



دانشکده علوم ریاضی

گروه علوم کامپیوتر و آمار

رشته آمار گرایش آمار ریاضی
پایان نامه کارشناسی ارشد

استنباط بیزی تقریبی مدل‌های رگرسیونی پویا با استفاده از تقریب لاپلاس آشیانه‌ای جمع بسته

مسعود میرزاجانی بختانی

استاد راهنما

دکتر حسین باغیشنی

تیر ۱۳۹۵

تقدیم به

روح پاک پدرم

که عالمانه به من آموخت تا چگونه در عرضه زندگی، ایستادگی را تجربه نمایم

و مادرم

دریای بی کران فداکاری و عشق که وجودش برایم همه مهر است.

سپاس گزارى...پ

سپاس بى کران پروردگار يکتاب را که هستى مان بخشيد و به طريق علم و دانش رهنمونمان شد و به همنشيني رهروان علم و دانش مفتخرمان نمود و خوشه چيني از علم و معرفت را روزيمان ساخت. با تقدير و تشکر شايسته از استاد فرهيخته جناب آقاى دکتر حسين باغيشنى که همواره راهنما و راهگشاي اينجانب در اتمام و اکمال پايان نامه بوده اند. با تشکر خالصانه از جناب آقاى دکتر محمدرضا ربيعى و دکتر احمد نزاكتى رضازاده که زحمت مطالعه و داورى پايان نامه اينجانب را تقبل کردند.

معود ميرزاچانى بختانى
سپريه ۱۳۹۵

تعمدنامه

اینجانب مسعود میرزاجانی بجمستانی دانشجوی کارشناسی ارشد رشته آمار ریاضی دانشکده علوم ریاضی دانشگاه صنعتی شاهرود، نویسنده پایان‌نامه با عنوان استنباط بیزی تقریبی مدل‌های رگرسیونی پویا با استفاده از تقریب لاپلاس آشیانه‌ای جمع‌بسته، 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، تاکنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود متعلق دارد، و مقالات مستخرج با نام “ دانشگاه صنعتی شاهرود “ یا “ Shahrood University “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده) شده است، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

مسعود میرزاجانی بجمستانی
سیر ۱۳۹۵

مالکیت نتایج و حق نشر

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

چکیده

مدل‌های پویا که دسته‌ی وسیعی از مدل‌های پارامتری هستند، کاربردهایی در مدل‌بندی و پیش‌بینی داده‌های سری زمانی و رگرسیون دارند. استنباط این مدل‌ها معمولاً بر پایه صورت‌های بازگشتی صافی کردن و هموارسازی بردارهای حالت با توجه به ساختار زمانی مشاهدات است، اما استنباط بیزی مدل‌های پیچیده آن با استفاده از الگوریتم‌های MCMC بسیار وقت‌گیر است و در اغلب مواقع مشکل آمیختگی و همگرایی را نیز دارد. حال در این پایان‌نامه به بیان روش تقریب لاپلاس آشیانه‌ای جمع‌بسته می‌پردازیم که استنباط بیزی کامل را برای مدل‌های خطی پویا و مدل‌های تعمیم‌یافته پویا اجرا می‌کند و همچنین مستقیماً چگالی‌های پسین را تقریب می‌زند. برای بیان این روش مثال‌های شبیه‌سازی شده از مدل خطی پویا و مدل‌های فضایی-زمانی را ارائه می‌کنیم. همچنین، دو مثال واقعی، داده‌های مصرف گاز کشور انگلستان و آب زیرزمینی استان گلستان، را به وسیله این روش استنباط می‌کنیم.

واژه‌های کلیدی: مدل‌های پویا، مدل تکمیل‌یافته، مدل‌های پویا فضایی-زمانی، تقریب لاپلاس آشیانه‌ای جمع‌بسته

پیشگفتار

موقعیت‌های کاربردی مختلفی در علوم هستند که داده‌های حاصل از مطالعه آن‌ها وابسته به زمان هستند. به این نوع داده‌ها در آمار، داده‌های سری زمانی گفته می‌شود. داده‌های سری زمانی گسترده‌ی وسیعی دارند و از ماهیتی پویا برخوردار هستند. بنابراین برای مدل‌بندی این نوع داده‌ها باید از رده مدل‌هایی استفاده کرد که ویژگی پویا بودن (وابستگی به زمان) داده‌ها را در نظر بگیرند. از چنین رده‌ای در متون آماری به مدل‌های پویا یاد می‌شود. مدل‌های خطی پویا، در مهندسی در دهه ۱۹۶۰ به منظور کنترل سیستم‌های پویا توسعه یافتند. اما در دهه‌های اخیر مدل‌های خطی پویا (DLMs) و گسترده‌تر از آن‌ها، مدل‌های فضای حالت کاربردهای فراوانی با دامنه وسیع پیدا کرده‌اند. همچنین در سال‌های اخیر رشد قابل توجهی در کاربردهای مدل‌های فضای حالت برای تحلیل سری زمانی ایجاد شده‌اند. شاید در ابتدا انجام استنباط بر اساس مدل‌های فضای حالت کمی دشوار به نظر برسد، اما این رده از مدل‌ها، یک رهیافت قدرتمند برای انجام تحلیل نزدیک‌تر به واقعیت پیشنهاد می‌کند. مدل‌های فضای حالت برتری‌های کاملاً آشکاری در مقایسه با مدل‌های ARIMA دارند. در دیدگاه بیزی همه استنباط‌ها مبتنی بر توزیع پسین به دست می‌آیند. از طرفی، توزیع پسین توأم پارامترها و فرآیند حالت در مدل‌های فضای حالت معمولاً شکل بسته‌ای ندارند و بسته به انتخاب توزیع‌های پیشین، پیچیده هستند. بنابراین برای استخراج استنباط‌های آماری باید از روش‌های تقریبی برای تقریب توزیع پسین استفاده کرد. یکی از رده‌های تقریب توزیع پسین استفاده از روش‌های شبیه‌سازی مونت کارلویی است. وظیفه الگوریتم‌های شبیه‌سازی مونت کارلویی، تولید نمونه‌هایی از یک توزیع احتمالی مفروض است. اما این روش شبیه‌سازی مشکلاتی مانند آمیختگی و همگرایی را دارد. جنبه بحرانی دیگری که نمی‌توان آن را نادیده گرفت، مشکل هزینه محاسباتی روش‌های MCMC است. زمانی که با مدل‌های پیچیده (مخصوصاً مدل‌های سلسله‌مراتبی) یا مجموعه داده‌های حجیم سروکار داریم، الگوریتم‌های MCMC ممکن است به شدت کند و حتی غیرعملی باشند. یکی از رهیافت‌های جانشین مناسب برای روش‌های MCMC که هزینه‌های محاسباتی استنباط بیزی را به شدت کاهش می‌دهد، روش INLA است. با توجه به این مقدمه، ساختار پایان‌نامه به صورت زیر است:

- در فصل اول، تعاریف و مفاهیم مقدماتی را بیان می‌کنیم.
- در فصل دوم، به‌طور مفصل روش INLA را تعریف می‌کنیم.
- در فصل سوم، به بیان داده‌های فضایی و فضایی-زمانی و روش SPDE می‌پردازیم.
- در فصل چهارم، با استفاده از مثال‌های شبیه‌سازی شده و داده‌های واقعی، استنباط بیزی تقریبی مدل‌های پویا به روش INLA را اجرا می‌کنیم.

لیست مقالات مستخرج از پایان نامه

۱. استنباط بیزی تقریبی مدل‌های پویا با استفاده از روش تقریب لاپلاس آشیانه‌ای جمع بسته-INLA، سیزدهمین کنفرانس آمار ایران، کرمان، پذیرش شده.

فهرست مطالب

ز	لیست تصاویر
۱	لیست جداول
۳	۱ استنباط بیزی در مدل‌های پویا
۳	۱.۱ مقدمه
۴	۲.۱ تعاریف و مفاهیم مقدماتی
۷	۱.۲.۱ مدل‌های خطی تعمیم‌یافته
۸	۲.۲.۱ رهیافت بیزی استنباط آماری
۱۰	۳.۱ مدل‌های آماری برای داده‌های سری زمانی
۱۲	۴.۱ مدل‌های فضای حالت
۱۴	۱.۴.۱ معرفی مدل
۱۷	۲.۴.۱ مدل مؤلفه‌های مشاهده‌نشده
۱۹	۳.۴.۱ مدل‌های خطی پویا
۲۱	۴.۴.۱ مدل‌های غیرخطی و غیرنرمال پویا
۲۳	۵.۱ استنباط بیزی تقریبی
۲۴	۱.۵.۱ تقریب لاپلاس
۲۶	۲.۵.۱ الگوریتم‌های MCMC
۳۲	۳.۵.۱ اجرای روش‌های MCMC
۳۲	۴.۵.۱ ضرورت معرفی رهیافت‌های جانشین
۳۵	۲ تقریب لاپلاس آشیانه‌ای جمع‌بسته
۳۵	۱.۲ مدل‌های گوسی پنهان
۳۷	۲.۲ تقریب لاپلاس آشیانه‌ای جمع‌بسته
۳۸	۱.۲.۲ تقریب چگالی پسین توأم بردار ابرپارامتر با راهبرد طرح مرکب مرکزی
۳۹	۲.۲.۲ روش طرح مرکب مرکزی
۴۲	۳.۲.۲ تقریب چگالی حاشیه‌ای پسین درایه‌های ابرپارامتر

۴۶	تقریب چگالی حاشیه‌ای پسین عناصر میدان پنهان	۴.۲.۲
۵۰	ملاک‌های ارزیابی	۳.۲
۵۰	تقریب حاشیه‌ای درست‌نمایی	۱.۳.۲
۵۱	ملاک انحراف اطلاع	۲.۳.۲
۵۲	ملاک پیش‌گویی شرطی مؤلفه‌ها	۳.۳.۲
۵۳	۳ داده‌های فضایی و فضایی-زمانی	
۵۳	مقدمه	۱.۳
۵۴	چرا از آمار فضایی و فضایی-زمانی استفاده می‌کنیم؟	۲.۳
۵۵	چرا از روش‌های بیزی برای مدل‌بندی فضایی و فضایی-زمانی استفاده می‌کنیم؟	۳.۳
۵۶	داده‌های فضایی	۴.۳
۵۸	روش معادله دیفرانسیل جزئی تصادفی	۵.۳
۶۰	SPDE با INLA - R	۶.۳
۶۱	تشریح روش SPDE با مثال شبیه‌سازی شده	۷.۳
۶۲	ساختن شبکه	۱.۷.۳
۶۳	ماتریس مشاهده	۲.۷.۳
۶۴	برازش مدل	۳.۷.۳
۶۵	استفاده از تابع inla.stack	۴.۷.۳
۶۷	پیش‌گویی فضایی	۵.۷.۳
۷۱	۴ استنباط بیزی تقریبی در مدل‌های پویا با روش INLA	
۷۱	مقدمه	۱.۴
۷۲	مثال ۱: مثالی بسیار ساده	۲.۴
۷۴	مثال ۲: مدل رگرسیون پویا تعمیم‌یافته	۳.۴
۷۵	مثال ۳: مدل پویای فصلی با همسازها	۴.۴
۷۷	مدل‌هایی با ساختار خطای همبسته	۵.۴
۷۷	مثال ۴: مدل پویای فضایی-زمانی مرتبه اول با متغیرهای تبیینی	۱.۵.۴
۷۸	مثال ۵: مدل پویای فضایی-زمانی مرتبه دوم	۲.۵.۴
۸۱	مطالعات موردی	۶.۴
۸۱	مثال ۶: مصرف گاز کشور انگلستان	۱.۶.۴
۸۳	مثال ۷: کیفیت آب‌های زیرزمینی استان گلستان	۲.۶.۴
۹۴	نتیجه‌گیری	۷.۴
۹۷		آ
۹۷	بسته INLA - R	۱.آ

لیست تصاویر

۵	۱۰.۱	چهار سری زمانی با الگوهای مختلف از داده‌های اقتصادی: الف) سود اوراق قرضه آمریکا، ب) حجم تولید برق خالص آمریکا، ج) تخودرو تولید شده انگلستان و د) بازدیدکنندگان از کشور استرالیا.
۱۳	۲۰.۱	هزینه مصرفی غذایی خانواده از سال ۱۹۹۶ تا ۲۰۰۵ در ایتالیا.
۱۳	۳۰.۱	مصرف گاز کشور انگلستان از سال ۱۹۶۰ تا ۱۹۸۶ بر حسب میلیون.
۱۴	۴۰.۱	میزان جریان سالیانه رودخانه نیل در آشوان از سال ۱۸۷۰ تا ۱۹۷۱.
۱۴	۵۰.۱	قیمت‌های روزانه گوگل.
۱۶	۶۰.۱	ساختار وابستگی مدل فضای حالت.
۱۸	۷۰.۱	نرخ تورم آمریکا از سال ۱۹۴۷ تا ۲۰۱۱.
۲۷	۸۰.۱	تابع چگالی توزیع گاما با پارامترهای $a = 5$ و $b = 1$ (نمودار بالا) و $a = 10$ و $b = 0.5$ (نمودار پایین). منحنی خط چین تقریب نرمال حاصل از روش لاپلاس است.
۴۲	۱۰.۲	موقعیت نقاط در طرح مرکزی مرکب
۴۳	۲۰.۲	مکان نقاط انتگرال‌گیری به وسیله راهبرد توری (نمودار سمت چپ) و CCD (نمودار سمت راست).
۴۳	۳۰.۲	توزیع نرمال استاندارد (خط ممتد) و چگالی‌های معادله (۱۴.۲) برای مقادیر مختلف پارامترهای مقیاس (خطوط خط چین).
۴۷	۴۰.۲	نمایش ماتریسی معادله تاکاهاشی
۵۷	۱۰.۳	نقشه گامبیا، آفریقا: نقاط نشان دهنده روستاهایی است که شیوع بیماری مالاریا در کودکان مشاهده شده است.
۵۷	۲۰.۳	نسبت کودکان یا بیماری‌های تنفسی در انگلیس (نمودار سمت چپ) و نرخ مرگ و میر بر اثر سرطان ریه در ۲۴ بخش شهر لندن (نمودار سمت راست).
۵۸	۳۰.۳	محل تخلفات رانندگی در یک مسابقه رالی
۵۹	۴۰.۳	تابع همبستگی مترن برای مقادیر مختلف k_i و فواصل: $\lambda = 0.5$ (نمودار بالا)، $\lambda = 1$ (نمودار مرکز) و $\lambda = 2$ (نمودار پایین).
۶۲	۵۰.۳	مکان ۲۰۰ نقطه فضایی داده‌های SPDEtoy.

۶۳	۶.۳	سه مثلث‌بندی برای داده‌های SPDEtoy با نقاط مختلف $max.edge : mesh0$ (نمودار بالا سمت چپ)، $mesh1$ (نمودار بالا سمت راست) و $mesh2$ (نمودار پایین). نقاط مکان‌های مشاهدات هستند.
۷۰	۷.۳	پیش‌گویی میانگین پسین (نمودار سمت چپ) و انحراف استاندارد پسین (نمودار سمت راست) برای میدان پنهان فضایی (نمودارهای بالا) و متغیر پاسخ (نمودارهای پایین) در سطح توری.
۷۳	۱.۴	مقادیر شبیه‌سازی شده و برآورد شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) برای مشاهدات (سمت چپ) و حالت‌ها (سمت راست) در مثال ساده.
۷۴	۲.۴	مقادیر شبیه‌سازی شده و برآورد شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) برای مشاهدات (سمت چپ) و حالت‌ها (سمت راست) در مثال ساده.
۷۵	۳.۴	مقادیر شبیه‌سازی شده و برآورد شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) برای مشاهدات و ضرایب رگرسیون $\beta_{0t}, \beta_{1t}, \beta_{2t}$.
۷۶	۴.۴	مقادیر شبیه‌سازی شده و برآورد شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) برای مشاهدات و حالت‌های a_t و b_t در مدل پویای فصلی. خط‌های قرمز برای ۱۰ ماه پیش‌بینی شده است.
۷۸	۵.۴	نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای مشاهدات y_t در زمان‌های ۱۸، ۵۱ و ۷۷ برای مثال ۴.
۷۹	۶.۴	نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای حالت x_t در زمان‌های ۱۸، ۵۱ و ۷۷ برای مثال ۴.
۸۰	۷.۴	نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای حالت z_t در زمان‌های ۱۸، ۵۱ و ۷۷ برای مثال ۴.
۸۱	۸.۴	نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای مشاهدات y_t در زمان‌های ۷، ۱۵ و ۲۹ برای مثال ۵.
۸۲	۹.۴	نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای بردار حالت x_1 در زمان‌های ۷، ۱۵ و ۲۹ برای مثال ۵.
۸۳	۱۰.۴	نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای بردار حالت x_2 در زمان‌های ۷، ۱۵ و ۲۹ برای مثال ۵.
۸۴	۱۱.۴	چگالی‌های حاشیه‌ای پسین برای پارامترهای دقت که به وسیله دو روش INLA (چپ) و MCMC (راست) به‌دست آمده‌اند.
۸۵	۱۲.۴	پیش‌بینی مصرف گاز کشور انگلستان برای ۱۲ دوره‌ی بعدی که با استفاده از روش INLA (نمودار بالا) و بسته $d1m$ (نمودار پایین) به‌دست آمده.
۹۱	۱۳.۴	نقشه استان گلستان (نمودار سمت چپ) و مثلث‌بندی این استان (نمودار سمت راست) با ۱۶۶ راس. نقاط قرمز ایستگاه‌ها، نقاط سبز ۱۶ گوشه ناحیه و خط سیاه مرزبندی ناحیه هستند.

- ۱۴۰۴ پیش‌گویی میانگین پسین (نمودار سمت چپ) و انحراف استاندارد پسین (نمودار
سمت راست) برای لگاریتم متغیر پاسخ (EC) در زمان‌های اول، دهم و بیستم. ۹۵
- ۱۵۰۴ نقشه‌های پهنه‌بندی برآوردشده از نمونه‌های زنجیر مارکوف MCMC در قالب میانه
(ستون سمت راست) و انحراف معیار (ستون سمت چپ) در ۳ زمان مختلف: یکم،
دهم و بیست و یکم. ۹۶

لیست جداول

۲۶	Gamma(a, b) توزیع چگالی تابع برای $\int_{\alpha}^{\beta} f(x)dx$ انتگرال دقیق تقریبی و	۱۰.۱
۴۰ نقاط طرح هامارد والش برای $k = 3$	۱.۲
۴۱ نقاط طرح هامارد والش برای $k = 3$	۲.۲
۴۱ تعداد اجرا در طرح عاملی کسری	۳.۲
	برآوردهای پسین (میانگین، انحراف معیار و چندکها) بردار ضریب β متغیرهای	۱.۴
۹۳ تعیینی.	
۹۴ برآوردهای پسین (میانگین، انحراف معیار و چندکها) برای $\rho, \sigma_{\omega}^2, \sigma_{\epsilon}^2$ و a .	۲.۴

فصل ۱

استنباط بیزی در مدل‌های پویا

۱.۱ مقدمه

موقعیت‌های کاربردی مختلفی در علوم هستند که داده‌های حاصل از مطالعه آن‌ها وابسته به زمان هستند. به عنوان چند نمونه می‌توان به داده‌های حاصل از مطالعه هواشناسی، در طول زمان، مانند فشار جو، بارش باران؛ داده‌های اقتصادی مانند قیمت سهام، نرخ ارز، میزان صادرات نفتی؛ و داده‌های پزشکی مانند تعداد مراجعان به یک کلینیک پزشکی، و نرخ مرگ و میر ناشی از یک بیماری اشاره کرد. به این نوع داده‌ها در آمار، داده‌های سری زمانی^۱ گفته می‌شود.

اگر داده‌های سری زمانی به‌طور پیوسته جمع‌آوری شوند، مثل ثبت سیگنال‌های یک دستگاه الکترونیکی مانند رادار، آن را یک سری زمانی پیوسته و اگر در فاصله‌های زمانی معین مشاهده شوند، مثل اندازه‌گیری فشار هوا در هر یک ساعت، نتیجه را یک سری زمانی گسسته می‌نامند. سری‌های زمانی گسسته نیز در دو گروه دسته‌بندی می‌شوند: (۱) مشاهدات با فاصله‌های زمانی برابر و (۲) مشاهدات با فاصله‌های زمانی نابرابر. اگر T مشاهده سری زمانی گسسته در فاصله‌های زمانی برابر به دست آیند، آن‌گاه مجموعه را می‌توان به صورت $\{y_t; t = 1, \dots, T\}$ نمایش داد. مشاهدات یک سری زمانی گسسته با فاصله‌های زمانی نابرابر را به صورت $\{y_{t_i}; t = 1, \dots, T\}$ نمایش می‌دهند. در واقعیت، معمولاً، با یک سری زمانی گسسته با فاصله‌های زمانی برابر مواجه می‌شویم.

یک سری زمانی می‌تواند یک یا چندمتغیره باشد. یعنی، برای هر نقطه زمانی دلخواه t ، y_t می‌تواند کمیتی عددی، مانند میزان کل بارندگی در روز t ، یا برداری k بعدی باشد که دارای k کمیت عددی است که هم‌زمان مشاهده شده‌اند. برای مثال اگر میزان کل بارندگی و میانگین دما در زمان t هم‌زمان اندازه‌گیری شده باشند، $k = 2$ کمیت عددی و برداری دو بعدی از مشاهدات داریم.

داده‌های سری زمانی گسترده‌ی وسیعی دارند و از ماهیتی پویا برخوردار هستند. بنابراین برای مدل‌بندی این نوع داده‌ها باید از رده مدل‌هایی استفاده کرد که ویژگی پویا بودن (وابستگی به زمان) داده‌ها را در

^۱Time series data

نظر بگیرند. از چنین رده‌ای در متون آماری به مدل‌های پویا^۲ یاد می‌شود. این مسئله بسیار مهم است که ویژگی‌های یک سری زمانی را به کمک ابزار گرافیکی داده‌ها، برای مدل‌بندی و تحلیل بهتر آن‌ها، استخراج کنیم. یک نمودار سری زمانی، اغلب، الگوهایی را مانند حرکتی رو به بالا یا رو به پایین (روند^۳)، یا تغییراتی دوره‌ای^۴ (فصلی^۵) نشان می‌دهد. شناخت رفتار چنین الگوهایی برای رسیدن به اهدافی مانند پیش‌بینی مقادیر آینده سری زمانی، ضروری است. نمودارهای پراکنش چهار سری زمانی مختلف در شکل ۱.۱ نشان داده شده‌اند. شکل ۱.۱ (الف)، سود ماهیانه حاصل از ۱۲۵ اوراق قرضه آمریکا از ژانویه ۱۹۹۴ تا می ۲۰۰۴ را نشان می‌دهد. این سری زمانی مثالی است برای روند رو به پایین؛ اما الگوی فصلی در آن مشاهده نمی‌شود. شکل ۱.۱ (ب)، حجم تولید برق خالص آمریکا (برحسب میلیارد کیلو وات بر ساعت) را در طول ۵۵ سال از ۱۹۴۹ تا ۲۰۰۳ را نشان می‌دهد. این سری زمانی روندی افزایشی نسبت به زمان دارد. شکل ۱.۱ (ج)، تعداد فصلی خودرو تولیدشده در انگلستان (برحسب هزار خودرو) برای ۱۱۳ نقطه زمانی از فصل اول سال ۱۹۷۷ تا فصل اول ۲۰۰۵ را نشان می‌دهد. برای این سری زمانی تغییر ثابتی حول یک سطح تغییرپذیر وجود دارد و روندی فصلی در آن نیز مشهود است. شکل ۱.۱ (د)، تعداد ماهیانه بازدیدکنندگان از کشور استرالیا را در ۲۴۰ ماه از می ۱۹۸۵ تا آوریل ۲۰۰۵ نشان می‌دهد (هایندمن و همکاران، ۲۰۰۸). یک روند افزایشی فصلی در این سری زمانی مقابل مشاهده است.

با این مقدمه، در این فصل به معرفی رده بزرگی از مدل‌های پویا برای مدل‌بندی داده‌های سری زمانی می‌پردازیم. قبل از این کار، ابتدا برخی از مفاهیم مورد نیاز را ارایه می‌کنیم. دیدگاه استنباطی مورد انتخاب ما در این پایان‌نامه، مبتنی بر مکتب بیزی است. به همین دلیل، به روش‌های استنباطی کلاسیک نخواهیم پرداخت.

۲.۱ تعاریف و مفاهیم مقدماتی

در ساخت و توسعه مدل‌ها و روش‌های آماری، پذیره استقلال بین مشاهدات، یک پذیره مهم و زیربنایی محسوب می‌شود. ماهیت وابسته به زمان بودن داده‌های سری زمانی باعث شده است نوع وابستگی (زمانی) بین مشاهدات سری زمانی القاء و نیاز به توسعه مدل‌های مناسب برای تحلیل آن‌ها احساس شود. آماردانان برای مدل‌بندی داده‌های سری زمانی از نظریه فرآیندهای تصادفی^۶ استفاده می‌کنند. فرآیندهای تصادفی یکی از شاخه‌های کاربردی نظریه آمار و احتمال است که در سایر علوم کاربردهای بسیاری دارند. فرآیندهای تصادفی در ابتدا در علم فیزیک و برای توصیف پدیده‌های تصادفی که حالت آن‌ها در طی زمان تغییر می‌کند، مطرح شدند. در مدل‌سازی هر سیستم تصادفی که حالت آن در طی زمان (فضا یا سایر شاخص‌ها) تغییر می‌کند، مدل باید قادر به توصیف حالت سیستم در طول زمان باشد.

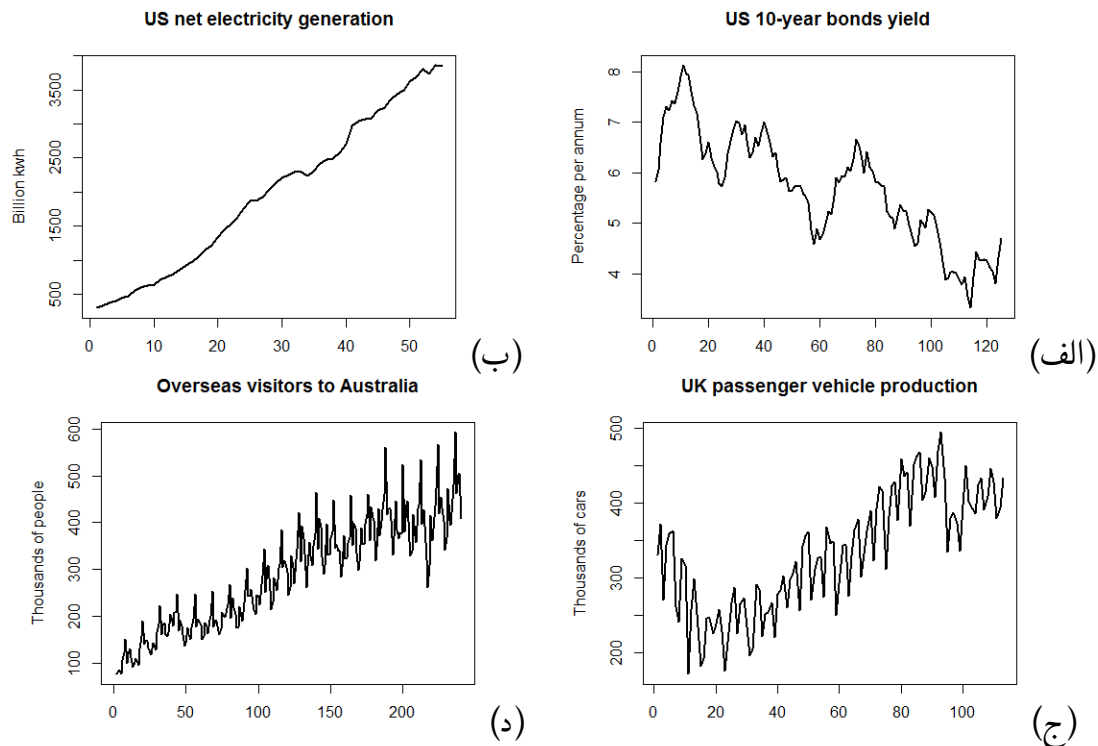
^۲Dynamic models

^۳Trend

^۴Circular

^۵Seasonal

^۶Stochastic process



شکل ۱.۱: چهار سری زمانی با الگوهای مختلف از داده‌های اقتصادی: الف) سود اوراق قرضه آمریکا، ب) حجم تولید برق خالص آمریکا، ج) تخودرو تولید شده انگلستان و د) بازدیدکنندگان از کشور استرالیا.

به عبارت دیگر مدل شامل دنباله‌ای از متغیرهای تصادفی است که پدیده تصادفی را توصیف می‌کنند. مجموعه‌ای از متغیرهای تصادفی با اندیس مرحله یا زمان که وضعیت یک پدیده یا آزمایش تصادفی را در طول یک دوره نمایش می‌دهند، تعریف یک فرآیند تصادفی است.

تعریف ۱.۲.۱ (فرآیند تصادفی). فرض کنید t اندیس زمان و $X(t)$ متغیر تصادفی متناظر با t باشد، در این صورت یک فرآیند تصادفی مجموعه‌ای از متغیرهای تصادفی $\{X(t); t \in \mathcal{I}\}$ است که در آن $X(t)$ به ازای هر $t \in \mathcal{I}$ یک متغیر تصادفی است. متغیرهای تصادفی و اندیس آن‌ها می‌توانند از انواع گسسته و پیوسته و همچنین چند بعدی باشند. مثلاً بررسی وضعیت آب و هوا، تعداد افراد و وضعیت یک بازی در مکان‌ها، زمان‌ها یا مراحل مختلف، نمونه‌هایی از یک فرآیند تصادفی هستند.

تعریف ۲.۲.۱ (فرآیند سری زمانی). یک فرآیند سری زمانی، فرآیندی تصادفی یا مجموعه‌ای از متغیرهای تصادفی y_t در زمان‌های $t \in \mathcal{I}$ است. توجه کنید که y_t در این پایان‌نامه یک متغیر تصادفی یا تحقیقی واقعی از فرآیند سری زمانی در زمان t می‌باشد. بنابراین $\{y_t, t \in \mathcal{I}\}$ یا به‌طور ساده‌تر $\{y_t\}$ یک فرآیند سری زمانی را مشخص می‌کند. اگر \mathcal{I} به‌صورت $\{t_i, i \in \mathcal{N}\}$ باشد، آن‌گاه این فرآیند یک فرآیند تصادفی گسسته تحت زمان است و اگر \mathcal{I} یک فاصله یا مجموعه‌ای از فواصل باشد، آن‌گاه این فرآیند یک فرآیند تصادفی پیوسته تحت زمان است. در این چارچوب، مجموعه داده‌های سری زمانی $\{y_t; t = 1, \dots, T\}$ ، که به‌صورت $y_{1:T}$ نیز مشخص می‌شود، مجموعه‌ای از T تحقق یک فرآیند سری زمانی در فاصله‌های برابر است.

برای استقلال یک تعریف یکتا وجود دارد، اما برای وابستگی نمی‌توان تعریف واحد و مشخصی ارائه کرد. آماردانان برای مدل‌بندی داده‌های وابسته، انواع وابستگی را، برحسب موقعیت تحلیل داده‌ها، تعریف کرده‌اند. یک تعریف از وابستگی که مدل‌بندی داده‌های سری زمانی را ساده‌تر می‌کند، ایستایی^۷ (مانایی) است که خود شامل انواع متفاوتی است.

تعریف ۳.۲.۱ (فرآیند سری زمانی ایستا). یک فرآیند سری زمانی، ایستا است اگر در زیرفاصله‌های جدا و مشخصی با طول یکسان، رفتار وابستگی سری ثابت بماند. اکثر مدل‌های سری زمانی بر مبنای پذیره ایستایی کار می‌کنند.

در ادامه به تعریف دو نوع ایستایی کامل یا قوی^۸ و ضعیف یا مرتبه‌ی دوم^۹، می‌پردازیم.

تعریف ۴.۲.۱ (ایستایی کامل یا قوی). یک فرآیند سری زمانی $\{y_t; t \in \mathcal{I}\}$ ایستای کامل یا قوی است اگر برای هر دنباله‌ی زمانی t_1, \dots, t_n و هر تأخیر h ، توزیع احتمالی بردار $(y_{t_1}, \dots, y_{t_n})$ با توزیع احتمالی $(y_{t_1+h}, \dots, y_{t_n+h})$ برابر باشد.

در عمل بررسی ایستای کامل یک فرآیند تصادفی بسیار مشکل است. بنابراین از نسخه‌های ضعیف‌تری مانند ایستای ضعیف یا مرتبه‌ی دوم استفاده می‌شود.

تعریف ۵.۲.۱ (ایستایی ضعیف یا مرتبه دوم). فرآیند $\{y_t; t \in \mathcal{I}\}$ ایستای ضعیف یا مرتبه دوم است، اگر

$$E(y_t) = \mu, \quad \text{Var}(y_t) = \nu, \quad \text{Cov}(y_t, y_s) = \gamma(t - s)$$

که در آن μ و ν ثابت و مستقل از t هستند و $\gamma(t - s)$ نیز مستقل از t و s بوده و تنها به طول فاصله زمانی وابسته است.

لازم به ذکر است که ایستایی قوی، ایستایی ضعیف را نتیجه می‌دهد اما عکس آن، به‌جز موارد خاص مثل سری زمانی با توزیع نرمال، برقرار نیست. تعریف بعدی یکی از فرآیندهای تصادفی کلیدی برای ساخت و توسعه مدل‌های پویا را معرفی می‌کند.

تعریف ۶.۲.۱ (فرآیند قدم زدن تصادفی). الف) فرآیند تصادفی $\{\omega_t\}$ یک فرآیند قدم زدن تصادفی مرتبه اول^{۱۰} است، هرگاه

$$\omega_t = \omega_{t-1} + \epsilon_t$$

که در آن ϵ_t مؤلفه اغتشاش محض است؛ یعنی ϵ_t ‌ها ناهمبسته هستند و از توزیعی با میانگین صفر و واریانس ثابت σ_ϵ^2 پیروی می‌کنند.

ب) فرآیند تصادفی $\{\omega_t\}$ را یک فرآیند قدم زدن تصادفی مرتبه دوم^{۱۱} گویند، هرگاه

$$\omega_t = 2\omega_{t-1} - \omega_{t-2} + \epsilon_t$$

^۷Stationarity

^۸Strong stationarity

^۹Second order stationarity

^{۱۰}First order random walk process

^{۱۱}Second order random walk process

به سادگی می توان نشان داد که این فرآیند، ایستا نیست و رفتاری به شدت ناپایدار دارد. فرآیند قدم زدن تصادفی مرتبه اول را با نماد $rw1$ و فرآیند قدم زدن تصادفی مرتبه دوم را با نماد $rw2$ نمایش می دهند. یکی دیگر از فرآیندهای سری زمانی پرمصرف در مدل بندی داده های وابسته به زمان، فرآیند خودبرگشت^{۱۲} (اتورگرسیو) است که در زیر تعریف شده است.

تعریف ۷.۲.۱ (مدل اتورگرسیو). در یک فرآیند اتورگرسیو مرتبه p ام، $AR(p)$ ، متغیر تصادفی در زمان t به طور خطی به مشاهدات قبلی Y_{t-1}, \dots, Y_{t-p} وابسته است. یعنی

$$Y_t = a_0 + a_1 Y_{t-1} + \dots + a_p Y_{t-p} + \epsilon_t, \quad t = 1, \dots, T$$

که در آن $\{\epsilon_t\}$ فرآیند اغتشاش محض با توزیع $N(0, \sigma^2)$ است.

در این پایان نامه توسعه مدل های پویای نرمال به مدل های پویای غیرنرمال، بر اساس مدل های خطی تعمیم یافته^{۱۳} (GLMs) صورت می گیرد که در زیر بخش بعدی، به اختصار، معرفی می شوند.

۱.۲.۱ مدل های خطی تعمیم یافته

مدل رگرسیون خطی

$$y = \mathbf{x}^T \boldsymbol{\beta} + \epsilon = \mu + \epsilon$$

را در نظر بگیرید، که در آن $\mathbf{x} = (x_1, \dots, x_k)^T$ بردار متغیرهای تبیینی، $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^T$ بردار پارامترهای رگرسیونی، و ϵ جمله خطا است که برای n مشاهده، اعضای آن ناهمبسته هستند و معمولاً فرض می شود دارای توزیع نرمال با میانگین صفر و واریانس σ_ϵ^2 می باشند. همچنین μ میانگین شرطی پاسخ به شرط متغیرهای تبیینی \mathbf{x} است.

زمانی مجاز به استفاده از مدل خطی هستیم که همه پذیره های آن برقرار باشند. یکی از این پذیره ها، توزیع نرمال برای متغیر پاسخ است. در موقعیت های مختلفی مانند پاسخ های گسسته، به وضوح، متغیر پاسخ غیرنرمال است و باید از تعمیمی از مدل خطی استفاده کرد. مدل های خطی تعمیم یافته (نلدن و ودربرن، ۱۹۷۲) رده ای از مدل ها هستند که برای مدل بندی پاسخ های غیرنرمالی که دارای توزیعی از خانواده توزیع های نمایی (یا شبیه به آن) هستند، معرفی شدند. میانگین پاسخ در GLM به صورت مستقیم مدل بندی نمی شود، بلکه تبدیلی از میانگین پاسخ با استفاده از تابع پیوند^{۱۴} به متغیرهای تبیینی مرتبط می شود. مدل های خطی تعمیم یافته دارای سه مؤلفه هستند:

۱. **مؤلفه تصادفی**: توزیع شرطی متغیر پاسخ را مشخص می کند. توزیع پاسخ، معمولاً، از خانواده توزیع های نمایی مانند نرمال، دو جمله ای، و پواسن انتخاب می شود. انتخاب مؤلفه تصادفی بنا به ساختار متغیر پاسخ، متفاوت است. برای مثال، اگر ساختار متغیر پاسخ به صورت متغیر شمارشی غیرمنفی باشد، می توان توزیع پواسن را به عنوان توزیع مؤلفه تصادفی انتخاب کرد.

^{۱۲}Autoregressive

^{۱۳}Generalized linear models

^{۱۴}Link function

۲. مؤلفه منظم: این مؤلفه نقش متغیرهای تبیینی را در قالب یک ترکیب خطی از متغیرهای تبیینی، η ، با نام پیش‌گوی خطی^{۱۵} مشخص می‌کند. در واقع پیش‌گوی خطی به صورت

$$\eta = \mathbf{x}^T \boldsymbol{\beta}$$

تعریف می‌شود.

۳. تابع پیوند: تابع معکوس‌پذیر پیوند $g(\cdot)$ نشان می‌دهد که μ چگونه با متغیرهای تبیینی، در پیش‌گوی خطی، ارتباط دارد. به عبارت دیگر، داریم

$$g(\mu) = \eta$$

یا به طور معادل

$$g^{-1}(\eta) = \mu.$$

تابع پیوند انواع مختلفی دارد، که می‌توان به تابع پیوند لگاریتمی $g(\mu) = \log(\mu)$ ، و تابع پیوند لجیت^{۱۶} $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ اشاره کرد. در این مدل فرض می‌شود که پاسخ‌های Y_1, \dots, Y_n از هم مستقل هستند.

۲.۲.۱ رهیافت بیزی استنباط آماری

روش‌های آماری را در یک دسته‌بندی کلی می‌توان به دو دسته بسامدی^{۱۷} و بیزی تقسیم کرد. هر کدام از این چارچوب‌ها منطق و هدف خاص خود را در استنباط آماری دنبال می‌کنند (رابرت، ۲۰۰۷). در دیدگاه بسامدی استنباط آماری، پارامترهای مدل ثابت ولی نامعلوم فرض می‌شوند و بر اساس روش‌هایی با استفاده از داده‌ها به دنبال شناخت آن‌ها هستیم. در حیطه مدل‌بندی، یکی از شاخه‌های پرطرفدار استنباط‌های بسامدی مبتنی بر درست‌نمایی است.

فرض کنید $y_{1:T}$ یک تحقق T بعدی از فرآیند سری زمانی یک متغیره $\{y_t; t \geq 1\}$ باشد به طوری که از یک توزیع احتمالی پارامتری با تابع (چگالی) احتمال توام $p(y_{1:T}|\boldsymbol{\psi})$ پیروی می‌کند و $\boldsymbol{\psi}$ بردار پارامترهای مدل است. تابع $p(y_{1:T}|\boldsymbol{\psi})$ به عنوان تابعی از پارامتر $\boldsymbol{\psi}$ ، تابع درست‌نمایی مدل سری زمانی است. روش‌های معمول استنباط بسامدی، روش‌های مبتنی بر درست‌نمایی هستند که برای محاسبه تابع درست‌نمایی در فرآیندهای سری زمانی، باید یک تابع توام T متغیره را محاسبه کنیم. اگر مشاهدات از هم مستقل باشند، آنگاه

$$p(y_{1:T}|\boldsymbol{\psi}) = \prod_{t=1}^T p(y_t|\boldsymbol{\psi})$$

که محاسبه آن کار ساده‌ای است. اما مشاهدات سری زمانی مستقل نیستند و کنار آمدن با توابع T متغیره کار راحتی نیست.

^{۱۵}Linear predictor

^{۱۶}Logit Link

^{۱۷}Frequency

در مقابل دیدگاه مبتنی بر درست‌نمایی، دیدگاه بیزی فرض می‌کند پارامترهای مدل تحقق‌هایی از توزیع‌های تصادفی هستند یا به عبارتی فرض می‌شود پارامترها خود متغیرهای تصادفی هستند. این دیدگاه، با هدف کمی کردن عدم قطعیت^{۱۸} در مورد پارامتر، یک توزیع احتمالی معروف به توزیع پیشین^{۱۹} را که برگرفته از اطلاعات و تجربیات پیشین محقق است، برای پارامتر مورد بررسی لحاظ می‌کند. دیدگاه بیزی با درآمیختن اطلاعات قبلی محقق در مورد پارامتر نامعلوم (توزیع پیشین $p(\psi)$) و اطلاعات موجود در نمونه (تابع درست‌نمایی) به استنباط در مورد پارامتر می‌پردازد. این ترکیب اطلاعات با محاسبه تابع (چگالی) احتمال توزیع پسین^{۲۰}، $p(\psi|y_{1:T})$ ، به کمک قضیه بیز انجام می‌شود. یعنی

$$p(\psi|y_{1:T}) = \frac{p(y_{1:T}|\psi)p(\psi)}{p(y_{1:T})} \propto c(y_{1:T})p(y_{1:T}|\psi)p(\psi) \quad (1.1)$$

که در آن $p(y_{1:T}) = \int p(y_{1:T}|\psi)p(\psi)d\psi$ و $c(y_{1:T}) = \frac{1}{p(y_{1:T})}$ ثابت نرمال‌ساز^{۲۱} نامیده می‌شود و در حالت گسسته، انتگرال به مجموع تبدیل می‌شود. در دیدگاه بیزی تمام استنباط‌ها مبتنی بر توزیع پسین انجام می‌شوند.

بدون افزودن پذیره‌هایی بر ساختار وابستگی فرآیند سری زمانی، محاسبه توزیع پسین بر حسب تابع درست‌نمایی T متغیره کار پیچیده‌ای است. برای مثال، فرض کنید هر مشاهده در زمان t تنها به ψ و مشاهده زمان $t-1$ وابسته است. در این صورت

$$p(\psi|y_{1:T}) \propto p(\psi)p(y_1|\psi) \prod_{t=2}^T p(y_t|y_{t-1}, \psi). \quad (2.1)$$

توزیع پسین در (۲.۱) خیلی ساده‌تر از (۱.۱) محاسبه می‌شود. مثال زیر این سادگی را تشریح می‌کند.

مثال ۸.۲.۱ (مدل $AR(1)$). فرآیند $AR(1)$ را در نظر بگیرید

$$y_t = \phi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \nu).$$

پارامترهای این مدل $\psi = (\phi, \nu)'$ می‌باشند. حال برای هر زمان $t > 1$ تابع درست‌نمایی شرطی

$$p(y_t|y_{t-1}, \psi) = N(y_t|\phi y_{t-1}, \nu)$$

است. به علاوه، به سادگی می‌توان نشان داد اگر فرآیند ایستا باشد، آنگاه $y_1 \sim N(0, \frac{\nu}{1-\phi^2})$ و بنابراین تابع درست‌نمایی به صورت

$$p(y_{1:T}|\psi) = \frac{(1-\phi^2)^{1/2}}{(2\pi\nu)^{T/2}} \exp \left\{ -\frac{Q^*(\phi)}{2\nu} \right\}$$

نتیجه می‌شود، که در آن

$$Q^*(\phi) = y_1^2(1-\phi^2) + \sum_{t=2}^T (y_t - \phi y_{t-1})^2.$$

^{۱۸}Uncertainty

^{۱۹}Prior distribution

^{۲۰}Posterior

^{۲۱}Normalizing constant

بنابراین، تابع چگالی احتمال پسین به صورت

$$p(\psi|y_{1:T}) \propto p(\psi) \frac{(1 - \phi^2)^{1/2}}{(2\pi\nu)^{T/2}} \exp\left\{-\frac{Q^*(\phi)}{2\nu}\right\}$$

حاصل می‌شود.

توزیع پیش‌گوی بیزی^{۲۲}

در دیدگاه بیزی برای پیش‌بینی مقدار جدید سری در نقطه زمانی آینده، $t' > T$ ، از توزیع پیش‌گوی بیزی استفاده می‌شود که در زیر تعریف شده است.

تعریف ۹.۲.۱ (توزیع پیش‌گوی بیزی). فرض کنید $\{Y_t; t \geq 1\}$ یک فرآیند سری زمانی از توزیعی با تابع (چگالی) احتمال $p(y|\psi)$ باشد، که در آن ψ پارامترهای مدل و دارای توزیع پیشین $p(\psi)$ هستند. در این صورت توزیع پسین ψ بر حسب مشاهدات $y_{1:T}$ ، متناسب است با

$$p(\psi|y_{1:T}) \propto p(y_{1:T}|\psi)p(\psi).$$

اگر توزیع مشاهده جدید $y_{t'}$ به صورت $p(y_{t'}|y_{1:T}, \psi)$ باشد، آنگاه توزیع پیش‌گو با میانگین‌گیری از این توزیع نسبت به توزیع پسین $p(\psi|y_{1:T})$ به صورت

$$p(y_{t'}|y_{1:T}) = \int p(y_{t'}|y_{1:T}, \psi)p(\psi|y_{1:T})d\psi$$

حاصل می‌شود که توزیع پیش‌گوی بیزی نامیده می‌شود.

۳.۱ مدل‌های آماری برای داده‌های سری زمانی

برای تحلیل داده‌های سری زمانی، مدل‌های مختلفی پیشنهاد شده‌اند که هر کدام صاحب مزایا و معایبی هستند. توسعه و بهبود این مدل‌ها و روش‌های استنباطی مورد استفاده از آن‌ها هنوز ادامه دارد و یکی از شاخه‌های داغ و پرطرفدار حوزه‌های آمار محسوب می‌شود.

مدل پایه برای تحلیل یک سری زمانی، یک مدل جمعی به صورت

$$y_t = \mu_t + \gamma_t + \epsilon_t, \quad t = 1, \dots, T \quad (3.1)$$

است، که در آن μ_t مؤلفه روند، γ_t مؤلفه دوره‌ای یا فصلی، و ϵ_t مؤلفه خطا یا اغتشاش نام دارند. به طور کلی، مشاهدات y_t و سایر مؤلفه‌ها در (۳.۱) می‌توانند برداری باشند، اما در این پایان‌نامه فرض می‌کنیم همه مؤلفه‌ها عددی هستند. در خیلی از کاربردها، به ویژه در اقتصاد، از مدل ضربی

$$y_t = \mu_t \gamma_t \epsilon_t, \quad t = 1, \dots, T$$

استفاده می‌شود. البته با لگاریتم‌گیری از این مدل، همان مدل (۳.۱) نتیجه می‌شود. مدل‌های سری زمانی مختلف ساخته شده بر اساس مدل پایه، در نحوه تدوین مؤلفه‌های روند، و روند فصلی با یکدیگر متفاوت عمل می‌کنند.

^{۲۲}Bayesian predictive distribution

یک رده بزرگ از مدل‌های سری زمانی، مبتنی بر نظریه باکس-جنکینز^{۲۳} (باکس و همکاران، ۱۹۹۴) ساخته شده‌اند که به مدل‌های خودبرگشت میانگین متحرک تجمیع‌شده^{۲۴} (ARIMA) معروف هستند. فرآیند سری زمانی اتورگرسیو یکی از زیررده‌های مدل‌های ARIMA است. برای مشاهده جزئیات مدل‌های ARIMA به باکس و همکاران (۱۹۹۴) مراجعه کنید. با این‌که مدل‌های ARIMA هنوز استفاده می‌شوند، اما از معایب جدی برخوردار هستند که به برخی از آن‌ها اشاره می‌کنیم:

۱. این مدل‌ها خطی هستند و در مواردی که ماهیت تغییرپذیری داده‌ها، بر حسب مشاهدات زمان‌های گذشته، غیرخطی باشد، استفاده از آن‌ها کارا نیست.

۲. یکی از شرایط استفاده از این مدل‌ها، برقراری پذیره توزیعی نرمال برای مشاهدات سری است. در حالی‌که در موارد متعددی، قبول این پذیره توزیعی چندان معقول نیست، به‌عنوان مثال فرض کنید سری زمانی، مربوط به یک پدیده شمارشی مانند تعداد بازدیدکنندگان کشور استرالیا باشد. در این موارد پذیرفتن توزیعی مثل پواسن مناسب است اما نرمال کاملاً غیرمعقول است.

۳. یکی دیگر از شرایط استفاده از مدل‌های ARIMA ایستا بودن سری زمانی است. بنابراین برای تحلیل یک سری زمانی نایستا با مدل‌های ARIMA، ابتدا باید مراحل (تبدیلاتی) را برای ایستا کردن سری اعمال کرد. در حالی‌که نتایج تحلیل مستقیم (بدون تبدیل) داده‌ها، در صورت امکان، تفسیرپذیری و کارایی بهتری خواهند داشت.

۴. امکان دخالت دادن سایر سری‌های زمانی در قالب جملات رگرسیونی در این مدل‌ها وجود ندارد. در حالی‌که، به دلیل محدود بودن اطلاعات موجود در داده‌های سری زمانی مورد نظر، ورود اطلاعات کمکی می‌تواند کارایی مدل و استفاده از آن برای اهدافی مانند پیش‌بینی مقادیر آینده سری را به‌طور فزاینده‌ای بهبود بخشد. البته نسخه‌های چندمتغیره این مدل‌ها وجود دارند، اما برآزش و تحلیل آن‌ها پیچیدگی‌های خود را داراست.

هر چهار عیب ذکرشده، در عمل، بسیار جدی و مورد توجه هستند. بنابراین برای مرتفع ساختن آن‌ها سایر رده‌های کارا از مدل‌های آماری معرفی شدند. یکی از رده‌های بزرگ و منعطف مدل‌های پویا، رده مدل‌های فضای حالت^{۲۵} است که مبتنی بر مدل‌های رگرسیونی پویا^{۲۶} ساخته می‌شوند. انعطاف مدل‌های فضای حالت بسیار زیاد است و شامل مدل‌های خطی و غیرخطی و همچنین با پذیره توزیعی نرمال و غیرنرمال می‌شوند. در بخش بعدی این مدل‌ها را معرفی می‌کنیم.

^{۲۳}Box-Jenkins theory

^{۲۴}Autoregressive Integrated Moving Average

^{۲۵}State space models

^{۲۶}Dynamic regression models

۴.۱ مدل‌های فضای حالت

مدل‌های خطی پویا، در مهندسی در دهه ۱۹۶۰ به منظور کنترل سیستم‌های پویا توسعه یافتند. البته استفاده پراکنده از این مدل‌ها پیشینه قدیمی‌تری دارد (تیل، ۱۸۸۰). اولین کاربردهای ملموس این مدل‌ها در مسایل هوافضا و فضاپیمای آپولو^{۲۷} دیده شد (هاتچینسون، ۱۹۸۴). اما در دهه‌های اخیر مدل‌های خطی پویا^{۲۸} (DLMs) و گسترده‌تر از آن‌ها، مدل‌های فضای حالت کاربردهای فراوانی با دامنه وسیع از زیست‌شناسی تا اقتصاد، از مهندسی و کنترل کیفیت تا مطالعات زیست محیطی و از زمین‌شناسی تا ژنتیک، پیدا کرده‌اند. همچنین در سال‌های اخیر رشد قابل توجهی در کاربردهای مدل‌های فضای حالت برای تحلیل سری زمانی ایجاد شده‌اند. برای مثال، هاروی (۱۹۸۹)، وست و هریسون (۱۹۹۷)، دوربین و کوپمن (۲۰۰۱) و میگون و همکاران (۲۰۰۵)، و مراجع آن‌ها را مشاهده کنید. دلیل عمده پیدایش این رشد چشم‌گیر کاربردها، امکان حل مشکلات محاسباتی در چارچوب استنباط بیزی به کمک روش‌های مونت کارلو می‌باشد.

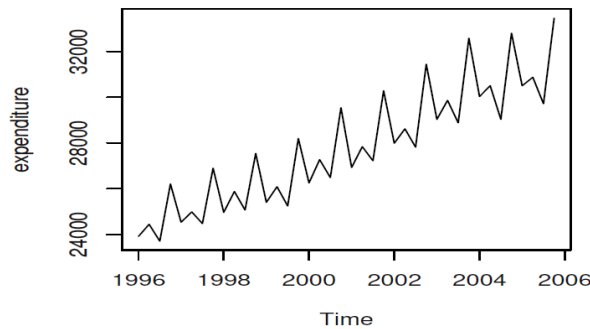
تحلیل سری‌های زمانی به وسیله مدل‌های خطی پویا در بین سال‌های ۱۹۷۰ تا ۱۹۸۰ بسیار گسترش یافت و امروزه مدل‌های فضای حالت بیشتر مورد توجه هستند. شاید در ابتدا انجام استنباط بر اساس مدل‌های فضای حالت کمی دشوار به نظر برسد، اما این رده از مدل‌ها، یک رهیافت قدرتمند برای انجام تحلیل نزدیک‌تر به واقعیت پیشنهاد می‌کند. مدل‌های فضای حالت برتری‌های کاملاً آشکاری در مقایسه با مدل‌های ARIMA دارند. به‌عنوان نمونه:

۱. انعطاف‌پذیری بالایی در تحلیل سری‌های زمانی ناپایستا یا مدل‌بندی تغییرات ساختاری دارند.
۲. از تفسیرپذیری راحت‌تری برخوردارند.
۳. مدل‌های غیرخطی را نیز شامل می‌شوند. همچنین از سایر پذیره‌های توزیعی (به‌جز نرمال) می‌توان استفاده کرد.
۴. به دلیل بهره‌مندی از چارچوب رگرسیونی، می‌توان سایر متغیرها (حتی سری‌های زمانی دیگر) را به‌عنوان اطلاعات کمکی، وارد مدل کرد.

برای تشریح این مزایا، ابتدا چند سری زمانی واقعی را مثال می‌زنیم. شکل ۲.۱ نمودار سری زمانی هزینه سبد غذایی خانوارهای ایتالیایی را در بین سال‌های ۱۹۹۶ تا ۲۰۰۵ نشان می‌دهد. این سری زمانی نسبتاً رفتاری منظم را طی زمان تکرار می‌کند و قابل پیش‌بینی است: حضور یک روند و یک مؤلفه فصلی مشهود است. برای تحلیل این نوع داده‌ها، می‌توان یک مدل سری زمانی ساده مثل یک ARIMA با یک روند و یک مؤلفه فصلی را در نظر گرفت. در حقیقت روش پایه‌ای تحلیل سری زمانی به مشاهده رفتار مناسب در پدیده مورد مطالعه متکی است؛ پیش‌بینی رفتار آینده بسیار ساده است اگر سری یک الگوی منظم را تحت زمان تکرار کند. شکل ۳.۱ میزان مصرف گاز در کشور انگلستان از سال ۱۹۶۰ تا ۱۹۸۶ را نشان می‌دهد که هر چهار ماه جمع‌آوری شده‌اند. تغییر سطح میانگین در مؤلفه

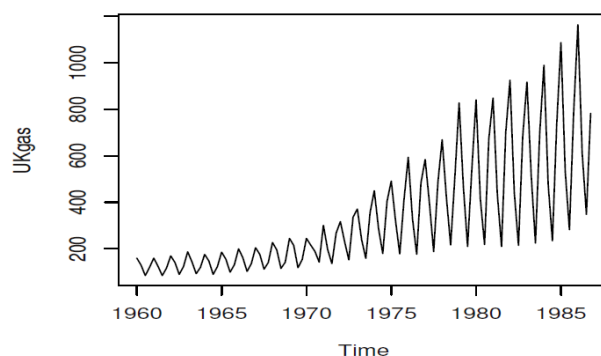
^{۲۷}Apollo spaceflight

^{۲۸}Dynamic Linear Models



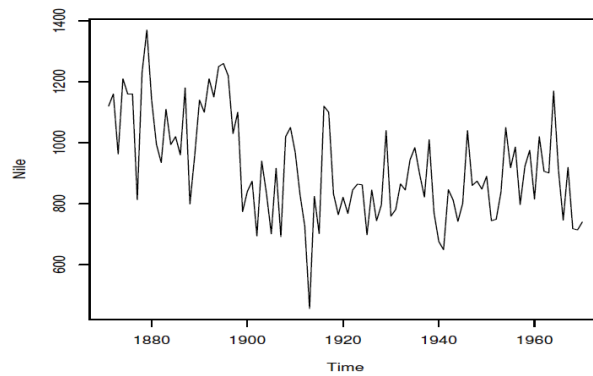
شکل ۲.۱: هزینه مصرفی غذایی خانواده از سال ۱۹۹۶ تا ۲۰۰۵ در ایتالیا.

فصلی در این سری زمانی کاملاً قابل مشاهده است. شکل ۴.۱ حجم جریان سالیانه رودخانه نیل در آشوان طی سال‌های ۱۸۷۱ تا ۱۹۷۰ را نشان می‌دهد. این سری نشان دهنده چندین سطوح انتقال در میانگین است. به این معنی که در بازه‌های متفاوتی از زمان مورد مطالعه، متوسط حجم جریان آب تغییر کرده است. در سال ۱۸۹۸ ساخت اولین سد در آشوان شروع شد و سد بزرگ دیگری در سال ۱۹۷۱ تکمیل شد. اگر این سدهای بزرگ را مد نظر قرار دهیم، می‌توانیم به راحتی تغییرات بزرگ ایجادشده را که این سدها در جریان رودخانه نیل و در محدوده اطراف آن ایجاد کرده‌اند، درک کنیم. پس برای این سری زمانی باید از مدلی با انعطاف‌پذیری بیشتر استفاده کرد که نیازی به یک الگوی منظم و پذیرش برقراری ایستایی سری ندارند. شکل ۵.۱ یک سری زمانی کاملاً نامنظم را از قیمت‌های روزانه شرکت گوگل از تاریخ ۲۰۰۴/۰۸/۱۹ تا ۲۰۰۶/۰۳/۳۱ نشان می‌دهد. واضح است که این سری ناپایدار است. برای تحلیل این سری زمانی ناپایستا با مدل‌های ARIMA، باید تبدیلاتی روی داده‌ها برای دستیابی به ایستایی انجام داد. اما می‌توان یک مدل فضای حالت را به‌کار برد که علاوه بر در نظر گرفتن بی‌ثباتی در سطح میانگین و واریانس، داده‌ها را مستقیم و بدون اعمال تبدیلی بر روی آن‌ها، تحلیل می‌کند.



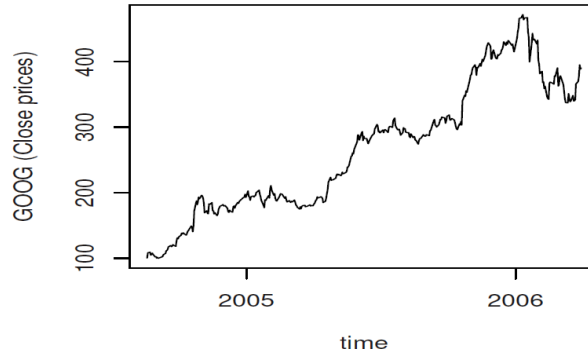
شکل ۳.۱: مصرف گاز کشور انگلستان از سال ۱۹۶۰ تا ۱۹۸۶ بر حسب میلیون.

زمانی که با سری‌های زمانی اقتصادی یا مالی یک متغیره سروکار داریم، اطلاعات موجود در آن‌ها اغلب محدود هستند و بررسی (پیش‌بینی) تغییرات ناگهانی در سری، بدون تلاش بیشتر در مطالعه عمیق‌تر سایر متغیرهایی که در بازار تأثیرگذار هستند، به‌سختی قابل تحلیل و مدل‌بندی خواهند بود. برای مثال یکی از متغیرهای مؤثر در اقتصاد کلان، سود حاصل از یک سیستم اقتصادی است و با توجه



شکل ۴.۱: میزان جریان سالانه رودخانه نیل در آشوان از سال ۱۸۷۰ تا ۱۹۷۱.

به وضعیت بازارهای جهانی امروز، به شدت ناپایدار (ناایستا) است (شکل ۵.۱ را ببینید). بنابراین، پیش‌بینی تغییرات ناگهانی ممکن در سود یک سیستم اقتصادی به هیچ وجه کار آسانی نیست. یک راه‌حل پیشنهادی برای برخورد با این مشکل، وارد کردن سایر متغیرهای مؤثر بر سود در نقش جملات رگرسیونی در مدل می‌باشد. این وظیفه توسط مدل‌های فضای حالت (به‌عنوان یک مدل رگرسیونی پویا) به‌خوبی اجرا می‌شود. مدل‌های فضای حالت را می‌توان برای مدل‌های سری زمانی یک و چندمتغیره به‌کار برد.



شکل ۵.۱: قیمت‌های روزانه گوگل.

افزون بر این اگر ماهیت سری زمانی گسسته (غیرنرمال) باشد، باید مدلی معرفی کرد که مبتنی بر توزیعی نرمال نباشد. مدل‌های فضای حالت برای حالتی که توزیع غیرنرمال و شکل رابطه غیرخطی باشد، نیز توسعه یافته است. این توسعه مبتنی بر مدل‌های خطی تعمیم‌یافته انجام گرفته است.

۱.۴.۱ معرفی مدل

سری زمانی $\{Y_t; t \geq 1\}$ را در نظر بگیرید. همان‌طور که قبلاً اشاره کردیم، با توجه به وابستگی به زمان Y_t ها، تعیین توزیع توأم (Y_1, \dots, Y_t) ، برای هر $t \geq 1$ ، کار آسانی نیست و برای فائق آمدن بر این

دشواری، اعمال شرایط اضافی بر ساختار وابستگی سری زمانی لازم است. وابستگی مارکوفی^{۲۹} یکی از ساده‌ترین صورت‌های بیان وابستگی میان Y_t هاست.

تعریف ۱۰۴.۱. فرآیند $\{Y_t; t \geq 1\}$ یک زنجیره مارکوف مرتبه اول است. اگر برای هر $t \geq 1$

$$\pi(y_t | y_{1:t-1}) = \pi(y_t | y_{t-1}).$$

در واقع در یک فرآیند مارکوف مرتبه اول، اطلاعات به دست آمده درباره y_t تمام مشاهدات تا زمان $t-1$ ، دقیقاً با اطلاعات حاصل از y_{t-1} یکسان است. به بیانی دیگر می‌توان گفت Y_t و $Y_{1:t-2}$ به شرط Y_{t-1} مستقل هستند. برای یک زنجیره مارکوف مرتبه اول، توزیع‌های توأم را می‌توان به صورت ساده

$$\pi(y_{1:t}) = \pi(y_1) \cdot \prod_{j=2}^t \pi(y_j | y_{j-1}).$$

نوشت. البته در نظر گرفتن ساختار وابستگی مارکوفی در بعضی از موقعیت‌ها نامناسب است (گیر، ۱۹۹۲). زنجیره‌های مارکوف مرتبه دوم و بالاتر به شیوه‌ای مشابه قابل تعریف هستند.

مدل‌های فضای حالت تحت ساختار وابستگی نسبتاً ساده مارکوفی، به منظور تعریف مدل‌های پیچیده‌تر برای مشاهدات ساخته می‌شوند. در یک مدل فضای حالت فرض می‌شود یک زنجیره مارکوف مشاهده نشده $\{\theta_t; t \geq 1\}$ وجود دارد، که به آن فرآیند حالت^{۳۰} یا فرآیند تکامل^{۳۱} می‌گویند. همچنین فرض می‌شود Y_t اندازه دقیقی از θ است. در کاربردهای مهندسی، θ_t معمولاً حالت یک سیستم فیزیکی قابل مشاهده می‌باشد که این سیستم پاسخ Y_t را تولید می‌کند. از طرفی دیگر در کاربردهای اقتصادی θ_t اغلب ساختاری پنهان دارد که ممکن است تفسیر مفیدی داشته باشد. در هر صورت می‌توان θ_t را به عنوان یک سری زمانی کمکی در نظر گرفت که تعیین توزیع احتمالی سری زمانی Y_t را آسان می‌کند. به طور رسمی، یک مدل فضای حالت شامل سری زمانی $\{\theta_t; t \geq 1\}$ در \mathbb{R}^p و سری زمانی $\{Y_t; t \geq 1\}$ در \mathbb{R}^m است که شرط‌های زیر را می‌پذیرد:

$$(1) \quad \{\theta_t; t \geq 1\} \text{ یک زنجیره مارکوف را تشکیل می‌دهد؛}$$

$$(2) \quad \text{به شرط } \theta_t, Y_t \text{ ها مستقلند و } Y_t \text{ فقط به } \theta_t \text{ وابسته است.}$$

از (۱) و (۲) می‌توان نتیجه گرفت که یک مدل فضای حالت توسط توزیع اولیه $\pi(\theta_0)$ و توزیع‌های شرطی $\pi(\theta_t | \theta_{t-1})$ و $\pi(y_t | \theta_t)$ ، $t \geq 1$ ، مشخص می‌شود. در واقع برای هر $t > 0$ داریم

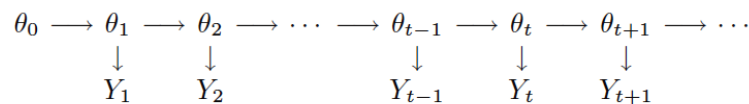
$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \prod_{j=1}^t \pi(\theta_j | \theta_{j-1}) \pi(y_j | \theta_j). \quad (4.1)$$

از (۴.۱) می‌توان هر توزیع مورد نظر دیگری را به دست آورد. برای مثال، توزیع توأم مشاهدات $Y_{1:t}$ را می‌توان به وسیله انتگرال‌گیری از این معادله بر حسب θ_j ها به دست آورد.

^{۲۹}Markov dependency

^{۳۰}State process

^{۳۱}Evolution process



شکل ۶.۱: ساختار وابستگی مدل فضای حالت.

ساختار وابستگی یک مدل فضای حالت در شکل ۶.۱ نشان داده شده است. توجه داشته باشید که در گراف شکل ۶.۱ کمان‌ها به صورت یال‌های بی‌سو^{۳۲} در نظر گرفته شده‌اند. نمایش گرافی این مدل می‌تواند به منظور نتیجه‌گیری ویژگی‌های استقلال شرطی متغیرهای تصادفی در یک مدل فضای حالت به کار برده شود. در واقع، در این نوع نمایش، دو مجموعه متغیر تصادفی A و B به شرط مجموعه متغیرهای سوم مانند C مستقل شرطی هستند، اگر و تنها اگر C از A و B مجزا باشد؛ یعنی اگر هر مسیر ارتباطی از یک متغیر در A به یک متغیر در B از طریق C عبور کند. به عنوان مثال، شکل ۶.۱ را می‌توان برای بیان استقلال شرطی Y_t و $(\theta_{0:t-1}, Y_{1:t-1})$ به شرط θ_t مورد استفاده قرار داد. به سادگی می‌توان دید که هر مسیر ارتباطی Y_t با یکی از مشاهده‌های قبلی Y_s ($s < t$) یا با یکی از حالت‌های θ_s ($s < t$) از طریق θ_t می‌گذرد؛ بنابراین $\{\theta_t\}$ ، $\{\theta_{0:t-1}, Y_{1:t-1}\}$ و $\{Y_t\}$ از هم مجزا هستند. بنابراین

$$\pi(y_t | \theta_{0:t}, y_{1:t-1}) = \pi(y_t | \theta_t).$$

به همین ترتیب می‌توان نشان داد که θ_t و $(\theta_{0:t-2}, Y_{1:t-1})$ به شرط θ_{t-1} مستقل شرطی هستند که در جملاتی از توزیع‌های شرطی به صورت زیر

$$\pi(\theta_t | \theta_{0:t-1}, y_{1:t-1}) = \pi(\theta_t | \theta_{t-1}).$$

ظاهر می‌شود.

یک مدل فضای حالت معمولاً دو سطح مدل‌بندی را در بر دارد: در سطح اول مشاهدات به متغیرهای پنهان^{۳۳} یا مشاهده نشده (حالت‌ها) براساس معادله مشاهده^{۳۴} مرتبط هستند؛ در سطح دوم تکامل حالت‌ها به وسیله معادله حالت^{۳۵} یا معادله تکامل^{۳۶} مدل‌بندی می‌شود. برای تشریح این دو مرحله، یک مدل AR مرتبه اول را در نظر بگیرید که در آن ضریب اتورگرسیو (متغیر حالت) تابعی از زمان است. به چنین مدلی یک مدل اتورگرسیو تغییرپذیر زمانی^{۳۷} می‌گویند. به عنوان مثال اگر معادله حالت یک rw1 باشد، دو معادله مشاهده و حالت به ترتیب به صورت

$$y_t = \phi_t y_{t-1} + \epsilon_t,$$

$$\phi_t = \phi_{t-1} + \nu_t$$

می‌باشند، که در آن $\epsilon_t \sim N(0, \nu)$ و $\nu_t \sim N(0, \omega)$ و ν_t و ϵ_t از هم مستقل هستند.

^{۳۲}Undirected edges

^{۳۳}Latent variables

^{۳۴}Observation equation

^{۳۵}State equation

^{۳۶}Evolution equation

^{۳۷}Time-varying Autoregressive

تعریف ۲.۴.۱ (مدل فضای حالت). در یک مدل فضای حالت، مشاهدات Y_t ، $t = 1, 2, \dots$ ، از توزیع شرطی $\pi(y_t | \theta_t, y_{1:t-1}; \psi)$ آمده‌اند که در آن θ_t حالت پنهان در زمان t است. این حالت‌های θ_t ، $t = 1, 2, \dots$ ، براساس یک زنجیره مارکوف با توزیع تکاملی $\pi(\theta_t | \theta_{t-1}, \psi)$ تکامل می‌یابند. در این جا ψ بردار پارامترهای مدل می‌باشد.

به دلیل پیچیدگی مدل‌های فضای حالت، برازش آن‌ها به وسیله الگوریتم‌های بازگشتی^{۳۸} صورت می‌پذیرد. این الگوریتم‌ها در هر دو دیدگاه بیزی و بسامدی پیاده‌سازی شده‌اند. به‌عنوان مثال، در دیدگاه بسامدی روش‌های کمترین توان‌های دوم خطای تعمیم‌یافته^{۳۹} یا درست‌نمایی ماکسیمم استفاده می‌شوند. اما معتقدیم که روش‌های بیزی برتری‌های بیشتری از نظر روش‌شناسی و هم از نظر محاسباتی دارد. ریشه این برتری‌ها نیز به همان محاسبات بازگشتی برمی‌گردد. از آن‌جا که مسئله برآورد و پیش‌بینی توسط محاسبات بازگشتی توزیع شرطی کمیت‌های مورد نظر به شرط اطلاعات موجود حل می‌شود، استفاده از یک چارچوب بیزی سازگاری بیشتری با استنباط در این مدل‌ها دارد. کالمن (۱۹۶۰) برخی مفاهیم ابتدایی مدل‌های خطی پویا را مطرح کرد که برای روش بیزی مناسب هستند. در زیربخش‌های بعدی برخی از زیررده‌های پرکاربرد مدل‌های فضای حالت را تشریح می‌کنیم.

۲.۴.۱ مدل مؤلفه‌های مشاهده‌نشده

یک زیررده مهم مدل‌های فضای حالت، مدل‌های مؤلفه‌های مشاهده‌نشده^{۴۰} است. در سطح اول این مدل‌ها، y_t در زمان t به منظور وابستگی به حالت یا مؤلفه مشاهده‌نشده τ_t به صورت زیر مدل‌بندی می‌شود:

$$y_t = \tau_t + \epsilon_t, t = 1, \dots, T$$

که در آن $\epsilon_t \sim N(0, \sigma^2)$. یعنی y_t به صورت مجموع مؤلفه مشاهده‌نشده τ_t و جمله خطا ϵ_t مدل‌بندی می‌شود. این نوع مدل، حالت خاصی برای تکامل سری‌های زمانی یک‌متغیره است. چون برای هر y_t یک متغیر پنهان τ_t داریم، تعداد پارامترها و متغیرهای پنهان (یعنی σ^2 و $\tau = (\tau_1, \dots, \tau_T)'$) نسبت به تعداد مشاهدات بیشتر است. بنابراین اگر فقط معادله مشاهده را داشته باشیم، برآورد (τ, σ^2) قابل دستیابی نیست. مثلاً برآوردگرهای ML^{۴۱} تعریف نمی‌شود. به طور دقیق‌تر تابع درست‌نمایی در (τ, σ^2) بی‌کران است (کروس و چان، ۲۰۱۴).

برای رفع این مشکل، یک راه‌حل ساختارمند کردن مؤلفه مشاهده‌نشده است به طوری که برازش مدل با داده‌های موجود امکان‌پذیر شود. چون با داده‌های سری زمانی سروکار داریم، به نظر می‌رسد که معقول باشد فرض کنیم این مؤلفه مشاهده نشده به تدریج تحت زمان تکامل می‌یابد. برای مثال داده‌های نرخ تورم آمریکا را در نظر بگیرید. این داده‌ها به صورت فصلی گردآوری و در شکل ۷.۱ نشان داده شده‌اند.

^{۳۸}Recursive algorithms

^{۳۹}Generalized least squared errors

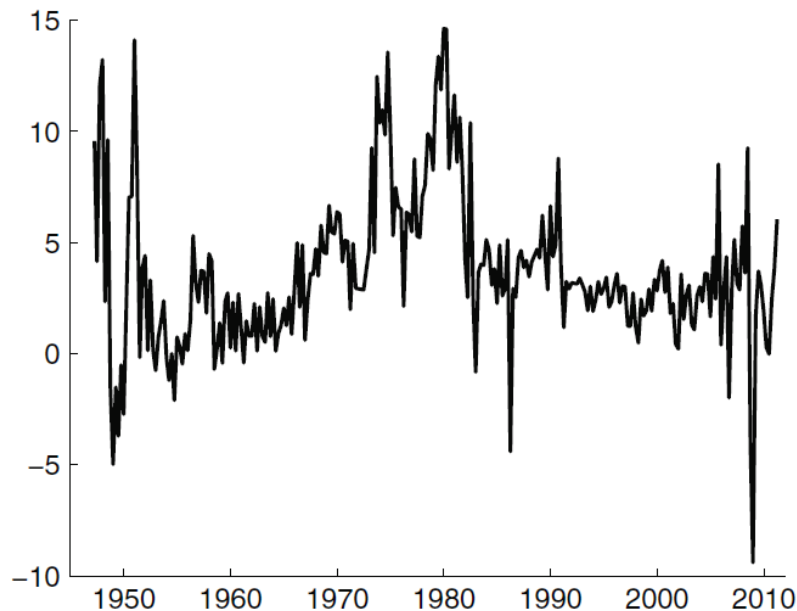
^{۴۰}Unobserved components models

^{۴۱}Maximum Likelihood

در این مثال، روندهای متوالی تورم به هم نزدیک هستند. بنابراین استفاده از مدل $rw1$ مناسب است. یعنی

$$\tau_t = \tau_{t-1} + u_t, \quad (5.1)$$

که در آن $u_t \sim N(0, \omega^2)$. یعنی توزیع شرطی τ_t به شرط τ_{t-1} و ω^2 ، $N(\tau_{t-1}, \omega^2)$ می‌باشد.



شکل ۷.۱: نرخ تورم آمریکا از سال ۱۹۴۷ تا ۲۰۱۱.

پارامتر ω^2 به‌عنوان یک پارامتر هموارساز، مقداری ثابت فرض می‌شود یا به‌عنوان یک پارامتر نامعلوم تعریف می‌شود که باید توسط داده‌ها برآورد شود. مقادیر بزرگ ω^2 ، τ_t را به‌سرعت تکامل می‌بخشد، اما مقادیر کوچک آن τ_t را به‌تدریج و کند تکامل می‌دهد. در چارچوب بیزی اغلب یک توزیع پیشین سلسله‌مراتبی^{۴۲} برای ω^2 انتخاب می‌شود که هموارسازی معادله تحول را انعکاس می‌دهد (کروس و چان، ۲۰۱۴). توجه داشته باشید که معادله (۵.۱) صریحاً توزیعی برای τ_t تعریف نمی‌کند. برای تکمیل این مدل، معمولاً فرض می‌شود که این فرآیند با $\tau_1 \sim N(\tau_0, \omega_0^2)$ برای ثابت‌های معلوم τ_0 و ω_0^2 مقداردهی اولیه شده است. این توزیع، شرط اولیه این مدل است. به‌طور خلاصه یک مدل مؤلفه‌های مشاهده‌نشده را می‌توان به‌صورت زیر بیان کرد:

تعریف ۳.۴.۱ (مدل مؤلفه‌های مشاهده‌نشده). در این مدل، معادله مشاهده به‌صورت

$$y_t = \tau_t + \epsilon_t$$

تعریف می‌شود، که در آن $\epsilon_t \sim N(0, \sigma^2)$. حالت‌ها با $\tau_1 \sim N(\tau_0, \omega_0^2)$ برای ثابت‌های معلوم τ_0 و ω_0^2 مقداردهی اولیه می‌شوند و براساس معادله تحول (مدل قدم زدن تصادفی)

$$\tau_t = \tau_{t-1} + u_t$$

^{۴۲}Hierarchical prior

تکامل می‌یابند، که در آن $u_t \sim N(0, \omega^2)$.

واضح است که مدل مؤلفه‌های مشاهده‌نشده در خانواده مدل‌های فضای حالت قرار می‌گیرد. یعنی با توجه به تعریف مدل فضای حالت، حالت θ_t در این مورد τ_t و $\psi = (\sigma^2, \omega^2)$ است. توزیع شرطی y_t به شرط τ_t ، $N(\tau_t, \sigma^2)$ و تابع چگالی تکاملی $\pi(\tau_t | \tau_{t-1}, \psi)$ متناظر با توزیع $N(\tau_{t-1}, \omega^2)$ است. علاوه بر این، چون هر دو معادله مشاهده و تکامل با خطاهای نرمال و خطی هستند، مدل مؤلفه‌های مشاهده‌نشده مثالی از یک مدل فضای حالت خطی نرمال است.

۳.۴.۱ مدل‌های خطی پویا

زیررده مهم دیگری از مدل‌های فضای حالت، مدل‌های خطی پویا هستند. این رده شامل دسته وسیعی از مدل‌های با پارامترهای وابسته به زمان است که برای مدل‌بندی پاسخ‌های وابسته به زمان و سری‌های زمانی در قالب مدل‌های رگرسیونی خطی، به‌کار گرفته می‌شوند. این مدل‌ها توسط هریسون و استیونز (۱۹۷۶) معرفی شدند. در این رده تغییرات پارامتر و داده‌های در دسترس، هر دو به‌صورت احتمالی بیان می‌شوند. یک مدل خطی پویا با توزیع پیشین نرمال برای بردار حالت p بعدی در زمان $t = 0$ به شکل

$$\theta_0 \sim N_p(m_0, c_0)$$

و دو معادله زیر، برای هر زمان $t \geq 1$ ، مشخص می‌شود:

$$y_t = F_t' \theta_t + \nu_t, \quad \nu_t \sim N_m(0, V_t) \quad (6.1)$$

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N_p(0, W_t) \quad (7.1)$$

که در آن $\{y_t\}$ دنباله‌ای از مشاهدات وابسته به زمان است که به‌طور شرطی از دنباله $\{\theta_t\}$ مستقل است. F_t ماتریس $m \times p$ از متغیرهای تیبینی، θ_t بردار p بعدی پارامترهای حالت، و G_t ماتریس $p \times p$ است که سیر تکاملی پارامترها را نشان می‌دهد. همچنین ν_t و ω_t دو دنباله مستقل از بردارهای نرمال با میانگین صفر و ماتریس‌های واریانس V_t و W_t هستند. معادله (۶.۱) معادله مشاهده و معادله (۷.۱) معادله تکامل یا حالت است. افزون بر این، فرض می‌شود که θ مستقل از ν_t و ω_t است. می‌توان نشان داد که یک DLM شرط‌های (۱) و (۲) بیان‌شده در بخش معرفی مدل فضای حالت را با $Y_t | \theta_t \sim N_m(F_t \theta_t, V_t)$ و $\theta_t | \theta_{t-1} \sim N_p(G_t \theta_{t-1}, W_t)$ می‌پذیرد. توجه کنید برای حالت یک‌متغیره، $m = 1$ است.

یک مدل خطی پویا به‌طور خلاصه به وسیله چهارگانه $\{F_t, G_t, V_t, W_t\}$ مشخص می‌شود. برای این مدل‌ها، دو حالت خاص قابل توجه هستند:

۱. اگر برای هر t ، $F_t = F$ و $G_t = G$ ، یعنی کمیت‌های ثابتی نسبت به زمان باشند، آنگاه نتیجه به مدل خطی پویای سری زمانی معروف است.

۲. اگر برای هر t ، G_t یک ماتریس همانی باشد، به مدل نتیجه‌شده مدل رگرسیون خطی پویا می‌گویند.

به دلیل کاربردهای فراوان مدل‌های خطی پویا و زیررده‌های آن، در ادامه چند مدل از این رده را که در تحلیل داده‌های سری زمانی مورد استفاده قرار می‌گیرند، معرفی می‌کنیم.

مدل قدم زدن تصادفی

ساده‌ترین مدل برای سری زمانی یک‌متغیره، مدل قدم زدن تصادفی یا همان مدل مؤلفه‌های مشاهده‌نشده است که در زیربخش قبلی معرفی شد. برای درک آن که این مدل حالت خاصی از مدل خطی پویا است، یک بار دیگر دو معادله زیر را در نظر بگیرید:

$$y_t = \mu_t + \nu_t, \quad \nu_t \sim N(0, V)$$

$$\mu_t = \mu_{t-1} + \omega_t, \quad \omega_t \sim N(0, W)$$

که در آن دنباله‌های ν_t و ω_t از هم مستقل هستند. این مدل یک مدل خطی پویا با $m = p = 1$ ، $\theta = \mu_t$ و $F_t = G_t = 1$ است. این مدل برای نشان دادن روند یا تغییرات فصلی در یک سری زمانی مناسب است. مشاهدات y_t به صورت مشاهدات یک سطح μ_t مدل‌بندی می‌شوند که این سطح، تغییرات تصادفی را تحت زمان نشان می‌دهد و توسط یک مدل RW1 بیان می‌شود. به این مدل، مدل سطح مقطعی^{۴۳} نیز گفته می‌شود. اگر $W = 0$ آن‌گاه به ازای هر t ، $\mu_t = \mu$. توجه کنید با توجه به نایستا بودن مدل RW1، مدل‌های خطی پویا را می‌توان برای مدل‌بندی سری‌های زمانی نایستا به کار برد.

مدل خطی مقطعی

یک مدل پیچیده‌تر نسبت به مدل سطح مقطعی، مدل رشد خطی یا روند خطی مقطعی است که در آن معادله مشاهده همانند مدل سطح مقطعی می‌باشد، اما شبیهی با تغییرپذیری در زمان برای μ_t دارد. در واقع مدل به صورت

$$y_t = \mu_t + \nu_t, \quad \nu_t \sim N(0, V)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{t,1}, \quad \omega_{t,1} \sim N(0, \sigma_\mu^2)$$

$$\beta_t = \beta_{t-1} + \omega_{t,2}, \quad \omega_{t,2} \sim N(0, \sigma_\beta^2)$$

با خطاهای ناهمبسته ν_t ، $\omega_{t,1}$ و $\omega_{t,2}$ است به طوری که خطاها از هم مستقل می‌باشند. این مدل یک مدل خطی پویا با مشخصات زیر است:

$$\theta_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}, \quad G_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad W = \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}, \quad F_t = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

در این مدل واریانس‌های σ_μ^2 و σ_β^2 می‌توانند صفر فرض شوند.

^{۴۳}Local level model

مدل رگرسیون خطی

مدل ARIMA، با توزیع نرمال برای جملات خطا، را می‌توان به‌عنوان حالت خاصی از مدل‌های خطی پویا به‌دست آورد (هانان و دیستلر، ۱۹۹۸).

صرف‌نظر از مدل‌های ARIMA، مدل‌های خطی پویا را می‌توان به‌عنوان تعمیمی از مدل رگرسیون خطی با فرض وابسته به زمان بودن ضرایب رگرسیون، در نظر گرفت. می‌دانیم مدل رگرسیون خطی رابطه بین پاسخ y و متغیر تبیینی x را به‌صورت

$$y_t = \theta_1 + \theta_2 x_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2)$$

بیان می‌کند. در این جا (y_t, x_t) تحت زمان مشاهده شده‌اند. با فرض وابسته به زمان بودن پارامترهای رگرسیونی، می‌توان یک مدل رگرسیون خطی پویا به صورت

$$y_t = \theta_{t,1} + \theta_{t,2} x_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2)$$

برای بیان رابطه بین y_t و x_t در نظر گرفت. همچنین این مدل می‌تواند دارای معادله حالت به‌صورت زیر باشد:

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t)$$

که در آن $\theta_t = (\theta_{t,1}, \theta_{t,2})'$. ترکیب این دو معادله، یک مدل خطی پویا با $F_t = [1, x_t]$ است. به‌سادگی می‌توان دید اگر G_t یک ماتریس همانی باشد و برای هر t ، $\sigma_t^2 = \sigma^2$ و $W_t = 0$ ، آنگاه مدل رگرسیون خطی ساده نتیجه می‌شود.

۴.۴.۱ مدل‌های غیرخطی و غیرنرمال پویا

یک مدل فضای حالت کلی، به همراه تعیین توزیع پیشین برای θ_0 ، با دو معادله مشاهده و حالت به‌صورت زیر مشخص می‌شود:

$$y_t = h_t(\theta_t, \nu_t)$$

$$\theta_t = g_t(\theta_{t-1}, \omega_t),$$

که در آن $h_t(\cdot)$ و $g_t(\cdot)$ توابعی معلوم هستند. برای مدل‌های فضای حالت خطی، $h_t(\cdot)$ و $g_t(\cdot)$ هر دو توابعی خطی هستند. مدل‌های خطی نرمال، فرضیات توزیع‌های نرمال را نیز در بر دارند. فرض نرمال بودن در اکثر کاربردها محسوس است و می‌توان آن را به وسیله قضیه حد مرکزی توجیه کرد. با این حال، مواردی هستند که فرض توزیع نرمال منطقی نیست و باید از توزیع‌های دیگر استفاده کرد. مثل پاسخ‌های شمارشی یا دودویی. البته با حذف فرض نرمال بودن، محاسبات پیچیده خواهند بود.

برخی کاربردها وجود دارند که در آن‌ها غیرخطی بودن لازمه مدل‌بندی فرآیند می‌باشد. مثالی روشن یک مدل سری زمانی چندفصلی است. در این مدل، اثر میانگین به‌صورت $\mu_t = \alpha_t(1 + \phi_t)$ بیان می‌شود که در آن α_t مؤلفه روند و ϕ_t اثر فصلی است. این مدل برای بیان رابطه غیرخطی پویا که شامل متغیر تبیینی x_t است، مفید می‌باشد. تابع پاسخ این مدل به وسیله $E_t = \phi E_{t-1} + \gamma x_t$ به‌دست آمده است که در آن $\gamma > 0$ و $|\phi| < 1$. میانگین پاسخ را می‌توان به‌صورت $\mu_t = \alpha_t + E_t$ نوشت.

در این زیربخش، به اختصار، رده‌های مهمی از مدل‌های فضای حالت غیرخطی و غیرنرمال را بیان می‌کنیم. اگرچه تمرکز اصلی این پایان‌نامه بر مدل‌های فضای حالت نرمال است، اما از معرفی این رده‌ها دو هدف داریم:

- مدل‌های فضای حالت با توزیع پاسخ از خانواده توزیع‌های نمایی، دامنه کاربرد زیادی در مدل‌بندی پاسخ‌های وابسته به زمان گسسته دارند و در این پایان‌نامه برخی از مثال‌های کاربردی را شامل می‌شوند.
- مشاهده این رده‌ها می‌بایست این ایده را برای خواننده فراهم آورد که می‌توان مدل‌های فضای حالت را برای مواردی که پذیره‌های مدل خطی نرمال برقرار نیستند نیز به‌کار گرفت.

مدل‌های خطی تعمیم‌یافته پویا

مدل‌های خطی پویا را می‌توان با فرض نداشتن توزیع‌های نرمال تعمیم داد. این تعمیم، به‌ویژه، برای مدل‌بندی سری‌های زمانی گسسته مورد نیاز است. برای مثال، اگر y_t حضور یا عدم حضور یک ویژگی را در مسئله تحت مطالعه برحسب زمان بیان کند، می‌توان از یک توزیع برنولی استفاده کرد؛ یا اگر y_t شمارشی باشد، می‌توان مدل پواسون را به‌کار برد.

مدل‌های خطی تعمیم‌یافته پویا (وست و همکاران، ۱۹۸۵) فرض می‌کند توزیع شرطی $p(y_t | \theta_t)$ ، عضوی از خانواده توزیع‌های نمایی با پارامتر $F_t \theta_t$ است. معادله حالت نیز مانند مدل‌های خطی نرمال به صورت $\theta_t = G_t \theta_{t-1} + \omega_t$ است. استنباط بیزی مدل‌های خطی تعمیم‌یافته پویا با مشکلات محاسباتی همراه است که معمولاً از روش‌های تقریبی بیزی برای رفع آن‌ها استفاده می‌کنند.

مدل‌های مارکوفی پنهان

مدل‌های فضای حالت که حالت θ_t در آن‌ها گسسته باشد را مدل‌های مارکوفی پنهان^{۴۴} می‌نامند. این مدل‌ها به‌طور گسترده در تشخیص گفتار^{۴۵} مورد استفاده قرار می‌گیرند (رابینر و جوآنگ، ۱۹۹۳). در علم اقتصاد و مالی اغلب این مدل‌ها برای مدل‌بندی یک سری زمانی با فروپاشی ساختاری به‌کار گرفته می‌شوند. پویایی این سری و نقاط تغییر توسط یک زنجیر مارکوفی پنهان (θ_t) با فضای حالت $\{\theta_1^*, \dots, \theta_k^*\}$ و احتمال‌های

$$\pi(i|j) = p(\theta_t = \theta_i^* | \theta_{t-1} = \theta_j^*)$$

تعیین می‌شوند. در نتیجه Y_t دارای توزیعی وابسته به حالت زنجیر در زمان t می‌باشد، یعنی

$$Y_t | \{\theta_t = \theta_j^*\} \sim p(y_t | \theta_j^*), \quad j = 1, \dots, k.$$

اگرچه مدل‌های فضای حالت با حالت پیوسته و مدل‌های مارکوفی پنهان برای اهدافی متفاوت توسعه یافته‌اند، اما پذیره‌های اساسی و محاسبات بازگشتی آن‌ها به هم مرتبط هستند. روش‌های شبیه‌سازی

^{۴۴}Latent Markov models

^{۴۵}Speech recognition

مبتنی بر نمونه‌گیری مونت کارلو^{۴۶} نیز برای استنباط در مدل‌های مارکوفی پنهان توسعه یافته‌اند؛ برای مثال، رایدن و تیتزینگتون (۱۹۹۸)، کیم و نلسون (۱۹۹۹)، کاپه و همکاران (۲۰۰۵) و مراجع موجود در آن‌ها را مشاهده کنید.

مدل‌های نوسانات تصادفی

مدل‌های نوسانات تصادفی^{۴۷} به‌طور گسترده‌ای در کاربردهای مالی به‌کار می‌روند. فرض کنید y_t لگاریتم یک دارایی در زمان t باشد. یعنی $y_t = \log \frac{P_t}{P_{t-1}}$ که در آن P_t قیمت دارایی در زمان t است. تحت فرض بازارهای کارآمد، میانگین شرطی این لگاریتم‌ها صفر است. یعنی $E(Y_{t+1}|y_{1:t}) = 0$. اما واریانس شرطی (نوسان) تحت زمان تغییر می‌کند.

دو رده اصلی از مدل‌ها برای تحلیل نوسانات، رده مدل‌های ARCH^{۴۸} و GARCH^{۴۹} (انگل، ۱۹۸۲؛ بولرسلف، ۱۹۸۶) هستند. این دو رده، نوسان را به‌عنوان تابعی از مقادیر قبلی تعریف می‌کنند. در حالی‌که، مدل‌های نوسانات تصادفی نوسان را به‌عنوان یک فرآیند تصادفی در نظر می‌گیرند. این مدل منجر به مدل فضای حالتی می‌شود که نوسان بخشی از بردار حالت آن است. برای مثال شفارد (۱۹۹۶) را مشاهده کنید. ساده‌ترین مدل نوسانات تصادفی به صورت

$$y_t = \exp\left\{\frac{1}{\sigma_t}\theta_t\right\}\omega_t, \quad \omega_t \sim N(0, 1)$$

$$\theta_t = \phi\theta_{t-1} + \nu_t, \quad \nu_t \sim N(0, \sigma^2)$$

تعریف می‌شود، که در آن θ_t یک مدل اتوگرسیو مرتبه اول است. این مدل‌ها غیرنرمال و غیرخطی هستند و معمولاً محاسبات بیشتری نسبت به مدل‌های ARCH و GARCH دارند. البته برای استنباط در این مدل‌ها نیز معمولاً از روش‌های مونت کارلویی استفاده می‌کنند (جاکوئر و همکاران، ۱۹۹۴). از سوی دیگر، مدل‌های نوسانات تصادفی برای تعمیم حالتی که مجموعه‌ای از دارایی‌ها مد نظر هستند، ساده‌تر است. در حالی‌که در مدل‌های ARCH و GARCH چندمتغیره تعداد پارامترها بسیار زیاد می‌شوند.

۵.۱ استنباط بیزی تقریبی

قبلاً متذکر شدیم که در دیدگاه بیزی همه استنباط‌ها مبتنی بر توزیع پسین به‌دست می‌آیند. از طرفی، توزیع پسین توأم پارامترها و فرآیند حالت در مدل‌های فضای حالت معمولاً شکل بسته‌ای ندارند و بسته به انتخاب توزیع‌های پیشین، پیچیده هستند. بنابراین برای استخراج استنباط‌های آماری باید از روش‌های تقریبی برای تقریب توزیع پسین استفاده کرد.

^{۴۶}Monte Carlo based simulation methods

^{۴۷}Stochastic perturbation models

^{۴۸}Autoregressive conditionally heteroscedastic

^{۴۹}Generalized ARCH

روش‌های تقریب توزیع پسین به دو دسته کلی تقسیم می‌شوند: (۱) روش‌های عددی و (۲) روش‌های مبتنی بر نمونه‌گیری. از جمله روش‌های تقریب عددی می‌توان به تقریب نیوتون-رافسون، روش‌های تربیع‌بندی گاوسی^{۵۰} (آبرامویتز و استگن، ۱۹۷۲)، و تقریب لاپلاس^{۵۱} (تیرنی و کدین، ۱۹۸۶) اشاره کرد. از معروف‌ترین روش‌های مبتنی بر نمونه‌گیری که به روش‌های شبیه‌سازی مونت کارلویی نیز معروف هستند، می‌توان روش‌های نمونه‌گیری رد و پذیرش^{۵۲} (کسلا و برگر، ۲۰۰۲)، نمونه‌گیری نقاط مهم^{۵۳} (هستریگ، ۱۹۸۷)، و رده بزرگ الگوریتم‌های نمونه‌گیری مونت کارلوی زنجیر مارکوفی^{۵۴} (MCMC) (هستینگز، ۱۹۷۰؛ رابرت و کسلا، ۲۰۰۴) را نام برد.

برای مسائلی که بعد توزیع پسین کمتر از 10^1 است، روش‌های عددی خطای قابل قبولی دارند. اما در مواردی که این بعد بالا است، تجمع خطاهای گرد کردن در روش‌های عددی بزرگ شده و استفاده از آن‌ها توصیه نمی‌شود. در این موقعیت‌ها، روش‌های مبتنی بر نمونه‌گیری برای تقریب توزیع پسین انتخاب معمول هستند. در روش‌های نمونه‌گیری MCMC، خطای تقریب تابعی از بعد توزیع پسین نیست و به ازای هر بعدی ثابت و از مرتبه $O(n^{-\frac{1}{2}})$ می‌ماند، که در آن n حجم نمونه مونت کارلویی است. البته سرعت محاسبات در روش‌های مبتنی بر نمونه‌گیری می‌تواند خیلی کندتر از روش‌های عددی باشد. در روش‌های مبتنی بر نمونه‌گیری، قانون قوی اعداد بزرگ^{۵۵} برای نمونه‌های تولیدشده مستقل (کسلا و برگر، ۲۰۰۲) و قضیه ارگودیک^{۵۶} برای نمونه‌های تولیدشده وابسته (زنجیر مارکوفی) (بیرخوف، ۱۹۴۲) ضامن همگرایی برآوردهای حاصل از نمونه تولیدشده به کمیت‌های مورد نظر از توزیع پسین هستند.

در ساخت روش تقریبی مورد نظر این پایان‌نامه، با عنوان تقریب لاپلاس آشیانه‌ای جمع‌بسته^{۵۷} (INLA)، تقریب لاپلاس یکی از اجزای زیربنایی است. بنابراین در ادامه، ابتدا تقریب لاپلاس را معرفی می‌کنیم. سپس الگوریتم‌های نمونه‌گیری MCMC را توضیح می‌دهیم.

۱.۵.۱ تقریب لاپلاس

تقریب لاپلاس یک روش عددی برای تقریب انتگرال‌های پیچیده است که در استنباط بیزی از آن زیاد استفاده می‌شود. دقت تقریب زمانی قابل قبول است که تابع زیر انتگرال را بتوان به خوبی با یک هسته توزیع نرمال تقریب زد. حتی تیرنی و کدین (۱۹۸۷) نشان دادند خطای تقریب می‌تواند، تحت شرایطی، از مرتبه $O(n^{-2})$ باشد. فرض کنید علاقه‌مند باشیم انتگرال

$$\int f(x)dx = \int \exp(\log f(x))dx$$

^{۵۰} Gauss quadrature methods

^{۵۱} Laplace approximation

^{۵۲} Reject-accept sampling

^{۵۳} Importance sampling

^{۵۴} Markov chain Monte Carlo

^{۵۵} Strong law of large numbers

^{۵۶} Ergodic theorem

^{۵۷} Integrated Nested Laplace Approximation

را محاسبه کنیم. بسط سری تیلور تابع $\log f(x)$ تا مرتبه دوم در $x = x_0$ به صورت زیر است:

$$\log f(x) \approx \log f(x_0) + (x - x_0) \frac{\partial \log f(x)}{\partial x} \Big|_{x=x_0} + \frac{(x - x_0)^2}{2} \frac{\partial^2 \log f(x)}{\partial x^2} \Big|_{x=x_0}.$$

اگر x_0 را برابر با مُد $x^* = \arg \max_x \log f(x)$ قرار دهیم، آنگاه $\frac{\partial \log f(x)}{\partial x} \Big|_{x=x^*} = 0$ و

$$\log f(x) \approx \log f(x^*) + \frac{(x - x^*)^2}{2} \frac{\partial^2 \log f(x)}{\partial x^2} \Big|_{x=x^*}.$$

پس انتگرال مورد نظر به صورت زیر تقریب می‌شود:

$$\begin{aligned} \int f(x) dx &\approx \int \exp \left(\log f(x^*) + \frac{(x - x^*)^2}{2} \frac{\partial^2 \log f(x)}{\partial x^2} \Big|_{x=x^*} \right) dx \\ &= \exp(\log f(x^*)) \int \exp \left(\frac{(x - x^*)^2}{2} \frac{\partial^2 \log f(x)}{\partial x^2} \Big|_{x=x^*} \right) dx \end{aligned}$$

که می‌توان آن را با تابع چگالی توزیع نرمال مرتبط ساخت. با قرار دادن $\sigma^{2*} = -1 / \frac{\partial^2 \log f(x)}{\partial x^2} \Big|_{x=x^*}$ داریم

$$\int f(x) dx \approx \exp(\log f(x^*)) \int \exp \left(-\frac{(x - x^*)^2}{2\sigma^{2*}} \right) dx$$

که انتگرال تابع هسته توزیع نرمال با میانگین x^* و واریانس σ^{2*} است. به طور دقیق‌تر، این انتگرال در فاصله (α, β) به صورت

$$\int_{\alpha}^{\beta} f(x) dx \approx f(x^*) \sqrt{2\pi\sigma^{2*}} (\Phi(\beta) - \Phi(\alpha)) \quad (۸.۱)$$

قابل تقریب است، که در آن $\Phi(\cdot)$ تابع توزیع تجمعی $N(x^*, \sigma^{2*})$ است.

برای مثال، توزیع گاما با تابع چگالی

$$f(x) = \frac{b^a}{\Gamma(a)} \exp(-bx) x^{a-1} \quad x, a, b > 0 \quad (۹.۱)$$

را در نظر بگیرید. برای محاسبه تقریب لاپلاس داریم

$$\begin{aligned} \log f(x) &= (a - 1) \log x - bx + \text{ثابت}, \\ \frac{\partial \log f(x)}{\partial x} &= \frac{a - 1}{x} - b \\ \frac{\partial^2 \log f(x)}{\partial x^2} &= -\frac{a - 1}{x^2}. \end{aligned}$$

با حل $\frac{\partial \log f(x)}{\partial x} = 0$ ، مُد برابر با $x^* = \frac{a-1}{b}$ به دست می‌آید که در آن $a > 1$. همچنین واریانس σ^{2*} که به صورت $-1 / \frac{\partial^2 \log f(x)}{\partial x^2}$ در مُد x^* به دست می‌آید، برابر با $\sigma^{2*} = \frac{a-1}{b^2}$ است. بنابراین تقریب لاپلاس توزیع گاما به صورت

$$\text{Gamma}(a, b) \approx N \left(\frac{a-1}{b}, \frac{a-1}{b^2} \right)$$

است. برای ارزیابی این تقریب، مقادیر مختلفی برای پارامترهای a و b در توزیع گاما برای انتگرال‌گیری

جدول ۱۰.۱: مقادیر تقریبی و دقیق انتگرال $\int_{\alpha}^{\beta} f(x)dx$ برای تابع چگالی توزیع $\text{Gamma}(a, b)$

(a, b)	(α, β)	مقدار دقیق	مقدار تقریبی
(۵, ۱)	(۲, ۶)	۰/۶۶۲۲۹۰۵	۰/۶۶۸۶۴۲۴
(۵, ۱)	(۱, ۱۰)	۰/۹۶۷۰۸۷۵	۰/۹۱۲۶۶۹۳
(۱۰, ۰/۵)	(۱۸, ۲۲)	۰/۲۴۶۸۹۷۶	۰/۲۴۵۲۲۷۲
(۱۰, ۰/۵)	(۱۰, ۴۰)	۰/۹۶۳۱۷۶۵	۰/۹۰۰۲۹۴۶

در فاصله (α, β) در نظر گرفتیم. نتایج را در جدول ۱۰.۱ مشاهده می‌کنید. به‌عنوان نمونه، مقدار دقیق $\int_{\alpha}^{\beta} f(x)dx$ برای $f(x)$ در (۹.۱)، اختلاف مقادیر تابع توزیع تجمعی در β و در α است، درحالی‌که مقدار تقریبی انتگرال با استفاده از معادله (۸.۱) به‌دست آمده است. شکل ۸.۱ (نمودار بالا) منحنی‌های توابع چگالی دقیق $\text{Gamma}(۵, ۱)$ (منحنی خط توپر) و تقریب لاپلاس (منحنی خط چین) با $N(۴, ۴)$ است.

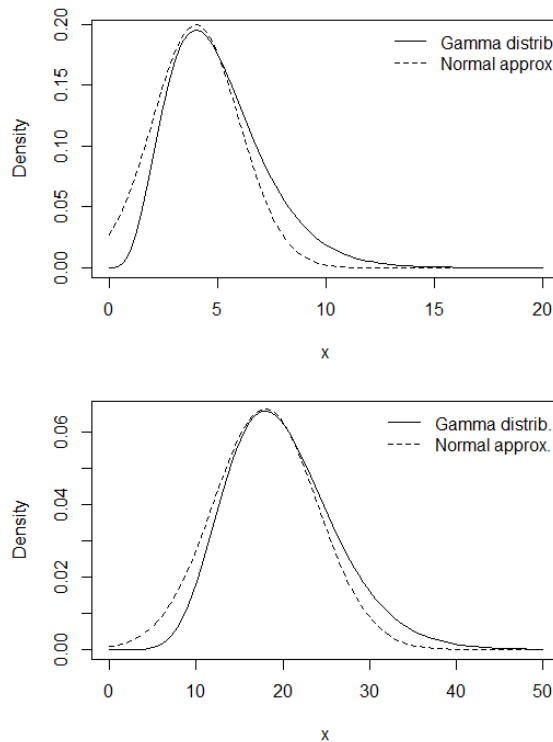
توجه داشته باشید برای $\alpha = ۱$ و $\beta = ۱۰$ ، فاصله انتگرال پهن‌تر و شامل دم‌های توزیع گاما است. در این حالت، خطای تقریب لاپلاس چندان ناچیز نیست (مقدار تقریبی ۰/۹۱۲۶۶۹۳ در مقابل مقدار دقیق ۰/۹۶۷۰۸۷۵). اتفاقی مشابه با $\text{Gamma}(۱۰, ۰/۵)$ رخ می‌دهد (نمودار پایین شکل ۸.۱). تقریب لاپلاس زمانی که انتگرال در بخش مرکزی توزیع (با $\alpha = ۱۸$ و $\beta = ۲۲$) محاسبه می‌شود، نزدیک به مقدار دقیق است و زمانی که فاصله افزایش می‌یابد (با $\alpha = ۱۰$ و $\beta = ۴۰$) از مقدار دقیق دور می‌شود.

تقریب لاپلاس برای حالت چندمتغیره به روشی مشابه قابل محاسبه است. تنها در بسط سری تیلور و انتگرال‌ها، کمیت‌ها به‌صورت بردار و ماتریس (دترمینان ماتریس) خواهند بود. برای دیدن جزئیات به کاس و استیفی (۱۹۸۹) مراجعه کنید.

۲.۵.۱ الگوریتم‌های MCMC

در مواردی که توزیع پسین به‌صورت بسته در دسترس نیست، باید از روش‌های تقریبی برای محاسبه (تقریبی) آن استفاده کرد. یکی از رده‌های تقریب توزیع پسین استفاده از روش‌های شبیه‌سازی مونت کارلویی است. وظیفه الگوریتم‌های شبیه‌سازی مونت کارلویی، تولید نمونه‌هایی از یک توزیع احتمالی مفروض است. زمانی که توزیع‌ها پیچیده یا بعد توزیع پسین (بعد پارامترها) بالا باشد، اجرای این وظیفه به‌صورت مستقیم به عملی سخت و پیچیده تبدیل می‌شود. برای رفع این دشواری، الگوریتم‌های MCMC (متروپولیس و همکاران، ۱۹۵۳؛ هستینگز، ۱۹۷۰) معرفی شدند که به‌طور غیرمستقیم اقدام به تولید نمونه می‌کنند.

مکانیسم الگوریتم‌های MCMC، مبتنی بر تولید نمونه وابسته از یک زنجیر مارکوف است که توزیع ایستای آن، همان توزیع پسین مورد نظر باشد. در این صورت، اگر حجم نمونه (وابسته) مونت کارلویی تولیدشده، مثلاً n ، به اندازه کافی بزرگ باشد، از جایی به بعد، مثل $n^{Burn} + ۱$ ، نمونه‌ها به‌عنوان یک نمونه (تقریبی) از توزیع پسین در نظر گرفته می‌شوند. نمونه‌های تولیدی از ابتدا تا نقطه n^{Burn} به دوره



شکل ۸.۱: تابع چگالی توزیع گاما با پارامترهای $a = 5$ و $b = 1$ (نمودار بالا) و $a = 10$ و $b = 0.5$ (نمودار پایین). منحنی خط چین تقریب نرمال حاصل از روش لاپلاس است.

داغیدن^{۵۸} معروف است که برای کم کردن اثر نقطه شروع الگوریتم از کل نمونه‌ها حذف می‌شوند. برای معرفی جزئیات الگوریتم‌های MCMC، ابتدا مساله انتگرال‌گیری مونت کارلویی را تشریح می‌کنیم.

انتگرال‌گیری مونت کارلویی

فرض کنید بخواهیم انتگرال

$$H = \int h(x)f(x)dx \quad (10.1)$$

را محاسبه کنیم، که در آن $h(\cdot)$ تابعی از متغیر تصادفی X با تابع چگالی $f(\cdot)$ است. چنان‌چه نتوان این انتگرال را با استفاده از محاسبات معمولی حل کرد، می‌توان از روش‌های عددی یا شبیه‌سازی برای تقریب H استفاده کرد. در این‌جا به معرفی روش‌های انتگرال‌گیری عددی نمی‌پردازیم. برای دیدن این روش‌ها و جزئیات آن‌ها می‌توانید به پرس و همکاران (۲۰۰۷) و لانگ (۲۰۱۰) مراجعه کنید.

با داشتن نمونه $x^{(1)}, \dots, x^{(n)}$ از توزیع $f(x)$ ، تقریب مونت کارلوی H به صورت

$$\hat{H} = \frac{1}{n} \sum_{i=1}^n h(x^{(i)})$$

^{۵۸}Burn in period

محاسبه می‌شود. با استفاده از قانون قوی اعداد بزرگ، زمانی که $n \rightarrow \infty$

$$\hat{H} \xrightarrow{a.s.} H.$$

دقت این تقریب زمانی که حجم نمونه n بزرگ شود، افزایش می‌یابد و می‌توان آن را به صورت زیر ارزیابی کرد:

$$se(\hat{H}) = \sqrt{\frac{\sum_{i=1}^n (h(x^{(i)}) - \hat{H})^2}{n(n-1)}}.$$

توجه داشته باشید که این دقت تقریب، معکوس نسبت ریشه دوم اندازه نمونه است. بنابراین، برای مثال اگر بخواهیم خطای استاندارد را نصف کنیم به چهار برابر n اولیه نیاز داریم. همچنین، بنا بر قضیه حد مرکزی (کسلا و برگر، ۲۰۰۲)، اگر واریانس X متناهی باشد، برای $n \rightarrow \infty$ ، \hat{H} در توزیع به یک متغیر تصادفی نرمال همگرا می‌شود. یعنی

$$\hat{H} \xrightarrow{d} N(H, se(\hat{H})^2).$$

در چارچوب استنباط بیزی، $f(x)$ در (۱۰.۱) نقش تابع چگالی توزیع پسین مدل، $p(\psi|y_{1:T})$ را دارد و برای انجام استنباط بیزی، باید انتگرال توابعی مانند $h(\psi)$ را بر حسب توزیع پسین به دست آوریم. یعنی باید کمیت‌هایی مانند

$$E(h(\psi)|y_{1:T}) = \int_{\psi \in \Psi} h(\psi)p(\psi|y_{1:T})d\psi$$

را محاسبه کنیم. بنابراین با داشتن نمونه $\{\psi^{(1)}, \dots, \psi^{(n)}\}$ از توزیع پسین، می‌توان شاخص‌های پسین مورد نیاز را به طور تقریبی محاسبه کرد. برای مثال احتمال پسینی مجموعه A که به صورت

$$p(\psi \in A|y_{1:T}) = \int_{\psi \in A} p(\psi|y_{1:T})d\psi$$

تعریف می‌شود، به وسیله نسبت تعداد (#) نمونه‌های $\psi^{(i)}$ که متعلق به A هستند به صورت

$$\frac{\#\{\psi^{(i)} \in A, i = 1, \dots, n\}}{n}$$

تقریب زده می‌شود. شاخص دلخواه دیگری از توزیع پسین، چندک مرتبه q (ψ_q) است که به شکل

$$\int_{-\infty}^{\psi_q} p(\psi|y_{1:T})d\psi = q$$

تعریف می‌شود و با چندک تجربی مرتبه q نمونه مونت کارلویی تقریب زده می‌شود. برای محاسبه این تقریب‌ها تنها کافی است نمونه‌های مستقل از توزیع پسین در اختیار باشند که اگر $p(\psi|y_{1:T})$ توزیع شناخته شده‌ای داشته باشد یا تولید مستقیم نمونه از آن راحت باشد، در دسترس نیست.

تولید غیرمستقیم نمونه

به جز مثال‌های ساده و استثنا، تولید مستقیم نمونه از توزیع‌های پسین معمولاً کار ساده‌ای نیست. همان‌طور که در بالا اشاره شد، یک راه‌حل معرفی شده تولید نمونه به صورت غیرمستقیم است. این راه‌حل به کمک نظریه زنجیرهای مارکوف ارایه شد. در واقع در رهیافت غیرمستقیم تولید نمونه به کمک زنجیرهای مارکوف، به جای شبیه‌سازی نمونه‌های مستقل از توزیع پسین، نمونه‌ای وابسته از زنجیر مارکوفی که توزیع

ایستای آن $p(\psi|y_{1:T})$ است، شبیه‌سازی می‌شود. سپس این نمونه می‌تواند برای محاسبه کمیت‌های پسینی مورد نظر مانند میانگین، چندک‌ها و احتمال‌ها، مورد استفاده قرار گیرد. این روش ترکیبی از نمونه‌گیری مونت کارلویی و زنجیرهای مارکوف است که به روش نمونه‌گیری MCMC مشهور شد.

با این‌که کشف MCMC به مطالعات متروپولیس و همکاران (۱۹۵۳) و هستینگز (۱۹۷۰) برمی‌گردد، اما استفاده شایع از آن به‌عنوان ابزاری برای اجرای استنباط‌های بیزی با کارگلفاند و اسمیت (۱۹۹۰) شروع شد. برای مشاهده تاریخچه پیدایش و گسترش الگوریتم‌های MCMC به رابرت و کسلا (۲۰۱۱) مراجعه کنید. امروزه به دلیل پیشرفت‌های اساسی ایجادشده در منابع محاسباتی ارزان توسط رایانه‌ها، الگوریتم‌های MCMC به هسته اصلی جعبه ابزار استنباط بیزی تبدیل شده‌اند.

برای تشریح الگوریتم‌های MCMC، فرض کنید مجموعه‌ای از متغیرهای تصادفی وابسته

$$\{X^{(0)}, X^{(1)}, \dots, X^{(t)}, \dots\}$$

داریم که هر یک روی فضای حالت \mathcal{X} تعریف شده است. ایده کلی الگوریتم‌های MCMC ساخت زنجیر مارکوفی است که توزیع ایستای آن توزیع هدف π باشد که مایل به تولید نمونه از آن هستیم؛ به این معنی که اگر $X^{(t)} \sim \pi$ آن‌گاه $X^{(t+1)} \sim \pi$. بنابراین، اگر حجم نمونه تولیدشده از زنجیر مارکوف به اندازه کافی بزرگ باشد، نمونه تولیدشده، تحقق‌هایی از توزیع هدف محسوب می‌شوند.

اطمینان از وجود چنین توزیع ایستای منحصر به فردی، مستلزم آن است که زنجیر مارکوف چند ویژگی را داشته باشد:

- تحویل‌ناپذیر^{۵۹} باشد. به این معنی که صرف نظر از مقدار اولیه $X^{(0)}$ ، با احتمال مثبت زنجیر به هر ناحیه از \mathcal{X} دسترسی داشته باشد.

- برگشت‌پذیر^{۶۰} باشد. به این معنی که زنجیر به هر حالتی از فضای \mathcal{X} به هر تعدادی که مایل باشد، بتواند منتقل شود. به عبارت دیگر تعداد دفعات مورد انتظار بازگشت زنجیر برای یک مجموعه $A \subseteq \mathcal{X}$ نامتناهی باشد.

- نامتناوب^{۶۱} باشد. به این معنی که زنجیر در جستجوی فضای حالت، گرفتار هیچ چرخشی نشود. به عبارت دیگر زنجیر در یک مجموعه موضعی گیر نکند.

تحت این شرایط، توزیع ایستای π یک توزیع حدی هم خواهد بود. به این معنی که $X^{(t)}$ در توزیع π همگرا می‌شود. یعنی برای $t \rightarrow \infty$

$$X^{(t)} \xrightarrow{d} \pi$$

و نقطه شروع تاثیری بر این همگرایی ندارد. بنابراین احتمال $\pi(A) = \int_A \pi(x) dx$ را می‌توان به وسیله $p(X^{(t)} \in A)$ برای $A \subset \mathcal{X}$ تقریب زد. نتیجه این ویژگی همگرایی، قضیه ارگودیک است که بیان

^{۵۹}Irreducible

^{۶۰}Recurrent

^{۶۱}Aperiodic

می‌کند میانگین تجربی تابع دلخواه $h(x)$ حاصل از نمونه‌های وابسته تولیدی از زنجیر مارکوف، با احتمال متمایل به ۱، به مقدار واقعی مورد انتظار $E(h(X)) = \int_{\mathcal{X}} h(x)\pi(x)dx$ همگرا می‌شود. یعنی

$$\frac{1}{n} \sum_{t=1}^n h(X^{(t)}) \xrightarrow{a.s.} E(h(X)).$$

این قضیه مشابه قانون قوی اعداد بزرگ برای نمونه‌های مستقل و هم‌توزیع است. اکنون سوالی که مطرح می‌شود، چگونگی ساخت چنین زنجیرهای مارکوف است. در واقع، چطور زنجیر مارکوفی را بسازیم که توزیع ایستای آن، توزیعی باشد که مورد دلخواه ما است؟ نظریه روش‌های MCMC به این سوال پاسخ می‌دهد. دو الگوریتم معروف از روش‌های MCMC، الگوریتم‌های متروپولیس-هستینگز^{۶۲} (MH) و نمونه‌گیر گیبز^{۶۳} هستند که به دلیل استفاده شایع از آن‌ها، در ادامه آن‌ها را به اختصار معرفی می‌کنیم. وجود این معرفی به درک بهتر الگوریتم‌های MCMC نیز کمک می‌کند.

الگوریتم متروپولیس-هستینگز

الگوریتم MH ابتدا توسط متروپولیس و همکاران (۱۹۵۳) پیشنهاد شد و سپس توسط هستینگز (۱۹۷۰) به‌طور کامل تدوین یافت. تعریفی از این الگوریتم در چیب و گرینبرگ (۱۹۹۵) آمده است. برای توضیح آن، پارامتر ψ را یک بعدی در نظر می‌گیریم. برای تعریف الگوریتم، مقدار اولیه (نقطه شروع زنجیر) را برابر $\psi^{(0)}$ قرار می‌دهیم. سپس t امین تکرار الگوریتم شامل مراحل زیر است:

۱. نمونه‌گیری یک مقدار منتخب ψ^* از توزیع پیشنهادی با تابع چگالی $q(\psi^*|\psi^{(t-1)})$.

۲. محاسبه نرخ پذیرش

$$r = \frac{p(\psi^*|y_{1:T})q(\psi^{(t-1)}|\psi^*)}{p(\psi^{(t-1)}|y_{1:T})q(\psi^*|\psi^{(t-1)})}.$$

۳. پذیرش ψ^* به‌عنوان مقدار $\psi^{(t)}$ با احتمال $\min(1, r)$ و رد آن با احتمال $1 - \min(1, r)$ و نسبت دادن مقدار قبلی $\psi^{(t-1)}$ برای $\psi^{(t)}$.

برای اجرای این الگوریتم، تابع چگالی پیشنهادی $q(\cdot|\cdot)$ طوری انتخاب می‌شود که تکیه‌گاه آن با تکیه‌گاه توزیع پسین یکی باشد، دم‌های آن از دم‌های توزیع پسین پهن‌تر و تولید نمونه از آن ساده باشد. معمولاً انتخاب یک تابع چگالی متقارن ارجحیت دارد که در این صورت $q(\theta^*|\theta^{(t-1)}) = q(\theta^{(t-1)}|\theta^*)$. برای مثال، یک توزیع یکنواخت یا توزیع نرمالی که حول مقدار قبلی مرکزی شده باشد، می‌تواند نامزد انتخاب توزیع پیشنهادی باشد. یعنی

$$q(\theta^*|\theta^{(t-1)}) \sim U(\theta^{(t-1)} - \gamma, \theta^{(t-1)} + \gamma) \quad \text{یا}$$

$$q(\theta^*|\theta^{(t-1)}) \sim N(\theta^{(t-1)}, \gamma^2)$$

^{۶۲}Metropolis-Hastings

^{۶۳}Gibbs sampler

که در آن پارامتر γ پارامتر تنظیم یا مقیاس است و با توجه به عملکرد الگوریتم انتخاب می‌شود. به الگوریتم MH با این انتخاب توزیع پیشنهادی، الگوریتم MH قدم زدن تصادفی^{۶۴} می‌گویند. برای مرحله تصمیم‌گیری برای پذیرش یا رد نمونه پیشنهادی، مرحله ۲، در عمل یک عدد تصادفی u از توزیع یکنواخت $U(0, 1)$ تولید می‌شود. آن‌گاه اگر $u < r$ تصمیم پذیرش و در غیر این صورت رد اتخاذ می‌شود. با اجرای این الگوریتم یک زنجیر مارکوف تولید می‌شود که توزیع ایستای آن $p(\psi|y_{1:T})$ خواهد بود (رابرت و کسلا، ۲۰۰۴).

نمونه‌گیر گیز

روش نمونه‌گیری گیز ابتدا توسط گمان و گمان (۱۹۸۴) در پردازش مدل‌های تصویر معرفی شد. سپس توسط گل‌فاند و اسمیت (۱۹۹۰) به عنوان روشی بر مبنای نمونه‌گیری برای محاسبه چگالی‌های حاشیه‌ای در چارچوب استنباط بیزی پیشنهاد شد. توضیح واضح و ساده این الگوریتم به وسیله کسلا و جورج (۱۹۹۲) بیان شده است. توجه کنید زمانی که توزیع شرطی کامل به عنوان توزیع پیشنهادی انتخاب شود، نمونه‌گیر گیز یک حالت خاص الگوریتم MH است.

برای تشریح این روش، فرض کنید علاقه‌مندیم از توزیع پسین $p(\psi|y_{1:T})$ که $\psi = \{\psi_1, \dots, \psi_p\}$ یک پارامتر p بعدی است، نمونه تولید کنیم. برای اجرای نمونه‌گیر گیز باید برای هر پارامتر ψ_i توزیع شرطی کامل یعنی توزیع پسین شرطی ψ_i به شرط همه پارامترهای دیگر را که با $p(\psi_i|\psi_{i \neq j}, y_{1:T})$ برای $i, j = 1, \dots, p$ مشخص می‌شود، به دست آوریم. برای شبیه‌سازی از توزیع توام $p(\psi|y_{1:T})$ ، نمونه‌گیر گیز مقادیر را به‌طور تکراری از تمام توزیع‌های شرطی کامل تولید می‌کند. برای شروع مقادیر اولیه $\psi^{(0)} = \{\psi_1^{(0)}, \dots, \psi_p^{(0)}\}$ را انتخاب می‌کنیم. آن‌گاه در t امین تکرار، به شرط داشتن نمونه $\psi^{(t-1)}$ که در گام قبلی به دست آمده است، به صورت زیر عمل می‌کنیم:

$$1. \text{ تولید نمونه } \psi_1^{(t)} \text{ از توزیع شرطی کامل } p(\psi_1|\psi_2^{(t-1)}, \dots, \psi_p^{(t-1)}, y_{1:T})$$

$$2. \text{ تولید نمونه } \psi_2^{(t)} \text{ از توزیع شرطی کامل } p(\psi_2|\psi_1^{(t)}, \psi_3^{(t-1)}, \dots, \psi_p^{(t-1)}, y_{1:T})$$

.

.

.

$$p. \text{ تولید نمونه } \psi_p^{(t)} \text{ از توزیع شرطی کامل } p(\psi_p|\psi_1^{(t)}, \dots, \psi_{p-1}^{(t)}, y_{1:T}) \text{ و قرار دادن } \psi^{(t+1)} = \{\psi_1^{(t+1)}, \dots, \psi_p^{(t+1)}\}.$$

تحت شرایط نظم، ثابت می‌شود t امین تحقق $\psi^{(t)} = \{\psi_1^{(t)}, \dots, \psi_p^{(t)}\}$ در توزیع به توزیع هدف $\pi = p(\psi|y_{1:T})$ همگرا است (رابرت و کسلا، ۲۰۰۴، فصل ۱۰).

^{۶۴}Random walk MH

۳.۵.۱ اجرای روش‌های MCMC

روش‌های MCMC ابزارهای محاسباتی قدرتمندی برای استنباط بیزی هستند. مثلاً نمونه‌گیر گیز و الگوریتم MH به شدت انعطاف‌پذیر هستند و برای برآورد مدل‌های پیچیده که شامل تعداد پارامترهای زیاد یا توزیع‌های پسین پیچیده هستند، قابل به‌کارگیری می‌باشند. این سهولت اجرا زمانی که به جای توسعه کد R، بسته‌های R به کار می‌روند، افزایش می‌یابد. کاربردی‌ترین نرم‌افزار مخصوصاً میان آماردان‌های کاربردی، WinBUGS (لون و همکاران، ۲۰۰۰؛ زوفراس، ۲۰۰۹) است، به طوری که بخشی از پروژه WinBUGS^{۶۵} برای تحلیل بیزی مدل‌های آماری پیچیده به وسیله روش‌های MCMC است. زمانی که WinBUGS به کار می‌رود، کاربر تنها باید مدل، داده‌ها و مقادیر اولیه را تعیین کند. سپس نرم‌افزار به طور خودکار روش‌های نمونه‌گیری را برای تولید مقادیر از توزیع پسین مدل مشخص شده، اجرا خواهد کرد. بسته‌ای بسیار مشابه با این نرم‌افزار برای اجرای استنباط بیزی به وسیله MCMC، بسته JAGS است. JAGS به عنوان موتوری برای زبان BUGS توسعه یافت و نقطه قوت آن این است که کاربران می‌توانند توابع، نمونه‌ها و توزیع‌های خودشان را تعریف کنند. نرم‌افزار دیگر BayesX است (بلیتز و همکاران، ۲۰۱۲) که استنباط بیزی دسته‌ای از مدل‌های رگرسیونی را بر مبنای MCMC اجرا می‌کند. WinBUGS، JAGS و BayesX را می‌توان به ترتیب با بسته‌های R2WinBUGS، R2JAGS و R2BayesX در نرم‌افزار R به کار برد. از طریق این بسته‌ها می‌توان از توابع و امکانات این سه نرم‌افزار در R استفاده کرد.

۴.۵.۱ ضرورت معرفی رهیافت‌های جانشین

در اجرای روش‌های MCMC، می‌بایست توجه خاصی به تحلیل خروجی MCMC داشت. اگرچه نظریه MCMC تضمین می‌کند با تعداد تکرار بی‌شمار، توزیع نمونه‌های شبیه‌سازی شده به توزیع هدف (یعنی توزیع پسین) همگرا است، اما اجرای نامتناهی یک زنجیر مارکوف عملی نیست. پس باید الگوریتم به تعدادی اجرا شود که همگرایی به توزیع ایستا نتیجه شود. بنابراین تشخیص همگرایی زنجیر تولیدشده یک مساله اساسی در کاربردهای الگوریتم‌های MCMC است. متأسفانه حتی زمانی که مشکلی مشاهده نمی‌شود، تضمینی از همگرایی زنجیر تولیدشده متناهی وجود ندارد. ابزارهایی برای تشخیص همگرایی، شامل نمودارهای خودهمبستگی و اثر^{۶۶}، نرخ پذیرش الگوریتم، و برخی آزمون‌ها مانند جی‌وک (جی‌وک، ۱۹۹۲)، گلن-روبین (گلن و روبین، ۱۹۹۲) و رفتری-لوئیس (رفتری و لوئیس، ۱۹۹۲) ارائه شده‌اند. آزمون‌های ذکرشده در بسته coda (پلامر و همکاران، ۲۰۰۶) در R در دسترس هستند.

خوب است بدانید نمودار اثر ابزاری ساده است که می‌تواند برای تشخیص آمیختگی^{۶۷} زنجیر نیز به کار رود. این ویژگی مربوط به سرعت پیمایش توزیع پسین توسط زنجیر مارکوف است. به عنوان مثال، اگر آمیختگی زنجیر ضعیف باشد، به معنی سرعت کند پیمایش زنجیر و در نتیجه همبستگی شدید نمونه‌های تولیدشده است. در این حالت زمان زیادی برای دستیابی به همگرایی صرف خواهد شد. علاوه بر این همبستگی روی دقت تقریب‌های کمیت‌های پسینی موثر است. گِیر (۱۹۹۲) نشان داد که همبستگی

^{۶۵}Bayesian inference using Gibbs sampling

^{۶۶}Trace

^{۶۷}Mixing

بالتر نمونه‌های تولیدشده، یک دلیل دقت پایین‌تر تقریب توزیع پسین است. خودهمبستگی را می‌توان به وسیله باریک کردن مجموعه نقاط شبیه‌سازی‌شده یا بهبود الگوریتم بهینه (مثلاً از طریق بازپارامتری کردن نمونه‌گیری گیبز یا به وسیله تنظیم توزیع پیشنهادی MH) کاهش داد. زمانی که استنباط به روش MCMC اجرا می‌شود، حجم بالایی از انرژی و زمان برای تنظیم و مشاهده همگرایی اختصاص می‌یابد. به منظور اطلاع از تنظیمات بهتر (پارامتری کردن، توزیع‌های پیشین، مقادیر اولیه و توزیع‌های پیشنهادی MH) که خروجی دقیق‌تر و مطمئن‌تری برای MCMC به وجود می‌آورد، به بروکس و همکاران (۲۰۱۱) و مراجع مربوط به اجرای MCMC مراجعه کنید.

جنبه بحرانی دیگری که نمی‌توان آن را نادیده گرفت، مشکل هزینه محاسباتی روش‌های MCMC است. زمانی‌که با مدل‌های پیچیده (مخصوصاً مدل‌های سلسه‌مراتبی) یا مجموعه داده‌های حجیم سروکار داریم، الگوریتم‌های MCMC ممکن است به شدت کند و حتی غیرعملی باشند. سنگینی این محاسبات مخصوصاً در مورد مدل‌های پویای زمانی، فضایی^{۶۸} و فضایی-زمانی^{۶۹} به وضوح مستند است و معمولاً به عنوان "مشکل n بزرگ" مشهور است (بنرجی و همکاران، ۲۰۰۴؛ جونا لاسینیو و همکاران، ۲۰۱۲). با توجه به مشکلات (محاسباتی) الگوریتم‌های مبتنی بر نمونه‌گیری MCMC، در سال‌های اخیر محققان متعددی به دنبال یافتن روش‌های جانشین بوده‌اند. یکی از رهیافت‌های جانشین مناسب برای روش‌های MCMC که مورد نظر ما در این پایان‌نامه است و هزینه‌های محاسباتی استنباط بیزی را به شدت کاهش می‌دهد، روش INLA است که توسط رو و همکاران (۲۰۰۹) معرفی شد و در فصل بعدی معرفی می‌شود.

^{۶۸}Spatial

^{۶۹}Spatio-temporal

فصل ۲

تقریب لاپلاس آشیانه‌ای جمع‌بسته

معمولاً برای استنباط بیزی مدل‌های رگرسیونی پویا که زیر مجموعه‌ای از مدل‌های گوسی پنهان هستند از الگوریتم‌های MCMC استفاده می‌کنیم، اما وجود پارامترهای زیاد در ساختار سلسله‌مراتبی این مدل‌ها، مخصوصاً زمانی که بُعد میدان پنهان^۱ بزرگ باشد، همچنین وجود همبستگی بین عناصر میدان پنهان ویژگی‌های همگرایی و آمیختگی این الگوریتم‌ها را با مشکل مواجه می‌کند. اگرچه استفاده از راهبرد هم‌گام‌سازی بلوکی (هلد و رو، ۲۰۰۲ و رو و هلد، ۲۰۰۵) می‌تواند تا حدی این مشکلات را مرتفع کند، اما مشکل زمان‌بر بودن محاسبات در این الگوریتم‌ها، کماکان باقی است. رو و همکاران (۲۰۰۹) روش تقریب لاپلاس آشیانه‌ای جمع‌بسته^۲ (INLA) را معرفی کردند، که در استنباط مدل‌های گوسی پنهان^۳ (LGMs) جایگزین مناسبی برای الگوریتم‌های MCMC است به طوری که در مدت زمان بسیار کوتاه، تقریبی قابل قبول را جایگزین شبیه‌سازی‌های سنگین می‌کند.

در این فصل ابتدا به معرفی مدل‌های گوسی پنهان می‌پردازیم، سپس استنباط بیزی تقریبی LGMs را با استفاده از تقریب لاپلاس آشیانه‌ای جمع‌بسته بیان می‌کنیم. در ادامه مراحل این روش را به همراه جزئیات محاسبات تقریب‌های چگالی حاشیه‌ای میدان پنهان و ابرپارامترها بیان می‌کنیم، همچنین ملاک‌های ارزیابی تقریب حاشیه‌ای درست‌نمایی، انحراف اطلاع و ملاک پیش‌گویی شرطی مؤلفه‌ها برای مقایسه‌ی مدل‌ها معرفی می‌شود و در پایان این فصل بسته INLA - R را برای استنباط بیزی معرفی می‌کنیم.

۱.۲ مدل‌های گوسی پنهان

اولین گام در تعریف یک مدل گوسی پنهان همراه با چارچوب بیزی، تشخیص توزیع داده‌های مشاهده‌شده $y = (y_1, \dots, y_n)$ است. پرکاربردترین روش تشخیص توزیع y_i با پارامتر ϕ_i (معمولاً میانگین $E(y_i)$ ،

^۱Latent field

^۲Integrated Nested Laplace Approximations

^۳Latent Gaussian Models

تابع پیش‌گوی جمعی η_i از طریق تابع پیوند $g(\cdot)$ ، مانند $g(\phi_i) = \eta_i$ ، است. پیش‌گوی خطی جمعی η_i به صورت زیر تعریف می‌شود:

$$\eta_i = \beta_0 + \sum_{m=1}^M \beta_m x_{mi} + \sum_{l=1}^L f_l(z_{li}), \quad i = 1, \dots, T \quad (1.2)$$

که در آن β_0 عرض از مبدا است؛ ضرایب $\beta = \{\beta_1, \dots, \beta_M\}$ اثرات ثابت خطی متغیرهای تبیینی $\mathbf{x} = (x_1, \dots, x_M)$ روی متغیر پاسخ است؛ و $f = \{f_1(\cdot), \dots, f_L(\cdot)\}$ مجموعه‌ای از توابع است که در جملات بردار متغیرهای تصادفی و پنهان $\mathbf{z} = (z_1, \dots, z_L)$ تعریف شده‌اند. این جملات $f_l(\cdot)$ را می‌توان به صورت‌های مختلفی، مانند هموارسازی و اثرات غیرخطی متغیرهای تبیینی، اثرات فصلی و روندهای زمان، عرض از مبدا تصادفی و همچنین اثرات تصادفی فضایی یا زمانی فرض کرد. به همین دلیل است که این دسته از مدل‌های گوسی پنهان بسیار انعطاف‌پذیر هستند و می‌توانند دسته وسیعی از مدل‌ها از مدل‌های خطی و تعمیم‌یافته پویا تا مدل‌های فضایی و فضایی-زمانی (مارتینز و همکاران، ۲۰۱۳) را شامل شوند.

کلیه مؤلفه‌های پنهان (غیرمشاهده‌شده) مورد نظر برای استنباط در مجموعه‌ی پارامترها با نام θ که به صورت $\theta = \{\beta_0, \beta, f\}$ تعریف شده، جمع‌آوری می‌شود. به علاوه، بردار m ابرپارامتر به صورت $\psi = \{\psi_1, \dots, \psi_m\}$ مشخص می‌شود. با فرض استقلال شرطی، توزیع n مشاهده (کلیه مشاهدات از خانواده توزیع مشابه‌ای می‌آیند) به وسیله درستی‌نمایی زیر به دست می‌آید

$$p(\mathbf{y}|\theta, \psi) = \prod_{i=1}^n p(y_i|\theta_i, \psi) \quad (2.2)$$

که در آن هر داده‌ی y_i تنها به یک عضو θ_i در میدان پنهان θ مرتبط است. در مارتینز و همکاران (۲۰۱۳) این فرض که هر مشاهده می‌تواند با ترکیب خطی عضوهای θ مرتبط باشد، بحث می‌شود؛ همچنین آن‌ها موردی را که داده‌ها به چندین توزیع یعنی درستی‌نمایی‌های چندگانه تعلق دارند، مد نظر قرار دادند.

فرض می‌کنیم پیشین θ دارای توزیع نرمال چندمتغیره با میانگین \circ و ماتریس دقت $\mathbf{Q}(\psi) = \Sigma_{\psi}^{-1}$ یعنی $\theta \sim N(\circ, \mathbf{Q}(\psi))$ با تابع چگالی زیر باشد

$$p(\theta|\psi) = (2\pi)^{-n/2} |\mathbf{Q}(\psi)|^{1/2} \exp\left(-\frac{1}{2} \theta' \mathbf{Q}(\psi) \theta\right). \quad (3.2)$$

مؤلفه‌های میدان گوسی پنهان θ فرض می‌شود به‌طور شرطی مستقل هستند، با این نتیجه که $\mathbf{Q}(\psi)$ ماتریس دقت تنک^۴ است. این ویژگی با عنوان میدان تصادفی مارکوف گوسی^۵ (GMRF)، شناخته

^۴ اگر θ_i و θ_j به‌طور شرطی به شرط مؤلفه‌های دیگر θ_{-ij} مستقل باشند، توزیع شرطی توأم به صورت زیر به دست می‌آید:

$$p(\theta_i, \theta_j|\theta_{-ij}) = p(\theta_i|\theta_{-ij})p(\theta_j|\theta_{-ij}).$$

و می‌نویسیم $\theta_i \perp \theta_j|\theta_{-ij}$. این ویژگی استقلال شرطی الگوی صفر ماتریس دقت $\mathbf{Q}(\psi)$ را بیان می‌کند، چون برای یک جفت i و j که $i \neq j$ ، داریم درایه ماتریس دقت متناظر با این زوج برابر صفر است:

$$\theta_i \perp \theta_j \iff Q_{ij}(\psi) = 0.$$

^۵Gaussian markov random field

می‌شود (رو و هلد، ۲۰۰۵). توجه کنید زمانی که استنباط با GMRF به‌کار می‌رود، تنک بودن ماتریس دقت باعث فواید محاسباتی می‌شود. در واقع، عملیات جبرخطی برای ماتریس‌های تنک به‌کار می‌روند که به محاسبات سرعت قابل توجهی می‌دهند (رو و هلد، ۲۰۰۵).

توزیع پسین توأم θ و ψ به وسیله ضرب درست‌نمایی (۲.۲) و چگالی GMRF (۳.۲) و توزیع پیشین ابرپارامتر $p(\psi)$ به‌دست می‌آید:

$$\begin{aligned} p(\theta, \psi | \mathbf{y}) &\propto p(\psi) \times p(\theta | \psi) \times p(\mathbf{y} | \theta, \psi) & (۴.۲) \\ &\propto p(\psi) \times p(\theta | \psi) \times \prod_{i=1}^n p(y_i | \theta_i, \psi) \\ &\propto p(\psi) \times |\mathbf{Q}(\psi)|^{1/2} \exp\left(-\frac{1}{\psi} \theta' \mathbf{Q}(\psi) \theta\right) \times \prod_{i=1}^n \exp(\log(p(y_i | \theta_i, \psi))) \\ &\propto p(\psi) \times |\mathbf{Q}(\psi)|^{1/2} \exp\left(-\frac{1}{\psi} \theta' \mathbf{Q}(\psi) \theta + \sum_{i=1}^n \log(p(y_i | \theta_i, \psi))\right). \end{aligned}$$

۲.۲ تقریب لاپلاس آشیانه‌ای جمع‌بسته

اهداف استنباط بیزی به‌دست آوردن توزیع‌های پسین حاشیه‌ای برای هر عضو بردار پارامتر

$$p(\theta_i | \mathbf{y}) = \int p(\theta_i, \psi | \mathbf{y}) d\psi = \int p(\theta_i | \psi, \mathbf{y}) p(\psi | \mathbf{y}) d\psi \quad (۵.۲)$$

و بردار ابرپارامتر

$$p(\psi_k | \mathbf{y}) = \int p(\psi | \mathbf{y}) d\psi_{-k} \quad (۶.۲)$$

هستند. بنابراین باید مراحل زیر را اجرا کنیم:

(۱) $p(\psi | \mathbf{y})$ را محاسبه کرده و سپس با استفاده از این چگالی، چگالی‌های حاشیه‌ای $p(\psi_k | \mathbf{y})$ را به‌دست می‌آوریم؛

(۲) $p(\theta_i | \psi, \mathbf{y})$ را محاسبه می‌کنیم که با استفاده از این چگالی، می‌توان پسین‌های حاشیه‌ای $p(\theta_i | \mathbf{y})$ را به‌دست آورد؛

(۳) انتگرال توزیع پسین میدان پنهان را با استفاده از جمع متناهی محاسبه می‌کنیم

$$\tilde{\pi}(\theta_i | \mathbf{y}) = \sum_k \tilde{\pi}(\theta_i | \psi_k, \mathbf{y}) \tilde{\pi}(\psi_k | \mathbf{y}) \Delta_k. \quad (۷.۲)$$

روش INLA فرض‌هایی از مدل را برای تقریب عددی پسین‌های مورد نظر بر مبنای روش تقریب لاپلاس که در زیربخش ۱.۵.۱ معرفی شد، به‌کار می‌برد (تیرنی و کدین، ۱۹۸۶).

۱.۲.۲ تقریب چگالی پسین توأم بردار ابرپارامتر با راهبرد طرح مرکب مرکزی

چگالی پسین توأم بردار ابرپارامترهای مدل به صورت

$$\begin{aligned}\pi(\boldsymbol{\psi}|\mathbf{y}) &= \frac{\pi(\boldsymbol{\psi}, \mathbf{y})}{\pi(\mathbf{y})} \\ &= \frac{\pi(\boldsymbol{\psi}, \mathbf{y}) \pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})}{\pi(\mathbf{y}) \pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \\ &= \frac{\pi(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y})}{\pi(\mathbf{y}) \pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \\ &= \frac{\pi(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \pi(\boldsymbol{\theta}|\boldsymbol{\psi}) \pi(\boldsymbol{\psi})}{\pi(\mathbf{y}) \pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \\ &\propto \frac{\pi(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \pi(\boldsymbol{\theta}|\boldsymbol{\psi}) \pi(\boldsymbol{\psi})}{\pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})}\end{aligned}\quad (۸.۲)$$

است. تقریب لاپلاس آشیانه‌ای جمع‌بسته معادله (۸.۲) را برای هر مقدار $\boldsymbol{\psi}$ به صورت

$$\tilde{\pi}(\boldsymbol{\psi}|\mathbf{y}) \propto \frac{\pi(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}) \pi(\boldsymbol{\theta}|\boldsymbol{\psi}) \pi(\boldsymbol{\psi})}{\tilde{\pi}_G(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\psi})} \quad (۹.۲)$$

تقریب می‌زند، که در آن مخرج کسر با استفاده از تقریب گوسی محاسبه شده است. در تقریب گوسی با استفاده از بسط تیلور، توزیع شرطی کامل

$$\pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y}) \propto \exp\left(-\frac{1}{\nu} \boldsymbol{\theta}^T \mathbf{Q} \boldsymbol{\theta} + \sum_{i \in \mathcal{I}} \log \pi(y_i|\theta_i, \boldsymbol{\psi})\right)$$

حول مُد توزیع شرطی کامل $\boldsymbol{\theta}$ یعنی $\boldsymbol{\theta} = (\theta_1^*(\boldsymbol{\psi}), \dots, \theta_n^*(\boldsymbol{\psi}))$ به صورت

$$\begin{aligned}\tilde{\pi}_G(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\psi}) &\propto \exp\left(-\frac{1}{\nu} \boldsymbol{\theta}^T \mathbf{Q} \boldsymbol{\theta} + \sum_{i \in \mathcal{I}} g_i(\theta_i)\right) \\ &\propto \exp\left(-\frac{1}{\nu} \boldsymbol{\theta}^T \mathbf{Q} \boldsymbol{\theta} + \sum_{i \in \mathcal{I}} (a_i + b_i \theta_i - \frac{1}{\nu} c_i \theta_i^2)\right) \\ &\propto \exp\left(-\frac{1}{\nu} \boldsymbol{\theta}^T (\mathbf{Q} + \text{diag}(\mathbf{c})) \boldsymbol{\theta} + \mathbf{b}^T \boldsymbol{\theta}\right)\end{aligned}\quad (۱۰.۲)$$

به‌دست می‌آید که در آن $c_i = -g''(\theta_i^*(\boldsymbol{\psi}))$ و $b_i = g'(\theta_i^*(\boldsymbol{\psi})) + \theta_i^*(\boldsymbol{\psi}) c_i$ است و a_i از تناسب حذف می‌شود (رو و هلد $2^\circ 5$)، سپس تقریب لاپلاس چگالی پسین $\pi(\boldsymbol{\psi}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y})}{\pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})}$ با بسط مخرج کسر حول $\boldsymbol{\theta} = \boldsymbol{\theta}^*(\boldsymbol{\psi})$ به صورت

$$\tilde{\pi}_{LA}(\boldsymbol{\psi}|\mathbf{y}) \propto \frac{\pi(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y})}{\tilde{\pi}_G(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\psi})} \quad (۱۱.۲)$$

محاسبه می‌شود.

تقریب $\tilde{\pi}_{LA}(\psi|y)$ در سه مرحله از فرآیند استنباطی روش INLA مورد استفاده قرار می‌گیرد. استفاده اصلی آن در محاسبه تقریب بیزی پسین میدان پنهان است. در مرحله دوم از (۹.۲) برای محاسبه تقریب حاشیه‌ای درست‌نمایی $\pi(y)$ استفاده می‌شود. مرحله سوم به منظور محاسبه چگالی حاشیه‌ای پسین برای برخی از بردار ابرپارامترها، ψ ، یعنی $\pi(\psi_j|y)$ به کار می‌رود. نحوه به کارگیری معادله (۹.۲) برای محاسبه سه مرحله‌ی بیان شده، شامل انتگرال‌گیری عددی روی محدوده‌ی چندبعدی از ψ است. لذا پیش از هر عملی ابتدا باید نقاط انتگرال‌گیری مناسب انتخاب شوند. رو و همکاران (۲۰۰۹) دو راهبرد توری و طرح مرکب مرکزی یا CCD^۶ را برای انتخاب نقاط انتگرال‌گیری معرفی کردند. راهبرد توری شامل انتخاب نقاطی از یک توری است که چگالی $\tilde{\pi}_{LA}(\psi|y)$ حجم بیشتری در آن نقاط دارد. زمانی که بُعد ψ بزرگ نباشد (کوچکتر از ۶) این شیوه بسیار دقیق و مناسب است (رو و همکاران، ۲۰۰۹). در مواردی که بُعد ψ بین ۶ تا ۱۲ باشد راهبرد CCD مفیدتر است در هر دو راهبرد فرض می‌شود $\tilde{\pi}_{LA}(\psi|y)$ تک مُدی است، چرا که در فرآیند انتخاب نقاط انتگرال‌گیری به مُد $\tilde{\pi}_{LA}(\psi|y)$ نیاز است که با ψ^* نمایش داده می‌شود و می‌توان آن را با استفاده از الگوریتم‌های بهینه‌سازی نظیر نیوتون-رافسون یا اگر تابع مورد نظر بسیار پیچیده باشد با استفاده از روش شبه نیوتون^۷ محاسبه کرد. در این بخش روش CCD برای انتخاب نقاط انتگرال‌گیری شرح داده می‌شود.

۲.۲.۲ روش طرح مرکب مرکزی

در این بخش به تعریف طرح مرکب مرکزی می‌پردازیم که هزینه‌ی محاسبات را حتی برای m های بزرگ کاهش می‌دهد. در روش CCD، انتگرال به عنوان مسئله طرح رویه‌ی پاسخ در نظر گرفته می‌شود و نقاطی روی فضای m بُعدی پخش و بر اساس پاسخ‌های اندازه‌گیری شده در آن نقاط، خمیدگی پسین حول مُد برآورد می‌شود. در این بخش تنها سطوح پاسخ از مرتبه دو را در نظر گرفته و از طرح درجه‌ی دو CCD استفاده می‌شود (باکس و ویلسون، ۱۹۵۱). قبل از ارائه‌ی جزئیات این روش به اختصار مسائل طرح معرفی می‌شود.

معمولاً در زمینه‌های تحقیقاتی مختلف، محققان آزمایش‌هایی برای کشف چیزی درباره‌ی فرآیند یا سیستمی خاص انجام می‌دهند. لازم است آزمایش‌گر عوامل متغیر در آزمایش، دامنه‌ی تغییرات هر عامل و سطوح خاصی که در هر اجرا باید اعمال شوند را انتخاب کند تا بتواند علل تغییرات در پاسخ را مشاهده و تشخیص دهد. در دنیای علم، طرح آزمایش ابزاری مهم برای اصلاح عملکرد فرآیندهای تولید است و منظور از طرح آزمایش آماری، طرح ریزی فرآیندی است که بتوان داده‌های مناسبی به روش‌های آماری جمع‌آوری و تحلیل کرد. طرح‌های k عاملی که هر عامل تنها دارای دو سطح کمی یا کیفی باشد را طرح عاملی 2^k گویند. با افزایش تعداد عامل‌های این طرح، تعداد اجراهای مورد نیاز برای پاسخ کامل طرح به سرعت افزایش می‌یابد. اگر بتوان به شیوه‌ای معقول ناچیز بودن بعضی اثرهای متقابل مراتب بالا را پذیرفت، آنگاه می‌توان اطلاعات مربوط به اثرهای اصلی و اثرهای متقابل مراتب پایین را تنها با اجرای کسری از طرح آزمایش عاملی کامل به دست آورد، به این آزمایش‌ها طرح‌های عامل

^۶Central Composite Design

^۷Quasi Newton

کسری^۸ گویند. اگر در این طرح‌ها اثرهای اصلی و اثرهای متقابل دو عاملی با اثرهای اصلی و اثرهای متقابل دو عاملی دیگر هم اثر نباشند به آن طرح V -تجزیه^۹ گویند. اگر تعداد عامل‌ها در این دو طرح زیاد باشد تعیین نقاط طرح و تعداد اجرا بسیار دشوار است. به عنوان مثال بزرگترین طرح عامل کسری V -تجزیه در 2^{10-3} و در 2^{11-4} است. سانچز و سانچز (۲۰۰۵) با به‌کار بردن ماتریس مرتب شده‌ی هادامارد^{۱۰} (بیوچمپ، ۱۹۸۴) و اندیس والش^{۱۱}، روشی دیگر برای تعیین طرح‌های عاملی کسری، حتی با 12^0 عامل، ارائه دادند که در ادامه به اختصار بیان می‌شود. ماتریس هادامارد H_k ، ماتریسی $2^k \times 2^k$ با درایه‌های $+1$ یا -1 است، که می‌توان آن را از روابط بازگشتی

$$H_0 = (1) \quad \text{و} \quad H_{k+1} = \begin{pmatrix} H_k & H_k \\ H_k & -H_k \end{pmatrix}$$

محاسبه کرد. اگر ستون‌های H_k با h_0, h_1, \dots, h_{N-1} مشخص شوند، که در آن $N = 2^k$ ، هر ستون h_i مربوط به اثر اصلی یا اثر متقابل دو عاملی است که اندیس والش آن i است. اندیس‌گذاری والش طرح عاملی کسری V -تجزیه با k عامل، با استفاده از الگوریتمی بازگشتی انجام می‌شود. این اندیس‌گذاری در k مرحله انجام می‌گیرد که در مرحله i ام آن، اندیس عامل X_i تعیین می‌شود. با فرض این‌که اندیس والش عامل‌های X_{j_1} و X_{j_2} به ترتیب a_1 و a_2 باشند، اندیس اثر متقابل بین این دو عامل، یعنی اثر $X_{j_1} X_{j_2}$ ، از رابطه $a_1 \oplus a_2$ به دست می‌آید، که در آن بعد از تغییر مبنای اعداد به دو، یای انحصاری \oplus ^{۱۲} با استفاده از جدول ۱.۲ محاسبه می‌شود. به‌عنوان مثال،

جدول ۱.۲: نقاط طرح هادامرد والش برای $k = 3$

A	B	$A \oplus B$
۰	۰	۰
۰	۱	۱
۱	۰	۱
۱	۱	۰

مجموعه A_i و T_i به صورت $6 \oplus 5 = (110)_2 \oplus (101)_2 = (011)_2 = 3$ در انتهای مرحله‌ی i ام، $i = 1, \dots, k$ ، دو

$$A_i = \{X_1, \dots, X_k \text{ های عامل اصلی اثرهای اصلی عامل‌های } X_1, \dots, X_k\}$$

$$T_i = \{A_i \text{ متناظر والش متناظر } A_i \text{ بین مولفه‌های اندیس والش متناظر } A_i\}$$

محاسبه می‌شوند که فرض می‌شود $A_0 = T_0 = \emptyset$. برای معتبر بودن اندیس والش، اندیس پیشنهادی عامل X_j که با b نمایش داده می‌شود باید در دو شرط زیر صدق کند:

^۸Fractional Factorial Design

^۹V-Resoulation

^{۱۰}Hadamard Ordered Matrix

^{۱۱}Walsh Index

^{۱۲}Exclusive Or

الف) b نباید متعلق به A_{j-1} و T_{j-1} باشد؛

ب) اندیس اثرهای متقابل دو عاملی بین X_j و X_1, \dots, X_{j-1} نباید متعلق به A_j و T_j باشد.

تعریف ۱.۰.۲.۲. (الگوریتم تعیین اندیس والش):

۱. $m = 1$ قرار می‌دهیم؛

۲. $b = \max\{i : i \in A_{m-1}\} + 1$ قرار داده می‌شود؛

۳. مادامی که $b \in T_{m-1}$ یا $\{\exists i \in A_{m-1} : b \oplus i \in A_{m-1} \cup T_{m-1}\}$ ، $b = b + 1$ قرار می‌دهیم؛

۴. مجموعه‌های $A_m = \{b\} \cup A_{m-1}$ و $T_m = T_{m-1} \cup \{i \oplus b : i \in A_m\}$ محاسبه می‌شوند؛

۵. قرار می‌دهیم $m = m + 1$ و مادامی که $m < k$ به گام ۲ برمی‌گردیم.

بعد از اندیس‌گذاری به روش والش هر ستون ماتریس هادمارد متعلق به یک اثر اصلی یا اثر متقابل دو عاملی خواهد بود و هر سطر ماتریس مشخص‌کننده سطوح عامل در یک اجرا خواهد بود که به آن نقاط طرح گویند. جدول ۲.۲.۲ نقاط طرح برای $k = 3$ را نشان می‌دهد. در این طرح‌ها به تعداد نقاط غیر تکراری، طرح نیاز به اجرا است. جدول ۲.۲.۲ تعداد اجرا در طرح‌های عاملی کسری را نشان می‌دهد.

جدول ۲.۲: نقاط طرح هادمارد والش برای $k = 3$

نقاط طرح	۱: X_1	۲: X_2	۳: X_1X_2	۴: X_3	۵: X_1X_3	۶: X_2X_3
۱	+	+	+	+	+	+
۲	-	+	-	+	-	+
۳	+	-	-	+	+	-
۴	-	-	+	+	-	-
۵	+	+	+	-	-	-
۶	-	+	-	-	+	-
۷	+	-	-	-	-	+
۸	-	-	+	-	+	+

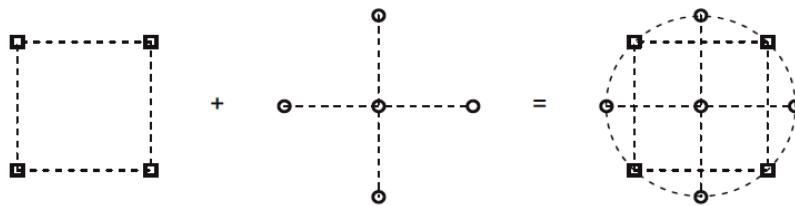
جدول ۳.۲: تعداد اجرا در طرح عاملی کسری

بعد ψ	۱-۲	۳	۴-۵	۶	۷-۸	۹-۱۱	۱۲-۱۷
تعداد نقاط	۴	۸	۱۶	۳۲	۶۴	۱۲۸	۲۵۶

تعیین نقاط و تعداد نقاط در طرح مرکب مرکزی

CCD شامل طرح عاملی کسری است که با یک نقطه مرکزی و یک گروه متشکل از $2m$ نقطه با عنوان نقاط ستاره‌ای ترکیب شده است. بنابراین، تعداد نقاط در این روش، یعنی k ، از مجموع تعداد طرح و $2m + 1$ نقطه جدید به دست می‌آید. به عنوان مثال، تعداد نقاط مورد نیاز در CCD برای $m = 5$ برابر $27 = 1 + (2 \times 5) + 16$ است که در مقایسه با تعداد نقاط در روش توری، یعنی $5^5 = 3125$ یا $243 = 3^5$ بسیار کم است.

در این روش تمام نقاط روی سطح کره m بُعدی به شعاع \sqrt{m} قرار می‌گیرند. نقاط ستاره‌ای در فاصله $\pm\sqrt{m}$ روی محورها و نقطه‌ی مرکزی در مبدأ قرار می‌گیرد. شکل ۱.۲ موقعیت نقاط در CCD برای



شکل ۱.۲: موقعیت نقاط در طرح مرکزی مرکب

$m = 2$ را نمایش می‌دهد. برای مشخص کردن وزن‌های انتگرال در (۷.۲)، فرض می‌شود که $\tilde{\pi}(\psi|y)$ گوسی استاندارد و انتگرال $\psi^T \psi$ برابر m است. با این شروط وزن انتگرال‌گیری نقاط روی کره با شعاع $f \cdot \sqrt{m}$ به صورت زیر محاسبه می‌شود

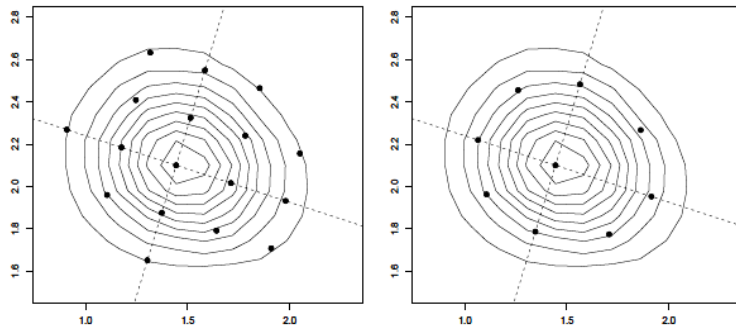
$$\Delta_k = \left[(k-1)(f_*^2 - 1) \left\{ 1 + \exp\left(-\frac{mf_*^2}{4}\right) \right\} \right]^{-1}$$

که در آن $f_* > 1$ مقداری ثابت است و وزن انتگرال‌گیری نقطه‌ی مرکزی برابر $(k-1)\Delta_k - 1$ است (باکس و ویلسون، ۱۹۵۱).

۳.۲.۲ تقریب چگالی حاشیه‌ای پسین درایه‌های ابرپارامتر

محاسبه چگالی حاشیه‌ای پسین ψ_k مستلزم انتگرال‌گیری از چگالی $\tilde{\pi}(\psi|y)$ نسبت به بردار θ_{-j} است (معادله (۶.۲) را مشاهده کنید)، اما اگر بُعد بردار ψ بزرگ باشد، محاسبه‌ی این انتگرال طاقت‌فرسا و زمان‌بر است. یک شیوه‌ی جایگزین انتگرال‌گیری از $\tilde{\pi}(\psi|y)$ با استفاده از نقاط انتگرال‌گیری است که حاصل از راهبرد توری یا CCD هستند. این روش نه تنها ساده‌تر است بلکه سرعت محاسبات را به مراتب افزایش می‌دهد.

در این بخش با بیان راهبرد توری، انتگرال‌گیری نسبت به ψ را برای محاسبه چگالی حاشیه‌ای پسین ابرپارامترها بیان می‌کنیم. همچنین دو الگوریتمی که برای محاسبه این چگالی‌ها با هزینه کمتری می‌توانند به کار روند را ارائه می‌کنیم، به طوری که به کمک نقاط توأم توزیع ابرپارامترها از طریق راهبرد توری، ارزیابی می‌شوند.



شکل ۲.۲: مکان نقاط انتگرال‌گیری به وسیله راهبرد توری (نمودار سمت چپ) و CCD (نمودار سمت راست).

راهبرد توری

در این بخش $\tilde{\pi}(\psi|\mathbf{y})$ برای انتگرال‌گیری نسبت به ψ به‌کار می‌رود، زمانی که $\tilde{\pi}(\theta_i|\mathbf{y})$ تقریب می‌شود. برای این عمل اگر نقاط ارزیابی مناسبی را برای حل عددی معادله (۶.۲) انتخاب کنیم، نیاز به کشف جزئیات $\tilde{\pi}(\psi|\mathbf{y})$ نداریم. رو و همکاران (۲۰۰۹) دو روش مختلف را برای دستیابی به این انتگرال پیشنهاد دادند.

هر دو روش نیاز به بازپارامتری کردن فضای ψ به منظور ساخت چگالی منظم‌تر دارد، این پارامتری کردن را به‌عنوان پارامتری کردن \mathbf{z} مشخص می‌کنیم. فرض کنید $\psi = (\psi_1, \dots, \psi_m) \in \mathbb{R}^m$ لذا داریم:

۱. مد ψ^* چگالی $\tilde{\pi}(\psi|\mathbf{y})$ را به‌دست آورده و ماتریس هسیان منفی \mathbf{H} را در این مد محاسبه می‌کنیم؛

۲. تجزیه مقدار ویژه $\Sigma = \mathbf{V}\Lambda^{1/2}\mathbf{V}^T$ را محاسبه می‌کنیم که در آن $\Sigma = \mathbf{H}^{-1}$.

۳. متغیر جدید \mathbf{z} را به صورت زیر تعریف می‌کنیم:

$$\psi(\mathbf{z}) = \psi^* + \mathbf{V}\Lambda^{1/2}\mathbf{z}$$

متغیر $\mathbf{z} = (z_1, \dots, z_m)$ استاندارد شده است و مؤلفه‌های آن دو به دو متعامدند.

اگر بعد ψ کوچک باشد، $m \leq 5$ ، رو و همکاران (۲۰۰۹)، این پارامتری کردن \mathbf{z} را برای ایجاد یک توری که بخش اعظم چگالی $\tilde{\pi}(\psi|\mathbf{y})$ پوشش می‌دهد، به‌کار بردند. این راهبرد هزینه محاسباتی دارد به‌طوری که با افزایش m بیشتر می‌شود.

اگر بعد ψ بزرگتر باشد از روش CCD که بیان شد، استفاده می‌کنیم. شکل ۲.۲ مکان نقاط انتگرال‌گیری با استفاده از راهبرد توری و CCD را نشان می‌دهد.

الگوریتم‌های محاسبه $\tilde{\pi}(\psi_k|\mathbf{y})$

اگر بعد ψ بزرگ نباشد، می‌توان $\tilde{\pi}(\psi|\mathbf{y})$ را روی یک توری منظم ارزیابی کرد و این نقاط برای محاسبه عددی انتگرال (۶.۲) به صورت جمع‌بسته روی متغیرهای ψ_{-k} استفاده می‌شود. البته این روش راه‌حلی

ساده است به طوری که هزینه دستیابی m چگالی حاشیه‌ای به طور نمایی تحت m افزایش می‌یابد. راه حل ساده‌تر استفاده از تقریب لاپلاس است:

$$\tilde{\pi}(\psi_k | \mathbf{y}) \approx \frac{\tilde{\pi}(\boldsymbol{\psi} | \mathbf{y})}{\tilde{\pi}_G(\boldsymbol{\psi}_{-k} | \psi_k, \mathbf{y})} \Big|_{\psi_{-k} = \psi_{-k}^*} \quad (12.2)$$

که در آن ψ_{-k}^* مد $\tilde{\pi}(\boldsymbol{\psi}_{-k} | \theta_k, \mathbf{y})$ است و $\tilde{\pi}_G(\boldsymbol{\psi}_{-k} | \psi_k, \mathbf{y})$ تقریبی گوسی می‌باشد. این تقریب نتایج دقیقی را به ما می‌دهد، اما برای این تقریب ماکسیمم تابع $(m - 1)$ بعدی $\pi(\boldsymbol{\psi}_{-k} | \psi_k, \mathbf{y})$ برای هر مقدار ψ_k مورد نیاز است به طوری که با m بعد این مسئله دوباره مقیاسی خوبی نیست. از سوی دیگر ماتریس هسیان محاسبه شده در مد $\pi(\boldsymbol{\psi}_{-k} | \psi_k, \mathbf{y})$ همواره مثبت نیست که مشکل اصلی می‌باشد. لازم به ذکر است که در مدل‌های گوسی پنهان مورد نظر، بعد میدان پنهان معمولاً بزرگ هستند که ارزیابی $\tilde{\pi}(\boldsymbol{\psi} | \mathbf{y})$ به وسیله معادله (۱۱.۲) هزینه سنگینی دارد. با این فرض، مفید است که الگوریتم‌های بسازیم و نقاط چگالی که در راهبرد توری $\tilde{\pi}(\boldsymbol{\psi} | \mathbf{y})$ ارزیابی شده‌اند، استفاده کنیم.

درون‌یابی گوسی نامتقارن

برخی چگالی‌های حاشیه‌ای $\pi(\psi_k | \mathbf{y})$ را می‌توان به وسیله تقریب توزیع توأم $\pi(\boldsymbol{\psi} | \mathbf{y})$ با یک توزیع نرمال چندمتغیره به کمک مد $\tilde{\pi}(\boldsymbol{\psi} | \mathbf{y})$ به دست می‌آیند. تقریب گوسی $\pi(\psi_k | \mathbf{y})$ محاسبات اضافی را ندارد، چون مد $\boldsymbol{\psi}^*$ و ماتریس هسیان منفی \mathbf{H} با استفاده از (۷.۲) تقریب شده است.

متأسفانه $\pi(\psi_k | \mathbf{y})$ ممکن است چوله باشد، در این صورت تقریب گوسی نامناسب است. می‌توانیم برای تصحیح تقریب گوسی با هزینه کمتر به صورت زیر عمل می‌کنیم.

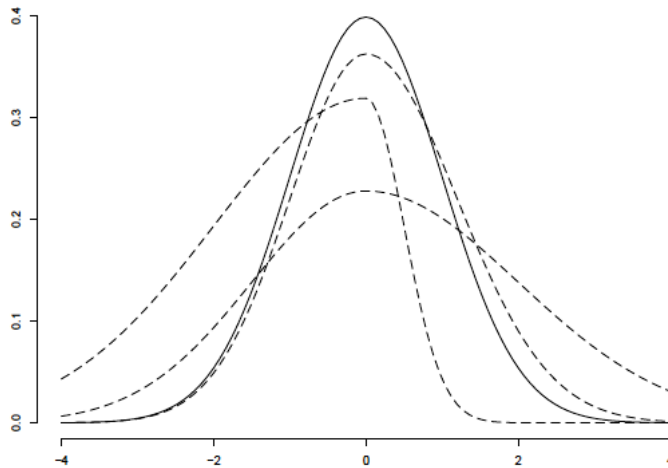
فرض کنید $\mathbf{z}(\boldsymbol{\psi}) = (z_1(\boldsymbol{\psi}), \dots, z_m(\boldsymbol{\psi}))$ نقاط در \mathbf{z} باشند. تابع $f(\boldsymbol{\psi})$ را به صورت زیر تعریف می‌کنیم:

$$f(\boldsymbol{\psi}) = \prod_{k=1}^m f_k(z_k(\boldsymbol{\psi})) \quad (13.2)$$

که در آن

$$f_k(z) \propto \begin{cases} \exp(-\frac{1}{\sqrt{(\sigma^{k+})^2}} z^2) & z \geq 0 \\ \exp(-\frac{1}{\sqrt{(\sigma^{k-})^2}} z^2) & z < 0 \end{cases} \quad (14.2)$$

به منظور غلبه بر نامتقارنی $\tilde{\pi}(\boldsymbol{\psi} | \mathbf{y})$ ، پارامترهای $(\sigma^{k+}, \sigma^{k-})$ ، $k = 1, \dots, m$ ، را طوری تعیین می‌کنیم که بر اساس m محور مختلف تغییر نکنند، اما بر طبق مسیر، مثبت و منفی، هر محور تغییر کنند. برای محاسبه این توابع ابتدا توجه داریم که در یک توزیع گوسی، در لگاریتم چگالی وقتی مد به ± 2 می‌رود، انحراف معیار -2 می‌شود. این پارامترها را در روشی که به طور دقیق برای کلیه مسیرها تقریب زده می‌شود، محاسبه می‌کنیم. سپس تقریب‌های $\pi(\psi_k | \mathbf{y})$ با انتگرال‌گیری عددی معادله (۱۳.۲) به دست می‌آیند، به طوری که محاسبه آن برای پارامترهای معلوم ساده است. شکل ۳.۲ انعطاف پذیری $f_k(z)$ را در معادله (۱۴.۲) برای نقاط مختلف σ^+ و σ^- نشان می‌دهد.



شکل ۳.۲: توزیع نرمال استاندارد (خط ممتد) و چگالی‌های معادله (۱۴.۲) برای مقادیر مختلف پارامترهای مقیاس (خطوط خط چین).

الگوریتم انتگرال‌گیری عددی

تقریب چگالی‌های حاشیه‌ای پسین $\tilde{\pi}(\psi_k | \mathbf{y})$ به وسیله الگوریتم انتگرال‌گیری عددی به صورت زیر است:

$$\tilde{\pi}(\psi_k | \mathbf{y}) = \begin{cases} N(0, \sigma_{k+}^2) & \psi_k > 0 \\ N(0, \sigma_{k-}^2) & \psi_k \leq 0 \end{cases} \quad (15.2)$$

و حال این معادله را برای محاسبه σ_{k+}^2 و σ_{k-}^2 ، $k = 1, \dots, m$ ، بدون محاسبه انتگرال‌گیری عددی به صورت الگوریتم قبلی به کار می‌رود. لم زیر برای این عمل مفید خواهد بود (رو و همکاران، ۲۰۰۹):

لم ۲.۲.۲. با فرض $\theta = (\theta_1, \dots, \theta_n)^T \sim N(0, \Sigma)$ برای θ_1 داریم

$$-\frac{1}{2}(\theta_1, E(\theta_{-1} | \theta_1)^T) \Sigma^{-1} \begin{pmatrix} \theta_1 \\ E(\theta_{-1} | \theta_1)^T \end{pmatrix} = -\frac{1}{2} \frac{\theta_1^2}{\Sigma_{11}}$$

لم بالا بیان می‌کند که توزیع توأم ψ به عنوان تابعی از ψ_i با ψ_{-i} ارزیابی شده در میانگین شرطی $E(\psi_{-i} | \psi_i)$ ، که به صورت چگالی حاشیه‌ای ψ_i رفتار می‌کند، این مطلب می‌تواند در هدف ما کاربرد داشته باشد. این محاسبه یک تقریب خواهد بود، چون θ گوسی نیست.

برای هر محور $k = 1, \dots, m$ این الگوریتم میانگین شرطی $E(\psi_{-k} | \psi_k)$ را با فرض این که ψ گوسی است، محاسبه خواهد شد به طوری که در ψ_j خطی است و تنها به مد ψ^* و کوواریانس Σ محاسبه شده در راهبرد توری، وابسته است و سپس ۲.۲.۲ را برای دستیابی به چگالی حاشیه‌ای پسین تقریب شده ψ_j در هر مسیر این محور استفاده می‌کنیم. برای هر مسیر این محور، فقط باید سه نقطه از این چگالی را که به وسیله ۲.۲.۲ تقریب شده‌اند، ارزیابی شوند به طوری که برای محاسبه مشتق دوم کافی است و انحراف معیارهای σ_k^- و σ_k^+ معادله (۱۵.۲) را به دست می‌دهد.

۴.۲.۲ تقریب چگالی حاشیه‌ای پسین عناصر میدان پنهان

پس از انتخاب نقاط انتگرال‌گیری $\{\psi_k\}$ باید چگالی حاشیه‌ای پسین θ_i ‌ها را محاسبه کنیم تا بتوان با استفاده از دو تقریب $\tilde{\pi}(\psi_k|y)$ و $\tilde{\pi}(\theta_i|\psi_k, y)$ چگالی (۵.۲) را تقریب زد. لذا در این بخش تقریب پسین $\pi(\theta_i|\psi_k, y)$ با استفاده از سه روش تقریب گوسی، لاپلاس و لاپلاس ساده شده^{۱۳} محاسبه و معایب و محاسن هریک بیان می‌شود.

الف) تقریب گوسی:

ساده‌ترین راه محاسبه تقریب $\pi(\theta_i|\psi_k, y)$ استفاده از $\tilde{\pi}_G(\theta|\psi_k, y)$ و استخراج حاشیه‌ای مورد نظر از آن می‌باشد. در مرحله انتخاب نقاط انتگرال‌گیری ψ_k و محاسبه‌ی $\tilde{\pi}(\psi|y)$ ، تقریب گوسی $\tilde{\pi}_G(\theta|\psi_k, y)$ محاسبه شد، بنابراین بردار میانگین معلوم است و تنها باید واریانس حاشیه‌ای را محاسبه کنیم. این واریانس را می‌توان از وارون معکوس ماتریس دقت Q به دست آورد. چون اغلب، بُعد میدان پنهان بزرگ است وارون کردن ماتریسی با ابعاد بزرگ دشوار است. رو و مارتینو (۲۰۰۷) رابطه بازگشتی را براساس معادلات تاکاهاشی^{۱۴} (تاکاهاشی و همکاران، ۱۹۷۳) معرفی کردند که واریانس حاشیه‌ای را برای GMRF‌های با ابعاد بزرگ بسیار سریع محاسبه می‌کند و نشان دادند تقریب

$$\tilde{\pi}_G(\theta_i|\psi_k, y) \sim \mathcal{N}(\mu_i(\psi_k), \sigma_i^2(\psi_k)) \quad (۱۶.۲)$$

اغلب برای $\pi(\theta_i|\psi_k, y)$ نتایج قابل قبولی را می‌دهد، که در آن $\sigma^2(\psi_k)$ بردار n بُعدی واریانس حاشیه‌ای است که در ادامه نحوه محاسبه‌ی آن را به تفصیل بیان می‌کنیم. اگر $Q = LL^T$ و $z \sim \mathcal{N}(0, I)$ باشد، آن‌گاه θ در معادله $L^T \theta = z$ دارای توزیع $\mathcal{N}(0, \Sigma)$ است. معادله‌ی $L^T \theta = z$ را می‌توان به صورت

$$L_{ii}\theta_i = z_i - \sum_{k=i+1}^n L_{ki}\theta_k, \quad i = 1, \dots, n$$

نوشت. با ضرب طرفین این معادله در θ_j به ازای $i \geq j$ و امید گرفتن از آن درایه‌های ماتریس Σ رابطه‌ی

$$\Sigma_{ij} = \frac{\delta_{ij}^2}{L_{ii}} - \frac{1}{L_{ii}} \sum_{k=i+1}^n L_{ki}\Sigma_{kj}, \quad j \geq i, i = n, \dots, 1 \quad (۱۷.۲)$$

محاسبه می‌شوند، که در آن $\delta_{ij} = 1$ ، اگر $i = j$ ؛ و $\delta_{ij} = 0$ ، اگر $i \neq j$ باشد. معادله‌ی (۱۷.۲) همان معادله تاکاهاشی است که به محاسبه‌ی سریع عناصر معینی از وارون یک ماتریس، با استفاده از تجزیه چولسکی آن کمک می‌کند، به این ترتیب که اگر $Q = LL^T = VDV^T$ باشد، که در آن $L = VD^{1/2}$ ، D یک ماتریس قطری و V یک ماتریس پایین مثلثی با درایه‌های قطری یک است، آن‌گاه

$$\Sigma = D^{-1}V^{-1} + (I - V^T)\Sigma \quad (۱۸.۲)$$

^{۱۳}Simplified Laplace Approximation

^{۱۴}Takahashi Equation

زیرا

$$\begin{aligned} Q\Sigma = I &\Rightarrow VDV^T\Sigma = I \\ &\Rightarrow (VD)^{-1}VDV^T\Sigma = (VD)^{-1} \\ &\Rightarrow IV^T\Sigma = D^{-1}V^{-1} \\ &\Rightarrow D^{-1}V^{-1} - V^T\Sigma = 0 \\ &\Rightarrow D^{-1}V^{-1} - V^T\Sigma + \Sigma = \Sigma \end{aligned}$$

شکل ۴.۲ نواحی غیرصفر را برای رابطه (۱۸.۲) نشان می‌دهد. همان‌طور که ملاحظه می‌کنید $D^{-1}V^{-1}$

$$\begin{aligned} \Sigma &= \begin{bmatrix} D^{-1} & & \\ & 0 & \\ & & 0 \end{bmatrix} * \begin{bmatrix} V^{-1} & & \\ & 1 & \\ & & 1 \end{bmatrix} + \left[\begin{bmatrix} I & & \\ & 1 & \\ & & 1 \end{bmatrix} - \begin{bmatrix} V^T & & \\ & 1 & \\ & & 1 \end{bmatrix} \right] * \Sigma \\ &= \begin{bmatrix} D^{-1} & & \\ & 0 & \\ & & 0 \end{bmatrix} + \begin{bmatrix} & & \\ & 0 & \\ & & \end{bmatrix} * \Sigma \end{aligned}$$

شکل ۴.۲: نمایش ماتریسی معادله تاکاهاشی

یک ماتریس پایین مثلثی است، به طوری که $(D^{-1}V^{-1})_{ii} = (D^{-1})_{ii}$. این امر موجب بی‌نیازی به محاسبه V^{-1} می‌شود، بنابراین درایه‌های ماتریس Σ به صورت

$$\Sigma_{ij} = D_{ij}^{-1} - \sum_{k>j}^n V_{kj} \Sigma_{ik}, \quad i \geq j$$

قابل محاسبه‌اند که پایه‌ای برای محاسبه واریانس‌های حاشیه‌ای θ_1 تا θ_n فراهم می‌کند.

(ب) تقریب لاپلاس:

خطای ناشی از تشخیص نادرست مُد یا در نظر نگرفتن چولگی توزیع، ممکن است تقریب گوسی را تحت تأثیر قرار دهد. برای رفع این مشکل تقریب لاپلاس به صورت

$$\tilde{\pi}_{LA}(\theta_i | \psi, \mathbf{y}) \propto \frac{\pi(\theta, \psi, \mathbf{y})}{\tilde{\pi}_{GG}(\theta_{-i} | \theta_i, \psi, \mathbf{y})} \Big|_{\theta_{-i} = \theta_{-i}^*(\theta_i, \psi)} \quad (19.2)$$

معرفی می‌شود، که در آن $\tilde{\pi}_{GG}$ تقریب گوسی و $\theta_{-i}^*(\theta_i, \psi)$ مُد چگالی شرطی $\pi(\theta_{-i} | \theta_i, \psi, \mathbf{y})$ است. توجه داشته باشید که $\tilde{\pi}_{GG}$ متفاوت از چگالی شرطی متناظر با $\tilde{\pi}(\theta | \psi, \mathbf{y})$ است. در $\tilde{\pi}$ ماتریس دقت نسبت به θ_i ثابت و میانگینش تابعی خطی از θ_i است. در حالی که در $\tilde{\pi}_{GG}$ ابتدا مُد $\theta_{-i}^*(\theta_i, \psi)$ و سپس لگاریتم درست‌نمایی حول آن محاسبه می‌شود که ماتریس دقت آن با تغییر θ_i تغییر می‌کند. از آن‌جا که چگالی $\tilde{\pi}_{GG}$ براساس شرطی کردن روی θ_i و سپس به‌کار بردن تقریب لاپلاس بر روی بقیه‌ی متغیرها محاسبه می‌شود، از تقریب (۱۶.۲) که تنها بر اساس تقریب گوسی است، دقیق‌تر است.

محاسبه‌ی عبارت (۱۹.۲) مستلزم محاسبه $\tilde{\pi}_{GG}$ برای θ_i و ψ است که بسیار زمان‌بر است. رو و همکاران (۲۰۰۹) دو تعدیل برای آسان کردن محاسبات ارائه کردند. در تعدیل اول به دلیل این‌که میانگین شرطی $E[\theta_{-i}|\theta_i, \psi, \mathbf{y}]$ و مُد شرطی $\theta_{-i}^*(\theta_i, \psi)$ به شرط گوسی بودن توزیع $(\theta_{-i}|\theta_i, \psi, \mathbf{y})$ بر هم منطبق خواهند بود و به دلیل در نظر گرفتن پیشین‌گوسی برای $\pi(\theta, \psi)$ ، گوسی بودن توزیع $(\theta_{-i}|\theta_i, \psi, \mathbf{y})$ خیلی دور از انتظار نیست. بنابراین با استفاده از

$$\theta_{-i}^*(\theta_i, \psi) \approx E_{\tilde{\pi}_G}(\theta_{-i}|\theta_i) \quad (20.2)$$

از میانگین به‌جای مُد استفاده و از گام بهینه‌سازی $\tilde{\pi}_{GG}(\theta_{-i}|\theta_i, \psi, \mathbf{y})$ صرف نظر می‌کنیم. هاسیو و همکاران (۲۰۰۴) نشان دادند که استفاده از $E_{\tilde{\pi}_G}(\theta_{-i}|\theta_i)$ به‌جای $\theta_{-i}^*(\theta_i, \psi)$ نتایجی تقریباً مشابه را می‌دهد.

در تعدیل دوم از این ایده استفاده می‌شود که تنها θ_j هایی که در همسایگی θ_i هستند بر متغیر حاشیه‌ای θ_i اثر می‌گذارند. اگر نزدیکی بین θ_i و θ_j به وابستگی بین دو رأس i و j تعبیر شود، فقط آن θ_j هایی که در شعاع مشخصی از رأس i هستند در محاسبه‌ی چگالی حاشیه‌ای θ_i لحاظ خواهند شد، که این شعاع همسایگی با $R(\psi)$ نمایش داده می‌شود. رو و همکاران (۲۰۰۹) نشان دادند که با استفاده از (۲۰.۲) رابطه

$$\frac{E_{\tilde{\pi}_G}(\theta_j|\theta_i) - \mu_j(\psi)}{\sigma_j(\psi)} = a_{ij}(\psi) \frac{\theta_i - \mu_i(\psi)}{\sigma_j(\psi)} \quad (21.2)$$

برای بعضی a_{ij} که $i \neq j$ برقرار است. از این رو یک روش برای تعیین شعاع همسایگی، ساخت مجموعه‌ی $R(\psi)$ به صورت $\{j : |a_{ij}| > 0.1\}$ است.

(ج) تقریب لاپلاس ساده شده:

در تقریب لاپلاس حتی اگر میانگین به جای مُد استفاده شود، کماکان این تقریب از نظر محاسبات سنگین و زمان‌بر است. با توجه به (۱۰.۲) و $\pi(\theta_{-i}|\theta_i) = \frac{\pi(\theta)}{\pi(\theta_i)} \propto \pi(\theta)$ مخرج کسر (۱۹.۲) به صورت

$$\begin{aligned} \tilde{\pi}_{GG}(\theta_{-i}|\theta_i, \psi, \mathbf{y})|_{\theta_{-i}=E_{\tilde{\pi}_G}(\theta_{-i}|\theta_i)} &\propto (2\pi)^{-\frac{n}{2}} |\mathbf{Q} + \text{diag}(\mathbf{c})|^{\frac{1}{2}} \\ &\times \exp\left(-\frac{1}{2}\boldsymbol{\theta}^T \mathbf{Q} \boldsymbol{\theta} + \mathbf{b}^T \boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}^T (\text{diag}(\mathbf{c}) \boldsymbol{\theta})\right) |_{\theta_{-i}=E_{\tilde{\pi}_G}(\theta_{-i}|\theta_i)} \end{aligned} \quad (22.2)$$

است، که در آن $c_i = -\frac{\partial^2 \pi(y_j|\theta_j, \psi)}{\partial \theta_j^2} |_{\theta_j=E_{\tilde{\pi}_G}(\theta_{-j}|\theta_j)}$. بنابراین برای هر $i = 1, \dots, n$ ماتریس دقت $\theta_{-i}|\theta_i$ به صورت زیرماتریس اصلی $Q_{[-i, -i]}$ است. از طرفی رابطه

$$\begin{aligned} \pi(\boldsymbol{\theta}) &= \pi(\theta_i)\pi(\boldsymbol{\theta}_{-i}|\theta_i) \\ &\propto \text{Var}(\theta_i)^{-1/2} |Q_{[-i, -i]}|^{1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\theta}^T \mathbf{Q} \boldsymbol{\theta}\right\} \end{aligned}$$

برقرار است. بنابراین رابطه‌ی (۲۲.۲) را می‌توان به صورت

$$\begin{aligned} & \tilde{\pi}_{GG}(\boldsymbol{\theta}_{-i}|\theta_i, \boldsymbol{\psi}, \mathbf{y})|_{\boldsymbol{\theta}_{-i}=E\tilde{\pi}_G(\boldsymbol{\theta}_{-i}|\theta_i)} \\ & \propto (\boldsymbol{\Upsilon}\pi)^{-\frac{n}{\Upsilon}} \text{Var}(\theta_i)^{-1/\Upsilon} |Q_{[-i,-i]} + \text{diag}(\mathbf{c})|^{1/\Upsilon} \exp\left(-\frac{1}{\Upsilon}(\theta_i, E(\boldsymbol{\theta}_{-i}|\theta_i))^T \mathbf{Q} \begin{pmatrix} \theta_i \\ E(\boldsymbol{\theta}_{-i}|\theta_i) \end{pmatrix}\right) \\ & \exp\left(\mathbf{b}^T \begin{pmatrix} \theta_i \\ E(\boldsymbol{\theta}_{-i}|\theta_i) \end{pmatrix} - \frac{1}{\Upsilon}(\theta_i, E(\boldsymbol{\theta}_{-i}|\theta_i))^T (\text{diag}(\mathbf{c})) \begin{pmatrix} \theta_i \\ E(\boldsymbol{\theta}_{-i}|\theta_i) \end{pmatrix}\right) \\ & \propto |Q_{[-i,-i]} + \text{diag}(\mathbf{c})|^{1/\Upsilon} \exp\left(-\frac{1}{\Upsilon}(\theta_i, E(\boldsymbol{\theta}_{-i}|\theta_i))^T \mathbf{Q} \begin{pmatrix} \theta_i \\ E(\boldsymbol{\theta}_{-i}|\theta_i) \end{pmatrix}\right) \end{aligned}$$

نوشت. رو و همکاران (۲۰۰۹)، نشان دادند که اگر $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{o}, \Sigma)$

$$-\frac{1}{\Upsilon}(\theta_i, E(\boldsymbol{\theta}_{-i}|\theta_i))^T \Sigma^{-1} \begin{pmatrix} \theta_i \\ E(\boldsymbol{\theta}_{-i}|\theta_i) \end{pmatrix} = -\frac{1}{\Upsilon} \frac{\theta_i^\Upsilon}{\Sigma_{ii}}, \quad i = 1, \dots, n.$$

بنابراین

$$\log\{\tilde{\pi}_{GG}(\boldsymbol{\theta}_{-i}|\theta_i, \boldsymbol{\psi}, \mathbf{y})\}|_{\boldsymbol{\theta}_{-i}=E\tilde{\pi}_G(\boldsymbol{\theta}_{-i}|\theta_i)} \propto c_0 + \frac{1}{\Upsilon} \log |\mathbf{Q}^* + \text{diag}(\mathbf{c})|. \quad (23.2)$$

در محاسبه‌ی دترمینان موجود در (۲۳.۲) باید برای هر θ_i یک ماتریس $(N-1) \times (N-1)$ محاسبه شود که بسیار زمان‌بر است. رو و همکاران (۲۰۰۹)، تقریب لاپلاس ساده‌شده را معرفی کردند که این مشکل را برطرف می‌کند. در این روش با تعریف $\theta_i^{(s)} = \frac{\theta_i - \mu_i(\boldsymbol{\psi})}{\sigma_i(\boldsymbol{\psi})}$ صورت کسر (۱۹.۲) حول $\theta_i = \mu_i(\boldsymbol{\psi})$ به صورت

$$\begin{aligned} & \log\{\pi(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y})\}|_{\boldsymbol{\theta}_{-i}=E\tilde{\pi}_G(\boldsymbol{\theta}_{-i}|\theta_i)} = \\ & -\frac{1}{\Upsilon}(\theta_i^{(s)})^2 + \frac{1}{\Upsilon}(\theta_i^{(s)})^3 \sum_{j \in \mathcal{I}/i} d_j^{(\Upsilon)} \{\mu_i(\boldsymbol{\psi}), \boldsymbol{\psi}\} \{\sigma_j(\boldsymbol{\psi}) a_{ij}(\boldsymbol{\psi})\}^3 + \dots \end{aligned} \quad (24.2)$$

بسط داده می‌شود. جملات مرتبه‌ی اول و دوم تقریب گوسی و جمله مرتبه‌ی سوم تصحیحی برای چولگی فراهم می‌کند. در بسط لگاریتم مخرج کسر (۱۹.۲) با توجه به این‌که برای هر ماتریس \mathbf{M} ، رابطه‌ی $\frac{\partial \log |\mathbf{M}|}{\partial m} = \text{Trace}(\mathbf{M}^{-1} \frac{\partial \mathbf{M}}{\partial m})$ برقرار است (ماردیا و همکاران، ۱۹۸۹)، رابطه

$$\begin{aligned} \frac{d \log |\mathbf{Q}^* + \text{diag}(\mathbf{c})|}{d\theta_i} &= \text{Trace}[\mathbf{Q}^* + \text{diag}(\mathbf{c})]^{-1} \frac{d[\mathbf{Q}^* + \text{diag}(\mathbf{c})]}{d\theta_i} \\ &= \text{Trace}[\mathbf{Q}^* + \text{diag}(\mathbf{c})]^{-1} \text{diag}[d^\Upsilon(\theta_i, \boldsymbol{\psi})] \\ &= \sum_j \text{Var}(\theta_j|\theta_i) d_j^\Upsilon(\theta_i, \boldsymbol{\psi}) \\ &= \sum_j \sigma_j(\boldsymbol{\psi}) \{1 - \text{corr}_{\tilde{\pi}_G}(\theta_i, \theta_j|\boldsymbol{\psi})\} d_j^\Upsilon(\theta_i, \boldsymbol{\psi}) \end{aligned} \quad (25.2)$$

به دست می‌آید، زیرا $Var(\theta_j|\theta_i) = \sigma_j(\boldsymbol{\psi})\{1 - \text{corr}_{\tilde{\pi}_G}(\theta_i, \theta_j|\boldsymbol{\psi})\}^2$ با استفاده از رابطه‌ی (۲۵.۲) بسط تیلور مخرج کسر (۱۹.۲) حول $\theta_i = \mu_i(\boldsymbol{\psi})$ به صورت

$$\begin{aligned} \log\{\pi(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y})\}_{\boldsymbol{\theta}_{-i}=E_{\tilde{\pi}_G}(\boldsymbol{\theta}_{-i}|\theta_i)} = \\ c_0 - \frac{1}{\varphi}(\theta_i^{(s)}) \sum_{j \in \mathcal{I}/i} \sigma_j(\boldsymbol{\psi})\{1 - \text{corr}_{\tilde{\pi}_G}(\theta_i, \theta_j|\boldsymbol{\psi})\}^2 d_j^{(r)}(\mu_i(\boldsymbol{\psi})_i, \boldsymbol{\psi}) \sigma_j(\boldsymbol{\psi}) a_{ij}(\boldsymbol{\psi}) + \dots \end{aligned} \quad (26.2)$$

حاصل می‌شود که در آن c_0 مقداری ثابت است. با توجه به روابط (۲۴.۲) و (۲۶.۲)

$$\begin{aligned} \gamma_i^{(1)}(\boldsymbol{\psi}) &= \frac{1}{\varphi}(\theta_i^{(s)}) \sum_{j \in \mathcal{I}/i} \sigma_j(\boldsymbol{\psi})\{1 - \text{corr}_{\tilde{\pi}_G}(\theta_i, \theta_j|\boldsymbol{\psi})\}^2 d_j^{(r)}(\mu_i(\boldsymbol{\psi})_i, \boldsymbol{\psi}) a_{ij}(\boldsymbol{\psi}) \\ \gamma_i^{(r)}(\boldsymbol{\psi}) &= \sum_{j \in \mathcal{I}/i} d_j^{(r)}\{\mu_i(\boldsymbol{\psi}), \boldsymbol{\psi}\} \{\sigma_j(\boldsymbol{\psi}) a_{ij}(\boldsymbol{\psi})\}^r \end{aligned}$$

تعریف می‌شود، بنابراین تقریب لاپلاس ساده شده به صورت زیر به دست می‌آید.

$$\tilde{\pi}_{SLA} = c_0 - \frac{1}{\varphi}(\theta_i^{(s)}) + \gamma_i^{(1)}(\boldsymbol{\psi})(\theta_i^{(s)}) + \frac{1}{\varphi}(\theta_i^{(s)})^r \gamma_i^{(r)}(\boldsymbol{\psi}) + \dots$$

۳.۲ ملاک‌های ارزیابی

برای مقایسه مدل‌ها باید ملاک‌هایی در اختیار باشد. در هیافت بیزی برای مقایسه مدل‌های سلسله‌مراتبی پیچیده، از ملاک‌هایی نظیر عامل بیزی و انحراف اطلاع، همچنین برای مقایسه تقریب‌های توزیع‌های پسین پارامترها با توزیع اصلی از ملاک کولبک-لیبلر، استفاده می‌شود. روش INLA ملاک‌های ذکر شده را به طور خودکار محاسبه و گزارش می‌کند که در ادامه به شرح آن‌ها خواهیم پرداخت.

۱.۳.۲ تقریب حاشیه‌ای درستنمایی

در تحلیل بیزی یک شیوه معمول مقایسه مدل‌ها، استفاده از عامل بیزی^{۱۵} است. فرض کنید بردار مشاهدات \mathbf{y} را داریم و هدف مقایسه دو مدل

$$M_1 : f_1(\mathbf{y}|\boldsymbol{\psi}_1), \quad M_2 : f_2(\mathbf{y}|\boldsymbol{\psi}_2)$$

باشد. توزیع‌های پیشین $p_1(\boldsymbol{\psi}_1)$ و $p_2(\boldsymbol{\psi}_2)$ را برای پارامترها و احتمالات پیشین $p(M_1)$ و $p(M_2)$ را برای مدل‌ها در نظر می‌گیریم. بر اساس قانون بیز، بخت پسین‌ها برای دو مدل به صورت

$$\begin{aligned} \frac{\pi(M_1|\mathbf{y})}{\pi(M_2|\mathbf{y})} &= \frac{p(M_1)}{p(M_2)} \cdot \frac{\int_{\boldsymbol{\psi}_1} f_1(\mathbf{y}|\boldsymbol{\psi}_1) p_1(\boldsymbol{\psi}_1) d\boldsymbol{\psi}_1}{\int_{\boldsymbol{\psi}_2} f_2(\mathbf{y}|\boldsymbol{\psi}_2) p_2(\boldsymbol{\psi}_2) d\boldsymbol{\psi}_2} \\ &= \text{عامل بیزی} \times \text{بخت پیشین مدل‌ها} \end{aligned}$$

^{۱۵}Bayes Factor

است. بعد از مرتب‌سازی عامل بیزی به صورت

$$\begin{aligned} \mathcal{B}(\mathbf{y}) &= \frac{\pi(M_1|\mathbf{y})}{\pi(M_2|\mathbf{y})} \cdot \frac{p(M_2)}{p(M_1)} \\ &= \frac{\pi(M_1|\mathbf{y})/\pi(M_2|\mathbf{y})}{p(M_1)/p(M_2)} \\ &= \frac{\text{بخت پسین مدل‌ها}}{\text{بخت پیشین مدل‌ها}} \end{aligned}$$

به دست می‌آید. اگر $p(M_1) = p(M_2)$ و فضای پارامتر ψ_1 و ψ_2 یکسان باشند، آنگاه

$$\begin{aligned} \mathcal{B}(\mathbf{y}) &= \frac{\pi(M_1|\mathbf{y})}{\pi(M_2|\mathbf{y})} \cdot \frac{p(M_2)}{p(M_1)} \\ &= \frac{\frac{\pi(M_1, \mathbf{y})}{p(\mathbf{y})p(M_1)}}{\frac{\pi(M_2, \mathbf{y})}{p(\mathbf{y})p(M_2)}} \\ &= \frac{\pi(\mathbf{y}|M_1)}{\pi(\mathbf{y}|M_2)} \end{aligned}$$

یعنی عامل بیزی با نسبت درست‌نمایی‌ها برابر می‌شود. بنابراین مدل‌ها را می‌توان با استفاده از درست‌نمایی‌های دو مدل، یعنی $\pi(\mathbf{y}|M_1)$ و $\pi(\mathbf{y}|M_k)$ مقایسه کرد. برای این منظور تقریب درست‌نمایی‌های $\pi(\mathbf{y}|M_k)$ با استفاده از (۹.۲)، به روش INLA به صورت

$$\tilde{\pi}(\mathbf{y}|M_k) = \int \frac{\pi_k(\boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{y})}{\tilde{\pi}_G(\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y})} d\boldsymbol{\psi}, \quad k = 1, 2 \quad (27.2)$$

محاسبه می‌شود.

۲.۳.۲ ملاک انحراف اطلاع

برای مقایسه مدل‌های سلسله‌مراتبی پیچیده از ملاک انحراف اطلاع^{۱۶} که اندازه‌ای از پیچیدگی^{۱۷} (تعداد مؤثر پارامترها) و برازش^{۱۸} مدل است، استفاده می‌شود. با توجه به این که افزایش تعداد پارامترها تأثیری مستقیم بر پیچیدگی و قدرت پیش‌گویی مدل دارد، بهتر است به هنگام مقایسه مدل‌ها تعداد پارامترهای مؤثر نیز مورد توجه قرار گیرد. دمپستر (۱۹۷۴)، ملاکی تحت عنوان انحراف را براساس لگاریتم درست‌نمایی توزیع پسین معرفی کرد که از آن برای استخراج کمیت‌های برازش و پیچیدگی استفاده می‌شود و نهایتاً با ترکیبی از این دو ملاک DIC تعریف می‌گردد.

فرض کنید بردار مشاهدات \mathbf{y} در اختیار باشد و هدف انتخاب بهترین مدل از بین مدل‌های برازش داده شده به این مجموعه باشد. با استفاده از رهیافت بیزی، توزیع توأم $p(\mathbf{y}, \boldsymbol{\phi})$ به صورت

$$p(\mathbf{y}, \boldsymbol{\phi}) = p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})$$

^{۱۶}Deviance Information Criterion

^{۱۷}Complexity

^{۱۸}Fit

است، که در آن $\phi = (\theta, \psi)$ و با شرط معلوم بودن θ و ψ ، y_i ها مستقل شرطی هستند. دمپستر (۱۹۷۴)، ملاک انحراف را به صورت

$$D(\theta) = -2 \log p(\mathbf{y}|\theta) + 2 \log f(\mathbf{y}) \quad (28.2)$$

تعریف کرد، که در آن $f(\mathbf{y})$ ، عبارت استاندارد است که فقط تابعی از داده‌ها است و بنابراین تأثیری در مقایسه مدل‌ها ندارد. لازم به ذکر است که عبارت (۲۸.۲) را انحراف بیزی^{۱۹} نیز می‌نامند. اشپیگل‌هالتر و همکاران (۲۰۰۲)، کمیت برازش را بر اساس میانگین پسین انحرافات به صورت

$$\bar{D} = E_{\theta|\mathbf{y}}[D]$$

و کمیت پیچیدگی را اختلاف بین میانگین پسین انحرافات و انحراف بر اساس میانگین پسین پارامترها به صورت

$$\begin{aligned} p_D &= E_{\theta|\mathbf{y}}[D] - D[E_{\theta|\mathbf{y}}(\theta)] \\ &= \bar{D} - D(\bar{\theta}) \end{aligned}$$

تعریف کردند. ملاک DIC با در نظر گرفتن دو مفهوم برازش و پیچیدگی به صورت

$$\begin{aligned} \text{DIC} &= \bar{D} + p_D \\ &= D(\bar{\theta}) + 2p_D \end{aligned}$$

تعریف می‌شود. مدل‌های با DIC کوچکتر داده‌ها را بهتر پشتیبانی می‌کنند.

۳.۳.۲ ملاک پیش‌گویی شرطی مؤلفه‌ها

برای مقایسه مدل‌ها، از نظر اعتبار پیش‌گویی می‌توان از ملاک پیش‌گویی شرطی مؤلفه‌ها^{۲۰} (CPO) برای y_i به صورت

$$\text{CPO}_i = \pi(y_i|\mathbf{y}_{-i}), \quad i = 1, \dots, n$$

استفاده کرد، که در آن \mathbf{y}_{-i} همان بردار y_i به جز درایه i ام است. مقادیر صفر و خیلی کوچک CPO بیانگر پرت بودن y_i متناظر است. از منهای میانگین لگاریتم CPO_i ها برای مقایسه مدل‌ها استفاده می‌شود، که ملاک نمره لگاریتمی اعتبارسنجی متقابل^{۲۱} (LogScore) (گنتینگ و رفتری، ۲۰۰۷) نامیده می‌شود. با توجه به این‌که مقادیر بزرگ CPO_i ها گویای پیش‌گویی برتر مدل است، بنابراین هرچه مقدار LogScore کوچک باشد مدل برای پیش‌گویی مناسب‌تر است.

^{۱۹}Bayesian Criterion

^{۲۰}Conditional Predictive Ordinates

^{۲۱}Cross Validated Logarithmic Score

فصل ۳

داده‌های فضایی و فضایی-زمانی

۱.۳ مقدمه

اغلب در مطالعات محیطی با داده‌هایی سر و کار داریم که مستقل از یکدیگر نیستند و وابستگی خاصی بین آن‌ها وجود دارد که ناشی از مکان یا موقعیت جغرافیایی آن در فضای تحت مطالعه است، به این داده‌ها، داده‌های فضایی^۱ می‌گویند. مثال‌هایی از این نوع داده‌ها، میزان شیوع یک بیماری در استان‌های مختلف، درصد مرگ و میر در مناطق مختلف شهری خاص و عوامل موثر بر آلودگی هوا در نقاط مختلف یک شهر هستند.

دو ویژگی کلیدی داده‌های فضایی، نمایش هر داده با مکان آن در فضای مورد مطالعه و وابستگی فضایی آن‌هاست. معمولاً داده‌های فضایی جمع‌آوری شده از مکان‌های مجاور، دارای وابستگی بیشتری هستند و با افزایش فاصله بین مکان داده‌ها، وابستگی کاهش می‌یابد. آمار فضایی، شاخه نسبتاً جدیدی از علم آمار است که داده‌های فضایی را تحلیل می‌کند و در واقع بین مقادیر مختلف یک متغیر، فاصله و جهت قرارگیری آن‌ها ارتباطی برقرار می‌کند. این ارتباط فضایی را ساختار فضایی^۲ می‌نامیم. بررسی و مطالعه این ساختار و برآورد آن، با استفاده از مشاهدات به دست آمده در بعضی از مکان‌های فضای مورد مطالعه، یکی از مسائل مهم آمار فضایی است.

اگرچه، آمار فضایی را عمدتاً آماردانان گسترش دادند، اما پیدایش و کاربرد عملی این شاخه از آمار به وسیله زمین‌شناسان و تحت عنوان زمین‌آمار^۳ صورت گرفت. نخستین بار به دنبال روند تکاملی ذخایر معدنی که از قبل از سال ۱۹۶۰ آغاز شده بود، ماترون با انتشار مقاله‌ای در سال ۱۹۶۲ اساس زمین‌آمار را بنا نهاد. زمین‌آمار شاخه‌ای از علم آمار است که مبتنی بر تحلیل متغیرهایی می‌باشد که دارای ساختار فضایی بوده و با داده‌ها یا متغیرهای فضایی سر و کار دارد. در این شاخه، ناحیه جغرافیایی یک ناحیه چگال^۴ است، به این معنی که بین هر دو مکان یک مکان دیگر نیز در نظر گرفت.

^۱ Spatial data

^۲ Spatial structure

^۳ Geostatistics

^۴ Dense

داده‌های فضایی در مطالعات متعدد علمی مانند هواشناسی، محیط زیست، همه‌گیرشناسی، زمین‌شناسی و اقیانوس‌شناسی در طول زمان نیز به یکدیگر وابسته‌اند. مشاهداتی که هم به مکان فضایی و هم به زمان وابسته باشند را داده‌های فضایی-زمانی^۵ می‌نامیم. تحلیل این‌گونه داده‌ها مستلزم تعیین دو نوع ساختار وابستگی، یعنی وابستگی فضایی و وابستگی زمانی، می‌باشند. به این نوع از ساختار وابستگی، ساختار وابستگی فضایی-زمانی^۶ می‌گوییم و معمولاً به‌وسیله تابع کوواریانس مدل‌بندی می‌کنیم. در سالیان اخیر، بررسی‌های زیادی برای تعیین این ساختار وابستگی، مدل‌سازی و تحلیل چنین داده‌هایی انجام شده‌اند. تابع کوواریانس که نقش به‌سزایی در پیش‌گویی مکان‌های فضایی یا زمانی مشاهده‌نشده دارد، معمولاً نامعلوم است و باید آن را بر اساس مشاهدات برآورد کنیم.

همچنین، در سالیان اخیر همراه با توسعه زمینه‌های مختلف آمار از جمله آمار بیزی، آمار فضایی نیز شاهد رشد و توسعه چشم‌گیری بوده است و به کمک این پیشرفت‌ها آمار فضایی به‌عنوان وسیله‌ای کارآمد برای مطالعه داده‌های فضایی و فضایی-زمانی شناخته می‌شود. نخستین بار کیتانیدیس (۱۹۸۶)، با اشاره به مسائل و دشواری‌های موجود در تحلیل به روش‌های کلاسیک داده‌های فضایی، پیشنهاد کرد از روش‌های بیزی برای تحلیل این نوع از داده‌ها استفاده کنیم. اما از آن‌جا که به‌کارگیری رهیافت بیزی برای تحلیل داده‌های فضایی دشوار و نیازمند محاسبات پیچیده است، در ابتدا چندان مورد توجه قرار نگرفت. اما با توسعه و گسترش آمار بیزی و ابداع روش‌های محاسبات بیزی پیشرفته مانند روش‌های مونت کارلوی زنجیر مارکوفی (MCMC)، تحلیل بیزی داده‌های فضایی نیز رشد چشم‌گیری یافته است.

۲.۳ چرا از آمار فضایی و فضایی-زمانی استفاده می‌کنیم؟

علت اصلی رشد قابل توجه داده‌های فضایی و فضایی-زمانی، ابزار محاسباتی است که ما را قادر می‌سازد داده‌های واقعی را از طریق GPS، ماهواره و ... جمع‌آوری کنیم. یعنی این‌که امروزه در بخش وسیعی از علوم، از اپیدمیولوژی تا بوم‌شناسی، هواشناسی و علوم اجتماعی، محققان با داده‌های از جغرافیا، یعنی شامل اطلاعاتی درباره فضا (و همچنین زمان) سروکار دارند.

به‌عنوان مثال، مطالعه‌ای در زمینه نوعی از بیماری‌های مسری را در نظر می‌گیریم، به‌طوری که علاقه‌مند به ارزیابی شیوع بیماری خاصی مانند سرطان ریه در یک کشور هستیم. داده‌ها معمولاً به‌صورت تعداد بیماران برای نواحی کوچک در چندین سال خواهند بود. چه نوعی از مدل‌ها محققان قادر می‌سازد که کلیه اطلاعات داده‌ها را به‌کار ببرند؟ در نظر گرفتن الگوی جغرافیایی بیماری مسئله‌ای مهم است: ناحیه‌هایی که به یکدیگر نزدیک هستند برخی ویژگی‌های جغرافیایی را به‌طور مشترک باهم دارند، به‌طوری که مربوط به بیماری هستند، پس شیوع مشابه‌ای دارند. همچنین، چگونه شیوع در زمان در حال تغییر است؟ این مسئله نیز معقول است که انتظار داشته باشیم اگر الگوی زمانی وجود داشت، این الگو برای سالیان مابعد نسبت به سالیان ماقبل قوی‌تر است.

در مثالی دیگر، فرض کنید که مشاهداتی از مقدار بارش روزانه در مکان‌های خاص یک محدوده

^۵Spatio-temporal data

^۶Spatio-temporal structure

پراکنده داریم: می‌خواهیم مقدار بارش را در مکان مشاهده نشده پیش‌بینی کنیم و باید همبستگی فضایی و وابستگی زمانی را به حساب آوریم.

در حال حاضر مدل‌های فضایی و فضایی-زمانی به‌طور گسترده‌ای مورد استفاده قرار می‌گیرند: مدل‌های آماری برای داده‌های فضایی که بیش از ۳ میلیون مقاله در مورد آن موجود است و مدل‌های آماری برای داده‌های فضایی-زمانی که حدود ۱۵۹۰۰۰ مقاله در این نوع از مدل‌ها وجود دارد. در هر مجله مقالات کمی وجود دارد که مدل‌های آماری ساختار فضایی و یا زمانی را با خلاقیت و پیچیدگی کم یا زیاد مورد استفاده قرار دهند که این نوع از مدل‌ها دسته وسیعی از کاربردها را پوشش می‌دهند. بر اساس لیست زیر فقط این مقاله‌ها هستند که محدوده‌ی اصلی آن‌ها این نوع مدل‌ها را به‌کار بردند: هاسلت و رفتری (۱۹۸۹)، هندکاک و والیس (۱۹۹۴) و یوهانسون و گلاس (۲۰۰۸)، در بخش هواشناسی کار کردند؛ شواسمیت (۲۰۱۳)، مدلی برای نرخ جرائم و سرقت‌ها ارائه کرده است، در حالی‌که پاپویا و همکاران (۲۰۰۸)، مدل‌های فضایی را برای پیش‌گویی نتیجه انتخابات مورد استفاده قرار دادند؛ در اپیدمیولوژی نور-هلد و ریچاردسون (۲۰۰۳)، در زمینه سرایت بیماری‌ها کار کردند، در حالی‌که والر و همکاران (۱۹۹۷) و الیوت و همکاران (۲۰۰۱)، مدل‌هایی برای بیماری‌های مزمن را ارائه کردند. در نهایت زیپرو و همکاران (۲۰۱۰)، تمرکز روی برآوردها و پیش‌گویی آلودگی هوا داشته‌اند.

۳.۳ چرا از روش‌های بیزی برای مدل‌بندی فضایی و فضایی-زمانی استفاده می‌کنیم؟

انواع بسیاری از مدل‌ها با داده‌های فضایی و فضایی-زمانی بسته به هدف مطالعه مورد استفاده قرار می‌گیرند. اگر علاقه‌مند باشیم که تغییر فضایی و فضایی-زمانی بین نواحی را با استفاده از ریسک‌ها و احتمالات خلاصه‌سازی کنیم، آن‌گاه می‌توانیم روش‌های آماری مانند نقشه بیماری برای مقایسه نقشه‌ها و شناسایی خوشه‌بندی‌ها به‌کار ببریم. شاخص موران^۷ برای بررسی خودهمبستگی فضایی مورد استفاده قرار می‌گیرد (موران، ۱۹۵۰)، در حالی‌که ابزار SaTScan برای کشف خوشه به‌کار می‌رود. انواع مدل‌های مشابه را نیز می‌توان در مطالعات به‌کار برد، که در آن روی پیشامدها نتیجه علت‌شناسی برای ارزیابی اثر عامل‌های ریسک، وجود دارد.

نوع دیگری از مطالعه، تعیین ریسک آزمایش یک پیشامد، به‌عنوان فاصله‌ای از منبع معلوم در حال افزایش است. این نوع معمولاً در چارچوب محیطی است، به‌طوری‌که منبع می‌تواند یک نقطه (مانند فرستنده رادیو) یا یک خط (مانند جاده) باشد. در این حالت، آزمون‌های ناپارامتری که توسط استون (۱۹۸۸) پیشنهاد شده تا روش پارامتری که توسط دایگل و همکاران (۱۹۹۸) بیان شده، متفاوت هستند.

روش‌های بیزی برای استنباط داده‌های فضایی و فضایی-زمانی در سال ۲۰۰۰ با توسعه روش‌های شبیه‌سازی مونت کارلوی زنجیر مارکوفی شروع شد (کسلا و جورج، ۱۹۹۲؛ گیلکس و همکاران، ۱۹۹۶). قبل از این سال، روش بیزی فقط برای مدل‌های نظری به‌کار می‌رفت و کاربردهای کمی در موارد

^۷Moran index

مطالعاتی واقعی، به دلیل عدم وجود ابزارهای شبیه‌سازی یا عددی برای محاسبه‌ی توزیع‌های پسین، داشت. پیدایش MCMC برای محققان این امکان را فراهم آورد که مدل‌های پیچیده را با داده‌های حجیم، بدون نیاز به ساختارهای ساده‌شده توسعه دهند. یکی از بخش‌های اصلی آمار فضایی و فضایی-زمانی را بسگ و همکاران (۱۹۹۱)، ارایه کردند که آن‌ها روش بسگ-یورک-مولیه^۸ (BYM) توسعه دادند، به طوری که برای نقشه بیماری مورد استفاده قرار می‌گیرد. بنرجی و همکاران (۲۰۰۴)، دایگل و ریبریو (۲۰۰۷) و کرسی و ویکل (۲۰۱۱)، روی مدل‌های بیزی زمین‌آماري کار کردند. مزیت اصلی روش بیزی در پیش‌گویی‌ها یا برآوردهای مقادیر نامعلوم به وسیله این روش و انعطاف‌پذیری و توانایی در مدل‌بندی با داده‌های گمشده است. برای جزئیات بیشتر در این مورد به مراجع زیر مراجعه کند: نور-هلد (۲۰۰۰) و بست و همکاران (۲۰۰۵) برای نقشه بیماری و دایگل و همکاران (۱۹۹۸) برای مدل‌بندی داده‌های فضایی پیوسته و پیش‌گویی.

روش‌های MCMC به طور گسترده برای استنباط بیزی به کار می‌روند، اما محاسبات این روش‌ها سنگین هستند. این مشکل زمانی پدیدار می‌شود که مجموعه داده‌های پیشرفته و حجیم را در نظر بگیریم که می‌توانند از داده‌های فضایی و زمانی با بعد بالا باشند. پیچیدگی مدل در ساختارهای فضایی و فضایی-زمانی با مجموعه داده‌های حجیم منجر به محاسبه چندین روز برای اجرای استنباط بیزی با روش MCMC می‌شود. برای غلبه بر این مشکل از تقریب لاپلاس آشیانه‌ای جمع‌بسته با الگوریتمی معین که در فصل قبل بیان شد، استفاده می‌کنیم.

۴.۳ داده‌های فضایی

داده‌های فضایی را به داده‌های وابسته‌ای می‌گوییم که وابستگی آن‌ها ناشی از موقعیت و مکان قرار گرفتن آن‌ها در فضای (جغرافیایی) مورد مطالعه باشد. داده‌های فضایی را بر اساس این که مکان‌های فضایی و متغیرهای تصادفی مرتبط به چه صورت باشند، به سه گروه داده‌های زمین‌آماري^۹، داده‌های مشبکه‌ای^{۱۰} و الگوهای نقطه‌ای^{۱۱} تقسیم‌بندی می‌کنیم.

داده‌های زمین‌آماري: این نوع داده‌ها در مکان‌های ثابت و مشخص در ناحیه‌ای چگال مشاهده می‌شوند؛ در این نوع داده‌ها متغیر مورد مطالعه می‌تواند پیوسته یا گسسته باشد. به عنوان مثال شیوع بیماری مالاریا در کودکان در کشور گامبیا (شکل ۱.۳) نمونه‌ای از این داده‌هاست. هدف از استنباط این داده‌ها، پیش‌گویی پاسخ در مکان‌های مشاهده‌نشده است.

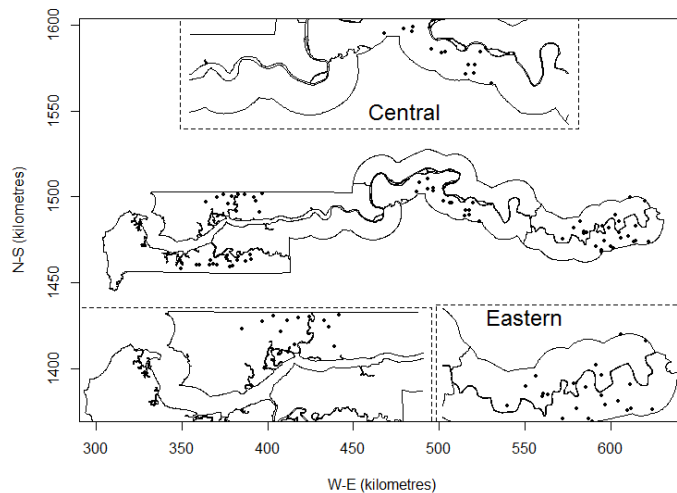
داده‌های مشبکه‌ای: این نوع داده‌ها مربوط به نواحی غیرچگال هستند، یعنی بین مکان‌های مشاهده‌شده مکان جدیدی نمی‌توان یافت. این مکان‌ها ممکن است منظم یا نامنظم باشند. نسبت کودکان یا بیماری‌های تنفسی (نمودار سمت چپ شکل ۲.۳) و نرخ مرگ و میر بر اثر سرطان در ۲۴ بخش شهر لندن (نمودار سمت راست شکل ۲.۳)، مثال‌هایی از این نوع داده‌هاست که به ترتیب دو

^۸ Besag-York-Mollié

^۹ Geostatistical data

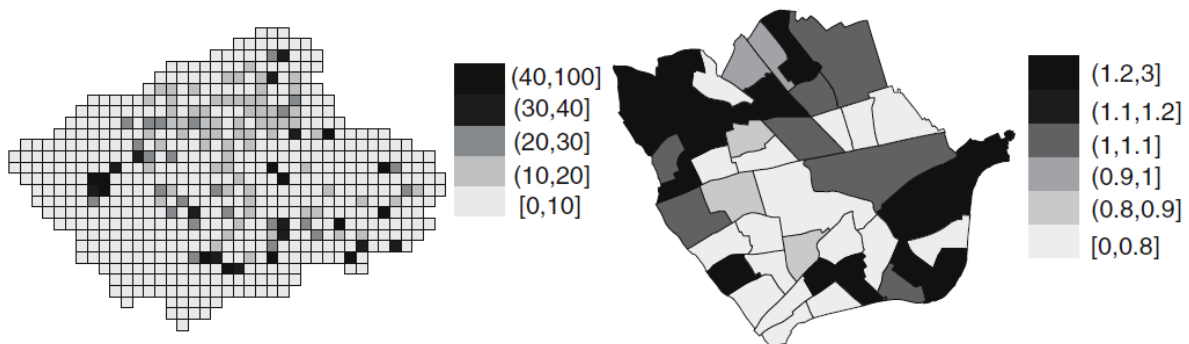
^{۱۰} Lattice data

^{۱۱} Point patterns



شکل ۱.۳: نقشه گامبیا، آفریقا: نقاط نشان دهنده روستاهایی است که شیوع بیماری مالاریا در کودکان مشاهده شده است.

تحقق فرآیند فضایی را تحت توری منظم و نواحی نامنظم ارایه می‌دهند. در این موارد اغلب علاقه‌مند به هموارسازی پاسخ هستیم.



شکل ۲.۳: نسبت کودکان یا بیماری‌های تنفسی در انگلیس (نمودار سمت چپ) و نرخ مرگ و میر بر اثر سرطان ریه در ۲۴ بخش شهر لندن (نمودار سمت راست).

الگوهای نقطه‌ای: در این نوع داده‌ها مکان مشاهده‌شده، خود یک متغیر تصادفی است. این داده‌ها را می‌توان به سه دسته به‌طور کامل تصادفی فضایی^{۱۲} (CSR)، منظم^{۱۳} و خوشه‌ای^{۱۴} تقسیم و آن‌ها مدل‌بندی کرد. یک مثال برای این نوع مشاهدات، محل تخلفات رانندگی در یک مسابقه رالی (شکل ۳.۳) است.

^{۱۲} Complete spatial randomness

^{۱۳} Regular

^{۱۴} Cluster



شکل ۳.۳: محل تخلفات رانندگی در یک مسابقه رالی

۵.۳ روش معادله دیفرانسیل جزئی تصادفی

روش SPDE توسط لیندگرن و همکاران (۲۰۱۱)، پیشنهاد شد. این روش فرآیند فضایی پیوسته (یعنی یک GF) را با استفاده از فرآیند تصادفی گسسته (یعنی یک GMRF) بیان می‌کند. نقطه شروع SPDE به صورت زیر است:

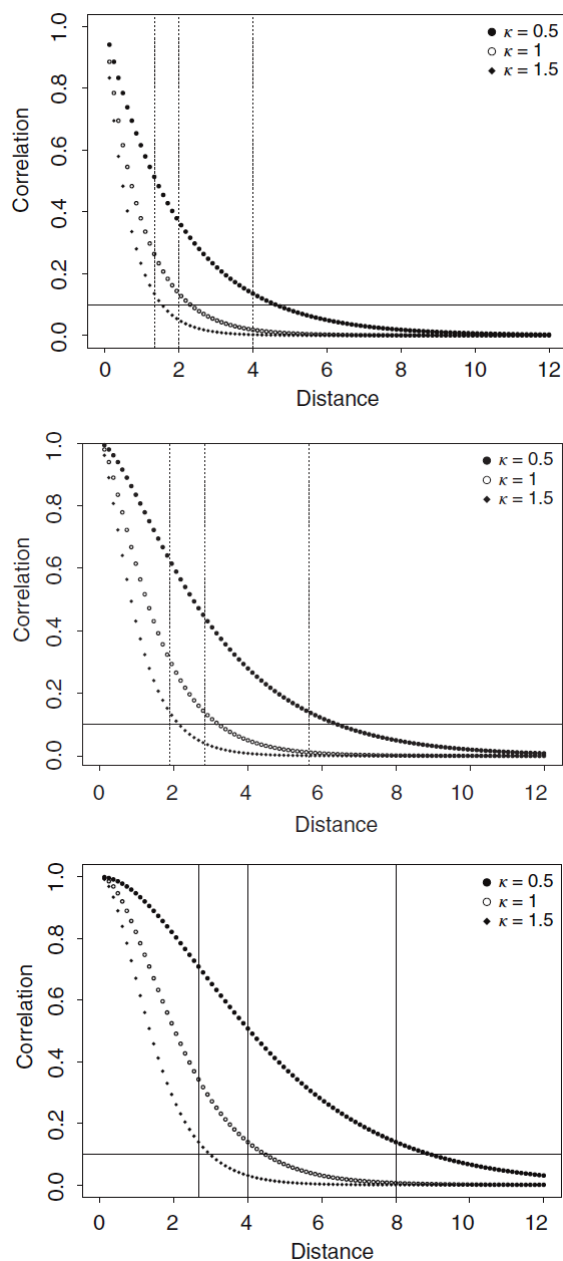
$$(\kappa^2 - \Delta)^{\alpha/2} (\tau \xi(s)) = \mathcal{W}(s) \quad (1.3)$$

که در آن $s \in \mathbb{R}^d$ ، $\kappa > 0$ پارامتر مقیاس، Δ لاپلاسیان و برابر با $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial \xi_i^2}$ است. α هموارسازی را کنترل و τ واریانس را کنترل می‌کند و $\mathcal{W}(s)$ فرآیند نوفه سفید فضایی گوسی است. راه‌حلی دقیق SPDE آن است که میدان ایستای گوسی $\xi(s)$ را به کمک تابع کوواریانس مترن به صورت زیر تعریف می‌کند:

$$\text{Cov}(\xi(s_i), \xi(s_j)) = \text{Cov}(\xi_i, \xi_j) = \frac{\sigma^2}{\Gamma(\lambda) 2^{\lambda-1}} (\kappa \|s_i - s_j\|)^\lambda K_\lambda(\kappa \|s_i - s_j\|) \quad (2.3)$$

که در آن $\|s_i - s_j\|$ فاصله اقلیدسی بین دو مکان $s_i, s_j \in \mathbb{R}^d$ است و σ^2 واریانس حاشیه‌ای می‌باشد. جمله K_λ تابع بسل نوع دوم را مشخص می‌کند و مرتبه $\lambda > 0$ را دارد (آبرامویتز و استگن، ۱۹۷۲) به طوری که این مقیاس درجه هموارسازی فرآیند را اندازه‌گیری می‌کند و معمولاً به دلیل تشخیص ضعیف آن، ثابت فرض می‌شود. $\kappa > 0$ پارامتر مقیاس مرتبط با درجه ρ است، یعنی فاصله در همبستگی فضایی تقریباً صفر است. معمولاً $\rho = \frac{\sqrt{\lambda \kappa}}{\kappa}$ فرض می‌شود (بخش ۲، لیندگرن و همکاران، ۲۰۱۱) را مشاهده کنید)، که این ρ فاصله‌ای است که همبستگی فضایی برای هر $\lambda \geq 1/2$ نزدیک به 0.1 است. به منظور ارزیابی خوبی این درجه، شکل ۴.۳ برای تابع کوواریانس مترن با مقادیر مختلف λ و κ را مشاهده کنید که خطوط عمودی این شکل مقادیر ρ هستند. ارتباط بین SPDE معادله (۱.۳) و پارامترهای مترن، معادلات زیر را که شامل پارامتر هموارسازی λ و واریانس حاشیه‌ای σ^2 می‌شود، بیان می‌کنند:

$$\begin{cases} \lambda = \alpha - d/2 \\ \sigma^2 = \frac{\Gamma(\lambda)}{\Gamma(\alpha) (2\pi)^{d/2} \kappa^{2\lambda} \tau^2} \end{cases}$$



شکل ۴.۳: تابع همبستگی مترن برای مقادیر مختلف κ و فواصل: $\lambda = 0.5$ (نمودار بالا)، $\lambda = 1$ (نمودار مرکز) و $\lambda = 2$ (نمودار پایین).

در $R - INLA$ مقدار پیش فرض برای پارامتر هموارسازی $\alpha = 2$ ($\lambda = 1$) است، می توان مقادیر $0 \leq \alpha < 2$ را نیز تعیین کرد (برای مقادیر ناصحیح α روش تقریب در لیندگرن و همکاران (۲۰۱۱) به کار رفته است). توجه کنید که با $\alpha = 2$ ، درجه ρ و واریانس σ^2 به صورت زیر به دست می آیند:

$$\rho = \frac{\sqrt{\lambda}}{\kappa}$$

$$\sigma^2 = \frac{1}{(4\pi\kappa^2\tau^2)}$$

به روش SPDE برای میدان گوسی و همسان‌گرد $\xi(s)$ ، می‌توان با استفاده از روش درایه متناهی تقریب زد که از طریق تابعی مبنا روی مثلث‌بندی^{۱۵} حوزه D تعریف می‌شود:

$$\xi(s) = \sum_{g=1}^G \phi_g(s) \hat{\xi}_g. \quad (3.3)$$

در این جا G تعداد کل رئوس مثلث‌بندی است، $\{\phi_g\}$ مجموعه‌ای از توابع مبنا و $\{\hat{\xi}_g\}$ وزن‌های دارای توزیع گوسی با میانگین صفر هستند. به‌منظور دستیابی به ساختار مارکوفی توابع مبنا ϕ_g طوری انتخاب می‌شوند که در رأس g ، ۱ و در مابقی رئوس صفر باشند. با استفاده از شرایط مرزی نیومن، داریم که (برای $\alpha = 2$) ماتریس دقت Q برای بردار وزنی گوسی $\xi = \{\hat{\xi}_1, \dots, \hat{\xi}_G\}$ به‌صورت زیر است:

$$Q = \tau^2 (\kappa^2 C + 2\kappa^2 G + GC^{-1}G),$$

که در آن درایه ماتریس قطری C به‌صورت $C_{ii} = \int \phi_i(s) ds$ است و درایه ماتریس تنک G ، $G_{ij} = \int \nabla \phi_i(s) \nabla \phi_j(s) ds$ است (∇ گرادیان). ماتریس دقت Q که درایه‌های آن به τ و κ وابسته‌اند، تنک است و در نتیجه ξ یک GMRF با توزیع نرمال $N(0, Q^{-1})$ است و این ماتریس راه‌حل تقریبی روش SPDE را بیان می‌کند.

در INLA - R پیش‌فرض پارامترهای SPDE، $\log(\tau) = \theta_1$ و $\log(\kappa) = \theta_2$ است که توزیع پیشین نرمال توأم دارند.

۶.۳ SPDE با INLA - R

همان‌طور که می‌دانیم در INLA - R پیش‌گوی خطی η_i به‌وسیله formula تعریف می‌شود که $f()$ شامل جملات اثرات غیرخطی متغیرهای تبیینی یا اثرات تصادفی است. پس، میدان گوسی مترن در formula با استفاده از مشخصه‌ای مناسب به‌کار خواهد رفت. برای ارتباطی واضح‌تر بین پیش‌گوی خطی η_i و formula، بهتر است با استفاده از لیندگرن و رو (۲۰۱۵)، پیش‌گوی خطی (۱.۲) را به‌صورت زیر بازنویسی کنیم:

$$\eta_i = \sum_k h_k(z_i^k) \quad (4.3)$$

که در آن k ، k امین جمله formula، z_i^k مقدار متغیر تبیینی برای اثر ثابت یا غیرخطی، اثر تصادفی و اندیس واحد مرتبه دوم (مثلاً ناحیه یا نقطه ID) هستند. تابع $h_k(\cdot)$ ، z_i^k را به مقدار واقعی میدان پنهان برای k امین مؤلفه این معادله پیوند می‌دهد.

به‌عنوان مثال، فرض کنید که پیش‌گوی خطی شامل اثر ثابت متغیر تبیینی (z_1 در R)، یک اثر غیرخطی (مانند rw2) متغیر time (دنباله‌ای از زمان) و اثر تصادفی با مؤلفه‌های هم‌توزیع و مستقل باشد که به‌وسیله متغیر index.random اندیس‌گذاری شده است. بنابراین

```
> formula <- -1 + z1 + f(time, model="rw2") +
```

^{۱۵} از طریق مثلث‌بندی، یک حوزه فضایی به مجموعه‌ای از مثلث‌های غیرمتقاطع تقسیم‌بندی می‌شود، که در آن هر دو مثلث لبه یا گوشه‌ای مشترک دارند.

f(index.random, model="iid")

با پیش‌گوی جدید (۴.۳)، $h_1(z_1^1)$ معادل با $z_1^1 \beta$ ، $h_2(z_1^2)$ اثر هموارسازی شده در z_1^2 (ایمین درایه بردار time) و $h_3(z_1^3)$ مؤلفه اثر تصادفی z_1^3 (ایمین درایه بردار index.random) می‌باشد. به‌طور معمول میدان پنهان θ بردار توأم متغیرهای گوسی پنهان در پیش‌گوی خطی می‌باشد. فرمول‌بندی معادله (۴.۳) باعث می‌شود که فقط هر مشاهده مستقیماً به مؤلفه z_i^k از هر مؤلفه $h_k(\cdot)$ وابسته باشد و این عمل زمانی که یک اثر تصادفی به‌صورت تابعی خطی از مقادیر فضایی یا زمانی، مانند SPDE در معادله (۳.۳)، تعریف شود، اثربخش نیست. در برخی موارد هر مشاهده y_i به ترکیب خطی مؤلفه‌های θ وابسته است و توزیع مشاهده به‌صورت زیر تعریف می‌شود:

$$y_i | \theta, \psi \sim p \left(y_i | \sum_j A_{ij} \theta_j, \psi \right)$$

A_{ij} مؤلفه ماتریس A است که به آن ماتریس مشاهده می‌گوییم. در بخش بعدی چگونگی ایجاد ماتریس A را به‌وسیله تابع inla.spde.matrix شرح خواهیم داد. ماتریس A نقشه‌ای بین میدان پنهان فضایی (تعریف شده روی شبکه) و مشاهدات (تعریف شده در مجموعه‌ای از مکان‌ها) تعریف می‌کند. INLA - R، پیش‌گوی خطی جدیدی η^* را به‌صورت ترکیب خطی η تولید می‌کند:

$$\eta^* = A\eta$$

و در این مورد درست‌نمایی از طریق η_i^* به جای η_i به میدان پنهان مرتبط می‌شود

$$p(y | \theta, \psi) = \prod_{i=1}^n p(y_i | \eta_i^*, \psi).$$

۷.۳ تشریح روش SPDE با مثال شبیه‌سازی شده

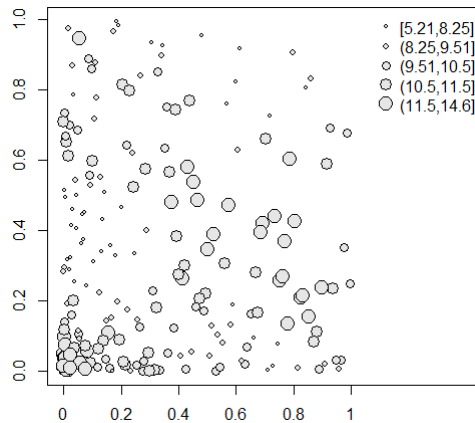
در این بخش مثالی ساده با داده‌های شبیه‌سازی شده را برای تشریح توابع INLA - R برای روش SPDE در نظر می‌گیریم و با استفاده از این توابع این مثال (داده‌های فضایی) و مثال واقعی (داده‌های فضایی و زمانی استان گلستان) در فصل بعد را استنباط می‌کنیم. برای این کار از مجموعه داده SPDEtoy در بسته INLA - R استفاده می‌کنیم. این مجموعه داده ۲۰۰ مقدار شبیه‌سازی شده برای متغیر y را دارد. نمودار این داده‌ها در شکل ۵.۳ نشان داده شده‌اند. مدل به‌کار رفته برای شبیه‌سازی داده‌های SPDEtoy فرض می‌کند که توزیع مشاهده y_i به‌صورت

$$y_i \sim N(\eta_i, \sigma_e^2), \quad i = 1, \dots, 200$$

است. که در آن σ_e^2 واریانس اندازه خطا e_i است که دارای توزیع نرمال و مستقل از e_i برای هر $i \neq j$ می‌باشد. میانگین پاسخ توسط پیش‌گوی خطی زیر تعریف می‌شود

$$\eta_i = b_0 + \xi_i \quad (5.3)$$

که b_0 عرض از مبدأ و ξ_i اثر تصادفی است که این اثر تصادفی تحقق از میدان گوسی پنهان $\xi(s) \sim MVN(0, \Sigma)$ می‌باشد. ماتریس کوواریانس Σ به‌وسیله تابع کوواریانس مترن معادله (۲.۳) تعریف



شکل ۵.۳: مکان ۲۰۰ نقطه فضایی داده‌های SPDEtoy.

می‌شود. مقادیر پارامتر برای این داده‌ها به صورت $\lambda = 1$ و $\kappa = 7$ ، $\sigma_e^2 = 5$ ، $\sigma_e^2 = 0.3$ ، $b_0 = 10$ انتخاب می‌شوند. برای جزئیات بیشتر شبیه‌سازی به کرانسکی و لیندگرن (۲۰۱۳) مراجعه کنید.

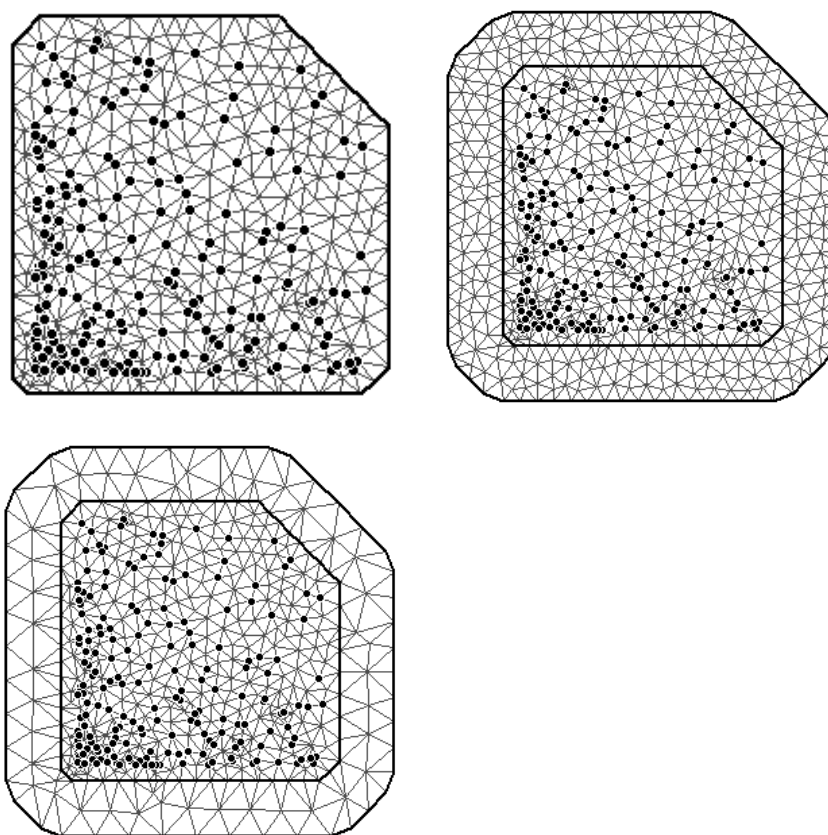
۱.۷.۳ ساختن شبکه

همان‌طور که گفته شد، روش SPDE بر مبنای مثلث‌بندی حوزه فضایی است. توجه کنید که در عمل تعریف این شبکه مبادله‌ای بین GMRF و هزینه‌های محاسباتی می‌باشد که هر دو به تعداد رئوس استفاده شده در مثلث‌بندی وابسته‌اند: تعداد مثلث‌های بزرگتر منجر به تقریب بهتر GF می‌شود، اما هزینه محاسباتی را بالا می‌برد.

در R-INLA از تابع تبیینی `inla.mesh.2d` استفاده می‌کنیم. این تابع برخی اطلاعات مورد نیاز حوزه فضایی را دارد که این اطلاعات یا به وسیله نقاط فضایی مرتبط یا بسط حوزه داده می‌شوند که در این تابع به وسیله گزینه‌های `loc` و `loc.domain` مشخص می‌شوند. گزینه دیگر این تابع `max.edge` است که بیشترین طول گوشه مثلث را تعیین می‌کند. اگر برداری از دو مقدار تولید کنیم، این حوزه فضایی به نواحی بیرونی و درونی تقسیم می‌شود که وضوح مثلث به وسیله `max.edge` تعیین می‌شود (مقدار بزرگتر `max.edge` منجر به دقت پایین‌تری می‌شود). با توجه به آن‌چه توسط لیندگرن و رو (۲۰۱۵) پیشنهاد شده، برای اجتناب از اثر مرز، این حوزه می‌بایست به وسیله فاصله‌ای حداقل برابر با محدوده ρ بسط داده شود.

در ادامه ۳ نوع مختلف مثلث‌بندی را برای داده‌های SPDEtoy شرح می‌دهیم. نمودار این نوع از مثلث‌بندی‌ها که با نقاط مختلف `max.edge` به دست آمده در شکل ۶.۳ نشان داده شده است. این سه نوع را با استفاده از کدهای زیر اجرا می‌کنیم:

```
> coords <- as.matrix(SPDEtoy[,1:2])
> mesh0 <- inla.mesh.2d(loc=coords, max.edge=0.1)
> mesh1 <- inla.mesh.2d(loc=coords, max.edge=c(0.1, 0.1))
> mesh2 <- inla.mesh.2d(loc=coords, max.edge=c(0.1, 0.2))
```



شکل ۶.۳: سه مثلث‌بندی برای داده‌های SPDEtoy با نقاط مختلف $\max.\text{edge}$: mesh0 (نمودار بالا سمت چپ)، mesh1 (نمودار بالا سمت راست) و mesh2 (نمودار پایین). نقاط مکان‌های مشاهدات هستند.

مشاهده می‌کنیم که mesh0 تنها با یک مقدار $\max.\text{edge}$ است که بیرون حوزه فضایی را بسط نداده است. همچنین اختلاف بین mesh1 و mesh2 آن است که برای mesh2 مقدار $\max.\text{edge}$ بزرگتری انتخاب شده است که همین مطلب باعث می‌شود بیرون حوزه فضایی به وسیله مثلث‌های بیشتری بسط داده شوند (یعنی دقت کمتر). این تابع ساخت شبکه، گزینه‌های دیگری دارد که به بیان آن‌ها نمی‌پردازیم و برای جزئیات بیشتر بلانگیاردو و کاملتی (۲۰۱۵) را پیشنهاد می‌کنیم.

۲.۷.۳ ماتریس مشاهده

با توجه به تابع مبنا معادله (۳.۲) پیش‌گوی خطی (۵.۲) را می‌توان به صورت زیر بازنویسی کرد:

$$\eta_i = b_0 + \sum_{g=1}^G \phi_g(s_i) \hat{\xi}_g. \quad (6.3)$$

همچنین، می‌توان پیش‌گوی خطی را به صورت زیر بیان کرد:

$$\eta_i = b_0 + \sum_{g=1}^G A_{ig} \hat{\xi}_g$$

با $A_{ig} = \phi_g(s_i)$ که درایه ماتریس تنک A است به طوری که GMRF، $\hat{\xi}$ ، G رأس مثلث را برای d مشاهده ترسیم می‌کند.

در $R - INLA$ تابع `inla.spde.make.A` ماتریس تنک A را به وسیله مکان‌های داده در شبکه ایجاد می‌کند. مثلاً کد زیر ماتریس A را برای `mesh1` (1×8 رأس) ایجاد می‌کند.

```
> A.est1 <- inla.spde.make.A(mesh=mesh1, loc=coords)
> dim(A.est1)
[1] 200 801
```

۳.۷.۳ برازش مدل

در این مثال، بردار پارامترها $\theta = \{\hat{\xi}, b_0\}$ و بردار ابرپارامترها $\psi = (\sigma_e^2, \kappa, \sigma^2)$ است. حال با استفاده از مثلث‌بندی `mesh1` و ماتریس مشاهده `A.est1` پارامترهای مدل را برآورد می‌کنیم. ابتدا باید SPDE مترن را بسازیم:

```
> spde <- inla.spde2.matern(mesh=mesh1, alpha=2)
```

سپس پیش‌گوی خطی را به صورت زیر تعریف می‌کنیم:

```
> formula <- y ~ -1 + intercept + f(spatial.field, model=spde)
```

اثر تصادفی فضایی با `spatial.field` مشخص شده که مدل `spde` را دارد. بنابراین مدل برازش یافته به وسیله تابع `inla` اجرا می‌شود:

```
> output1 <- inla(formula,
data = list(y=SPDEtoy$y, intercept=rep(1,spde$n.spde),
spatial.field=1:spde$n.spde),
control.predictor=list(A=A.est1,compute=TRUE))
```

توجه کنید که لیست `data` نیز شامل متغیر `spatial.feild` برای اثر تصادفی، تعریف شده به صورت دنباله‌ای از ۱ تا تعداد رئوس شبکه، است. همچنین ماتریس مشاهده از طریق گزینه `compute = TRUE` در `control.predictor` تولید شده است که توزیع حاشیه‌ای پسین‌ها را برای پیش‌گوی خطی به دست می‌دهد.

خلاصه‌های پسین عرض از مبدا و دقت برای مشاهدات گوسی را می‌توان با نوشتن دستورات زیر مشاهده کرد:

```
> round(output1$summary.fixed,3)
      mean      sd 0.025quant 0.5quant 0.975quant  mode kld
```



```

intercept 9.509 0.819      7.779      9.526      11.138 9.555  0
> round(output1$summary.hyperpar[1,],3)
              mean      sd 0.025quant 0.5quant
Precision for the Gaussian observations 3.732 0.783      2.423      3.654
              0.975quant  mode
Precision for the Gaussian observations      5.488 3.507

```

اگر مایل باشیم واریانس σ_e^2 را به دست آوریم، می‌توانیم میانگین پسین حاشیه‌ای تبدیل شده را از طریق

```

> inla.emarginal(function(x) 1/x, output1$marginals.hyper[[1]])
[1] 0.2819983

```

محاسبه کنیم. خلاصه‌های پسین پارامترهای فضایی با استفاده از تابع `inla.spde2.result` در `inla` به دست می‌آیند که کلیه اطلاعات مربوط به لیست `output1` را استخراج می‌کند. همچنین، این تابع نتایج را از مقیاس‌های اولیه پارامترها تبدیل می‌کند و توزیع‌های پسین را برای واریانس σ^2 و درجه ρ فراهم می‌آورد. نتایج اولیه برای $\theta_1 = \log(\tau)$ و $\theta_2 = \log(\kappa)$ هستند. در کد

```

> output1.field <- inla.spde2.result(inla=output1,
name="spatial.field", spde=spde, do.transf=TRUE)

```

نام `name` اثر SPDE در `formula` را مشخص می‌کند. گزینه `do.transf` به منظور محاسبه حاشیه‌ای‌های پارامترها روی مقیاس تبدیل شده، به کار می‌رود. میانگین پسین برای τ ، σ^2 و ρ را می‌توان به صورت زیر به دست آورد:

```

> inla.emarginal(function(x) x, output1.field$marginals.kappa[[1]])
[1] 6.955692
> inla.emarginal(function(x) x, output1.field$marginals.variance.nominal[[1]])
[1] 3.543791
> inla.emarginal(function(x) x, output1.field$marginals.range.nominal[[1]])
[1] 0.4064151

```

۴.۷.۳ استفاده از تابع `inla.stack`

زمانی که پیچیدگی مدل افزایش می‌یابد، مثلاً اثرهای تصادفی در پیش‌گوی خطی زیاد باشد، ساخت بردار و ماتریس روش SPDE می‌تواند طاقت فرسا باشد و منجر به خطا شود. برای رفع این مشکل، تابع `inla.stack` در `R-INLA` برای بهینه و ساده کردن روش SPDE معرفی می‌شود. در ادامه با استفاده از نکته مشابه در لیندگرن و رو (۲۰۱۵)، به معرفی این تابع می‌پردازیم.

کلیه جملات z_i^k پیش‌گوی خطی (۴.۳) را در ماتریسی به شکل زیر قرار می‌دهیم:

$$\mathbf{Z} = [z^1 \dots z^k].$$

این ماتریس باعث می‌شود که بتوان پیش‌گوی خطی با K اثر را برای کلیه مشاهدات به صورت

$$\boldsymbol{\eta} = H(\mathbf{Z}) = \sum_{k=1}^K h_k(z^k),$$

نوشت، در حالی که $R - INLA$ پیش‌گوی خطی جدید $\boldsymbol{\eta}^* = \mathbf{A}H(\mathbf{Z})$ را به عنوان ترکیب خطی از مؤلفه‌های $\boldsymbol{\eta}$ و ماتریس مشاهده \mathbf{A} تولید می‌کند.

در مدل‌های پیچیده‌تر می‌توان مجموع پیش‌گویی‌های خطی را به صورت

$$\boldsymbol{\eta}^* = \boldsymbol{\eta}_1^* + \boldsymbol{\eta}_2^* + \dots = \mathbf{A}_1 H(\mathbf{Z}_1) + \mathbf{A}_2 H(\mathbf{Z}_2) + \dots$$

تعریف کرد. اجرای تابع `inla.stack` به صورت زیر است:

```
stack <- inla.stack(data = list(...),
A = list(A1, A2, ...),
effects = list(list(...), list(...), ...),
tag = "...")
```

که در آن هر ماتریس \mathbf{A} مرتبط با یک اثر مشخص شده در `list` است که در بردارنده متغیر تبیینی یا اندیس اثر تصادفی مربوط به جمله \mathbf{Z} است. لیست `data` متغیر پاسخ را دارد و باید طول آن برابر با پیش‌گو باشد.

همچنین می‌توان پیش‌گوها را به صورت توأم قرار داد که در تابع `inla` ایجاد شدند. این عمل را می‌توان با استفاده از گزینه `tag` اجرا کرد:

```
stack <- inla.stack(stack1, stack2, ...)
```

که در آن `stack1` و `stack2` خروجی تابع `inla.stack` با `tag` متفاوت هستند. این نوع توأم زمانی کاربرد دارد که علاقه‌مند به پیش‌گوی فضایی باشیم.

حال با استفاده از مثال ساده `SPDEtoy`، چگونگی دستیابی به برآوردهای پارامتر را به وسیله تابع `inla.stack` نشان می‌دهیم. با توجه به تعریف تابع `inla.stack`، عرض از مبدأ را شامل نمی‌شود و باید آن را مشخص کنیم. همچنین تابعی جدیدی با نام `inla.spde.make.index` که اندیس‌های مورد نیاز برای مدل `SPDE` ایجاد می‌کند، را تعریف می‌کنیم.

```
> s.index <- inla.spde.make.index(name="spatial.field",
n.spde=spde$n.spde)
```

که در آن `name` نام اثری است که در `formula` به کار خواهد رفت. خروجی `inla.spde.make.index` لیستی از مؤلفه‌های زیر را دارد:

```
> names(s.index)
[1] "spatial.field" "spatial.field.group" "spatial.field.repl"
```

و در این مثال ساده سه بردار به صورت زیر تعریف می‌شوند:

```
s.index$spatial.field = seq(1,spde$n.spde)
s.index$spatial.field.group = rep(1,spde$n.spde)
s.index$spatial.field.repl = rep(1,spde$n.spde)
```

در ادامه تابع `inla.stack` را به کار می‌بریم:

```
> stack.est <- inla.stack(data=list(y=SPDEtoy$y),
A=list(A.est1),
effects=list(c(s.index, list(intercept=1))),
tag="est") #Estimation
```

که در آن `A` ماتریس `A.est1` را دربر دارد و `effects` لیستی از عرض از مبدأ و اندیس میدان تصادفی فضایی را دارد. توجه کنید که اگر `s.index` به صورت یک لیست از قبل تعریف شده باشد، در گزینه `effects` از `list(...)` استفاده نمی‌کنیم. در پایان داریم:

```
> output1.stack <- inla(formula,
data=inla.stack.data(stack.est, spde=spde),
family="gaussian",
control.predictor=list(A=inla.stack.A(stack.est),
compute=TRUE))
```

در این جا `inla.stack.A` و `inla.stack.data` به ترتیب برای استخراج داده‌ها و ماتریس مشاهده از `stack.est` به کار رفته‌اند.

۵.۷.۳ پیش‌گویی فضایی

به منظور اجرای پیش‌گویی فضایی در `INLA - R`، نقاط یک توری منظم 50×50 را در حوزه فضایی داده‌های `SPDEtoy` ایجاد می‌کنیم:

```
> grid.x <- 50
> grid.y <- 50
> pred.grid <- expand.grid(x = seq(0, 1, length.out = grid.x),
y = seq(0, 1, length.out = grid.y))
> dim(pred.grid)
[1] 2500 2
```

در $R - INLA$ ساده‌ترین روش برای پیش‌گویی فضایی، به‌وسیله پیش‌گوهای خطی توأم است (یکی برای برآورد پارامتر و یکی برای پیش‌گویی) که از طریق تابع `inla.stack` اجرا می‌شود. برای اجرای این روش، ابتدا یک ماتریس مشاهده جدید برای 25° مکان توری ایجاد می‌کنیم:

```
> A.pred1 <- inla.spde.make.A(mesh=mesh1, loc=as.matrix(pred.grid))
> dim(A.pred1)
[1] 2500 801
```

و سپس تابع `inla.stack` را با استفاده از `tag = "pred.latent"` تولید می‌کنیم:

```
stack.pred.latent <- inla.stack(data=list(xi=NA),
  A=list(A.pred1),
  effects=list(s.index),
  tag="pred.latent")
```

توجه کنید، چون می‌خواهیم پیش‌گویی را انجام دهیم، باید `xi = NA` قرار دهیم؛ همچنین، چون عرض از مبدا در لیست اثرات وجود ندارد، فقط برای میدان پنهان فضایی $\xi(s)$ پیش‌گویی را محاسبه می‌کنیم. اگر علاقه‌مند باشیم پیش‌گویی متغیر پاسخ را انجام می‌دهیم، به‌صورت زیر استفاده خواهیم کرد:

```
> stack.pred.response <- inla.stack(data=list(y=NA),
  A=list(A.pred1),
  effects=list(c(s.index, list(intercept=1))),
  tag="pred.response")
```

حال کلیه سه پیش‌گویی خطی (برآورد پارامتر، پیش‌گویی میدان پنهان و پیش‌گویی پاسخ) را با هم به‌صورت توأم قرار می‌دهیم:

```
> join.stack <- inla.stack(stack.est, stack.pred.latent,
  stack.pred.response)
```

و لذا داریم:

```
> join.output <- inla(formula,
  data=inla.stack.data(join.stack),
  control.predictor=list(A=inla.stack.A(join.stack),
  compute=TRUE))
```

به‌منظور ارزیابی توزیع‌های حاشیه‌ای پسین پیش‌گویی‌ها در مکان‌های توری، از تابع `inla.stack.index` استفاده می‌کنیم:

```
> index.pred.latent <- inla.stack.index(join.stack,
tag="pred.latent")$data
> index.pred.response <- inla.stack.index(join.stack,
tag="pred.response")$data
```

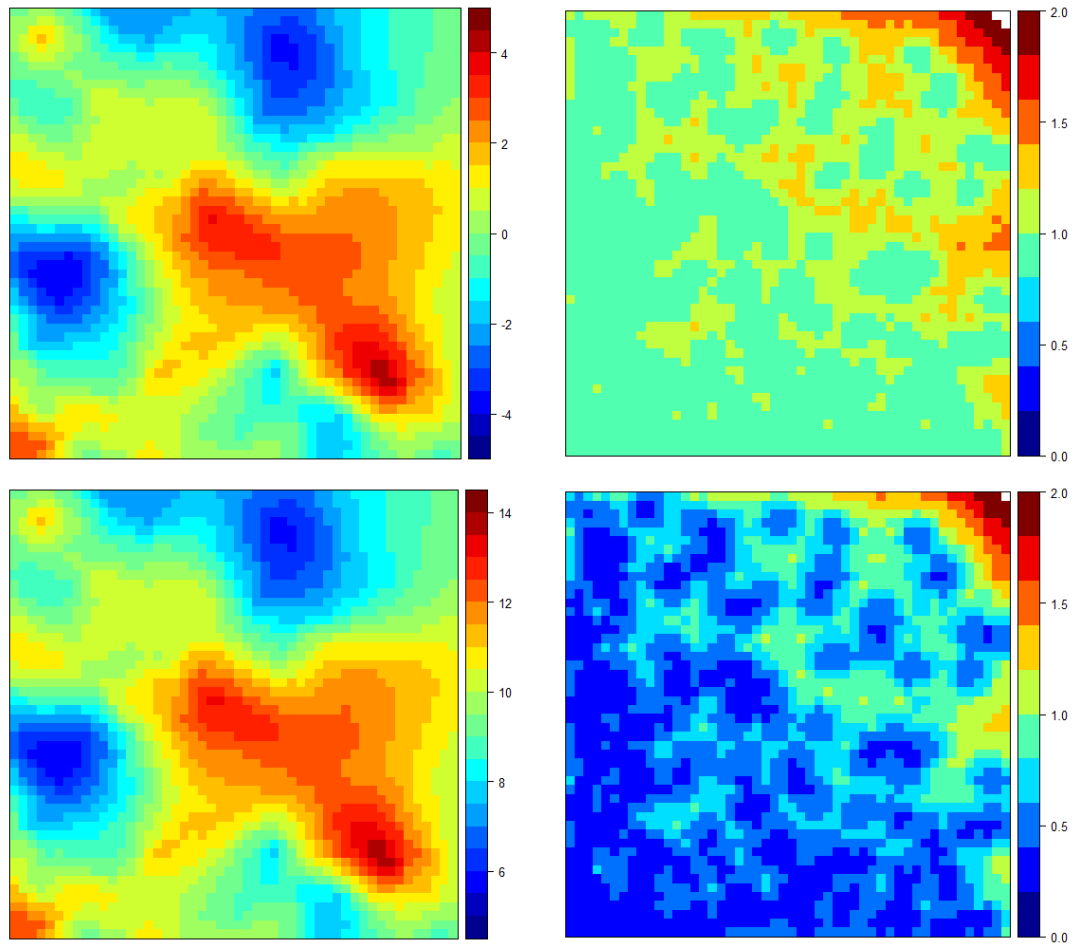
خلاصه‌های پسین برای میدان پنهان را می‌توان به صورت زیر به دست آورد:

```
> round(head(
join.output$summary.linear.predictor[index.pred.latent,1:5],n=3),3)
              mean      sd 0.025quant 0.5quant 0.975quant
Apredictor.0201 2.598 0.856      0.937   2.572   4.415
Apredictor.0202 2.829 0.854      1.134   2.815   4.609
Apredictor.0203 2.787 0.884      1.034   2.775   4.608
> round(head(
join.output$summary.fitted.values[index.pred.latent,1:5],n=3),3)
              mean      sd 0.025quant 0.5quant 0.975quant
fitted.Apredictor.0201 2.598 0.856      0.937   2.572   4.415
fitted.Apredictor.0202 2.829 0.854      1.134   2.815   4.609
fitted.Apredictor.0203 2.787 0.884      1.034   2.775   4.608
```

به طوری که با هم برابرند، چون در درست‌نمایی گوسی مقادیر پیش‌گویی خطی با مقادیر برازش یافته برهم منطبق هستند. میانگین پسین میدان پنهان به صورت زیر به دست می‌آید:

```
> post.mean.pred.latent <-
join.output$summary.linear.predictor[index.pred.latent,"mean"]
```

و نمودار آن در شکل ۷.۳ آمده است. این روش را نیز می‌توان برای انحراف استاندارد و برای خلاصه‌های پسین پاسخ به کار برد. در شکل ۷.۳ میانگین پسین میدان پنهان و پاسخ تنها در مقدار عرض از مبدأ متفاوت‌اند و تفاوتی در الگوی فضایی ندارند. برای انحراف استاندارد به نظر می‌رسد که میدان پنهان تغییرپذیری بیشتری نسبت به پاسخ دارد و این مطلب می‌تواند به وسیله یک همبستگی منفی بین عرض از مبدأ و میدان فضایی بیان شود.



شکل ۷.۳: پیش‌گویی میانگین پسین (نمودار سمت چپ) و انحراف استاندارد پسین (نمودار سمت راست) برای میدان پنهان فضایی (نمودارهای بالا) و متغیر پاسخ (نمودارهای پایین) در سطح توری.

فصل ۴

استنباط بیزی تقریبی در مدل‌های پویا با روش INLA

۱.۴ مقدمه

در این فصل می‌خواهیم با استفاده از روش INLA به کمک مثال‌های شبیه‌سازی شده و واقعی، انواع مختلف مدل‌های پویا را استنباط کنیم. ابتدا گام‌های این روش را برای اجرای سریع استنباط بیزی کامل با استفاده از یک مدل خطی پویا یک‌متغیره بیان می‌کنیم. سپس مثال‌هایی از مدل‌های فضای حالت را در نظر می‌گیریم که می‌توان مستقیماً با استفاده از مدل‌های موجود در بسته $R - INLA$ ، یعنی مدل‌هایی با سیر تکاملی قدم زدن تصادفی و جملات خطا مستقل، استنباط کرد. در نهایت این روش را روی مدل‌هایی که نمی‌توان به‌طور معمول استنباط کرد به‌کار می‌بریم، یعنی مدلی با چندین بردار حالت وابسته که جملات خطای آن‌ها همبسته‌اند. خواهیم دید که نتایج، توانایی INLA را برای استنباط مدل‌های فضای حالت پیچیده گواهی می‌دهد.

در کلیه مثال‌های این فصل پیشین‌های مختلف لگاریتم گاما برای لگاریتم دقت‌های مشاهدات و حالت‌ها به صورت، پیشینی آگاهی بخش با میانگین برابر مقدار واقعی و ضریب تغییرات برابر 0.5 ، پیشینی مبهم که روی مقادیر شبیه‌سازی شده مرکزی شده با ضریب تغییرات برابر با 10 ، مشخص شده‌اند و پیش فرض INLA پیشین لگاریتم گاما به صورت $(1, 0.00005)$ $\log \Gamma$ است.

۲.۴ مثال ۱: مثالی بسیار ساده

بحث را با استنباط مثال شبیه‌سازی شده بسیار ساده از مدل خطی پویا یک‌متغیره ادامه می‌دهیم. مدل معادلات مشاهده و سیستم را به صورت زیر دارد:

$$y_t = \theta_t + \nu_t, \quad \nu_t \sim N(0, V), \quad t = 1, \dots, T \quad (1.4)$$

$$\theta_t = \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W), \quad t = 2, \dots, T \quad (2.4)$$

یعنی فرض کردیم که $F_t = G_t = 1$ و $V_t = V$ و $W_t = W$ به ازای همه t ها است. بنابراین مدل $T + 2$ پارامتر نامعلوم دارد. بردار ابرپارامترها با $\psi = \{V, W\}$ مشخص می‌شود، درحالی‌که میدان پنهان متناظر با $\theta = \{\theta_1, \dots, \theta_T\}$ است. چون تکامل حالت‌ها در این مدل ساده یک فرآیند قدم زدن تصادفی مرتبه اول را دارد، به راحتی می‌توان با استفاده از مدل موجود در بسته R-INLA این مثال را استنباط کرد. اما از این مثال برای تشریح روش استنباط بیزی تقریبی مدل‌های پویا با INLA بر اساس ساختار یک مدل تکمیل یافته^۱ استفاده خواهیم کرد. این مثال ساده توانایی استنباط مستقیم مدل‌های فضای حالت پیچیده را که در مثال‌های بعدی به آن‌ها اشاره شده، به ما نشان می‌دهد. ویژگی کلیدی این روش، برابر با صفر قرار دادن معادله یا معادلات سیستم مدل فضای حالت است، پس در این مثال معادله (۲.۴) را به صورت زیر بازنویسی می‌کنیم:

$$0 = \theta_t - \theta_{t-1} - \omega_t, \quad t = 2, \dots, T \quad (3.4)$$

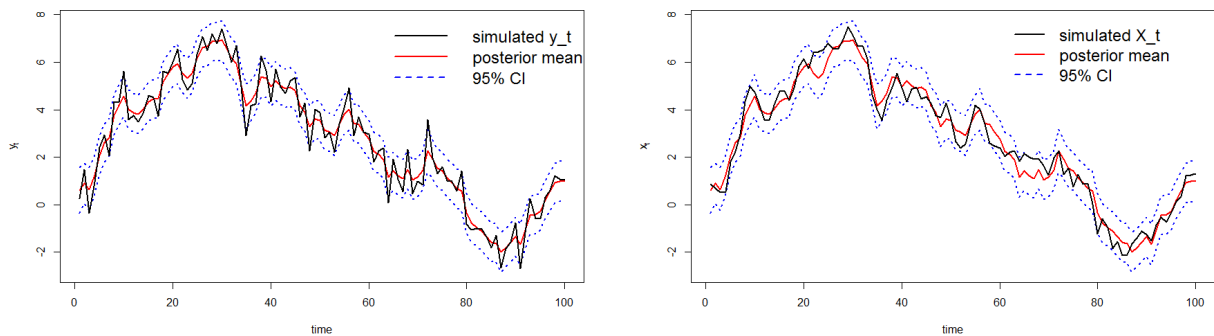
سپس یک مدل تکمیل یافته را می‌سازیم که در این مثال $T + (T - 1)$ بُعد خواهد داشت، به طوری که با ادغام کردن صفرهای مشاهدات ساختگی از معادله سیستم (۳.۴)، با مشاهدات واقعی از معادله (۱.۴) به این مدل تکمیل یافته می‌رسیم که در شکل ۴.۴ نشان داده شده است.

$$\begin{bmatrix} y_1 & NA \\ \vdots & \vdots \\ y_T & NA \\ NA & 0 \\ \vdots & \vdots \\ NA & 0 \end{bmatrix} \quad (4.4)$$

ستون اول این ساختار مربوط به مشاهدات واقعی $y_t = \{y_1, \dots, y_T\}$ است که T عضو اول را اشغال کرده است. ستون دوم مربوط به معادله سیستم است که $T - 1$ عضو آخر را دارد که متناظر با تعداد پارامترهای حالت معادله (۳.۴) است و برابر صفر هستند. کلیه عضوهای دیگر این ساختار مقداری ندارند (NA). این ساختار تکمیل یافته زمانی ستون‌های دیگری خواهد داشت که مدل فضای حالت بیش از یک معادله سیر تکاملی را دربر داشته باشد.

استنباط این مدل تکمیل یافته با استفاده از روش INLA با در نظر گرفتن درست‌نمایی‌های مختلف برای هر ستون این ساختار (دو ستون در این مثال) اجرا می‌کنیم. بنابراین، حالت‌های θ_t و داده‌های

^۱Augmented model



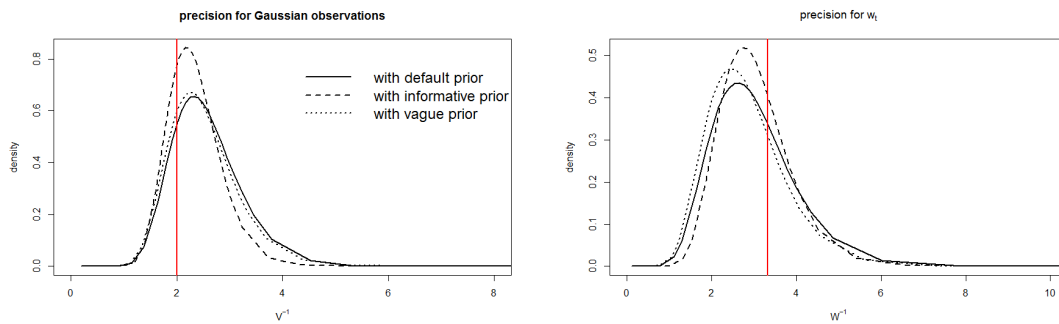
شکل ۱.۴: مقادیر شبیه‌سازی شده و برآورد شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) برای مشاهدات (سمت چپ) و حالت‌ها (سمت راست) در مثال ساده

واقعی در ستون اول این ساختار دارای توزیع گوسی با دقت نامعلوم V^{-1} هستند. مشاهدات ساختگی در ستون دوم معلومند (با واریانس صفر) که روی مقادیر θ_t ، θ_{t-1} و ω_t در (۲.۴) شرطی شده‌اند. این شرط در INLA با فرض این‌که این مشاهدات ساختگی دارای توزیع گوسی با یک دقت ثابت و بالا هستند، ارائه می‌شود.

به عبارت دیگر برای برآورد حالت‌های θ_t در این مدل، تنها نیاز به دانستن ω_t ، $t = \{2, \dots, T\}$ داریم. بنابراین، در نظر می‌گیریم که θ_t در پس سیر تکاملی زمانی‌اش، اطلاعاتی وجود ندارد. در فرمول‌بندی مدل با این اطلاع، از اندیس‌های i ، j و l که به ترتیب مربوط به θ_t ، θ_{t-1} و ω_t هستند، استفاده می‌کنیم. جملات θ_{t-1} به عنوان یک کپی از جملات θ_t مدل‌بندی می‌شوند، به طوری که مقدار موقعیت $k+1$ در اندیس j برابر با مقدار موقعیت k در اندیس i است؛ این تأخیر به سیر تکاملی زمانی θ_t منتهی می‌شود. برای فرمول‌بندی مدل چون حالت‌ها را به طور تصادفی در نظر نمی‌گیریم، باید مقادیر جملات θ_t را در یک فضای پارامتری قرار دهیم، یعنی x را به عنوان یک بردار مستقل تعریف کرده و این متغیرها دارای توزیع گوسی با یک دقت ثابت و پایین (واریانس زیاد) هستند. در $t = 1$ ، θ_t مستقیماً از این توزیع غیرآگاهی بخش می‌آید، چون θ_t ، $t = 2, \dots, T$ ، از معادله (۲.۴) پیروی می‌کند.

مدل تکمیل یافته، اندکی ساختار استاندارد مدل فضای حالت را که مشاهدات ساختگی با درست‌نمایی‌ها (یک در این مثال ساده) بیان شده‌اند، تغییر می‌دهد که منجر به مدل‌بندی حالت‌ها به وسیله‌ی INLA می‌شود. دقت‌های ثابت مربوط به مشاهدات ساختگی جدید و حالت‌ها نیز به مدل اضافه می‌شوند.

مقادیر سری شبیه‌سازی شده و برآورد شده برای بردارهای مشاهدات و حالت‌ها در شکل ۱.۴ آمده‌اند. دقت‌های دو جمله اغتشاش نیز به خوبی برآورد شده است که می‌توان در شکل ۲.۴ به همراه فواصل اعتبار آن‌ها مشاهده نمود.



شکل ۲.۴: مقادیر شبیه‌سازی شده و برآورد شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) برای مشاهدات (سمت چپ) و حالت‌ها (سمت راست) در مثال ساده.

۳.۴ مثال ۲: مدل رگرسیون پویا تعمیم‌یافته

در این بخش می‌خواهیم با استفاده از مدل‌هایی که در تابع INLA وجود دارد، مستقیماً استنباط مدل‌های فضای حالت (مدل رگرسیون پویا تعمیم‌یافته) را اجرا کنیم. ویژگی‌های اصلی این مدل استقلال جملات خطا و سیرتکاملی ساده برای حالت‌ها هستند.

داده‌ها را از مدل رگرسیون پواسون چندگانه با دو رگرسور Z_{1t} و Z_{2t} شبیه‌سازی می‌کنیم. بنابراین، پیش‌گوی خطی با $\lambda_t = \mathbf{F}_t \boldsymbol{\theta}_t$ به دست می‌آید که در آن $\mathbf{F}_t = (1, Z_{1t}, Z_{2t})$ و ضرایب رگرسیون $\boldsymbol{\theta}_t = (\beta_{0t}, \beta_{1t}, \beta_{2t})$ هستند که این ضرایب دارای مدل قدم زدن تصادفی می‌باشند. مدل دارای معادلات مشاهده و سیستم به صورت زیر

$$y_t | \mu_t \sim \text{Poi}(\mu_t)$$

$$\log(\mu_t) = \lambda_t = \beta_{0t} + \beta_{1t}Z_{1t} + \beta_{2t}Z_{2t}, \quad t = 1, \dots, T$$

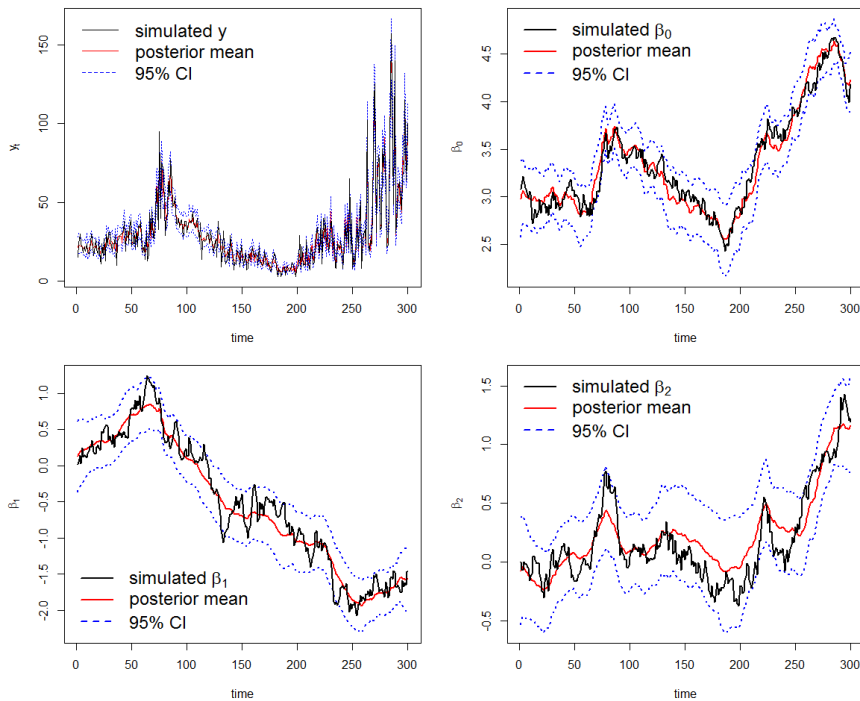
$$\beta_{0t} = \beta_{0,t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \quad t = 2, \dots, T$$

$$\beta_{1t} = \beta_{1,t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), \quad t = 2, \dots, T$$

$$\beta_{2t} = \beta_{2,t-1} + \omega_{3t}, \quad \omega_{3t} \sim N(0, W_3), \quad t = 2, \dots, T$$

است. علاقه‌مند به برآورد بیزی ضرایب رگرسیون پویا، $\beta_{it} = \{\beta_{i1}, \dots, \beta_{iT}\}$ و واریانس‌ها، W_i ، $i \in \{1, 2, 3\}$ هستیم.

توجه داشته باشید که روش‌های شناخته شده‌ای که استنباط بازگشتی مدل‌های خطی پویا تعمیم‌یافته را انجام می‌دهند، مانند آن‌چه در وست و همکاران (۱۹۸۵) با استفاده از تقریب بیزی خطی آمده یا بسته‌های *sspir* و *KFAS* که در R آمده است، قادر به استنباط این مدل چندگانه نیستند یا حتی نمی‌توانند برآوردهای دقیق پارامترهای مورد نظر را به دست آورند. اما تابع *INLA* برآوردهای کامل بیزی پارامترها را برای این مدل و مدل‌های دیگر به روشی آسان و زمانی کوتاه (در این مثال ۲/۷۹ ثانیه) با استفاده از مدل قدم زدن تصادفی مرتبه اول که برای سیرتکاملی ضرایب است، به دست می‌دهد. مقایسه بین



شکل ۳.۴: مقادیر شبیه‌سازی شده و برآورد شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) برای مشاهدات و ضرایب رگرسیون $\beta_{0t}, \beta_{1t}, \beta_{2t}$.

ضرایب رگرسیون شبیه‌سازی شده و پیش‌گو شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) در شکل ۳.۴ نشان داده شده است.

۴.۴ مثال ۳: مدل پویای فصلی با همسازها

در این مثال، مدل پویای فصلی با دو همساز را در نظر می‌گیریم، به طوری که ساختار سیر تکاملی حالت‌ها در این دسته از مدل‌های فضای حالت، برای بیش از یک بردار حالت است.

در این مثال یک سری زمانی ماهانه را با الگوی فصلی سالیانه با استفاده از مجموع جملات سینوس و کسینوس که بسامدی مشابه دارند، شبیه‌سازی می‌کنیم. این مدل دو پارامتر حالت دارد و

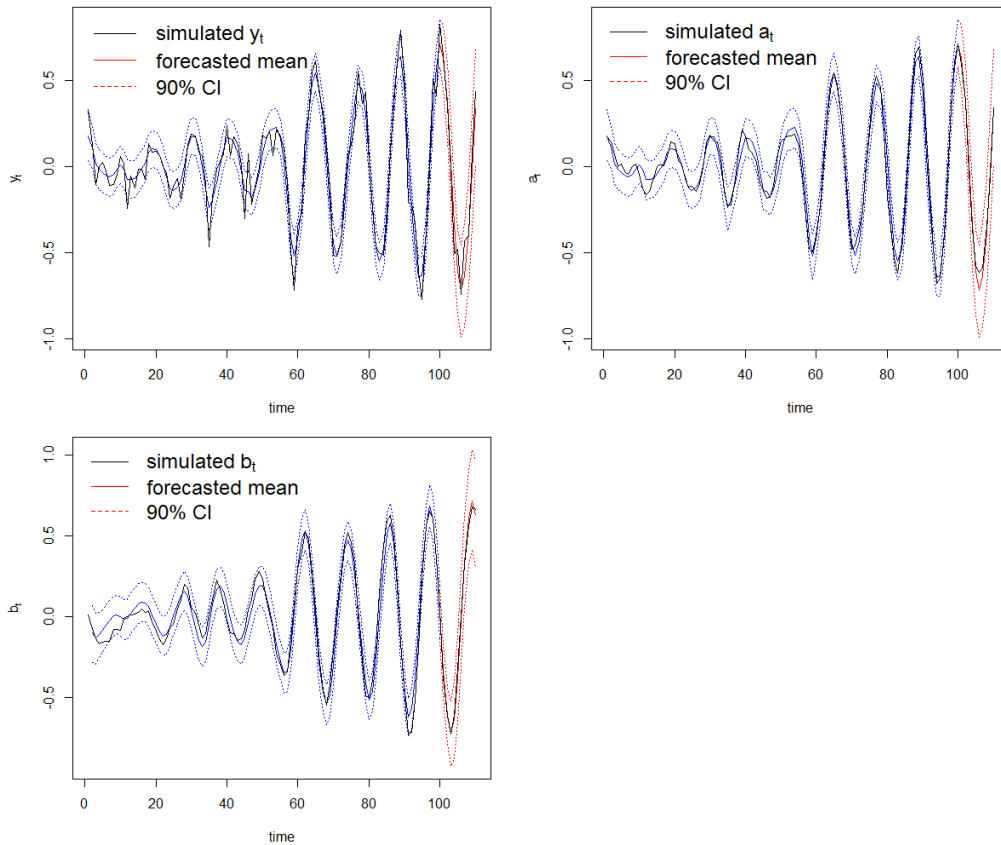
$$\theta_t = \begin{pmatrix} a_t \\ b_t \end{pmatrix}; \quad F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \quad G_t = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}$$

به طوری که $\phi = \frac{\pi}{6}$. بنابراین معادلات مشاهده و سیستم به صورت زیر هستند

$$y_t = a_t + \nu_t, \quad \nu_t \sim N(0, V), \quad t = 1, \dots, T \quad (5.4)$$

$$a_t = \cos(\phi)a_{t-1} + \sin(\phi)b_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), \quad t = 2, \dots, T \quad (6.4)$$

$$b_t = -\sin(\phi)a_{t-1} + \cos(\phi)b_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), \quad t = 2, \dots, T. \quad (7.4)$$



شکل ۴.۴: مقادیر شبیه‌سازی شده و برآورد شده (میانگین پسین و فاصله اعتبار ۹۵ درصد) برای مشاهدات و حالت‌های a_t و b_t در مدل پویای فصلی. خط‌های قرمز برای ۱۰ ماه پیش‌بینی شده است.

توجه داشته باشید که سیرتکاملی برای a_t در معادله (۶.۴) هم به a_{t-1} و هم به b_{t-1} وابسته است. این ساختار سیرتکاملی را نمی‌توان مستقیماً در تابع INLA محاسبه نمود. بنابراین، از ساختار مدل تکمیل یافته (مثال ۱) استفاده می‌کنیم. پس معادلات (۶.۴) و (۷.۴) را برابر با صفر قرار داده و سپس این مشاهدات ساختگی را با مشاهدات معادله (۵.۴) ادغام می‌کنیم. ساختار تکمیل یافته این مدل $T + 2(T - 1)$ بُعد دارد که عضوهای این ساختار در ستون اول دارای توزیع گوسی با دقت نامعلوم V^{-1} هستند، در حالی که مشاهدات ساختگی در دو ستون دیگر با توزیع گوسی که دقت بالا و ثابتی دارد مدل‌بندی شده‌اند. طول سری زمانی این مثال ۱۱۰ ماه است، اما برازش مدل را با ۱۰۰ ماه ابتدایی اجرا کرده و ۱۰ ماه آخر را برای مقایسه مقادیر شبیه‌سازی شده و پیش‌بینی شده به‌کار می‌بریم. برآوردها با مقادیر شبیه‌سازی شده و حتی با مقادیر پیش‌گو شده به خوبی مطابقت دارند. زمان صرف شده برای استنباط این مدل ۲/۵۶ ثانیه است که باز هم تأکیدی بر سریع بودن روش INLA دارد. میانگین‌های پسین برای مشاهدات و حالت‌ها، ناحیه پیش‌بینی و فواصل اعتبار ۹۰ درصد در شکل ۴.۴ آمده است.

۵.۴ مدل‌هایی با ساختار خطای همبسته

در این بخش با استفاده از دو مثال شبیه‌سازی شده از مدل‌های پویای فضایی-زمانی گوسی برای داده‌های فضایی (ویوار و فریرا، ۲۰۰۹)، اجرای روش INLA را برای این مدل‌ها شرح می‌دهیم. همچنین، خواهیم دید که این روش می‌تواند به راحتی مدل‌های پویا فضایی-زمانی پیچیده را با ساختارهای خطای همبسته استنباط کند.

۱.۵.۴ مثال ۴: مدل پویای فضایی-زمانی مرتبه اول با متغیرهای تبیینی

در این مثال می‌خواهیم یک مدل گوسی مرتبه اول نایستا را که متغیر تبیینی آن ساختار فضایی دارد شبیه‌سازی کرده و سپس به کمک روش INLA استنباط کنیم. در این مدل برای هر زمان $t = \{1, \dots, T\}$ و هر مکان $s = \{1, \dots, d\}$ ، متغیر پاسخ y_{ts} به صورت زیر مدل‌بندی می‌شود

$$\mathbf{y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \boldsymbol{\gamma}_t \mathbf{z}_t + \boldsymbol{\omega}_{1t}, \quad \boldsymbol{\omega}_{1t} \sim \text{PGMRF}(\circ \mathbf{s}, \mathbf{W}_1^{-1}) \quad (۸.۴)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_{2t}, \quad \boldsymbol{\omega}_{2t} \sim \text{PGMRF}(\circ \mathbf{s}, \mathbf{W}_2^{-1}) \quad (۹.۴)$$

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \boldsymbol{\omega}_{3t}, \quad \boldsymbol{\omega}_{3t} \sim \text{PGMRF}(\circ \mathbf{s}, \mathbf{W}_3^{-1}) \quad (۱۰.۴)$$

که در آن $\mathbf{y}_t = (y_{t1}, \dots, y_{ts})'$ و $\boldsymbol{\gamma}_t = (\gamma_{t1}, \dots, \gamma_{ts})'$ به ترتیب مشاهدات و بردار متغیرهای تبیینی در زمان t هستند. $\mathbf{G}_t = \rho \mathbf{I}_s$ ، $\mathbf{F}_t = \mathbf{I}_s$ و \mathbf{I}_s یک ماتریس مشخصه $d \times d$ است. بردار خطاها $\omega_{i1}, \dots, \omega_{iT}$ ، $i \in \{1, 2, 3\}$ فرض می‌شود که مستقل هستند و با میدان‌های تصادفی مارکوف گوسی مناسب^۲ مدل‌بندی شده‌اند. ماتریس‌های \mathbf{W}_i بیانگر ساختار کوواریانس فضایی $\mathbf{W}_i^{-1} = \tau_i (\mathbf{I}_s - \frac{\phi_i}{\lambda_{\max}} \mathbf{C})$ به صورت دقت \mathbf{W}_i^{-1} می‌باشند. $\omega_{it} = (\omega_{it1}, \dots, \omega_{its})'$ مدل‌بندی شده‌اند، به طوری که ماتریس \mathbf{C} به صورت

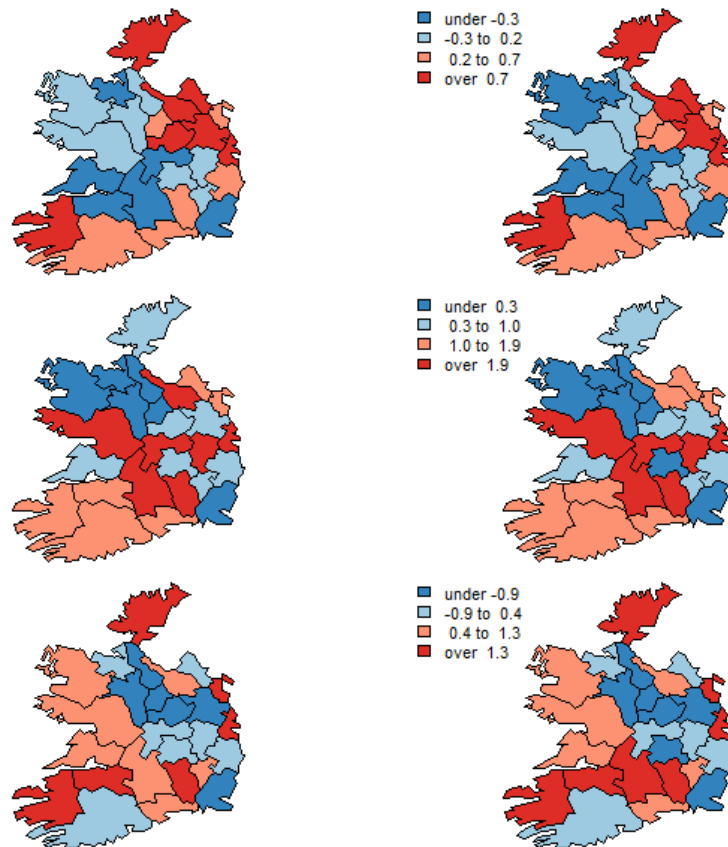
$$C_{k,l} = \begin{cases} \circ & \text{اگر } k = l \\ h_{k,l} & \text{اگر } k \in d_l \\ \circ & \text{جاهای دیگر} \end{cases}$$

است، که d_l مجموعه‌ی همسایگی مکان l ، $h_{k,l} > \circ$ اندازه شباهت بین مکان k و l (در این جا فرض می‌کنیم $h_{k,l} = 1$) است. λ_{\max} ماکسیمم مقدار ویژه ماتریس \mathbf{C} است؛ τ_i ها پارامترهای مقیاس هستند و $\circ \leq \phi_i < 1$ درجه همبستگی فضایی را کنترل می‌کنند.

یک سری زمانی با $T = 100$ برای $d = 26$ ناحیه در نقشه Eire شبیه‌سازی می‌کنیم. این نقشه در بسته spdep وجود دارد (بایوند، ۲۰۱۰). استنباط برای بردار حالت‌های $\boldsymbol{\theta}_t$ و \mathbf{z}_t و همچنین برای مقیاس τ_i و پارامترهای همبستگی ϕ_i ، $i = \{1, 2, 3\}$ ، اجرا می‌کنیم، اما چون مقادیر ρ به دلیل فرآیند نایستایی این مدل ثابت فرض شدند، استنباط را برای آن اجرا نمی‌کنیم.

مشاهدات و حالت‌های پیش‌گو شده به همراه مقادیر شبیه‌سازی شده آن‌ها برای کلیه مکان‌ها در نقشه‌های ۵.۴، ۶.۴ و ۷.۴ نشان داده شده‌اند.

^۲proper Gaussian Markov Random Fields



شکل ۵.۴: نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای مشاهدات y_t در زمان‌های ۱۸، ۵۱ و ۷۷ برای مثال ۴.

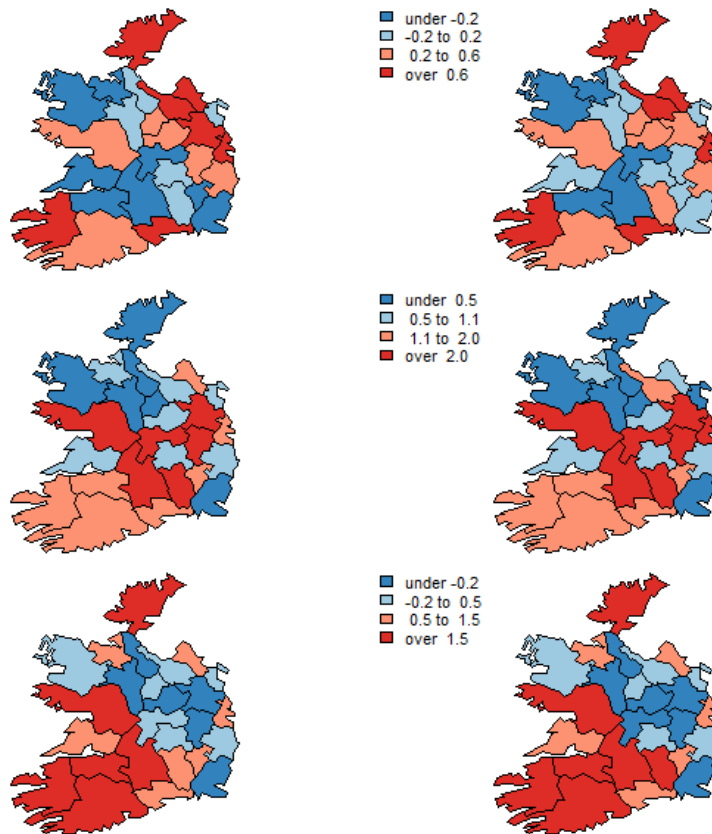
با استفاده از چارچوب روش INLA، استنباط بیزی کامل این مدل پیچیده تنها با نوشتن چند خط در R ممکن شده است. زمان صرف شده برای استنباط این مدل ۵۱۴۰/۵۲ ثانیه است.

۲.۵.۴ مثال ۵: مدل پویای فضایی-زمانی مرتبه‌ی دوم

در این مثال داده‌ها را از یک مدل گوسی مرتبه‌ی دوم نایستا، بدون متغیر تبیینی شبیه‌سازی می‌کنیم (ویوار و فریرا، ۲۰۰۹). برای این مدل داریم

$$y_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \omega_{1t}, \quad \omega_{1t} \sim \text{PGMRF}(\circ s, \mathbf{W}_1^{-1})$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \omega_{23t}, \quad \omega_{23t} \sim \text{PGMRF}(\circ s, \mathbf{W}_{23}^{-1})$$



شکل ۶.۴: نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای حالت x_t در زمان‌های ۱۸، ۵۱ و ۷۷ برای مثال ۴.

که در این معادلات

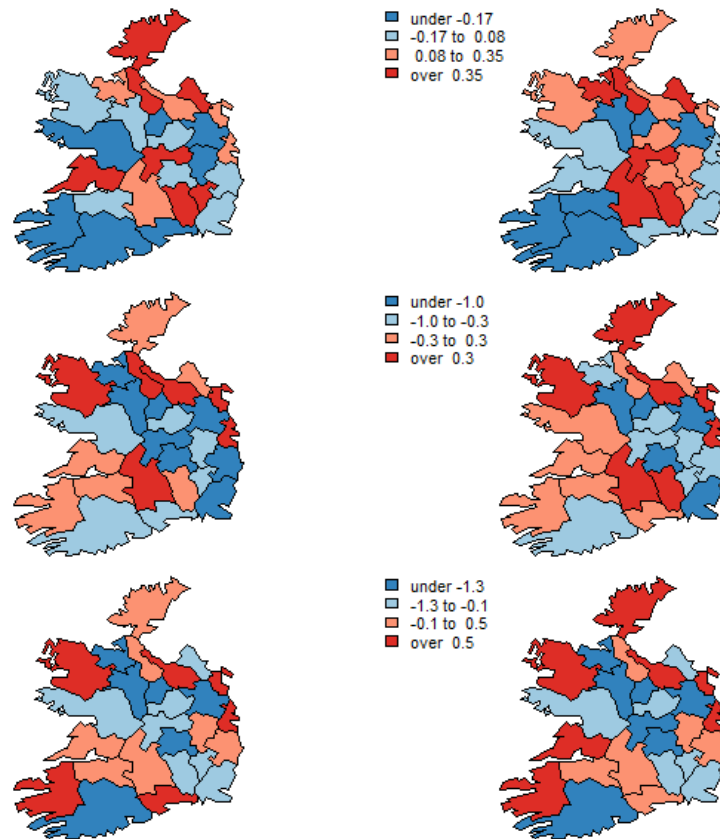
$$\theta_t = \begin{pmatrix} \theta_{1t} \\ \theta_{2t} \end{pmatrix}, \quad \mathbf{F}_t = \begin{pmatrix} \text{Is} \\ \circ \text{s} \end{pmatrix}, \quad \mathbf{G}_t = \begin{pmatrix} \rho_1 \text{Is} & \rho_1 \text{Is} \\ \circ \text{s} & \rho_2 \text{Is} \end{pmatrix},$$

$$\omega_{23t} = \begin{pmatrix} \omega_{2t} \\ \omega_{3t} \end{pmatrix}, \quad \mathbf{W}_{23}^{-1} = \begin{pmatrix} \mathbf{W}_2^{-1} & \circ \text{s} \\ \circ \text{s} & \mathbf{W}_3^{-1} \end{pmatrix}$$

هستند. در این جا سری زمانی برای ۱۰۰ ناحیه و ۳۰ زمان در نقشه‌ی کارولینای شمالی شبیه‌سازی شده است. استنباط را برای حالت‌های θ_1 و θ_2 و همچنین برای پارامترهای مقیاس τ_i و همبستگی ϕ_i ، اجرا می‌کنیم. فرض می‌کنیم که در این فرآیند نایبستا $\rho_1 = \rho_2 = 1$ است.

مقادیر پیش‌گو شده و شبیه‌سازی شده مشاهدات و حالت‌ها به خوبی مطابقت دارند که می‌توان در نقشه‌های ۸.۴، ۹.۴ و ۱۰.۴ مشاهده نمود. برای استنباط این مدل که یکی از پیچیده‌ترین نوع مدل‌های فضای حالت می‌باشد، حدود ۵۲ دقیقه به وسیله روش INLA زمان صرف شده است.

به منظور بررسی واکنش این مدل به تغییرات مقادیر اولیه ابرپارامترها و همچنین پیشین‌های مختلف آن‌ها، تحلیل با حساسیت را اجرا کرده‌ایم. ابتدا مدلی با پیشین‌های آگاهی‌بخش، مبهم و لگاریتم گاما برای $\log(\tau_i)$ ، $i \in \{1, 2, 3\}$ ، برازش می‌دهیم؛ در این حالت، پیشین برای پارامترهای همبستگی ϕ_i



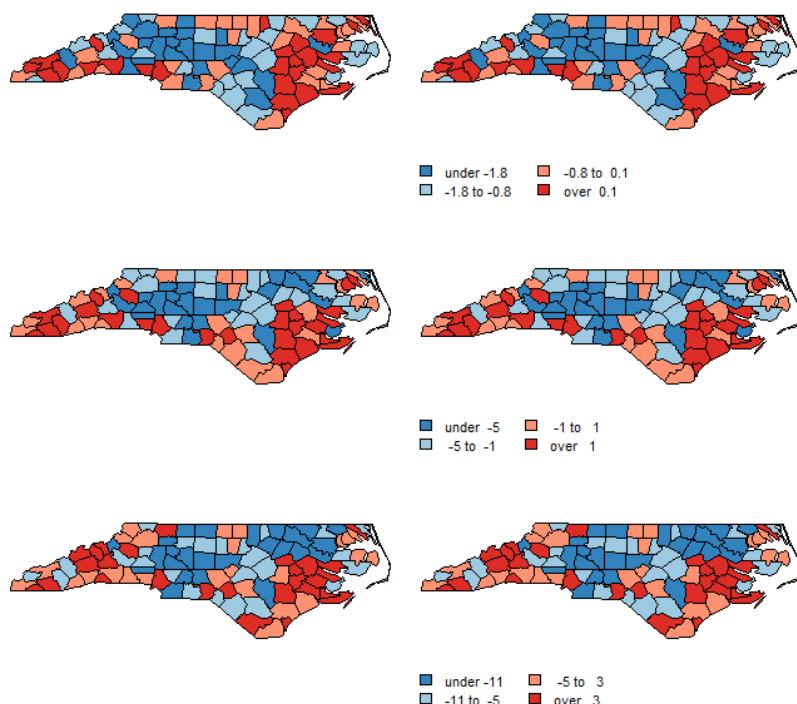
شکل ۷.۴: نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای حالت z_t در زمان‌های ۱۸، ۵۱ و ۷۷ برای مثال ۴.

و مقادیر اولیه ابرپارامترها به صورت پیش‌فرض تعیین شده‌اند. علاوه بر این، برای کلیه ابرپارامترها پیشین پیش‌فرض را مشخص کرده و مدل تحت دو موقعیت برازش می‌دهیم:

۱. مقادیر اولیه مختلف برای $\log(\tau_i)$ در مجموعه‌ی $\{\log(10), \log(20), \dots, \log(90), \log(100)\}$ انتخاب می‌شوند که مقادیر اولیه برای پارامترهای همبستگی پیش‌فرض هستند؛

۲. مقادیر اولیه مختلف ϕ_i در مجموعه‌ی $\{0/1, 0/2, \dots, 0/8, 0/9\}$ انتخاب می‌شوند که پارامترهای لگاریتم دقت مقادیر اولیه پیش‌فرض دارند.

در هر زمان، ۳ پارامتر لگاریتم دقت یا ۳ پارامتر همبستگی مدل از مجموعه مقادیر بالا مقداردهی اولیه شده‌اند. به‌طور کلی، مدل به تعیین پیشین لگاریتم دقت‌ها، $\log(\tau_i)$ ، با چگالی‌های حاشیه‌ای پسین پراکنده حساسیتی ندارد، اما برعکس، انتخاب مقادیر اولیه $\log(\tau_i)$ تأثیری مستقیم روی تقریب چگالی‌های حاشیه‌ای پسین برای این ابرپارامترها دارد که با توجه به این مطلب بهترین نتایج برای مقادیر اولیه در محدوده‌ی $\log(40)$ تا $\log(60)$ به‌دست آوردیم، به‌طوری که شامل مجموعه مقادیر اولیه پیش‌فرض تابع INLA می‌باشد. نتایج پسین برای پارامترهای همبستگی حساسیت کمی در انتخاب مقادیر اولیه داشت.



شکل ۸.۴: نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای مشاهدات y_t در زمان‌های ۷، ۱۵ و ۲۹ برای مثال ۵.

۶.۴ مطالعات موردی

در این بخش، برخی از مدل‌های فضای حالت با داده‌های واقعی را به کمک روش INLA استنباط می‌کنیم. همچنین، به مقایسه روش INLA با روش‌های استنباطی دیگر نیز می‌پردازیم.

۱.۶.۴ مثال ۶: مصرف گاز کشور انگلستان

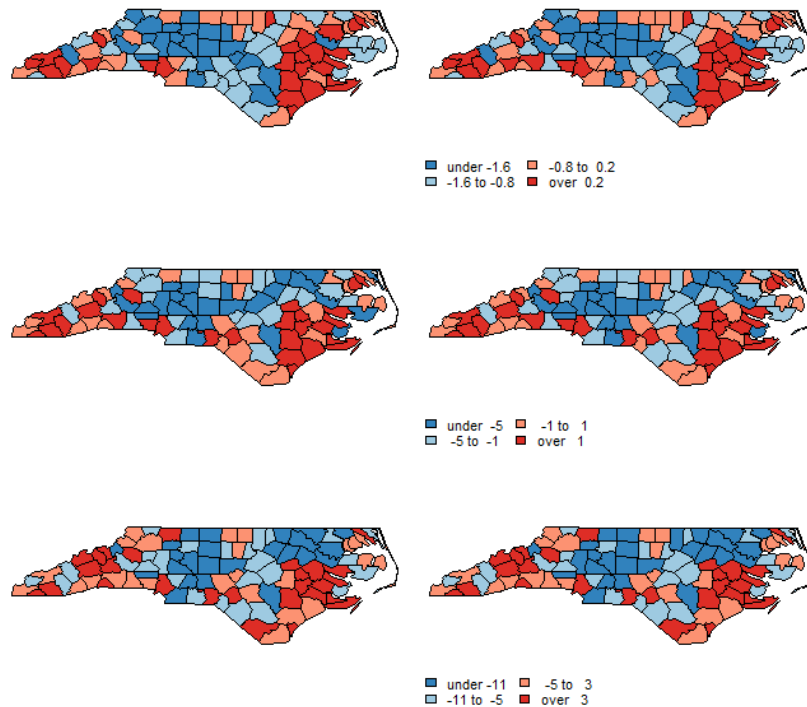
اولین مثال واقعی میزان مصرف گاز انگلستان است که در فصل اول به بیان آن پرداختیم. بر اساس دلفسن و لوندبای-کریستنسن (۲۰۰۶)، از لگاریتم میزان مصرف گاز به‌عنوان متغیر پاسخ استفاده می‌کنیم، به طوری که فرض می‌کنیم دارای توزیع نرمال هستند و مدلی با روند چندجمله‌ای مرتبه اول (T_t) با ضرایب تغییرپذیر زمان و مؤلفه فصلی بدون ساختار (S_t) با تغییرپذیری روی زمان، برازش می‌دهیم. بنابراین معادلات سیستم و مشاهده به صورت زیر آمده‌اند

$$y_t = \log_{10}(\text{gas UK})_t = T_t + S_t + \nu_t, \quad \nu_t \sim N(0, V), t = 1, \dots, T \quad (11.4)$$

$$T_t = T_{t-1} + \beta_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, W_1), t = 2, \dots, T \quad (12.4)$$

$$\beta_t = \beta_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, W_2), t = 2, \dots, T \quad (13.4)$$

$$S_t = -(S_{t-1} + S_{t-2} + S_{t-3}) + \omega_{3t}, \quad \omega_{3t} \sim N(0, W_3), t = 4, \dots, T \quad (14.4)$$

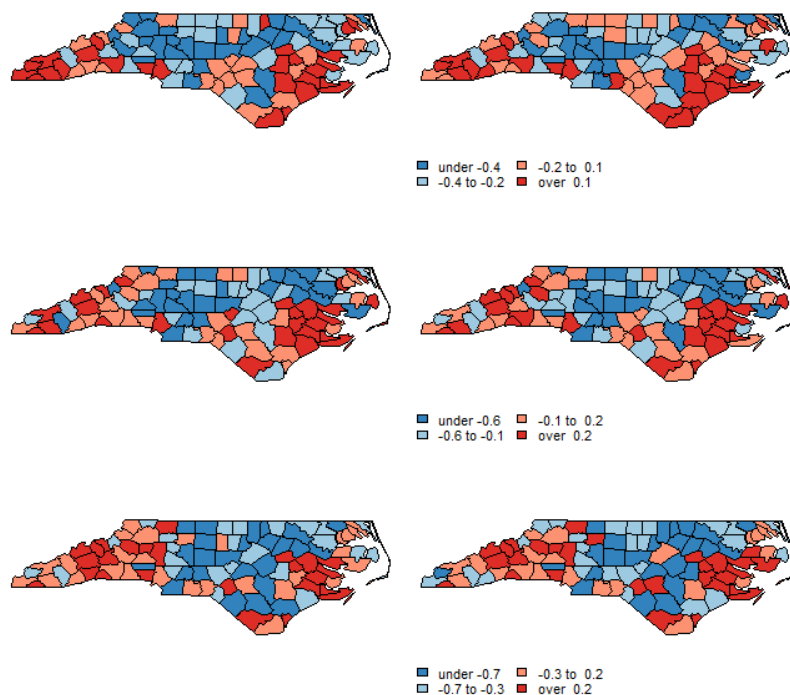


شکل ۹.۴: نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای بردار حالت x_1 در زمان‌های ۷، ۱۵ و ۲۹ برای مثال ۵.

استنباط بیزی تقریبی این مدل به کمک روش ترکیب شده، اجرا می‌شود. ابتدا ساختار تکمیل یافته‌ای را به کار می‌بریم که مشاهدات معادله (۱۱.۴) را با روند چندجمله‌ای معادله (۱۲.۴) ادغام می‌کند. جملات شیب و مؤلفه فصلی در معادلات (۱۳.۴) و (۱۴.۴) با مدل‌های استاندارد که در تابع INLA وجود دارد، مدل‌بندی شده‌اند، به طوری که برای β_t مدل قدم زدن تصادفی مرتبه اول و برای S_t مدل فصلی INLA را به کار می‌بریم.

به کمک روش MCMC و با استفاده از الگوریتم نمونه‌گیر گیبز که در بسته dlm وجود دارد، داده‌های این مثال را نیز استنباط کرده‌ایم. به منظور دستیابی به برآوردهای مناسب (دقیق) پارامترهای مورد نظر برای مقایسه‌ی بین روش‌های MCMC و INLA، ۱۰۰۰۰۰ نمونه‌ی پسین از زنجیره‌ی MCMC با دوره‌ی داغیدن ۱۰۰۰ و دوره‌ی نازکیدن ۲۰ به کار برده‌ایم. بنابراین، ۴۹۵۰ نمونه مؤثر برای برآورد کمیت‌های پسین باقی مانده‌اند. پیشین‌ها برای پارامترهای دقت مشاهدات و سیر تکاملی خطاها مشابه آن‌چه که در پتریس (۲۰۱۰) آمده، می‌باشند. بسته CODA (پلامر و همکاران، ۲۰۰۶) برای تشخیص همگرایی بر اساس آزمون‌های تشخیص گوتیک و رفتی-لوئیز مورد استفاده قرار گرفته است. شکل ۱۱.۴ چگالی‌های حاشیه‌ای پسین ابرپارامترها که با دو روش MCMC و INLA به دست آمده، مقایسه می‌کند. پیش‌بینی برای ۱۲ دوره (چهار ماه) بعدی این سری زمانی نیز در این شکل با این دو روش به دست آمده است.

بر اساس نمودارهای ۱۱.۴ و ۱۲.۴ هر دو روش هم در برآورد پارامتر و هم در پیش‌بینی عملکرد مشابه‌ای دارند. هرچند که اختلاف این دو روش در زمان صرف شده برای اجرای استنباط است. زمان

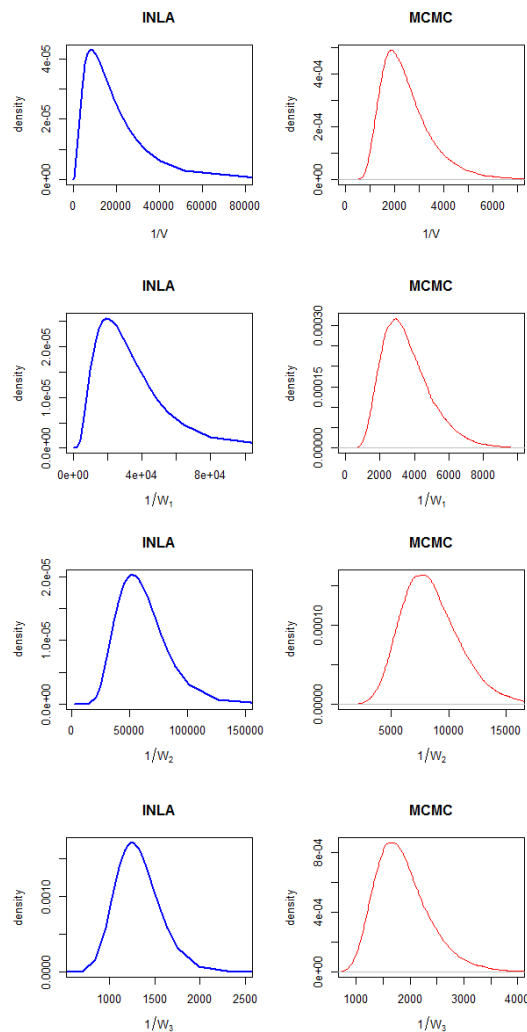


شکل ۱۰.۴: نقشه‌های مقادیر شبیه‌سازی شده (چپ) و پیش‌گو شده (راست) برای بردار حالت x_2 در زمان‌های ۷، ۱۵ و ۲۹ برای مثال ۵.

صرف شده برای اجرای الگوریتم MCMC، بسته‌ی dlm، حدود ۳۱۳۹ ثانیه بود، اما روش INLA برای استنباط این مدل فقط ۲/۹ ثانیه را صرف کرد (تقریباً ۱۰۸۰ برابر سریع‌تر است). شاید در نگاه اول این مقایسه‌ی زمان حائز اهمیت نباشد، چون روش INLA به شدت بهینه‌سازی شده است و بسته‌ی dlm کاملاً براساس R است و بنابراین، الگوریتم MCMC کندتر می‌باشد.

۲.۶.۴ مثال ۷: کیفیت آب‌های زیرزمینی استان گلستان

آب زیرزمینی به‌طور کامل خالص و تمیز نیست، و همیشه دارای مقداری مواد محلول است. مقادیر این مواد محلول، غلظت و نوع آن‌ها، شایستگی کاربرد آب برای اهداف مختلف را تعیین خواهند کرد. از این‌رو، مطالعات ژئوشیمیایی آب زیرزمینی، با توجه به کاربرد آب، اهمیت دارد. این مطالعات، به درک و فهم عمیق‌تر و بهتر فرآیندهای تکاملی و کیفی آب در منطقه مورد مطالعه، کمک خواهد کرد و می‌تواند اطلاعاتی در مورد محدودیت‌های کلی توسعه یا برنامه‌ریزی برای تصفیه مناسب آب فراهم آورد. این اطلاعات با توجه به تغییرات آبی در کیفیت منابع آبی، مورد نیاز است. زمانی که آب زیرزمینی در محیط زیر زمین جریان پیدا می‌کند، تغییرات شیمیایی آن با افزایش مواد جامد محلول و یون‌های اصلی، همراه است. هرچه مدت ماندگاری آب در زیر زمین زیاد باشد، کیفیت آن نامناسب‌تر خواهد بود. همچنین سنجش و مطالعه خواص شیمیایی آبخوان‌های ساحلی و حاشیه دریاچه، به دلیل تهدید آلودگی آب زیرزمینی با آب دریا و دریاچه، پیچیده است.



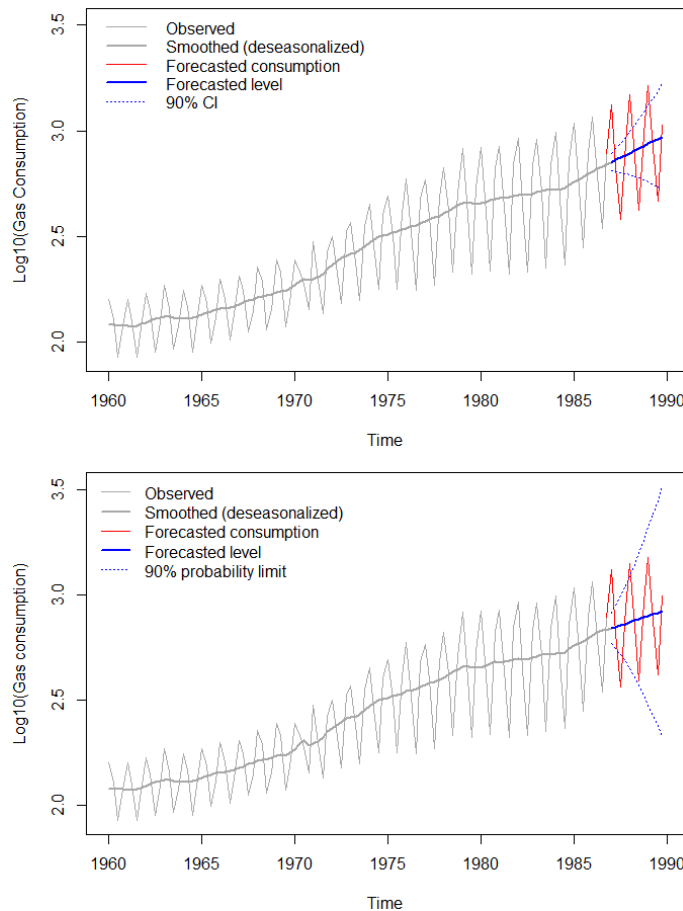
شکل ۱۱.۴: چگالی‌های حاشیه‌ای پسین برای پارامترهای دقت که به وسیله دو روش INLA (چپ) و MCMC (راست) به دست آمده‌اند.

آب‌های زیرزمینی، در منطقه مورد مطالعه، برای اهداف مختلفی مانند شرب، کشاورزی و صنعت مورد استفاده قرار می‌گیرند. به دلیل اهمیتی که آب‌های زیرزمینی در منطقه مورد مطالعه دارند، پایش، مطالعه و شناخت فرآیندها و پارامترهای حاکم و مؤثر بر کیفیت آب زیرزمینی، از نظر شرب، کشاورزی، و صنعت، ضروری است. کیفیت آب زیرزمینی اساساً به کیفیت کاتیون‌ها و آنیون‌های موجود در آن بستگی دارد. تحلیل‌های شیمیایی بر اساس تفسیر کیفیت آب در ارتباط با منشا، زمین‌شناسی، اقلیم و استفاده آب انجام می‌گیرند.

توانایی اندازه‌گیری آب برای عبور جریان الکتریکی، هدایت الکتریکی^۳ (EC) نامیده می‌شود. هدایت الکتریکی تابعی از غلظت آنیون‌ها، کاتیون‌ها و دمای محلول بوده و با مقدار کل جامدات محلول^۴ (TDS) نیز وابستگی دارد و یکی از سریع‌ترین روش‌های ارزیابی کلی کیفیت آب‌های زیرزمینی می‌باشد. قابلیت

^۳Electric conductivity

^۴Total dissolved solution



شکل ۱۲.۴: پیش‌بینی مصرف گاز کشور انگلستان برای ۱۲ دوره‌ی بعدی که با استفاده از روش INLA (نمودار بالا) و بسته dlm (نمودار پایین) به‌دست آمده.

هدایت الکتریکی یا عکس آن، یعنی مقاومت الکتریکی، یکی از عوامل مهم در تحقیقات هیدروژئولوژی می‌باشد و در اندازه‌گیری آن احتیاجی به دستگاه‌های پیچیده نبوده و مشکلات نمونه‌برداری وجود ندارد. همچنین اندازه‌گیری آن سریع و دقیق است. مقدار استاندارد تعیین شده برای EC توسط سازمان بهداشت جهانی 500 میکروموس است. با توجه به این ویژگی‌های EC، بر آن شدیم تا کیفیت آب‌های زیرزمینی بخشی از استان گلستان را مورد مطالعه قرار دهیم.

داده‌های موجود شامل اندازه‌های ثبت شده در 150 ایستگاه از آب‌های زیرزمینی استان گلستان است که از آبان‌ماه سال ۸۲ تا آبان‌ماه سال ۹۲ گردآوری شده‌اند. منظور از ایستگاه‌های آب زیرزمینی، چشمه‌ها و چاه‌های عمیق و نیمه‌عمیق است و زمان‌های نمونه‌گیری نیز در دو ماه سال یعنی اردیبهشت و آبان انجام گرفته‌اند. بنابراین مجموعه داده‌ها شامل 150 نقطه مکانی و 21 نقطه زمانی و در مجموع 3150 مشاهده فضایی-زمانی است. متغیر پاسخ EC و متغیرهای تبیینی شامل سختی کل^۵ (th)، پتاسیم (k)، سدیم (na)، سولفات (so4)، کلر (cl)، بیکربنات (hco3)، pH و کل جامدات محلول هستند. حال با استفاده از مدل فضایی-زمانی به روش SPDE به استنباط بیزی این داده‌ها می‌پردازیم.

^۵Total hardness

معرفی داده‌های زمین‌آماري

بسیاری از پدیده‌های زیست محیطی، حتی اگر به‌طور پیوسته تحت یک منطقه و در زمان معرفی شوند، می‌توانند تنها در تعداد مکان‌های فضایی محدود شده اندازه‌گیری و مشاهده شوند. برای مثال، بخش‌های هواشناسی (دما و بارش، سرعت باد و غیره) و همچنین متغیرهای آب‌شناسی و اقیانوس‌شناسی (رطوبت خاک، ارتفاع موج و غیره) از این دسته‌اند. در زمین‌آمار داده‌ها از ایستگاه‌ها گرفته شده‌اند که فرض شده تحقیقاتی از فرآیند فضایی (میدان تصادفی) به‌طور پیوسته با تغییر در زمان (کرسی، ۱۹۹۳؛ گل‌فاند و همکاران، ۲۰۱۰؛ کرسی و ویکل، ۲۰۱۱) را مشاهده کنید) به‌صورت زیر هستند:

$$Y(s, t) = \{y(s, t) : (s, t) \in \mathcal{D} \subseteq \mathbb{R}^2 \times \mathbb{R}\}.$$

این تحقیقات برای ساختن استنباط فرآیند و پیش‌گویی آن در مکان‌ها به‌کار می‌رود. معمولاً با یک میدان گوسی (GF) سروکار داریم که به‌طور کامل به‌وسیله میانگینش و تابع کوواریانس فضایی-زمانی $\text{Cov}(y(s, t), y(s', t')) = \sigma^2 \mathcal{C}((s, t), (s', t'))$ مشخص می‌شود که برای هر (s, t) و (s', t') در $\mathbb{R}^2 \times \mathbb{R}$ بیان شده است. هرچند که فرآیند ایستای مرتبه دوم است، اگر میانگین آن ثابت باشد و تابع کوواریانس فضایی-زمانی از طریق بردار فاصله فضایی $h = (s - s') \in \mathbb{R}^2$ و تأخیر زمان $l = (t - t') \in \mathbb{R}$ به نقاط مکان و زمان بستگی داشته باشد.

داده‌های زمین‌آماري تحقیقاتی از یک فرآیند فضایی (یا میدان تصادفی) $\{y(s), s \in \mathcal{D}\}$ می‌باشند که s اندیس فضایی است و در حوزه ثابت \mathcal{D} تغییر می‌کند. این فرآیند فضایی یک میدان گوسی است، اگر برای هر $n \geq 1$ و هر مجموعه مکان‌های (s_1, \dots, s_d) ، بردار $(y(s_1), \dots, y(s_d))$ دارای توزیع نرمال چندمتغیره با میانگین $\mu = (\mu(s_1), \dots, \mu(s_d))$ و ماتریس کوواریانس فضایی Σ باشد. درایه ماتریس Σ به کمک تابع کوواریانس $\mathcal{C}(\cdot, \cdot)$ تعریف می‌شود، به‌طوری که $\Sigma_{ij} = \text{Cov}(y(s_i), y(s_j)) = \mathcal{C}(\cdot, \cdot)$. این فرآیند فضایی ایستای مرتبه‌ی دوم است، اگر $\mu(s_i) = \mu$ برای هر i و تابع کوواریانس فضایی $\text{Cov}(y(s_i), y(s_j)) = \mathcal{C}(s_i - s_j)$ باشد. علاوه بر این، اگر تابع کوواریانس به مسیر وابسته نباشد، اما فقط به فاصله اقلیدسی $\|s_i - s_j\|$ بستگی داشته باشد، آن‌گاه این فرآیند را همسان‌گرد^۶ می‌نامیم. گام اول در تعریف مدلی برای یک میدان تصادفی در چارچوب سلسله‌مراتبی آن است که توزیع احتمال مشاهدات را در d موقعیت فضایی مشخص کنیم. در سطح اول مدل سلسله‌مراتبی، معمولاً توزیعی از خانواده نمایی برای مجموعه پارامترهای θ انتخاب می‌کنیم که شامل یک GF به وسیله $\xi(s)$ نیز می‌باشد که دارای همبستگی فضایی از طریق تابع کوواریانس $\mathcal{C}(\cdot, \cdot)$ است. گام آخر مدل سلسله‌مراتبی تعیین توزیع‌های پیشین $p(\psi)$ برای ابرپارامترها است.

نقطه ضعف این روش مدل‌بندی که شامل تابع کوواریانس فضایی می‌باشد، به‌عنوان ”مشکل n بزرگ” شناخته می‌شود (بنرجی و همکاران، ۲۰۰۴؛ جونا لازینیو و همکاران، ۲۰۱۲) و هزینه محاسباتی برای عملیات جبری ماتریس‌های کوواریانس چگال (مانند Σ) مشکل به‌وجود می‌آورد. مثلاً عملیات ماتریس چگال، توان دوم اندازه ماتریس می‌باشد که تعداد مکان‌های مشاهده‌شده در این فرآیند است. برای غلبه بر مشکل n بزرگ اخیراً برخی راه‌حل‌ها را پیشنهاد دادند، مانند کاهش تدریجی کوواریانس، مدل‌های فرآیند پیش‌گو و کریگینگ رتبه پایین (بنرجی و همکاران، ۲۰۰۸؛ کرسی و یوهانسون، ۲۰۰۸)

^۶Isotropic

را مشاهده کنید). این روش‌های پیشنهادی معمولاً سعی به کاهش بُعد یا ساده کردن ساختار ماتریس کوواریانس چگال GF دارند. اما ما روشی متفاوت را در نظر می‌گیریم که شامل بیان پیوستگی GF با تابع کوواریانس مترن به‌عنوان یک فرآیند گسسته تصادفی مانند میدان تصادفی مارکوف گوسی (GMRF) است. این روش محاسباتی مؤثر، روش معادله دیفرانسیل جزئی تصادفی (SPDE) است (لیندگرن و همکاران، ۲۰۱۱)، که در آن ارتباطی بین GF ها و GMRF ها به‌عنوان تابعی پایه فرمول‌بندی شده، از طریق روش معادلات دیفرانسیل جزئی تصادفی به وجود آمده است. نکته کلیدی آن است که تابع کوواریانس فضایی-زمانی و ماتریس کوواریانس چگال GF به ترتیب با یک ساختار همسایگی و ماتریس تُنک جایگزین می‌شوند که با هم یک GMRF را تشکیل می‌دهند. مزیت اصلی GMRF نسبت به GF ویژگی‌های محاسباتی خوب آن است. توجه کنید یک GMRF در \mathbb{R}^2 معمولاً هزینه محاسبات مرتبه $n^{3/2}$ دارد، به طوری که سرعت محاسبات را در مقایسه با ماتریس کوواریانس چگال GF بهبود می‌بخشد. در حقیقت GMRF ها به وسیله ماتریس‌های تُنک بیان می‌شوند که استفاده از روش‌های مؤثر محاسباتی مخصوصاً برای تجزیه سریع ماتریس را امکان‌پذیر می‌کند (رو و هلد، ۲۰۰۵). به علاوه زمانی که با استنباط بیزی برای GMRF ها سروکار داریم، استفاده از روش تقریب لاپلاس آشیانه‌ای جمع‌بسته امکان‌پذیر می‌شود.

می‌توان کاربردهای روش SPDE را برای داده‌های زمین‌آماري در سیمپسون و همکاران (۲۰۱۱)، بولین (۲۰۱۲)، سیمپسون و همکاران (۲۰۱۲)، کاملتی و همکاران (۲۰۱۳)، اینگریگستون و همکاران (۲۰۱۳) و موسنژ و همکاران (۲۰۱۳) مشاهده کرد.

معرفی مدل

فرض کنید $y(s_i, t)$ تحقق‌ی از فرآیند فضایی-زمانی $Y(\cdot, \cdot)$ باشد که متغیر پاسخ در ایستگاه s_i ، $i = 1, \dots, d$ و روز $t = 1, \dots, T$ اندازه‌گیری شده است. فرض می‌کنیم معادله سیستم به صورت زیر است:

$$y(s_i, t) = \mathbf{z}(s_i, t)\beta + \theta(s_i, t) + \epsilon(s_i, t) \quad (15.4)$$

که در آن $\mathbf{z}(s_i, t) = (z_1(s_i, t), \dots, z_p(s_i, t))$ بردار p متغیر تبیینی برای مکان s_i و زمان t و $\beta = (\beta_1, \dots, \beta_p)'$ بردار ضرایب است. به علاوه، $\epsilon(s_i, t) \sim N(0, \sigma_\epsilon^2)$ اندازه خطا است که به وسیله یک فرآیند نوفه سفید گوسی بیان شده و به طور سریالی و فضایی ناهمبسته است. توجه کنید که در ادبیات زمین‌آماري جمله $\mathbf{z}(s_i, t)\beta$ مؤلفه بزرگ مقیاس است، در حالی که واریانس اندازه خطا σ_ϵ^2 به‌عنوان اثر قطعه‌ای است (کرسی، ۱۹۹۳) را مشاهده کنید). در نهایت، $\theta(s_i, t)$ فرآیند حالت است؛ این فرآیند دارای میدان تصادفی گوسی فضایی-زمانی در زمان با اتورگرسو مرتبه اول با ضریب a به صورت زیر است:

$$\theta(s_i, t) = a\theta(s_i, t-1) + \omega(s_i, t) \quad (16.4)$$

برای $t = 1, \dots, T$ که در آن $|a| < 1$ و $\theta(s_i, 1)$ از توزیع ایستای $N(0, \frac{\sigma_\omega^2}{1-a^2})$ آمده است. علاوه بر این، $\omega(s_i, t)$ دارای توزیع گوسی با میانگین صفر است که استقلال زمانی دارد و کوواریانس فضایی-

زمانی آن به صورت زیر است:

$$Cov(\omega(s_i, t), \omega(s_j, t')) = \begin{cases} 0 & t \neq t' \\ \sigma_\epsilon^2 C(h) & t = t' \end{cases} \quad (17.4)$$

برای $i \neq j$. تابع همبستگی فضایی $C(h)$ تنها به فاصله فضایی اقلیدسی $h = \|s_i - s_j\| \in \mathbb{R}$ بستگی دارد. پس فرآیند ایستای مرتبه دوم و همسان‌گرد می‌باشد (کرسی، ۱۹۹۳). این ویژگی‌ها این مطلب را به دنبال دارد که $Var(\omega(s_i, t)) = \sigma_\omega^2$ برای هر s_i و t است. تابع همبستگی فضایی $C(h)$ به وسیله تابع مترن و به صورت زیر معرفی می‌شود:

$$C(h) = \frac{1}{\Gamma(\lambda) 2^{\lambda-1}} (\kappa h)^\lambda K_\lambda(\kappa h) \quad (18.4)$$

که در آن K_λ تابع بسل نوع دوم با مرتبه $\lambda > 0$ است. پارامتر λ که معمولاً ثابت است، درجه هموارسازی فرآیند را اندازه‌گیری کرده و مقدار آن میانگین مربع مشتق فرآیند را تعیین می‌کند. $\kappa > 0$ پارامتر مقیاس است که مرتبط با نرخ ρ است، یعنی فاصله در همبستگی فضایی تقریباً پوچ می‌شود. از تعریف $\rho = \frac{\sqrt{\lambda \kappa}}{\kappa}$ استفاده می‌کنیم که ρ متناظر با فاصله همبستگی فضایی نزدیک به ۱/۰ برای هر λ است.

مشاهدات در زمان t در بردار $\mathbf{y}_t = (y(s_1, t), \dots, y(s_d, t))'$ جمع‌آوری شده و با این بردار معادلات (۱۵.۴) و (۱۶.۴) را به صورت زیر می‌نویسیم:

$$\mathbf{y}_t = \mathbf{z}_t \boldsymbol{\beta} + \boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2 I_d), \quad (19.4)$$

$$\boldsymbol{\theta}_t = a \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \Sigma = \sigma_\omega^2 \tilde{\Sigma}) \quad (20.4)$$

که در آن I_d ماتریس مشخصه با بُعد d ، $\mathbf{z}_t = (\mathbf{z}(s_1, t)', \dots, \mathbf{z}(s_d, t)')$ و $\boldsymbol{\theta}_t = (\theta(s_1, t), \dots, \theta(s_d, t))'$ است که θ_1 از فرآیند $AR(1)$ ، با توزیع $N(\mathbf{0}, \frac{\Sigma}{1-a^2})$ آمده است. علاوه بر این، $\tilde{\Sigma}$ ماتریس کوواریانس چگال بُعد d با درایه‌های $C(\|s_i - s_j\|)$ است، به طوری که $C(\cdot)$ تابع مترن معادله (۱۸.۴) با پارامتر κ و λ است.

فرض کنید $\boldsymbol{\psi} = \{\boldsymbol{\beta}, \sigma_\epsilon^2, a, \sigma_\omega^2, \kappa\}$ بردار ابرپارامتر باشد. توزیع پسین توأم به صورت زیر است:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y}) \propto \pi(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\psi}) \pi(\boldsymbol{\theta} | \boldsymbol{\psi}) \pi(\boldsymbol{\psi}) \quad (21.4)$$

که در آن $\mathbf{y} = \{\mathbf{y}_t\}$ و $\boldsymbol{\theta} = \{\boldsymbol{\theta}_t\}$ با $t = 1, \dots, T$ هستند. معمولاً توزیع‌های پیشین مستقل برای پارامترها انتخاب می‌شوند، یعنی $\pi(\boldsymbol{\psi}) = \prod_{i=1}^{dim(\boldsymbol{\psi})} \pi(\psi_i)$. با در نظر گرفتن این که مشاهدات \mathbf{y}_t به طور شرطی و سریالی به شرط $\boldsymbol{\theta}$ مستقلند و فرآیند حالت خاصیت پویای زمانی مارکوفی را دارد، معادله (۲۱.۴) را به صورت زیر بازنویسی می‌کنیم:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y}) \propto \left(\prod_{t=1}^T \pi(\mathbf{y}_t | \boldsymbol{\psi}, \boldsymbol{\theta}_t) \right) \left(\pi(\boldsymbol{\theta}_1 | \boldsymbol{\psi}) \prod_{t=2}^T \pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \boldsymbol{\psi}) \right) \pi(\boldsymbol{\psi}). \quad (22.4)$$

از توزیع‌های گوسی معادله (۱۹.۴) و (۲۰.۴) توزیع پسین توام (۲۲.۴) به صورت زیر به دست می‌آید:

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y}) &\propto (\sigma_\epsilon^2)^{-\frac{dT}{2}} \exp\left(-\frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{z}_t \boldsymbol{\beta} - \boldsymbol{\theta}_t)' (\mathbf{y}_t - \mathbf{z}_t \boldsymbol{\beta} - \boldsymbol{\theta}_t)\right) \\ &\times \left(\frac{\sigma_\omega^2}{1-a^2}\right)^{-\frac{d}{2}} |\tilde{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1-a^2}{2\sigma_\omega^2} \boldsymbol{\theta}'_1 \tilde{\Sigma}^{-1} \boldsymbol{\theta}_1\right) \\ &\times (\sigma_\omega^2)^{-\frac{d(T-1)}{2}} |\tilde{\Sigma}|^{-\frac{T-1}{2}} \exp\left(-\frac{1}{2\sigma_\omega^2} \sum_{t=2}^T (\boldsymbol{\theta}_t - a\boldsymbol{\theta}_{t-1})' \tilde{\Sigma}^{-1} (\boldsymbol{\theta}_t - a\boldsymbol{\theta}_{t-1})\right) \\ &\times \prod_{i=1}^{\dim(\boldsymbol{\psi})} \pi(\boldsymbol{\psi}_i). \end{aligned}$$

در چارچوب بیزی روش متداول برای استنباط این مدل (یعنی برآورد پارامتر و پیش‌گویی فضایی) نمونه‌گیری MCMC است، می‌توانید برای جزئیات روش‌های استنباطی به ساهو و بیکر (۲۰۱۱)، مراجعه کنید.

بازسازی مدل

معادلات (۱۵.۴) و (۱۶.۴) را به صورت زیر می‌نویسیم:

$$\mathbf{y}_t = \mathbf{z}_t \boldsymbol{\beta} + \boldsymbol{\theta}_t + \epsilon_t, \quad \epsilon_t \sim N(\mathbf{0}, \sigma_\epsilon^2 I_d), \quad (23.4)$$

$$\boldsymbol{\theta}_t = a\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \Sigma = \sigma_\omega^2 \tilde{\Sigma}). \quad (24.4)$$

برای هر زمان $t = 1, \dots, T$ میدان مترن ω در معادله (۲۴.۴) از طریق میدان تصادفی مارکوف گوسی $\tilde{\omega} \sim N(\mathbf{0}, \mathbf{Q}_S^{-1})$ ارائه شده است که در آن ماتریس دقت \mathbf{Q}_S از SPDE مطرح شده در بخش ۵.۳ آمده است. ماتریس \mathbf{Q}_S طی زمان تغییر نمی‌کند، چون فرض استقلال سریالی به وسیله معادله (۱۷.۴) بیان شده و بُعد آن تعداد رئوس حوزه مثلث‌بندی شده، است. بنابراین معادله (۲۴.۴) را می‌توان به صورت زیر بازنویسی کرد:

$$\boldsymbol{\theta}_t = a\boldsymbol{\theta}_{t-1} + \tilde{\omega}_t, \quad \tilde{\omega}_t \sim N(\mathbf{0}, \mathbf{Q}_S^{-1}) \quad (25.4)$$

برای $t = 1, \dots, T$ و با $\boldsymbol{\theta}_1 \sim N(\mathbf{0}, \mathbf{Q}_S^{-1}/(1-a^2))$ این مطلب بیان می‌کند که توزیع توأم Tn -بعدی GMRF، $\boldsymbol{\theta} = (\theta'_1, \dots, \theta'_T)'$ برابر است با:

$$\boldsymbol{\theta} \sim N(\mathbf{0}, \mathbf{Q}^{-1}) \quad (26.4)$$

با $\mathbf{Q} = \mathbf{Q}_T \otimes \mathbf{Q}_S$ که در آن:

$$\mathbf{Q}_T = \begin{pmatrix} \sigma_\omega^2 & -a/\sigma_\omega^2 & & & \\ -a/\sigma_\omega^2 & (1+a^2)/\sigma_\omega^2 & & & \\ & & \dots & & \\ & & & (1+a^2)/\sigma_\omega^2 & -a/\sigma_\omega^2 \\ & & & -a/\sigma_\omega^2 & (1+a^2)/\sigma_\omega^2 \end{pmatrix},$$

ماتریس دقت T -بعدی فرآیند اتورگرسیو زمانی مرتبه اول معادله (۲۵.۴) می‌باشد. به‌علاوه، معادله (۲۳.۴) را به صورت زیر بازنویسی می‌کنیم:

$$\mathbf{y}_t = \mathbf{z}_t \boldsymbol{\beta} + \mathbf{B} \boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_d) \quad (27.4)$$

که در آن ماتریس $(d \times n)$ -بعدی \mathbf{B} ، مقدار $\boldsymbol{\theta}_t$ را برای هر مشاهده بردار \mathbf{y}_t انتخاب می‌کند، به‌طوری که \mathbf{B} یک ماتریس تنک با تنها یک درایه واحد برای هر سطر است و داریم:

$$y(s_i, t) = z(s_i, t) \boldsymbol{\beta} + \sum_{j=1}^n B_{ij} \boldsymbol{\theta}_t + \epsilon(s_i, t),$$

که در آن $B_{ij} = 1$ اگر راس مثلث j در مکان s_i باشد و در غیر این صورت صفر است.

برآورد پارامتر و پیش‌گویی بیزی

مدل سلسله‌مراتبی معرفی شده در معادلات (۲۵.۴) و (۲۷.۴) زیرمجموعه‌ای از مدل‌های گوسی پنهان است، بنابراین، می‌توان این مدل را به کمک الگوریتم INLA برآورد کرد.

فرض کنید $\boldsymbol{\xi} = \{\boldsymbol{\theta}, \boldsymbol{\beta}\}$ میدان پنهان با پیشینی مستقل باشد. پیشین گوسی مبهم با دقت معلوم برای $\boldsymbol{\beta}$ و توزیع GMRF معادله (۲۶.۴) برای $\boldsymbol{\theta}$ تعیین می‌کنیم. پس چگالی $\pi(\boldsymbol{\xi} | \boldsymbol{\psi})$ گوسی با میانگین صفر و ماتریس دقت $\mathbf{Q}(\boldsymbol{\psi}_1)$ با بردار ابرپارامترهای $(\sigma_\omega^2, \alpha, \kappa)$ $\boldsymbol{\psi}_1$ است. همچنین، مشاهدات توزیع نرمال دارند و به شرط $\boldsymbol{\xi}$ و $\sigma_\epsilon^2 = \boldsymbol{\psi}_2$ مستقلند. بنابراین با توجه به بردار ابرپارامتر $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$ توزیع پسین توام به شکل زیر است:

$$\pi(\boldsymbol{\psi}, \boldsymbol{\xi} | \mathbf{y}) = \pi(\boldsymbol{\psi}) \pi(\boldsymbol{\xi} | \boldsymbol{\psi}) \prod_{t=1}^T \pi(\mathbf{y}_t | \boldsymbol{\xi}, \boldsymbol{\psi}),$$

که در آن $\pi(\mathbf{y}_t | \boldsymbol{\xi}, \boldsymbol{\psi}) \sim N(\mathbf{z}_t \boldsymbol{\beta} + \mathbf{B} \boldsymbol{\theta}_t, \sigma_\epsilon^2 \mathbf{I}_d)$.

علاقه‌مندیم چگالی‌های حاشیه‌ای پسین میدان پنهان و ابرپارامترها را به‌دست آوریم:

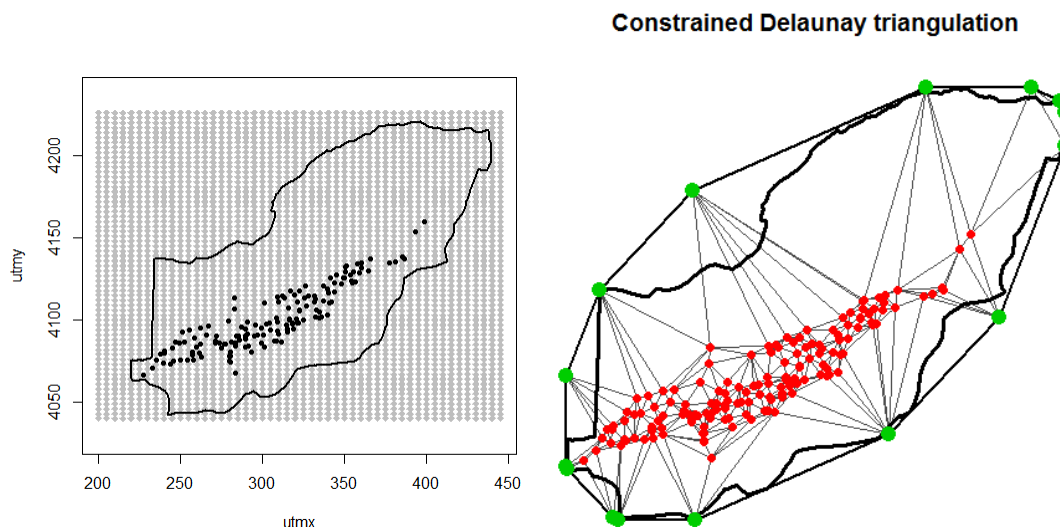
$$\pi(\xi_i | \mathbf{y}) = \int \pi(\xi_i | \mathbf{y}, \boldsymbol{\psi}) \pi(\boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi}, \quad i = 1, \dots, T + p$$

$$\pi(\psi_j | \mathbf{y}) = \int \pi(\boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi}_{-j}, \quad j = 1, \dots, 4.$$

توجه داشته باشید که مدلی با پاسخ‌های گوسی داریم که $\tilde{\pi}(\xi_i | \mathbf{y})$ گوسی و دقیق است و تقریب $\tilde{\pi}(\psi_j | \mathbf{y})$ تنها مستلزم انتگرال‌گیری عددی است.

برای پیش‌گویی فضایی، به این نکته توجه داریم که الگوریتم INLA توزیع شرطی پسین $\boldsymbol{\theta}$ را برای تمام n رأس مثلث به وجود می‌آورد. به شرط $\boldsymbol{\theta}$ پیش‌گویی \mathbf{y}_t را برای حوزه مثلث‌بندی شده در نقشه به‌دست می‌آوریم. مزیت این روش در زمان محاسبه با توجه به روش‌های MCMC است که ابتدا برای دستیابی به توزیع شرطی کامل پارامترها مورد نیاز است و سپس از چگالی توزیع پسین $y(s_0, t)$ برای هر $s_0 \in D$ شبیه‌سازی می‌کند.

حال با توجه به آن‌چه در فصل قبل، مثال شبیه‌سازی شده به روش SPDE، شرح دادیم، داده‌های گلستان را استنباط می‌کنیم.



شکل ۱۳.۴: نقشه استان گلستان (نمودار سمت چپ) و مثلث‌بندی این استان (نمودار سمت راست) با ۱۶۶ رأس. نقاط قرمز ایستگاه‌ها، نقاط سبز ۱۶ گوشه ناحیه و خط سیاه مرزبندی ناحیه هستند.

چون محدوده $p = 10$ متغیر تبیینی کاملاً متفاوتند، آن‌ها را استاندارد می‌کنیم، به‌علاوه، توزیع داده‌های EC را به‌طور تقریبی نرمال می‌کنیم، تبدیل لگاریتم را نیز برای آن به‌کار می‌بریم و متغیر جدیدی با نام \log_{ec} به داده‌ها اضافه می‌کنیم. حال مثلث‌بندی استان گلستان را مانند فصل قبل، می‌سازیم. توجه کنید که این مثلث‌بندی، ابتدا این منطقه را در رئوس اولیه مثلث‌های واقع در $d = 150$ ایستگاه تقسیم و سپس رئوس دیگر را اضافه می‌کند. پس شبکه مثلث‌بندی نهایی استان گلستان را با استفاده از مکان‌های ایستگاه و مرز ساخته شده، می‌سازیم. شکل ۱۳.۴ مثلث‌بندی تشکیل شده با ۱۶۶ رأس را برای منطقه استان گلستان نشان می‌دهد.

مدل SPDE را به‌وسیله تابع $\text{inla.spde2.matern}(\cdot)$ با مثلث‌بندی به‌دست آمده و مدل مترن را برای تابع کوواریانس فضایی، ایجاد می‌کنیم. یک آرگومان تابع $\text{inla.spde2.matern}(\cdot)$ پارامتر α است که به‌طور پیش‌فرض برابر با ۲ انتخاب می‌شود. همان‌طور که می‌دانیم، $\alpha = \lambda + 1$ و بنابراین پارامتر هموارساز λ از تابع کوواریانس مترن برابر با ۱ است. سپس از آن ویژگی $R - INLA$ که می‌توان معادله مشاهده را به صورت ماتریسی $y = A\eta + \epsilon$ بازنویسی کرد (فصل ۳)، استفاده می‌کنیم، که در آن y مشاهدات، η پیش‌گوی خطی، ϵ نوفه مشاهدات، و A یک ماتریس مشاهده است. حال می‌توانیم از تابع $\text{inla.stack}(\cdot)$ که در فصل قبل شرح داده شد، استفاده می‌کنیم. در هر تابع inla.stack ، effects لیستی از گروه‌های مؤلفه پیش‌گوی خطی است، به‌طوری که هر گروه ماتریس مشاهده خودش را دارد. در نتیجه مدل پیش‌گوی خطی، مجموعی از گروه‌های مؤلفه مشاهده شده به‌عنوان مقدار نهایی آن و پیش‌گوی توأم برای مؤلفه‌های فضایی روی روز پیش‌گو شده، است.

قبل از اجرای تابع $\text{inla}(\cdot)$ برای استنباط مدل، باید formula را که شامل $p = 10$ متغیر تبیینی (اثر ثابت) به همراه اثرات تصادفی است، تعریف کنیم. در تابع $f(\cdot)$ ، به کمک گزینه‌های group و control.group در هر زمان مشخص می‌کنیم که مکان‌های فضایی به‌وسیله مدل spde مرتبط

می‌شوند، در حالی‌که زمان فرآیند $AR(1)$ را شامل می‌شود.

```
formula <- logec ~ -1 + utmx + utmy + th + k +
  na + so4 + cl + hco3 + ph + tds + f(spatial.field, model=Golestan.spde,
  group=spatial.field.group, control.group=list(model="ar1"))
```

چون روش `inla.stack` اجرا شده است، به طور خودکار نمی‌توان عرض از مبداء را به کار برد و باید یک متغیر تبیینی `intercept` را مشخص می‌کنیم، در این جا مدل را بدون عرض از مبداء در نظر می‌گیریم. در نهایت مدل مشخص شده با پاسخ گوسی را می‌توان با استفاده از تابع `inla(.)` به صورت زیر اجرا کرد:

```
result <- inla(formula,
  data=inla.stack.data(join.stack),
  control.predictor=list(A=inla.stack.A(join.stack),
  compute=TRUE))
```

آماره‌های پسین (میانگین، چندک‌ها و انحراف معیار) اثرات ثابت یعنی ضرایب متغیر تبیینی β ، گزارش می‌شود که در جدول ۱.۴ آمده است.

آماره‌های توزیع پسین ضریب a در مدل اتورگرسو مرتبه اول از ماتریس `result$summary.hyperpar` با اجرای دستور زیر به دست می‌آید:

```
result$summary.hyperpar["GroupRho for field",]
```

و در جدول ۲.۴ گزارش شده است. توجه داشته باشید که تابع `inla(.)` میانگین، چندک‌ها و انحراف معیار پارامتر دقت $1/\sigma^2$ را گزارش می‌کند. لذا، چون علاقه‌مند به محاسبه واریانس σ^2 هستیم باید چگالی حاشیه‌ای دقت را به وسیله تابع `inla.tmarginal(.)` تبدیل کنیم:

```
sigma2eps.marg <- inla.tmarginal(function(x) 1/x,
  result$marginals.hyperpar$"Precision for the Gaussian observations")
```

سپس با استفاده از توابع `inla.emarginal(.)` و `inla.qmarginal(.)` می‌توانیم به راحتی میانگین، انحراف معیار و چندک‌های σ^2 محاسبه کنیم که در جدول ۲.۴ آمده است:

```
sigma2eps.m1 <- inla.emarginal(function(x) x, sigma2eps.marg)
sigma2eps.m2 <- inla.emarginal(function(x) x^2, sigma2eps.marg)
sigma2eps.stdev <- sqrt(sigma2eps.m2 - sigma2eps.m1^2)
sigma2eps.quantiles <- inla.qmarginal(c(0.025, 0.5, 0.975), sigma2eps.marg)
```

برآوردهای پارامتر برای مدل‌های فضایی SPDE را می‌توان با اجرای کد زیر به دست آورد:

جدول ۱۰.۴: برآوردهای پسین (میانگین، انحراف معیار و چندک‌ها) بردار ضریب β متغیرهای تبیینی.

متغیر تبیینی	مرتبه چندک				
	میانگین	انحراف معیار	۰/۰۲۵	۰/۵	۰/۷۵
UTMX	-۰/۶۶۱	۱/۶۲۷	-۳/۸۴۸	-۰/۶۶۹	۲/۵۶۹
UTMY	۱/۵۹۰	۱/۳۵۵	-۱/۱۰۸	۱/۵۹۹	۴/۲۳۸
TH	۰/۰۲۴	۰/۰۱۱	۰/۰۰۳	۰/۰۲۴	۰/۰۴۵
K	-۰/۰۰۳	۰/۰۰۵	-۰/۰۱۲	-۰/۰۰۳	۰/۰۰۵
NA	۰/۰۵۸	۰/۰۰۹	۰/۰۴۰	۰/۰۵۸	۰/۰۷۵
SO4	-۰/۰۰۴	۰/۰۰۹	-۰/۰۲۳	-۰/۰۰۴	۰/۰۱۴
CL	-۰/۰۲۰	۰/۰۰۸	-۰/۰۳۵	-۰/۰۲۰	-۰/۰۰۴
HCO3	۰/۰۵۷	۰/۰۰۴	۰/۰۴۹	۰/۰۵۷	۰/۰۶۴
PH	-۰/۰۱۲	۰/۰۰۳	-۰/۰۱۹	-۰/۰۱۲	-۰/۰۰۵
TDS	۰/۴۱۰	۰/۰۱۱	۰/۳۸۸	۰/۴۱۰	۰/۴۳۲

```
result.field = inla.spde.result(result, "spatial.field",
                                golestan_spde, do.transform=TRUE)
```

که تمام اطلاعات مرتبط با `result` و همچنین تبدیلات نتایج مقیاس‌های پارامترها، با توزیع‌های پسین برای واریانس با $\theta_1 = \log(\tau)$ و $\theta_2 = \log(\kappa)$ را می‌دهد. با توجه به آن چه در بخش ۲ لیندگرن و همکاران (۲۰۱۱) آمده، واریانس σ_ω^2 به صورت زیر به دست می‌آید:

$$\sigma_\omega^2 = \frac{1}{4\pi\kappa^2\tau^2}$$

که در آن پارامتر τ پارامتر مقیاس است که برآورد آن در درایه `Theta1 for spatial.field` ماتریس `result$summary.hyperpar` ذخیره شده است. میانگین پسین σ_ω^2 از طریق دستور زیر محاسبه می‌شود:

```
inla.emarginal(function(x) x, result.field$marginals.variance.nominal[[1]])
```

و برابر با $۱۸/۰۷۲$ است. برآوردهای پسین (میانگین، چندک‌ها و انحراف معیار) برای ابرپارامترهای σ_ω^2 ، ρ و a در جدول ۲.۴ آمده‌اند.

برای پیش‌گویی نیز یک توری $۵^\circ \times ۶۳$ با محدوده‌ی ۲۰°km تا ۴۴۵km در غرب و محدوده‌ی ۴۰۴°km تا ۴۲۲۶km در شمال در نظر می‌گیریم. هدف از دستیابی به نقشه‌ی یک روز خاص، بررسی کیفیت آب‌های زیرزمینی (در مقیاس لگاریتمی) می‌باشد. اگر بخواهیم پیش‌گویی EC بدون اثر قطعه‌ای به حساب آوریم، به راحتی مؤلفه بزرگ مقیاس $\beta z(s_0, t)$ را به مقدار میدان پنهان $\psi(s_0, t)$ اضافه می‌کنیم که $s_0 \in D$ و $۱ \leq t \leq ۲۱$.

جدول ۲.۴: برآوردهای پسین (میانگین، انحراف معیار و چندک‌ها) برای σ_ω^2 ، σ_ϵ^2 ، ρ و a .

پارامتر	مرتبۀ چندک				
	میانگین	انحراف معیار	۰/۰۲۵	۰/۵	۰/۷۵
σ_ϵ^2	۰/۰۱۱	۰/۰۰۰	۰/۰۱۰	۰/۰۱۱	۰/۰۱۲
σ_ω^2	۱۸/۰۷۲	۲/۶۳۱	۱۲/۷۰۴	۱۸/۰۷۱	۲۵/۳۷۶
ρ	۶۱/۳۸۸	۴/۸۶۶	۵۰/۵۷۹	۶۱/۳۹۰	۷۳/۷۹۳
a	۰/۹۹۹	۰/۰۰۰	۰/۹۹۹	۰/۹۹۹	۰/۹۹۹

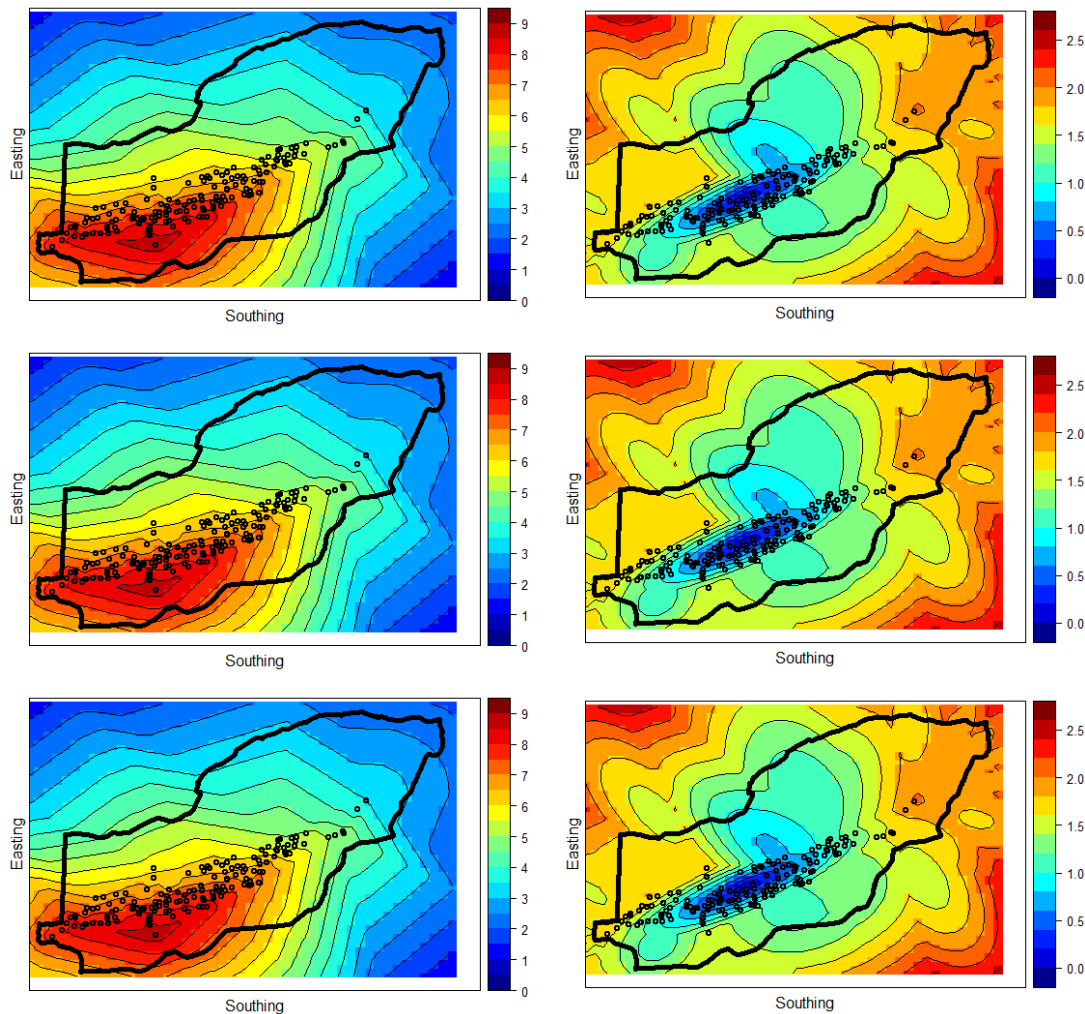
نقشه‌های پیش‌گویی میانگین و انحراف معیار پاسخ که در شکل ۱۴.۴ برای روزهای اول، دهم و بیستم گزارش شده است.

مقایسه با نتایج تحلیل حاصل از مدل‌های رتبه کاهشی

حمیدیان (۱۳۹۴) با استفاده از مدل‌های رتبه کاهشی نشان داد که متغیر سختی کل، پتاسیم و سدیم در اکثر زمان‌ها تأثیر مثبتی بر متغیر پاسخ، هدایت الکتریکی، دارند، اما متغیر سولفات، در زمان‌های مختلف، یک موج کسینوسی دارد و دارای هر دو اثر مثبت و منفی نسبت به متغیر پاسخ است. همچنین متغیر pH، در اکثر زمان‌ها، تأثیری معناداری ندارد و حتی در مواردی تأثیر عکس بر رفتار متغیر پاسخ را داراست. متغیر کل جامدات محلول نیز بیشترین تأثیر مثبت بر هدایت الکتریکی را دارد. در تحلیل این داده‌ها به روش INLA مشاهده می‌کنیم (جدول ۱۰.۴) که متغیر کل جامدات محلول بعد از UTMV بیشترین تأثیر مثبت را بر متغیر پاسخ دارد. همچنین متغیر pH بر هدایت الکتریکی تأثیرگذار است. نقشه‌های پهنه‌بندی پاسخ‌های برآوردشده، EC، برای مدل فرآیند پیش‌گو با ۵° گره، به عنوان نمونه، در شکل ۱۵.۴ به همراه انحراف معیار متناظر، در ۳ نقطه زمانی مختلف، نمایش داده شده‌اند. برای مدل‌های با سایر تعداد گره، این نتایج مشابه هستند. به همین دلیل از گزارش آن‌ها صرف‌نظر کردیم. برآورد رویه پاسخ در این شکل، میانه توزیع پسین است. در نمودارهای سمت چپ مشاهده می‌شود که نقاط قرمز از کیفیت آب پایین‌تری برخوردار هستند. همچنین کیفیت آب در جنوب و شرق این منطقه بهتر است. اما در تحلیل به روش INLA (شکل ۱۴.۴) جنوب غربی این منطقه کیفیت آب کمتری دارد و هرچه به سمت شمال می‌رویم کیفیت آب بهتر می‌شود.

۷.۴ نتیجه‌گیری

در این پایان‌نامه یک چارچوب محاسباتی برای اجرای مستقیم استنباط بیزی تقریبی در مدل‌های خطی و خطی تعمیم‌یافته پویا بر مبنای روش INLA معرفی کردیم. این روش را از طریق مثال‌های شبیه‌سازی شده و واقعی از مدل‌های ساده یک متغیره تا مدل‌های پیچیده پویای فضایی-زمانی تشریح کردیم. مشاهده کردیم که به راحتی می‌توان مدل‌های پویای پیچیده را به کمک این روش استنباط کرد. این چارچوب پیشنهادی در مدل‌های پویا جنبه‌های مهم دیگری نیز دارد:



شکل ۱۴.۴: پیش‌گویی میانگین پسین (نمودار سمت چپ) و انحراف استاندارد پسین (نمودار سمت راست) برای لگاریتم متغیر پاسخ (EC) در زمان‌های اول، دهم و بیستم.

- پارامترهای نامعلوم دقت و فواصل اعتبار آن‌ها همراه با پارامترهای حالت، حتی برای مدل‌های با مشاهدات غیرگوسی، مستقیماً به کمک INLA برآورد می‌شوند، برخلاف روش‌های دیگر که پارامترهای نامعلوم واریانس برآورد نمی‌شوند یا فقط برای مورد گوسی برآورد صورت می‌پذیرد. برای مثال، بسته `sspir` باید ترکیبی از الگوریتم‌های ماکسیم سازی عددی را با خروجی هموارساز کالمن بسط داده شده اجرا کند. بسته `d1m` برآورد بیزی ابرپارامترهای نامعلوم را به دست می‌دهد، اما این بسته به مدل‌های خطی پویا گوسی محدود شده است.
- روش INLA قادر است مشاهدات فضایی-زمانی پیچیده (مثال ۴ و ۵) را در راهبردی آسان برازش دهد. به بیانی دیگر ابزار محاسباتی دیگری در دسترس نیستند که با این نوع از داده‌ها در چارچوب مدل‌های پویا سروکار داشته باشند.
- روش INLA در مقایسه با روش MCMC (مثال ۶) استنباط را در کمترین زمان اجرا می‌کند.

شکل ۱۵.۴: نقشه‌های پهنه‌بندی برآوردشده از نمونه‌های زنجیر مارکوف MCMC در قالب میانه (ستون سمت راست) و انحراف معیار (ستون سمت چپ) در ۳ زمان مختلف: یکم، دهم و بیست و یکم.

انتظار داریم که این برتری در مورد مدل‌های پویای فضایی و فضایی-زمانی نیز صدق کند؛ در این موارد که پیچیده‌ترین آن‌ها در این پایان‌نامه در نظر گرفته شده است، روش INLA تنها دقایقی را برای اجرا استنباط صرف می‌کند و اکثر مثال‌ها در تنها چندین ثانیه برازش شده‌اند.

تقریب مستقیم چگالی‌های پسین برای حالت‌ها و ابرپارامترها به وسیله روش INLA، براساس چارچوب پیشنهادی در این پایان‌نامه، حتی برای مدل‌های پیچیده فضایی-زمانی در روشی سریع و راحت به دست می‌آیند. توجه داشته باشید که زمانی که مشاهدات دارای توزیع گوسی هستند (مانند اکثر مثال‌ها)، این روش خطای انتگرال‌گیری را کاهش می‌دهد. این نکته با روش‌های دنباله‌ای (برمبنای کالمن) یا الگوریتم‌های MCMC برای استنباط مدل‌های فضای حالت قابل بیان نیست. روش INLA روش‌های استنباطی دنباله‌ای را که در آن‌ها برآورد با گام‌های صافی کردن و هموارسازی برمبنای ساختار زمانی مشاهدات اجرا می‌شود، بهبود می‌بخشد به طوری که این روش‌ها برای استنباط مدل‌های پیچیده‌تر دشوار هستند.

پیوست آ

آ.۱ بسته INLA – R

بسته INLA – R در R، برای استنباط تقریبی بیزی به وسیله روش INLA به کار می‌رود (مارتینو و رو، ۲۰۱۰). این بسته که برنامه مستقل INLA را در بسته GMRFLib جایگزین می‌کند (مارتینو و رو، ۲۰۱۰)، برای سیستم‌های لینوکس، مک و ویندوز در دسترس است و می‌توان این بسته را با نوشتن دستور زیر در R دانلود و نصب کرد

```
> source("http://www.math.ntnu.no/inla/givemeINLA.R")
```

این بسته را می‌توان به صورت زیر فراخوانی کرد

```
> library(INLA)
```

دستور

```
> inla.upgrade(testing=TRUE)
```

برای به روزرسانی این بسته به آخرین نسخه به کار می‌رود. توجه کنید که به روزرسانی‌های مکرری وجود دارند و پیشنهاد می‌کنیم با دستور `inla.upgrade(testing = TRUE)` این بسته را دانلود کنید. وب سایت `http://www.r-inla.org/` توضیحات در مورد این بسته و همچنین مثال‌های به کار رفته و گروه پرسش و پاسخ را دارد.

برای برنامه‌نویسی با INLA – R، مثالی با دو متغیر تبیینی $x = (x_1, x_2)$ و تابع $f(\cdot)$ برای متغیر تبیینی سوم z_1 را در نظر می‌گیریم؛ پیش‌گوی خطی معادله (۱.۲) از طریق دستور زیر ایجاد می‌شود:

```
> formula <- y ~1 + x1 + x2 + f(z1, model="...")
```

که در آن y ، x_1 و x_2 ستون‌های ساختار داده‌ای هستند که داده‌ها مورد نظر را در خود دارند (نام این ساختار داده data است). 1 عبارت اجباری نیست و به عنوان عرض از مبدا می‌باشد. در

تابع $f(\cdot)$ ، $\text{model} = \dots$ نوع تابع $f(\cdot)$ را مشخص می‌کند. پیش‌فرض آن $\text{model} = \text{"iid"}$ است که متغیر تصادفی مستقل توزیع گوسی را برای z_1 تعیین می‌کند. لیست مدل‌های دیگر به صورت زیر آمده‌اند

```
> names(inla.models())$latent
 [1] "linear"      "iid"          "mec"          "meb"
 [5] "rgeneric"    "rw1"          "rw2"          "crw2"
 [9] "seasonal"    "besag"        "besag2"       "bym"
[13] "bym2"        "besagproper" "besagproper2" "ar1"
[17] "ar"          "ou"           "generic"       "generic0"
[21] "generic1"    "generic2"     "generic3"     "spde"
[25] "spde2"       "spde3"        "iid1d"         "iid2d"
[29] "iid3d"       "iid4d"        "iid5d"         "2diid"
[33] "z"           "rw2d"         "rw2diid"       "slm"
[37] "matern2d"    "copy"         "clinear"       "sigm"
[41] "revsigm"     "log1exp"
```

و تعریف هر یک از آن‌ها در <http://www.r-inla.org/models/latent-models> موجود است. می‌توان در `formula`، چندین جمله $f(\cdot)$ را به طور جداگانه به کار برد، برای مثال در

```
> formula <- y ~ x1 + x2 + f1(z1) + f2(z2) + f3(z3)
```

مجموعه داده ۳ ستون با نام z_1 ، z_2 و z_3 مقادیر متغیرهای تبیینی را دارد. تنظیمات دیگری نیز وجود دارند که می‌توان در جملات $f(\cdot)$ به کار برد، مانند `replicate`، `hyper` و `constr` (صفحه `?f` را مشاهده کنید). در نهایت الگوریتم INLA به وسیله تابع `inla` به صورت زیر اجرا می‌شود:

```
> inla(formula, family = "...", data)
```

که در آن `formula` از قبل تعیین می‌شود، `data` ساختار داده‌ای است که شامل کلیه متغیرهای موجود در `formula` می‌باشد و `family` توزیع داده‌ها (درست‌نمایی) را مشخص می‌کند. برخی از توزیع‌های موجود

```
> names(inla.models())$likelihood) #output partially omitted
 [1] "poisson"          "cenpoisson"
 [3] "gpoisson"         "binomial"
 [5] "testbinomial1"   "gamma"
```

[7]	"gammacount"	"kumar"
[9]	"beta"	"betabinomial"
[11]	"cbinomial"	"nbinomial"
[13]	"simplex"	"gaussian"
[15]	"normal"	"circularnormal"
[17]	"wrappedcauchy"	"iidgamma"
[19]	"iidlogitbeta"	"loggammafrailty"
[21]	"logistic"	"skewnormal"
[23]	"sn"	"sn2"
[25]	"gev"	"laplace"

هستند و توضیحات کامل با مثال در <http://www.r-inla.org/models/likelihoods> است. تابع inla تنظیمات دیگری هم دارد؛ لیست کامل آن را در inla? مشاهده کنید.

مراجع

- [1] Abramowitz, M. and Stegun, I. (1972), *Handbook of Mathematical Functions*, Courier Dover Publications.
- [2] Banerjee, S., Carlin, B., and Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall.
- [3] Banerjee, S., Gelfand, A. E., Finely, A.O., and Sang, H. (2008), Gaussian predictive process models for large spatial data sets, *Journal of the Royal Statistical Society*, **70**(4), 825-848.
- [4] Beauchamp, K. G. (1984), *Applications of Walsh and Related Functions, With an Introduction to Sequency Theory*, Academic Press, London.
- [5] Belitz, C., Brezger, A., Kneib, T., Lang, S., and Umlauf, N. (2012), BayesX – Software for Bayesian inference in structured additive regression model, Version 2.1. Available from <http://www.stat.uni-muenchen.de/bayesx>.
- [6] Besag, J., York, J., and Mollie, A. (1991), Bayesian image restoration, with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.
- [7] Best, N., Richardson, S., and Thompson, A. (2005), A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, **14**(1), 35–59.
- [8] Birkhoff, G.D. (1942), What is the ergodic theorem?, *American Mathematical Monthly*, **49**(4), 222-226.
- [9] Bivand, R. (2010), spdep: Spatial dependence: weighting schemes, statistics and models, R package version 0.5-9.
- [10] Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013), Spatial and spatio-temporal models with R-INLA, *Spatial and Spatio-temporal Epidemiology*, **4**, 33–49.
- [11] Bolin, D. (2012), *Models and Methods for Random Fields in Spatial Statistics with Computational Efficiency from Markov Properties*, Ph.D. thesis, Lund University.

-
- [12] Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327
- [13] Box, G.E.P., Jenkins, G., and Reinsel, G. C. (1994), *Time Series Analysis*, Forecasting and Control (3rd ed.). San Francisco: Holden-Day.
- [14] Box, G. E. P. and Wilson, K.G. (1951), On the Experimental Attainment of Optimum Conditions, *Journal of the Royal Statistical Society (series B)*, **13**, 1-45.
- [15] Brooks, S., Gelman, A., Jones, G., and Meng, X. (2011), *Handbook of Markov Chain Monte Carlo*, CRC Press, Taylor & Francis Group.
- [16] Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013), Spatio-temporal modeling of particulate matter concentration through the SPDE approach, *Advances in Statistical Analysis*, **97**(2), 109–131.
- [17] Cappe, O., Moulines, E., and Ryden, T., (2005), *Inference in Hidden Markov Models*. New York: Springer-Verlag.
- [18] Casella, G. and Berger, R. (2002), *Statistical Inference*, Duxbury, Thomson Learning.
- [19] Casella, G. and George, E. (1992), Explaining the Gibbs Sampler, *The American Statistician*, **46**, 167–174.
- [20] Chib, S. and Greenberg, E. (1995), Understanding the Metropolis–Hastings algorithm, *The American Statistician*, **49**, 327–335.
- [21] Cressie, N. (1993), *Statistics for Spatial Data*. Wiley.
- [22] Cressie, N. and Johannesson, G. (2008), Fixed rank kriging for very large data sets, *Journal of the Royal Statistical Society: Series B*, **70**(1), 209–226.
- [23] Cressie, N. and Wikle, C. (2011), *Statistics for Spatio-Temporal Data*, Wiley.
- [24] Dethlefsen, C. and Lundbye-Christensen, S. (2006), Formulating State Space Models in R with Focus on Longitudinal Regression Models, *Journal of Statistical Software*, **16**, 1-15.
- [25] Diggle, P. and Ribeiro, J.P. (2007), *Model-based Geostatistics*, Springer.
- [26] Diggle, P., Moyeed, R., and Tawn, J. (1998), Model-based geostatistics, *Journal of the Royal Statistical Society (Series C)*, **47**, 299–350.
- [27] Durbin, J. and Koopman, S. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.
- [28] Elliott, N., Wakefield, J., Best, N., and Briggs, D. (2001), *Spatial Epidemiology*, Oxford University Press.

- [29] Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.
- [30] Fruhwirth-Schnatter, S. (1994), Data augmentation and dynamic linear models, *Time Series Analysis*, **15**, 183–202.
- [31] Gamerman, D. and Migon, H. (1993), Dynamic hierarchical models, *Journal of the Royal Statistical Society (Series B)*, **55**, 629–642.
- [32] Gelfand, A., Diggle, P., Fuentes, M., and Guttorp, P. (2010), *Handbook of Spatial Statistics*, Chapman & Hall.
- [33] Gelfand, A. and Smith, A. (1990), Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.
- [34] Gelman, A. and Rubin, D. (1992), Inference from Iterative Simulation using Multiple Sequences, *Statistical Science*, **7**, 457–511.
- [35] Geman, S. and Geman, D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- [36] Geyer, C.J. (1992), Practical Markov chain Monte Carlo, *Statistical Science*, **7**(4), 473–483.
- [37] Gilks, W., Richardson, S., and Spiegelhalter, D. (1996), *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC.
- [38] Geweke, J. (1992), Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments, *In Bayesian Statistics 4*, Clarendon Press, Oxford, UK.
- [39] Gneiting, T. and Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.
- [40] Handcock, M. and Wallis, J. (1994), An approach to statistical spatial-temporal modelling of meteorological fields, *Journal of the American Statistical Association*, **86**, 368–390.
- [41] Hannan, E. and Deistler, M. (1988), *The Statistical Theory of Linear Systems*, Wiley, New York.
- [42] Harrison, P. and Stevens, C. (1976), Bayesian forecasting (with discussion), *Journal of the Royal Statistical Society (Series B)*, **38**, 205–247.
- [43] Harvey, A. (1989), *Forecasting, Structural Time Series Models and the Kalman filter*, Cambridge University Press, Cambridge.

- [44] Haslett, J. and Raftery, A. (1989), Space time modelling with long memory dependence: Assessing Ireland's wind power resource. *Journal of Applied Statistics*, **38**(1), 1–50.
- [45] Hastings, W. (1970), Monte Carlo sampling methods using Markov chains and their application, *Biometrika*, **57**, 97–109.
- [46] Hesterberg, T.C. (1987), Importance Sampling in Multivariate Problems, *Proceedings of the Statistical Computing Section of the American Statistical Association*, 412–417.
- [47] Hsiao, C.K., Huang, S.Y., and Chang, C.W. (2004), Bayesian marginal inference via candidate's formula, *Statistics and Computing*, **14**(1):59–66.
- [48] Hutchinson, C. (1984), The Kalman filter applied to aerospace and electronic systems, Aerospace and Electronic Systems, *IEEE Transactions on Aerospace and Electronic Systems*, **AES-20**, 500–504.
- [49] Hyndman, J. R., Koehler B. A, Ord J. K, and Snyder D. R. (2008), *forecasting with exponential smoothing*, Springer.
- [50] Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2013), Spatial models with explanatory variables in the dependence structure, *Spatial Statistics*, **8**, 20–38.
- [51] Jacquier, E., Polson, N., and Rossi, P. (1994), Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business and Economic Statistics*, **12**, 371–417.
- [52] Jona Lasinio, G., Mastrantonio, G. and Pollice, A. (2012), Discussing the "big n problem", *Statistical Methods & Applications*, 1–16.
- [53] Jonhansson, M. and Glass, G. (2008), High-resolution spatio-temporal weather models for climate studies. *International Journal of Health Geographics*, 7(52), 1–8.
- [54] Kalman, R. (1960), A new approach to linear filtering and prediction problems, *Trans. of the AMSE - Journal of Basic Engineering (Series D)*, **82**, 35–45.
- [55] Kass, R.E. and Steffey, D. (1989), Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, **84**(407): 717–726.
- [56] Kim, C.J. and Nelson, C. (1999), *State Space Models with Regime Switching*, MIT Press, Cambridge.
- [57] Kitanidis, K.P. (1986), Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resources Research*, **22**(4), 499–507.

- [58] Knorr-Held, L. (2000), Bayesian modelling of inseparable space-time variation in disease risk, *Statistics in Medicine*, **19**(17–18), 2555–2567.
- [59] Knorr-Held, L. and Richardson, S. (2003), A hierarchical model for space-time surveillance data on meningococcal disease incidence, *Journal of the Royal Statistical Society (Series C)*, **52**(2), 169–183.
- [60] Knorr-Held, L. and Rue, H. (2002), On block updating in Markov random field models for disease mapping, *shape Scandinavian Journal of Statistics*, **29**(4), 597–614.
- [61] Krainski, E. and Lindgren, F. (2013), The R-INLA tutorial: SPDE models.
- [62] Kroese, D.P. and Chan, J.C.C. (2014), *Statistical Modeling and Computation*, Springer.
- [63] Lange, K. (2010), *Numerical Analysis for Statisticians*, Springer.
- [64] Lin, X. and Zhang, D. (1999), Inference for Generalized Additive Mixed Models By Using Smoothing Splines, *Journal of the Royal Statistical Society*, **61**, 381–400.
- [65] Lindgren, F. and Rue, H. (2015), Bayesian spatial modelling with R-INLA, *Journal of Statistical Software*, **63**(29), 1–25.
- [66] Lindgren, F., Rue, H., and Lindström, J. (2011), An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion), *Journal of Royal Statistical Society (Series B)*, **73**(4), 423–498.
- [67] Lindley, D. (1980), Approximate Bayesian methods, *Springer*, **31**, 223–245.
- [68] Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000), WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility, *Statistics and Computing*, **10**, 325–337.
- [69] Martino, S. and Rue, H. (2010), Implementing approximate Bayesian inference using integrated nested Laplace approximation: a manual for the inla program, *Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway*. Compiled on April 8, 2010.
- [70] Martins, T.G., Simpson, D., Lindgren, F., and Rue, H. (2013), Bayesian computing with INLA: New features, *Computational Statistics & Data Analysis*, **67**, 68–83.
- [71] McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, Chapman and Hall, London.

- [72] Metropolis, N. Rosenbluth, A. Teller, A. and Teller, E. (1953), Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**(6), 1087–1092.
- [73] Migon, H., Gamerman, D., Lopez, H., and Ferreira, M. (2005), Bayesian dynamic models, in D. Day and C. Rao (eds), *Handbook of Statistics*, **25**, Elsevier B.V., chapter 19, 553–588.
- [74] Moran, P. (1950), Notes on continuous stochastic phenomena, *Biometrika*, **37**(1), 17–23.
- [75] Musenge, E., Chirwa, T.F., Kahn, K., and Vounatsou, P. (2013), Bayesian analysis of zero inflated spatio-temporal HIV/TB child mortality data through the {INLA} and {SPDE} approaches: Applied to data observed between 1992 and 2010 in rural northeast South Africa, *International Journal of Applied Earth Observation and Geoinformation*, **22**(0), 86–98.
- [76] Nelder, J.A., and Wedderburn, R.W.M. (1972), Generalized linear models. *Journal of the Royal Statistical Society, A*, **135**, 370-84.
- [77] Ntzoufras, I. (2009), *Bayesian Modeling Using WinBUGS*, John Wiley and Sons.
- [78] Pavia, J., Larraz, B., and Montero, J. (2008), Election forecasts using spatio-temporal models, *Journal of the American Statistical Association*, **103**(483), 1050–1059.
- [79] Petris, G. (2010), dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models, R package version 1.1-1.
- [80] Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic Linear Models with R*, Springer.
- [81] Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), Coda: Convergence diagnosis and output analysis for MCMC, *R News*, **6**, 7–11.
- [82] Prado, R. and West, M. (2010), *Time series Modelling Computation and Inference*, Chapman & Hall.
- [83] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (2007), *Numerical Recipes. The Art of Scientific Computing*, Cambridge University Press.
- [84] Rabiner, L. and Juang, B. (1993), *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs.
- [85] Raftery, A. E. and Lewis, S.M. (1992), One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo, *Statistical Science*, **7**, 493-497.

- [86] Robert, C. and Casella, G. (2004), *Monte Carlo Statistical Methods*, Springer.
- [87] Robert, C. and Casella, G. (2011), A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data, *Statistical Science*, **26**, 102–115.
- [88] Rue, H. and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, London: Chapman and Hall/CRC Press.
- [89] Rue, H. and Martino, S. (2007), Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *J. Statist. Planning Inf.*, **137**, 3177–3192.
- [90] Rue, H., Martino, S., and Chopin, N. (2009), Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society (series B)*, **71**, 319–392.
- [91] Ryden, T. and Titterton, D. (1998), Computational Bayesian analysis of hidden Markov models, *Comput. Graph. Statist.*, **7**, 194–211.
- [92] Sahu, S.K. and Bakar, K.S. (2011), Comparison of Bayesian models for daily Ozone concentration levels, *Statistical Methodology*, **9**(1-2), 144-157.
- [93] Sanchez, S.M. and Sanchez, P.J. (2005), Very Large Fractional Factorial and Central Composite Designs, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **15**(4), 362-377.
- [94] Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility, in D. Cox, D. Hinkley and O. Barndorff-Nielsen (eds), *Time Series Models with Econometric, Finance and other Applications*, Chapman and Hall, London, 1–67. measurement time series, *Biometrika*, **84**, 653–67.
- [95] Shoesmith, G. (2013), Space time autoregressive models and forecasting national, regional and state crime rates, *International Journal of Forecasting*, **29**(1), 191–201.
- [96] Simpson, D., Illian, J., Lindgren, F., Sørbye, S., and Rue, H. (2011), Going off grid: Computationally efficient inference for log-Gaussian Cox processes.
- [97] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. (2002), Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society (series B)*, **64**, 583–639.
- [98] Stone, R. (1988), Investigations of excess environmental risks around putative sources: Statistical problems and a proposed test, *Statistics in Medicine*, **7**, 649–660.
- [99] Szpiro, A., Sampson, P., Sheppard, L., Lumley, T., Adar, S., and Kaufman, J. (2010), Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies, *Environmetrics*, **21**(6), 606–631.

- [100] Takahashi, K., Fagan, J., and Chen, M.S. (1973). Formation of a sparse bus impedance matrix and its application to short circuit study. In 8th PICA Conference proceedings, pages 63–69. IEEE Power Engineering Society. Papers presented at the 1973 Power Industry Computer Application Conference in Minneapolis, Minnesota.
- [101] Thiele, T. (1880), Om anvendelse af mindste kvadraters metode i nogle tilfælde, hvor en komplikation af visse slags uensartede tilfældige fejlkilder giver fejlene en “systematisk” karakter, *Det Kongelige Danske Videnskabernes Selskabs Skrifter – Naturvidenskabelig og Matematisk Afdeling*, 381–408. English Translation in: Thiele: Pioneer in Statistics, S. L. Lauritzen, Oxford University Press (2002).
- [102] Tierney, L. and Kadane, J.B. (1986), Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, **81**, 82–86.
- [103] Vivar, J.C. and Ferreira, M.A.R. (2009), Spatiotemporal models for gaussian areal data. *Journal of Computational and Graphical Statistics*, **18**, 658-674.
- [104] Waller, L., Carlin, B., Xia, H., and Gelfand, A. (1997), Hierarchical spatio-temporal mapping of disease rates, *Journal of the American Statistical Association*, **92**(438), 607–617.
- [105] West, M., Harrison, J., and Migon H.S., (1985), Dynamic generalised linear models and Bayesian forecasting (with discussion), *Journal of the American Statistical Association*, **80**, 73–97.
- [106] West, M. and Harrison J., (1989), *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag.
- [107] West, M. and Harrison J., (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York: Springer-Verlag.
- [108] Zellner, A. (1996), *An introduction to Bayesian inference in econometrics*, New York: John Wiley & Sons.

Aabstract

Inference in state-space models usually relies on recursive forms for filtering and smoothing of the state vectors regarding the temporal structure of the observations, an assumption that is, from our view point, unnecessary if the data set is fixed, that is, completely available before analysis. In this paper we propose a computational framework to perform approximate full Bayesian inference in linear and generalized dynamic linear models based on the Integrated Nested Laplace Approximation (INLA) approach. The proposed framework directly approximates the posterior marginals of interest disregarding the assumption of recursive updating/estimation of the states and hyperparameters in the case of fixed data sets and, therefore, enable us to do fully Bayesian analysis of complex state-space models more easily and in a short computational time. The proposed framework overcomes some limitations of current tools in the dynamic modeling literature and is vastly illustrated with a series of simulated as well as worked real-life examples, including realistically complex models with correlated error structures and models with more than one state vector, being mutually dependent on each other.

Keywords: Dynamic models, augmented model, spatio-temporal dynamic models, Integrated Nested Laplace Approximation



Shahrood University of Technology

**Faculty Of Mathematical Sciences
MSc thesis in Mathematical Statistics**

**Approximate Bayesian inference of dynamic
regression models using integrated nested
Laplace approximation**

By: Masoud Mirzajani Bajestani

Supervisor

Dr. Hossein Baghishani

June 2016