Probit		Logit		Cloglog		GEV	
	میانگین	SE	میانگین	SE	میانگین	SE	میانگین	SE
β_1	۰/۵۷۶۷	۰/۰۳۶۶	۰/۹۳۷۶	۰/۰۶۱۶	۰/۲۳۳۹	۰/۰۳۴۰	۰/۲۵۳۸	۰/۰۴۲۷
β_2	−۰/۲۲۰۸	۰/۱۳۵۴	−۰/۳۸۰۲	۰/۲۲۰۹	−۰/۲۳۶۷	۰/۱۴۹۳	−۰/۱۸۹۹	۰/۱۲۵۸
β_3	−۰/۰۱۱۰	۰/۰۱۹۳	−۰/۰۱۹۷	۰/۰۳۴۱	−۰/۰۱۰۲	۰/۰۱۷۴	−۰/۰۱۲۰	۰/۰۲۰۹
β_4	۰/۲۸۳۲	۰/۰۴۳۹	۰/۴۸۱۳	۰/۰۸۶۳	۰/۲۵۵۳	۰/۰۳۹۸	۰/۳۰۹۵	۰/۰۶۶۱
β_5	−۰/۱۹۸۱	۰/۱۴۸۲	−۰/۳۲۰۶	۰/۲۴۰۷	−۰/۲۰۷۰	۰/۱۵۰۳	−۰/۲۱۷۵	۰/۱۵۳۷
β_6	۰/۶۲۲۶	۰/۰۸۸۳	۱/۱۱۱۱	۰/۱۶۷۲	۰/۵۳۳۶	۰/۰۷۱۴	۰/۷۰۰۳	۰/۱۶۷۵
β_7	−۰/۵۷۴۶	۰/۲۱۱۵	−۰/۹۳۷۶	۰/۳۴۱۷	−۰/۶۴۴۷	۰/۲۴۹۶	−۰/۵۷۱۷	۰/۲۲۴۴
ξ							−۰/۴۹۱۶	۰/۴۰۷۷

با توجه به ضرایب رگرسیونی برآوردشده و فاصله‌های اعتبار آن‌ها می‌توان نتیجه‌گیری کرد، متغیر طول مدت اشتراک تاثیر معنی‌داری بر خوش‌حسابی یا بدحسابی مشترکین ندارد؛ زیرا این ضریب هم خیلی به صفر نزدیک است و هم فاصله اعتبار متناظرش مقدار صفر را در بر دارد. برای متغیر تبیینی متوسط مصرف، علی‌رغم آن‌که فاصله اعتبار اثر این متغیر شامل صفر است، اما کران بالای فاصله خیلی نزدیک به صفر است. بنابراین با توجه به بزرگی ضریب رگرسیونی آن و با علامت منفی، می‌توان گفت هر چه متوسط مصرف بیشتر می‌شود، شانس خوش‌حساب بودن مشترک کمتر خواهد شد.

جدول ۱۱.۴: فاصله‌های اعتبار ۹۵٪ پارامترها برای مدل‌های Probit، Logit، Cloglog و GEV

پارامترها	Probit		Logit		Cloglog		GEV	
	٪۲/۵	٪۹۷/۵	٪۲/۵	٪۹۷/۵	٪۲/۵	٪۹۷/۵	٪۲/۵	٪۹۷/۵
β_1	۰/۵۰۴۵	۰/۶۴۸۵	۰/۸۱۲۲	۱/۰۵۳۷	۰/۱۷۱۰	۰/۳۰۴۴	۰/۱۶۹۹	۰/۳۳۶۵
β_2	−۰/۴۷۰۹	۰/۰۵۷۴	−۰/۷۹۰۸	۰/۰۷۳۴	−۰/۵۲۳۶	۰/۰۵۰۱	−۰/۴۳۸۳	۰/۰۵۲۵
β_3	−۰/۰۴۷۴	۰/۰۲۶۸	−۰/۰۸۳۵	۰/۰۴۹۶	−۰/۰۴۳۷	۰/۰۲۵۳	−۰/۰۵۸۱	۰/۰۲۶۳
β_4	۰/۲۰۰۷	۰/۳۶۹۴	۰/۳۳۷۹	۰/۶۳۴۷	۰/۱۷۶۲	۰/۳۳۳۱	۰/۱۷۹۳	۰/۴۳۷۶
β_5	−۰/۴۹۴۳	۰/۰۷۷۷	−۰/۷۹۱۱	۰/۱۴۴۱	−۰/۵۰۵۷	۰/۰۷۹۲	−۰/۵۱۹۴	۰/۰۸۰۹
β_6	۰/۴۴۰۰	۰/۷۸۶۴	۰/۷۹۸۲	۱/۴۴۹۲	۰/۳۹۲۰	۰/۶۷۰۲	۰/۳۷۰۵	۱/۰۲۲۵
β_7	−۰/۹۹۹۳	−۰/۱۸۱۱	−۱/۵۹۷۰	−۰/۲۵۶۹	−۱/۱۲۸۰	۰/۱۷۴۹	−۰/۹۸۰۶	−۰/۱۲۳۹
ξ							−۰/۹۹۹۷	۰/۴۳۲۹

در مورد متغیر تبیینی نوع مصرف، با توجه به این‌که از چهار متغیر ظاهری، با مبنا قرار دادن رده سایر مصارف، استفاده کردیم، تعبیر ضرایب در مقایسه با رده مبنا صورت می‌گیرد. به‌طور واضح‌تر ضریب β_7 (تقابل رده تولید با سایر مصارف) معنی‌دار نیست؛ این به آن معنی است که شانس خوش‌حساب یا بدحساب بودن مشترکین مصرف تولید با سایر مصارف، یکسان است. اما، با توجه به مثبت بودن ضرایب β_4 (تقابل رده نوع مصرف خانگی با سایر مصارف) و β_6 (تقابل رده نوع مصرف عمومی با سایر مصارف)، شانس خوش‌حساب بودن مشترکین خانگی و عمومی در مقایسه با سایر مصارف بیشتر است. در مقابل، با توجه به منفی بودن ضریب β_5 (تقابل رده نوع مصرف کشاورزی با سایر مصارف)، احتمال وجود مشترکین کشاورزی خوش‌حساب نسبت به سایر مصارف کمتر است. این تعبیرها با تغییر رده مبنا از سایر مصارف به انواع دیگر مصرف تغییر می‌کنند.

با این نتایج، می‌توان در مورد مشترکین خوش‌حساب و بدحساب با دانش از اطلاعات نهفته در متغیرهای تبیینی آن‌ها شامل متوسط مصرف، و نوع اشتراک، برنامه‌ریزی و تصمیم‌گیری کرد. به‌عنوان نمونه، مشتریان خوش‌حساب با نوع مصرف خانگی و عمومی بیشتر از سایر مشتریان هستند. پس اعمال سیاست‌های تشویقی برای این نوع مشترکین، در کنار در نظر گرفتن متوسط مصرف آن‌ها، می‌تواند بر افزایش درصد مشتریان خوش‌حساب تاثیرگذار باشد. از طرف دیگر، مشترکین کشاورزی مشتریان خوش‌حسابی نیستند. بنابراین شاید تغییر سیاست‌های قیمت‌گذاری بر برق مصرفی آن‌ها و نحوه وصول هزینه آن، برای افزایش خوش‌حسابی در این مشترکین لازم باشد.

۴.۴ نتیجه‌گیری و پیشنهادات برای آینده تحقیق

در این پایان‌نامه مدل رگرسیون مقادیر فرین تعمیم‌یافته برای پاسخ‌های دو سطحی نامتعادل معرفی شد. از آن‌جایی که توزیع مقادیر فرین تعمیم‌یافته دارای توزیع پیشین مزدوج نیست، توزیع پسین پارامترها با توزیع‌های پیشین در نظر گرفته‌شده، شکل بسته‌ای ندارد. بنابراین، برازش آن در دیدگاه بیزی با کمک روش‌های نمونه‌گیری MCMC صورت گرفت. پیچیدگی توزیع‌های شرطی کامل پارامترها موجب شد تا از ترکیب دو روش نمونه‌گیری متروپلیس-هستینگز و گیبز استفاده شود. عملکرد این مدل در کنار

سه مدل پرابیت، لجیت و Cloglog، با دو مطالعه شبیه‌سازی مورد ارزیابی قرار گرفت و کاربرد آن در یک مثال واقعی، در مورد مشترکین برق استان گیلان، نشان داده شد. با توجه به نتایج به‌دست آمده از شبیه‌سازی‌ها و پیاده‌سازی مدل بر روی مثال واقعی، به‌طور خلاصه می‌توان نتایج زیر را فهرست کرد:

- از مزیت‌های اساسی مدل رگرسیونی مقادیر فرین تعمیم‌یافته، انعطاف‌پذیری بالای آن در تشخیص و کنترل چولگی است؛ زیرا تابع چگالی این مدل، با مقادیر متفاوت پارامتر شکل می‌تواند برای مدل‌بندی داده‌های متقارن، چوله به راست و چوله به چپ به کار رود.
- شبیه‌سازی‌های انجام‌شده نشان دادند که انعطاف‌پذیری این مدل حتی با حجم نمونه کوچک نیز مناسب است.
- تقریباً در هر دو شبیه‌سازی و مثال واقعی، با توجه به معیارهای انتخاب مدل، مدل رگرسیونی مقادیر فرین تعمیم‌یافته با سایر مدل‌ها رقابت نزدیکی دارد.

با توجه به مطالب نظری و شبیه‌سازی در این پایان‌نامه، می‌توان از موارد زیر به‌عنوان آینده تحقیق نام برد:

- پارامتر شکل ξ که چولگی این مدل رگرسیونی را کنترل می‌کند، بر روی سنگینی دنباله‌ها اثر می‌گذارد. اگر بتوان با طراحی مکانیزمی توزیع مقادیر فرین تعمیم‌یافته را به گونه‌ای تغییر داد که این پارامتر فقط چولگی را کنترل کند، می‌توان انعطاف‌پذیری آن را بهبود بخشید.
- در این پایان‌نامه از این مدل برای متغیرهای پاسخ دو سطحی مستقل استفاده کردیم. استفاده از آن برای پاسخ‌های دو سطحی با ساختارهای وابستگی مختلف مانند وابستگی فضایی، زمانی، فضایی-زمانی، طولی و خوشه‌ای و استفاده از مدل‌های آمیخته در این مدل می‌تواند جالب توجه باشد.

مراجع

[۱] اعتمادی، علی. (۱۳۹۰)، استنباط آماری برای مدل‌های آمیخته، پایان‌نامه کارشناسی ارشد، دانشگاه مازندران، گروه آمار.

[۲] نیرومند، ح. (۱۳۸۴)، تحلیل رگرسیون خطی، چاپ اول، موسسه چاپ و انتشارات دانشگاه فردوسی مشهد.

[۳] نیرومند، ح. (۱۳۸۷)، الگوهای خطی تعمیم‌یافته با کاربردهای آن در مهندسی و علوم، چاپ اول، موسسه چاپ و انتشارات دانشگاه فردوسی مشهد.

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

[5] Albert, J.H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, **88**(422), 669-679.

[6] Aranda-Ordaz, F.J. (1981). On two families of transformations to additivity for binary response data, *Biometrika*, **68**(2), 357-363.

[7] Arellano-Valle, R.B. & Bolfarine, H. (1995). On some characterizations of the t-distribution, *Journal of the Statistics and Probability Letters*, **25**(1), 79-85.

[8] Arnold, B.C. & Groeneveld, R.A. (1995). Measuring skewness with respect to the mode, *the American Statistician*, **49**(1), 34-38.

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

[10] Bickel, P.J. & Doksum, K.A. (2001). *Mathematical statistics: basic ideas and selected topics*, CRC Press, New York.

[11] Birkhoff, G.D. (1942). What is the ergodic theorem?, *American Mathematical Monthly*, **49**(4), 222-226.

[12] Borsboom, D., Mellenbergh, G.J., & Van Heerden, J. (2003). The theoretical status of latent variables, *Psychological Review*, **110**(2), 203.

[13] Bowley, A.L. (1920). Elements of statistics, *Journal of the Royal Economic Society*, **31**(122), 220-224.

[14] Branscum, A.J., Jonson, W.O. & Thurmond, M.C. (2007). Bayesian beta regression: application to household data and genetic distanse between foot-and-mouth disease viruses, *Australian & New Zealand Journal of Statistics*, **49**(3), 287-301.

[15] Caers, J., Beirlant, J., & Maes, M.A. (1999). Statistics for Modeling HeavyTailed Distributions in Geology: Part II. Applications, *Journal of the International Association for Mathematical Geology*, **31**(4), 411-434.

[16] Casella, G. & Berger, R.L. (2002). *Statistical Inference*, Duxbury, Thomson Learning, New York.

[17] Charlier, C.V.L. (1905). Uber das fehlergesetz, *Arkiv For Mathematik, Astronomi och Fysik*, **2**, 1-8.

[18] Chen, M.H., Dey, D.K., & Shao, Q.M. (1999). A new skewed link model for dichotomous quantal response data, *Journal of the American Statistical Association*, **94**, 1172–1186.

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

[20] Chen, M.H. & Shao, Q.M. (2001). Propriety of posterior distribution for dichotomous quantal response models, *Proceedings of the American Mathematical Society*, **129**(1), 293–302.

[21] Chib, S. (1995). Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**, 1313–1321.

[22] Chib, S. & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output, *Journal of the American Statistical Association*, **96**, 270–281.

[23] Coles, S., Bawa, J., Trenner, L. & Dorazio, P. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

[24] Conti, E., Muller, C.W., & Stewart, M. (2006). Karyopherin flexibility in nucleocytoplasmic transport, *Current Opinion in Structural Biology*, **16**(2), 237-244.

[25] Cooley, D., Nychka, D., & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels, *Journal of the American Statistical Association*, **102**, 824-840.

[26] Czado, C. & Santner, T.J. (1992). The effect of link misspecification on binary regression inference, *Journal of Statistical Planning and Inference*, **33**(2), 213-231.

[27] Dayton, C.M. & Macready, G.B. (1988). Concomitant-variable latent-class models, *Journal of the American Statistical Association*, **83**, 173-178.

[28] Dryden, I.L. & Zempleni, A. (2006). Extreme shape analysis, *Journal of the Royal Statistical Society (Series C)*, **55**(1), 103-121.

[29] Edgeworth, F.Y. (1904). The law of error, *Transactions of the Cambridge Philosophical Society*, **20**, 36-65 and 113-114.

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

[31] Fisher, R.A. & Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *In Mathematical Proceedings of the Cambridge Philosophical Society*, **24**, 180-190.

[32] Galton, F. (1885). Photographic composites, *The Photographic News*, **29**, 243-245.

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

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

[35] Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2003). *Bayesian Data Analysis*, Chapman and Hall, New York.

[36] Gelman, A. & Hill, J. (2006). *Data Analysis Regression and Multilevel/ Hierarchical Models*, Cambridge University Press, New York.

[37] Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457-472.

[38] Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Byesian restoration of image, *IEEE Transactions on Pattern Analysis and Machine Inteligence*, **6**, 721-741.

[39] Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, Federal Reserve Bank of Minneapolis and Univercity of Minnesota, New York.

[40] Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Journal of the Econometric Society*, **57**(6), 1317-1339.

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

[42] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire, *Annals of Mathematics*, **44**, 423-453.

[43] Groeneveld, R.A., & Meeden, G. (1977). The mode, median, and mean inequality, *The American Statistician*, **31**(3), 120-121.

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

[45] Guerrero, V.M., & Johnson, R.A. (1982). Use of the Box-Cox transformation with binary response models, *Biometrika*, **69**(2), 309-314.

[46] Gupta, M., & Ibrahim, J.G. (2009). An information matrix prior for Bayesian analysis in generalized linear models with high dimensional data, *Statistica Sinica*, **19**(4), 1641.

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

[48] Heidelberger, P. & Welch, P.D. (1983). Simulation run length control in the presence of an initial transient, *Operations Research*, **31**(6), 1109-1144.

[49] Hosking, J.R.M. (1985). Maximum likelihood estimation of the parameters of the generalized extreme value distributions, *Journal of the Royal Statistical Society (Series C)*, **34**(3), 301-310.

[50] Jackman, S. (1989). *Generalized Linear Models*, Stanford University, New York.

[51] Jeffreys, H. (1936). Further significance tests, *Mathematical Proceedings of the Cambridge Philosophical Society*, **32**(3), 416-445.

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

[53] Kass, R.E. & Raftery, A.E. (1995). Bayes factors, *Journal of the American Statistical Association*, **90**, 773–795.

[54] Kim, S., Chen, M.H., & Dey, D.K. (2008). Flexible generalized t-link models for binary response data, *Biometrika*, **95**(1), 93–106.

[55] Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London.

[56] Laird, N.M. & Ware, J.H. (1982). Random-effects models for longitudinal data, *Biometrics*, **38**(4), 963-974.

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

[58] Laurmann, J.A., & Gates, W.L. (1977). Statistical Considerations in the Evaluation of Climatic Experiments with Atmospheric General Circulation Models, *Journal of the Atmospheric Sciences*, **34**(8), 1187-1199.

[59] Li, X. & Morris, J.M. (1991). On Measuring asymmetry and the reliability of the skewness measure, *Journal of the Statistics and Probability Letters*, **12**(3), 267-271.

[60] Liu, C. (2004). Robit regression: a simple robust alternative to logistic and probit regression, Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives, 227-238.

[61] McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*, Chapman and Hall, London.

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

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

[64] Montgomery, D.C., Peck, E.A. & Vining, G.G. (2015). *Introduction to Linear Regression Analysis*, John Wiley & Sons, Hoboken, New Jersey.

[65] Morgan, B.J.T. (1983). Observations on quantit analysis, *Biometrics*, **39**(4), 879–886.

[66] Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions, *Scandinavian Journal of Statistics*, **8**(3), 154-168.

[67] Pearson, K. (1895). Contributions to the mathematical theory of evolution, *Philosophical Transactions of the Royal Society of London (Series A)*, **186**, 343-414.

[68] Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). Coda: Convergence diagnosis and output analysis for MCMC, R News, **6**(1), 7-11.

[69] Prescott, P. & Walden, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme value distributions, *Biometrika*, **67**(3), 723-724.

[70] Reiss, R.D., Thomas, M., & Reiss, R.D. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance Hydrology and Other Fields*, Springer, New York.

[71] Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

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

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

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

[75] Rudin, W. (1964). *Principles of Mathematical Analysis*, McGraw-Hill, New York.

[76] Schwarz, G. (1987). Further analysis of the data by Akaike's information criterion and the finite corrections, *Communications in Statistics, Theory and Methods*,**7**(1), 13-26.

[77] Shmueli, G. & Koppius, O. (2009). The challenge of prediction in information systems research, *Robert H. Smith School Research Paper No. RHS*, 06-152.

[78] Shoemaker, A.L. (1996). What's Normal? temperature, gender, and heart rate, *Journal of Statistics Education*, **4**.

[79] Smith, R.L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone, *Statistical Science*, **4**, 367–393.

[80] Smith, R.L. (2003). *Statistics of Extremes, with Applications in Environment, Insurance and Finance*, Chapman and Hall/CRC Press, London.

[81] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Lind, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society (Series B)*, **64**(4), 583-639.

[82] Stukel, T. (1988). Generalized logistic models. *Journal of the American Statistical Association*, **83**, 426–431.

[83] Team, R.C. (2016), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

[84] Thompson, M.L., Reynolds, J., Cox, L.H., Guttorp, P., & Sampson, P.D. (2001). A Review of Statistical Methods for the Meteorological Adjustment of Tropospheric Ozone, *Atmospheric Environment*, **35**(3), 617-630.

[85] Van Zwet, W.R. (1964). *Convex Transformations of Random Variables*, Mathematisch Centrum Amsterdam.

[86] Vehtari, A. & Ojanen ,J. (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison, *Statistics Surveys*, **6**, 142-228.

[87] Von Mises, R. (1931). *Wahrscheinlichkeitschrechnung*, Deutiche, Vienna.

[88] Von Mises, R. (1936). La distribution de la plus grande de n valeurs, *Journal of the American Mathematical Society*, **11**, 271-294.

[89] Wang, X. & Dey, D.K. (2010). Generalized extreme value regression for binary response data: an application to B2B electronic payments system adoption, *The Annals of Applied Statistics*, **4**(4), 2000-2023.

[90] Watanabe, S. (2013). A widely applicable Bayesian information criterion, *Journal of Machine Learning Research*, **14**, 867-897.

[91] Yildirim, I. (2012). *Bayesian Inference: Metropolis-Hastings Sampling*, Rochester, New York.

پیوست آ

دستورهای R برای بازتولید قسمتی از نتایج پایان‌نامه

نمونه‌ای از کدهای R برای شبیه‌سازی‌های فصل چهارم و همچنین نمودارهای پایان‌نامه در این پیوست گزارش شده‌اند.

۱) دستورات لازم برای تولید شکل ۱.۱ به صورت زیر می‌باشند:

برای حالت (آ):

```
f=function(x){
alpha=0
beta=1
y = alpha + beta*x;
Y= 1/(1+ exp(-y));
}
curve(f,xlim=c(-5,5),ylim=c(0,1), lwd = 2,main=' ',xlab=' ',ylab=' ' )
```

برای حالت (ب):

```
f=function(x){
alpha =0
```

```
beta=1
y <- alpha + beta*x;
Y= exp(y)/((1+ exp(y)))^2;
}
curve(f,xlim=c(-5,5),ylim=c(0,0.4), lwd = 2,main=' ',xlab=' ',ylab=' ' )
```

۲) دستورات لازم برای تولید شکل ۲.۱ به صورت زیر می‌باشند:
برای حالت (آ):

```
f=function(x){
pnorm(x, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
}
curve(f,xlim=c(-5,5),ylim=c(0,1), lwd = 2,main=' ',xlab=' ',ylab=' ' )
```

برای حالت (ب):

```
f=function(x){
dnorm(x,mean = 0, sd = 1,log = FALSE)
}
curve(f,xlim=c(-5,5),ylim=c(0,0.4), lwd = 2,main=' ',xlab=' ',ylab=' ' )
```

۳) دستورات لازم برای تولید شکل ۳.۱ به صورت زیر می‌باشند:
برای حالت (آ):

```
f=function(x){
alpha =0
beta=1
y <- alpha + beta*x; Y= 1-(exp(-(exp(y))))
}
curve(f,xlim=c(-5,5),ylim=c(0,1), lwd = 2,main=' ',xlab=' ',ylab=' ' )
```

برای حالت (ب):

```
f=function(x){
alpha =0
beta=1
y <- alpha + beta*x; c=(beta*exp(y));Y=c*(exp(-(exp(y))))
}
curve(f,xlim=c(-5,5),ylim=c(0,0.4), lwd = 2,main=' ',xlab=' ',ylab=' ' )
```

۴) دستورات لازم برای تولید شکل ۲.۲ به صورت زیر می‌باشند:

```
dt<-function(v1,v2,x){
a<-gamma((v1+1)/2)
b<-gamma(v1/2)
c<-sqrt(pi*v2)
d<-(1+((x^2)/v2))^(-((v1+1)/2))
k<-(a/(b*c))*d
k
}
dt(v1,v2,x)
curve(dt(v1,v2,x),0,4,ylim=c(0,0.5),ylab="Probability Density Function",xlab="w")
curve(dnorm(x),lty=5,col="red",add=T)
```

۵) دستورات لازم برای تولید شکل ۴.۲ به صورت زیر می‌باشند:

```
f=function(x){
dnweibull(x, loc=0, scale=1, shape=0.5, log = FALSE)
}
curve(f,xlim=c(-4,6),ylim=c(0,1), main='  ',xlab='  ',ylab='  ' )

f=function(x){
dfrechet(x, shape=0.5, log = FALSE)
}
curve(f,xlim=c(-4,6),ylim=c(0,1),lty = 3,type = "l",col="red", main='  ',
xlab='  ',ylab='  ',add=T )

f=function(x){
dgumbel(x)
}
curve(f,xlim=c(-4,6),ylim=c(0,1),lty = 5,type = "l",col="blue", main='  ',
xlab='  ',ylab='  ',add=T )
```

۶) دستورات لازم برای تولید شکل ۵.۲ به صورت زیر می‌باشند:

```
f=function(x){
dnweibull(x, loc=0, scale=1, shape=7, log = FALSE)
```

```
}
curve(f,xlim=c(-4,6),ylim=c(0,3),type = "l", main=' ',xlab=' ',ylab=' ' )

f=function(x){
dnweibull(x, loc=0, scale=1, shape=3.6, log = FALSE)
}
curve(f,xlim=c(-4,6),ylim=c(0,3),lty = 3,type = "l",col="red", main=' ',
xlab=' ',ylab=' ',add=T )

f=function(x){
dnweibull(x, loc=0, scale=1, shape=2, log = FALSE)
}
curve(f,xlim=c(-4,6),ylim=c(0,3),lty = 5,type = "l",col="blue", main=' ',
xlab=' ',ylab=' ',add=T )
```

۷) دستورات لازم برای تولید شکل ۶.۲ به صورت زیر می‌باشند:

```
f=function(x){1-pgev(-x, loc=0, scale=1, shape=-0.5, lower.tail = TRUE)}
curve(f(x),xlim=c(-4,4),ylim=c(0,1),col="blue",lty=2,xlab="eta"
,ylab="Cumulative Distribution Fanction")

f=function(x){1-pgev(-x, loc=0, scale=1, shape=0.5, lower.tail = TRUE)}
curve(f(x),xlim=c(-4,4),ylim=c(0,1),col="red",lty=3,add=T)

f=function(x){1-pgev(-x, loc=0, scale=1, shape=0, lower.tail = TRUE)}
curve(f(x),xlim=c(-4,4),ylim=c(0,1),add=T,)
```

۸) دستورات لازم برای تولید شکل ۷.۲ به صورت زیر می‌باشند:

```
qqplot(1-qnorm(p),1-qgev(p, loc=-0.35579, scale=0.99903, shape=-0.27760,
lower.tail =TRUE),type="l",lwd=2,xlab="probit quantaile",ylab="GEV quantaile")
abline(0,1,lty = 5)
```

۹) دستورات R برای اجرای الگوریتم نمونه‌گیری MCMC در مدل GEV به صورت زیر می‌باشند:

```
#### MCMC algorithm (Metropolis-Hasting-within-Gibbs) for gev link model
mcmc.gev <- function(burn.N, mc.N){
```

```
# tuning parameters for the proposals
sigma.beta1 <- 1
sigma.beta2 <- 0.2
sigma.beta3 <- 0.3
sigma.beta4 <- 0.3
sigma.beta5 <- 0.3
sigma.xi <- 0.2
# uniform random variable for parameters
U.beta1 <- runif(burn.N + mc.N)
U.beta2 <- runif(burn.N + mc.N)
U.beta3 <- runif(burn.N + mc.N)
U.beta4 <- runif(burn.N + mc.N)
U.beta5 <- runif(burn.N + mc.N)
U.xi <- runif(burn.N + mc.N)
##
accept.beta1 <- 0
accept.beta2 <- 0
accept.beta3 <- 0
accept.beta4 <- 0
accept.beta5 <- 0
accept.xi <- 0
## space for the parameters and likelihood
beta1.star <- rep(NA, burn.N + mc.N)
beta2.star <- rep(NA, burn.N + mc.N)
beta3.star <- rep(NA, burn.N + mc.N)
beta4.star <- rep(NA, burn.N + mc.N)
beta5.star <- rep(NA, burn.N + mc.N)
sai.star <- rep(NA, burn.N + mc.N)
dev.star <- rep(NA, burn.N + mc.N)
#
out.glm <- glm(cbind(y, 1-y)~ x2 + x3 + x4 + x5,
    family=binomial(link="cloglog"), data=data.probit)
beta1.star[1] <- coef(out.glm)[1]
beta2.star[1] <- coef(out.glm)[2]
```

```
beta3.star[1] <- coef(out.glm)[3]
beta4.star[1] <- coef(out.glm)[4]
beta5.star[1] <- coef(out.glm)[5]
sai.star[1]  <- 0
dev.star[1] <- -2*like.gev(beta1.star[1], beta2.star[1], beta3.star[1],
beta4.star[1], beta5.star[1], sai.star[1])
##### starting updating states of the chain
for(r in 2:(burn.N + mc.N)){
print(r)
## updating the shape parameter: sai
sai.can <- rnorm(1,sai.star[r-1],sigma.xi)
fraction <- exp((like.gev(beta1.star[r-1], beta2.star[r-1],
beta3.star[r-1], beta4.star[r-1], beta5.star[r-1], sai.can) +
log(dnorm(sai.can,0,100))) - (like.gev(beta1.star[r-1],
beta2.star[r-1], beta3.star[r-1], beta4.star[r-1], beta5.star[r-1],
sai.star[r-1]) + log(dnorm(sai.star[r-1],0,100))))
metro.hasting.frac <- min(1,fraction)
if(U.xi[r] < metro.hasting.frac) {
accept.xi <- accept.xi + 1
sai.star[r] <- sai.can
} else {
sai.star[r] <- sai.star[r-1]
}
## updating the parameter beta1
beta1.can <- rnorm(1, beta1.star[r-1], sigma.beta1)
fraction <- exp((like.gev(beta1.can, beta2.star[r-1], beta3.star[r-1],
beta4.star[r-1], beta5.star[r-1], sai.star[r]) + log(dnorm(beta1.can, 0, 10^2)))-
(like.gev(beta1.star[r-1], beta2.star[r-1], beta3.star[r-1], beta4.star[r-1],
beta5.star[r-1], sai.star[r]) + log(dnorm(beta1.star[r-1], 0, 10^2))))
metro.hasting.frac <- min(1,fraction)
if(U.beta1[r] < metro.hasting.frac) {
accept.beta1 <- accept.beta1 + 1
beta1.star[r] <- beta1.can
} else {
```

```
beta1.star[r] <- beta1.star[r-1]
}
## updating the parameter beta2
beta2.can <- rnorm(1, beta2.star[r-1], sigma.beta2)
fraction <- exp((like.gev(beta1.star[r], beta2.can, beta3.star[r-1], beta4.star[r-1],
beta5.star[r-1], sai.star[r]) + log(dnorm(beta2.can, 0, 10^2)))-
(like.gev(beta1.star[r], beta2.star[r-1], beta3.star[r-1], beta4.star[r-1],
beta5.star[r-1], sai.star[r]) + log(dnorm(beta2.star[r-1], 0, 10^2))))
metro.hasting.frac <- min(1,fraction)
if(U.beta2[r] < metro.hasting.frac) {
accept.beta2 <- accept.beta2 + 1
beta2.star[r] <- beta2.can
} else {
beta2.star[r] <- beta2.star[r-1]
}
## updating the parameter beta3
beta3.can <- rnorm(1, beta3.star[r-1], sigma.beta3)
fraction <- exp((like.gev(beta1.star[r], beta2.star[r], beta3.can, beta4.star[r-1],
beta5.star[r-1], sai.star[r]) + log(dnorm(beta3.can, 0, 10^2)))-
(like.gev(beta1.star[r], beta2.star[r], beta3.star[r-1], beta4.star[r-1],
beta5.star[r-1], sai.star[r]) + log(dnorm(beta3.star[r-1], 0, 10^2))))
metro.hasting.frac <- min(1,fraction)
if(U.beta3[r] < metro.hasting.frac) {
accept.beta3 <- accept.beta3 + 1
beta3.star[r] <- beta3.can
} else {
beta3.star[r] <- beta3.star[r-1]
}
## updating the parameter beta4
beta4.can <- rnorm(1, beta4.star[r-1], sigma.beta4)
fraction <- exp((like.gev(beta1.star[r], beta2.star[r], beta3.star[r], beta4.can,
beta5.star[r-1], sai.star[r]) + log(dnorm(beta4.can, 0, 10^2)))-
(like.gev(beta1.star[r], beta2.star[r], beta3.star[r], beta4.star[r-1],
beta5.star[r-1], sai.star[r]) + log(dnorm(beta4.star[r-1], 0, 10^2))))
```

```
metro.hasting.frac <- min(1,fraction)
if(U.beta4[r] < metro.hasting.frac) {
accept.beta4 <- accept.beta4 + 1
beta4.star[r] <- beta4.can
} else {
beta4.star[r] <- beta4.star[r-1]
}
## updating the parameter beta5
beta5.can<- rnorm(1, beta5.star[r-1], sigma.beta5)
fraction <- exp((like.gev(beta1.star[r], beta2.star[r], beta3.star[r],
beta4.star[r],beta5.can, sai.star[r]) + log(dnorm(beta5.can, 0, 10^2)))-
(like.gev(beta1.star[r], beta2.star[r], beta3.star[r], beta4.star[r],
beta5.star[r-1], sai.star[r]) + log(dnorm(beta5.star[r-1], 0, 10^2))))
metro.hasting.frac <- min(1,fraction)
if(U.beta5[r] < metro.hasting.frac) {
accept.beta5 <- accept.beta5 + 1
beta5.star[r] <-beta5.can
} else {
beta5.star[r] <- beta5.star[r-1]
}
## updating the deviance
dev.star[r] <- -2*like.gev(beta1.star[r], beta2.star[r], beta3.star[r],
beta4.star[r],beta5.star[r], sai.star[r])
    }
    # end for loop
     result <- list("beta1.star"=beta1.star, "beta2.star"=beta2.star,
 "beta3.star"=beta3.star,"beta4.star"=beta4.star, "beta5.star"=beta5.star,
"sai.star"=sai.star,"dev.star"=dev.star, "accept.beta1"=accept.beta1,
"accept.beta2"=accept.beta2,"accept.beta2"=accept.beta2,
"accept.beta3"=accept.beta3, "accept.beta4"=accept.beta4,
"accept.beta5"=accept.beta5,"accept.sai"=accept.xi)
return(result)
}
```

Aabstract

For many applied situations, in a regression model, the response is a binary variable. The general models for analysis of such responses are logistic, probit and Cloglog. Logistic and probit symmetric link models are not appropriate choices when the number of observations in the two response categories is unbalanced. Further, the positively skewed Cloglog link model will lose its efficiency when responses are negatively skewed. The generalized t-link model is a good alternative in such situations, though the constraint on the shape parameter greatly reduces the possible range of skewness provided by this model. Wang and Dey (2010) proposed a new extremely flexible model based on generalized extreme value distribution. In this thesis, we reintroduce generalized extreme value link model and discuss its fitting from Bayesian framewrok. Due to the complexity of posterior distribution of the model we use Markov chain Monte Carlo methods. We explore the efficiency of generalized extreme value link model by a simulation study and demonstrate its applicability in the analysis of a real example.

Keywords: Generalized extreme value distribution; latent variable; Markov chain Monte Carlo algorithms; posterior distribution; rare events; skewness.

Shahrood University of Technology

Faculty Of Mathematical Sciences

MSc Thesis in: Statistics

Analysis of unbalanced binary responses by using generalized extreme value regression model

By: Solmaz Bolouri kalourazi

Supervisor

Dr. Hossein Baghishani

January 2017