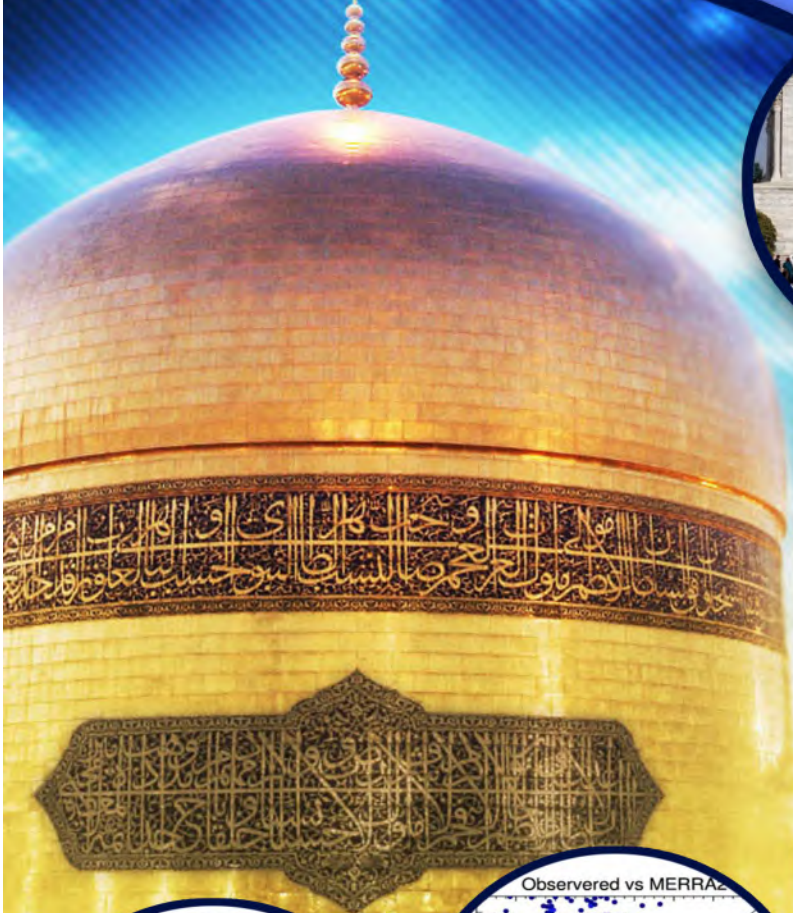
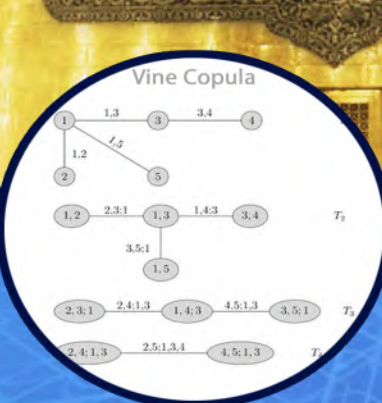
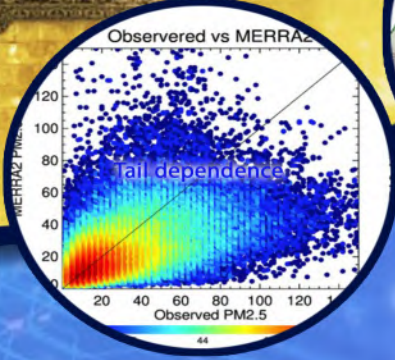


$$W(u, v) = \max(u + v - 1, 0)$$

$$\Pi(u, v) = uv$$

$$M(u, v) = \min(u, v)$$



هفتمین سمینار

نظریه مفصل و کاربردهای آن

دانشکده علوم ریاضی، دانشگاه فردوسی مشهد، ۱۹ و ۲۰ بهمن ماه ۱۴۰۱

The 7th Seminar on Copula Theory and Its Applications

Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, 8 - 9 Feb. 2023

نشانی دبیرخانه: استان خراسان رضوی، مشهد، میدان آزادی، پردیس دانشگاه فردوسی مشهد، دانشکده علوم ریاضی
صندوق پستی: ۱۱۵۹ - ۹۱۷۷۵ تلفن (گروه آمار): ۰۵۱-۳۸۸۰۶۲۰۶ وبگاه: scta7.um.ac.ir ایمیل: scta7@um.ac.ir

باسمه تعالی



مجموعه مقالات فارسی

هفتمین سمینار تخصصی

نظریه مفصل و کاربردهای آن

گروه آمار، دانشگاه فردوسی مشهد

با همکاری

قطب علمی داده های ترتیبی، قابلیت اعتماد و وابستگی

۱۹ و ۲۰ بهمن ماه ۱۴۰۱

این کتاب شامل مجموعه مقالات فارسی هفتمین سمینار تخصصی نظریه مفصل و کاربردهای آن است. بدیهی است که نویسندگان محترم مسئول صحت مطالب، اصالت و ویرایش علمی و ادبی مقالات خود بوده و کمیته های علمی و اجرایی سمینار مسئولیتی در این موارد ندارند.

عنوان: مجموعه مقالات فارسی هفتمین سمینار تخصصی نظریه مفصل و کاربردهای آن
تنظیم: دکتر هادی جباری نوقابی و خانم ریحانه احمدی
ویراستار: دکتر محمد امینی و دکتر هادی جباری نوقابی
طراح جلد: دکتر حسینعلی محتشمی برزادران
تاریخ انتشار: اسفندماه ۱۴۰۲

مقدمه

در ادامه برگزاری سلسله کارگاه‌های آموزشی، پژوهشی دو سالانه « نظریه مفصل »، در دانشگاه فردوسی مشهد (۱۳۸۸، ۱۳۹۱)، دانشگاه کرمان (۱۳۹۳)، دانشگاه یزد (۱۳۹۵)، با تأکید بر کاربردهای نظریه مفصل در علوم مهندسی، کشاورزی، بیمه و مالی، قابلیت اعتماد و تحلیل بقا، پزشکی، اقتصاد و...، که ذیل برنامه‌های قطب علمی داده‌های ترتیبی، قابلیت اعتماد و وابستگی می‌باشد، به پیشنهاد کمیته علمی کارگاه و نظر مساعد اعضای قطب مقرر گردید تا این کارگاه به صورت سمینارهای دوسالانه برگزار گردد. این سمینار دو روزه با حمایت قطب علمی داده‌های ترتیبی، قابلیت اعتماد و وابستگی، مرکز اطلاع‌رسانی و پایگاه استنادی جهان اسلام، همکاری انجمن آمار ایران، و انجمن بیمه و مالی با هدف فراهم نمودن امکانات مناسب برای تبادل اطلاعات پژوهش‌های اعضای هیأت علمی، دانشجویان تحصیلات تکمیلی و کارشناسان و کاربران در محورهای سمینار، روزهای چهارشنبه و پنجشنبه ۱۹ و ۲۰ بهمن ماه ۱۴۰۱ به صورت مجازی در دانشگاه فردوسی مشهد برگزار می‌شود. امیدواریم که تلاش کمیته‌های مختلف برگزاری سمینار بتواند رضایت خاطر میهمانان و شرکت‌کنندگان گرامی را فراهم نماید.

لازم به ذکر است که پس از فراخوان سمینار، مقالات دریافت شده توسط اعضای کمیته علمی و کمیته داوران سمینار مورد ارزیابی و در نهایت ۳۰ مقاله به صورت ارائه شفاهی مورد پذیرش قرار گرفت. لازم به ذکر است که شرکت‌کنندگان این سمینار ۴۰ نفر از اعضای هیئت علمی و دانشجویان تحصیلات تکمیلی دانشگاه‌های معتبر کشور هستند. در پایان لازم می‌دانیم از کمیته‌های علمی، اجرایی و کمیته داوران سمینار نهایت تشکر و سپاسگزاری را داشته باشیم. همچنین مراتب قدردانی خود را از حوزه معاونت پژوهشی دانشگاه فردوسی مشهد و دانشکده علوم ریاضی، ریاست محترم دانشکده علوم ریاضی، قطب علمی داده‌های ترتیبی، قابلیت اعتماد و وابستگی، پایگاه استنادی علوم جهان اسلام، انجمن آمار ایران، انجمن بیمه و مالی و همکاران بخش آمار دانشگاه فردوسی مشهد اعلام نماییم.

از خداوند منان آرزوی توفیق تمامی شرکت‌کنندگان محترم در این سمینار را داریم.

محمد امینی (دبیر همایش)

بهمن ماه ۱۴۰۱

محورهای سمینار

- روش‌های ساخت تابع مفصل
- توابع مفصل و مفاهیم وابستگی
- مدل‌سازی با استفاده از تابع مفصل
- استنباط آماری بر اساس مفصل
- کاربردهای تابع مفصل در تحلیل داده‌های فضایی، تحلیل داده‌های بقا، قابلیت اعتماد، مهندسی، هواشناسی، هیدرولوژی، پزشکی، کشاورزی، بیمه، اقتصاد، و سایر زمینه‌ها

اعضای مشاورین علمی افتخاری (به ترتیب حروف الفبا):

- دکتر محمد امینی، دانشگاه فردوسی مشهد
- دکتر ابوالقاسم بزرگ نیا، دانشگاه فردوسی مشهد
- دکتر غلامعلی پرهام، دانشگاه شهید چمران اهواز
- دکتر صدیقه شمس، دانشگاه الزهرا (س)
- دکتر قوشه عابدهدتنی، دانشگاه فردوسی مشهد
- دکتر محمد حسین علامت ساز، دانشگاه اصفهان
- دکتر غلامرضا محتشمی برزادران، دانشگاه فردوسی مشهد
- دکتر محسن محمدزاده درودی، دانشگاه تربیت مدرس

اعضای کمیته علمی (به ترتیب حروف الفبا):

- دکتر کامبیز احمدی انگالی، دانشگاه علوم پزشکی جندی شاپور
- دکتر محمد امینی، دانشگاه فردوسی مشهد (دبیر همایش)

- دکتر امیر تیمور پاینده، دانشگاه شهید بهشتی
- دکتر هادی جباری نوقابی، دانشگاه فردوسی مشهد
- دکتر علی دولتی، دانشگاه فردوسی مشهد
- دکتر محمد ذونعمت کرمانی، دانشگاه شهید باهنر کرمان
- دکتر مصطفی رزمخواه، دانشگاه فردوسی مشهد
- دکتر یوسف رضانی، دانشگاه بیرجند
- دکتر ایوب شیخی، دانشگاه شهید باهنر کرمان
- دکتر منصوره کوهی، پژوهشکده اقلیم شناسی مشهد
- دکتر سید محسن میرحسینی، دانشگاه یزد
- دکتر حمیدرضا نیلی ثانی، دانشگاه بیرجند

اعضای کمیته برگزار کننده (به ترتیب حروف الفبا):

- دکتر جعفر احمدی، دانشگاه فردوسی مشهد
- دکتر مجتبی اصفهانی، دانشگاه ولایت ایرانشهر
- دکتر محمد امینی، دانشگاه فردوسی مشهد (دبیر همایش)
- دکتر هادی جباری نوقابی، دانشگاه فردوسی مشهد (دبیر کمیته اجرایی)
- دکتر علی دولتی، دانشگاه فردوسی مشهد
- دکتر منصور زرگر، دانشگاه ولی عصر (عج) رفسنجان
- دکتر حسینعلی محتشمی برزادران، دانشگاه فردوسی مشهد
- دکتر مرتضی محمدی، دانشگاه زابل

کادر اجرایی سمینار:

رضا احمدی (امور رایانه)، مہری باقدان (کارشناس دفتر گروه)، سیدحسن احمدی (مسئول امور عمومی دانشکده)، ریحانہ احمدی (دانشجو)، ریحانہ ذاکری (دانشجو)، محدثہ رحمانی نیا (دانشجو)، فاطمہ رنجبر (خدمات)

حامیان سمینار

برگزاری ہفتمین سمینار تخصصی نظریہ مفصل و کاربردهای آن، علاوہ بر تلاش‌های خستگی‌ناپذیر اعضای کمیٹہ‌های علمی و اجرایی، مرہون پشتیبانی و حمایت‌های بی‌دریغ مسئولان دانشگاه فردوسی مشهد و تعدادی از سازمان‌های دولتی و خصوصی است. کمیٹہ برگزاری سمینار از سازمان‌های فہرست شدہ در این صفحہ کہ حمایت خود را از برگزاری سمینار اعلام کردہ‌اند، قدردانی می‌کند.

جدول زمان بندی ارائه سخنرانی ها

صبح روز چهارشنبه: ۱۴۰۱/۱۱/۱۹

| | |
|----------|---------------|
| افتتاحیه | ۰۸:۳۰ - ۰۸:۱۵ |
|----------|---------------|

| ردیف | عنوان سخنرانی | ارائه دهنده | زمان | مسئول جلسه |
|------|--|--------------------------------|---------------|--------------------------|
| ۱ | نگاه تاریخی به وابستگی منفی | ابوالقاسم بزرگ نیا، محمد امینی | ۰۸:۳۰ - ۰۹:۱۵ | غلامرضا محتشمی بزراداران |
| ۲ | Analysis of spatial survival data using copula functions | محسن محمدزاده | ۰۹:۱۵ - ۱۰:۰۰ | |
| | استراحت | | | |
| ۳ | کاربردهای متنوع نظریه مفصل در انواع مسائل مهندسی و علوم | قوشه عابد هدتنی | ۱۰:۱۰ - ۱۱:۰۰ | |
| ۴ | نگاهی به سیر تکاملی نظریه مفصل و کاربردهای آن | علی دولتی | ۱۱:۰۰ - ۱۱:۴۵ | |
| | نماز - ناهار - استراحت | | | ۱۱:۴۵ - ۱۴:۰۰ |

عصر روز چهارشنبه: ۱۴۰۱/۱۱/۱۹

| ردیف | عنوان سخنرانی | ارائه دهنده | زمان | مسئول |
|------|--|------------------------|----------------|---------------|
| ۵ | مدل های وابسته با ساختار موازی - سری و سری - موازی با تعداد زیر سیستم های تصادفی | وحیده محتشمی بزراداران | ۱۴:۰۰ - ۱۴:۲۰ | ایوب شیخی |
| ۶ | Stochastic comparisons of extreme order statistics from the generalized Gompertz distribution | اسماعیل بشکار | ۱۴:۲۰ - ۱۴:۴۰ | |
| ۷ | Copula-based reliability analysis of a complex system subject to Wiener degradation process | زهرا صابریزاده | ۱۴:۴۰ - ۱۵:۰۰ | |
| ۸ | روش جدیدی برای تعمیم مفصل های دومتغیره | علی دست برآورده | ۱۵:۰۰ - ۱۵:۲۰ | علی دولتی |
| ۹ | Skew-Elliptical distribution of copula related random variables | فرشته آراد | ۱۵:۲۰ - ۱۵:۴۰ | |
| ۱۰ | Semiparametric estimation of mutual information for elliptical copulas | مرتضی محمدی | ۱۵:۴۰ - ۱۶:۰۰ | |
| | نماز و استراحت | | | ۱۶:۰۰ - ۱۶:۲۰ |
| ۱۱ | مدل سازی وابستگی نرخ جرم و شاخص فلاکت در ایران: رویکرد مفصل مبنا | یاسمن خلیلی | ۱۶:۲۰ - ۱۶:۴۰ | صدیقه شمس |
| ۱۲ | برآورد کسری مورد انتظار پرتفوی بر اساس تابع مفصل و رویکرد POT-GARCH | فاطمه علیزاده | ۱۶:۴۰ - ۱۷:۰۰ | |
| ۱۳ | پیش بینی بارش پاییزه شمال غرب با استفاده از مدل های آماری مفصل مبنا | منصوره کوهی | ۱۷:۰۰ - ۱۷:۲۰ | |
| ۱۴ | Sharpe ratio analysis: a copula approach | سمانه السادات موسوی | 17:20 -- 17:40 | |
| | استراحت | | | ۱۷:۴۰ - ۱۷:۵۰ |
| ۱۵ | FGM copula based analysis of wireless communication performances for multi-user channels | منا سادات محسن زاده | ۱۷:۵۰ - ۱۸:۱۰ | محمد امینی |
| ۱۶ | مفصل چندخطی تعمیم یافته به کمک ماتریس آماری مضاعف | محمد اعتصام | ۱۸:۱۰ - ۱۸:۳۰ | |
| ۱۷ | On the copula-based time between events control chart | حسام احمد | ۱۸:۳۰ - ۱۸:۵۰ | |
| ۱۸ | Investigating the effect of dependence of fading coefficients and its modeling with copula theory in non-orthogonal multiple access (NOMA) channels with physical layer security | فرامرز عجمی | ۱۸:۵۰ - ۱۹:۱۰ | |

صبح روز پنجشنبه: ۱۴۰۱/۱۱/۲۰

| ردیف | عنوان سخنرانی | ارائه دهنده | زمان | مسئول |
|------|--|---------------------|---------------|--------------|
| ۱۹ | On the use of time-varying vine copulas in multivariate time series analysis | ایوب شیخی | ۰۸:۰۰ - ۰۸:۳۰ | علی دولتی |
| ۲۰ | مفصل های پویا در مدل بندی سری های زمانی | حمید رضا نیلی ثانی | ۰۸:۳۰ : ۰۹:۰۰ | |
| ۲۱ | برآورد ذخیره خسارت های معوق با داده های سانسور شده با استفاده از تابع مفصل | صدیقه شمس | ۰۹:۰۰ - ۰۹:۳۰ | |
| | استراحت | | ۰۹:۳۰ - ۰۹:۴۰ | |
| ۲۲ | مدل سازی توام داده های طولی و بقا با استفاده از توابع مفصل عاملی | محمد رضا آخوند | ۰۹:۴۰ - ۱۰:۰۰ | احمدی انگالی |
| ۲۳ | مدل سازی داده های چندمتغیره طولی با استفاده از توابع مفصل جفتی و این | محمدصادق لؤلؤ | ۱۰:۰۰ - ۱۰:۲۰ | |
| ۲۴ | Analysis of diabetes using copula generalized additive regression model | مهدی امیدی | ۱۰:۲۰ - ۱۰:۴۰ | |
| ۲۵ | Application of d-vine regression copula in COVID-19 data | ندا دودمان | ۱۰:۴۰ - ۱۱:۰۰ | |
| | استراحت | | ۱۱:۰۰ - ۱۱:۱۰ | |
| ۲۶ | قانون تفکیک دو مقدم برای استلزام های احتمالاتی مفصل مینا | الناز معدنی خوش بخت | ۱۱:۱۰ - ۱۱:۳۰ | محمد امینی |
| ۲۷ | Ordering of conditional asymmetry | الهام مختاری | ۱۱:۳۰ - ۱۱:۵۰ | |
| ۲۸ | Kolmogorov-Feller weak law of large numbers for FGM random sequences | ام البنین هادیان فر | ۱۱:۵۰ - ۱۲:۱۰ | |
| ۲۹ | A new family of copulas based on distortion function and its properties | مجتبی اصفهانی | ۱۲:۱۰ - ۱۲:۳۰ | |
| ۳۰ | آزمون استقلال برای خانواده ای از مفصل های آمیخته | هادی جباری نوقابی | ۱۲:۳۰ - ۱۲:۵۰ | |
| | نماز و استراحت | | ۱۲:۵۰ - ۱۳:۱۰ | |
| | میزگرد و اختتامیه | | ۱۳:۱۰ - ۱۴:۰۰ | |
| | کارگاه آموزشی مدلسازی وابستگی با استفاده از نرم افزار R | | ۱۴:۰۰ - ۱۶:۰۰ | |

• نشانی کلاس مجازی سخنرانی ها:

<https://vroom.um.ac.ir/scta1>

فهرست مقالات فارسی

| | |
|--|----|
| مدل سازی توام داده‌های طولی و بقا با استفاده از توابع مفصل عاملی | |
| آخوند، م.ر. | ۹ |
| مفصل چندخطی تعمیم‌یافته به کمک ماتریس آماری مضاعف | |
| اعتصام، م.، محتشمی برزادران، غ. | ۲۱ |
| نگاه تاریخی به وابستگی منفی | |
| امینی، م.، بزرگنیا، ا. | ۲۹ |
| آزمون استقلال برای خانواده‌ای از مفصل‌های آمیخته | |
| زرگر، م.، جباری، ه.، امینی، م. | ۳۵ |
| مدل سازی وابستگی نرخ جرم و شاخص فلاکت در ایران: رویکرد مفصل مبنا | |
| خلیلی، ی.، دولتی، ع.، دست برآورده، ع.، رحمانی، ج. | ۴۱ |
| برآورد کسری مورد انتظار پرتفوی بر اساس تابع مفصل و رویکرد <i>POT - GARCH</i> | |
| علیزاده، ف.، محتشمی برزادران، غ. | ۵۱ |
| مدل‌های وابسته با ساختار موازی-سری و سری-موازی با تعداد زیر سیستم‌های تصادفی | |
| محتشمی برزادران، و.، امینی، م.، احمدی، ج. | ۶۳ |
| قانون تفکیک دو مقدم برای استلزام‌های احتمالاتی مفصل مبنا | |
| معدنی خوش‌بخت، ا.، دولتی، ع. | ۷۲ |

روش جدیدی برای تعمیم مفصل‌های دو متغیره

میرحسینی، م.، دست‌برآورده، ع. ۷۹

مدل‌سازی داده‌های چندمتغیره طولی با استفاده از توابع مفصل جفتی واین

لؤلؤ، م.ص.، آخوند، م.ر.، احمدی انگالی، ک.، برازجانی، ف. ۸۸



هفتمین سمینار
نظریه مفصل و کاربردهای آن

۱۹ و ۲۰ بهمن ۱۴۰۱



مدل سازی توام داده های طولی و بقا با استفاده از توابع مفصل عاملی

آخوند، م. ر.^۱

^۱ گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز

چکیده

از جمله روش هایی که در سال های اخیر توجه محققان را برای مدل سازی داده های چندمتغیره به خود جلب کرده است تابع مفصل می باشد. در این مقاله قصد داریم مدلی رگرسیونی را برای مدل سازی پاسخ های چندمتغیره آمیخته با استفاده مفصل های عاملی معرفی کنیم. مدل های مفصل عاملی، مفصل های جفتی و این هستند که شامل متغیرهای پنهان می باشند و بنابراین اجازه می دهند که وابستگی دمی، غیر خطی و نامتقارن میان متغیرها وجود داشته باشد. در این مقاله فرض می شود که متغیر پاسخ چندمتغیره بوده و به صورت آمیخته ای از پاسخ های گسسته و پیوسته باشد. همچنین علاوه بر پاسخ های آمیخته گسسته و پیوسته فرض می شود که یک متغیر پاسخ بقا نیز داشته باشیم که امکان وقوع سانسور در آن وجود دارد. پس از معرفی مدل در حالت کلی برای حالتی که مدل توام طولی گسسته و بقا داشته باشیم شبیه سازی انجام و کارایی مدل مورد را مورد بررسی قرار می دهیم.

کلمات کلیدی: مفصل عاملی، مدل سازی توام، تحلیل بقا، پاسخ های آمیخته

¹Mr.Akhood@scu.ac.ir

۱ پیش‌گفتار

بسیاری از مطالعات علوم مختلف مستلزم جمع‌آوری داده‌هایی هستند که آمیخته‌ای از داده‌های گسسته و پیوسته هستند. ساده‌ترین روش برای مدل‌سازی چنین داده‌هایی تجزیه و تحلیل متغیرهای گسسته و پیوسته به صورت جداگانه است و سپس دو مجموعه از نتایج را با هم ترکیب کنیم. روش ذکر شده شامل از دست دادن برخی اطلاعات می‌باشد و نتایج رضایت بخشی ندارد، زیرا هرگونه ارتباطی که بین متغیرهای آمیخته وجود دارد را نادیده می‌گیرد. از اولین پیشنهادات برای مدل‌سازی داده‌های چندمتغیره آمیخته روش فاکتورگیری است. این روش دو رویکرد برای توزیع‌های توام با پاسخ‌های آمیخته پیشنهاد می‌کند: در رویکرد اول توزیع‌های حاشیه‌ای برای پاسخ‌های گسسته و توزیع شرطی برای پاسخ‌های پیوسته به شرط پاسخ‌های گسسته در نظر گرفته می‌شوند و در رویکرد دوم، توزیع حاشیه‌ای برای پاسخ‌های پیوسته و توزیع شرطی برای پاسخ‌های گسسته به شرط پاسخ‌های پیوسته در نظر گرفته می‌شود. حالت اول در داده‌های آمیخته توجه زیادی در تحقیقات به خود جذب کرده است. این مدل توزیع نرمال چندمتغیره را برای پاسخ‌های پیوسته به شرط پاسخ‌های گسسته فرض می‌کند و یک توزیع چندجمله‌ای حاشیه‌ای برای پاسخ‌های گسسته فرض می‌کند. نمونه‌هایی از رویکرد دوم عبارت‌اند از مطالعات کاکس و ورموت (۱۹۹۲) که در این روش توزیع شرطی لجستیک یا پروبیت را برای متغیرهای پیوسته باینری به شرط پاسخ پیوسته و یک توزیع حاشیه‌ای نرمال چندمتغیره برای پاسخ پیوسته عرضه می‌کنند. از ایرادات روش فاکتورسازی این است که، وقتی پاسخ‌های چندمتغیره آمیخته باشند به راحتی توسعه نمی‌یابند، هم‌چنین چون پارامترها با توجه به تجزیه مورد استفاده تفسیرهای متفاوتی دارند، مدل‌های به‌دست آمده قابل قیاس نیستند. فیس و همکاران (۲۰۰۷) و جفا و همکاران (۲۰۱۶) مدل‌هایی با اثر تصادفی را در یک چارچوب چندمتغیره ارائه و یک تابع درست‌نمایی کلی برای تمام متغیرهای پاسخ با مشخص کردن یک توزیع توام برای اثرات تصادفی تشکیل دادند. این مدل‌ها جذاب‌اند زیرا می‌توان آنها را بدون در نظر گرفتن توزیع حاشیه‌ای پاسخ‌ها و همگن بودن داده‌ها به کار گرفت. اما انتگرال‌هایی با ابعاد زیاد با افزایش تعداد متغیرهای پاسخ ایجاد می‌کنند. روش دیگر برای مدل‌سازی داده‌های چندمتغیره استفاده از مفصل است.

توابع مفصل توابعی هستند که توزیع‌های چندمتغیره را به تابع توزیع حاشیه‌ای تک متغیره پیوند می‌دهند. مفصل‌ها در مدل‌سازی توزیع‌های چندمتغیره به ویژه تجزیه و تحلیل با ابعاد بالا کاربرد دارند. (جو ۱۹۹۷؛ نلسن ۲۰۰۶) در هنگام مدل‌سازی داده‌های چندمتغیره زمانی که پاسخ‌ها از نوع آمیخته هستند موضوع پیچیده‌تر است. با توجه به ماهیت برخی از داده‌های گسسته مدل‌سازی برای داده‌های آمیخته دشوار می‌شود. در این حالت برای تابع‌های مفصلی که فرم آن‌ها بسته است تابع جرم احتمال می‌تواند با گرفتن تفاوت متناهی از تابع مفصل برای حاشیه‌های گسسته محاسبه شود.

اسمیت و خالد (۲۰۱۲)، استوبر و همکاران (۲۰۱۵) و زیلکو و کوروویکا (۲۰۱۶) از توابع مفصل جفتی و این برای مدل‌سازی داده‌های آمیخته استفاده کردند. توابع مفصل جفتی و این دارای دو مزیت نسبت به توابع مفصل نرمال چند متغیره هستند، همانطوریکه که پاناجیتلیس و همکاران (۲۰۱۷) تاکید کرده‌اند. اولین مزیت پیچیدگی محاسباتی تابع توزیع توام به صورت توانی از d تغییر می‌کند در حالیکه در توزیع نرمال چندمتغیره پیچیدگی محاسباتی به صورت نمایی از مرتبه d افزایش می‌یابد. دوم اینکه توابع مفصل جفتی و این انعطاف‌پذیری بالایی از طریق تخصیص توابع مفصل پارامتری دومتغیره با وابستگی‌های جفتی یا خواص عدم تقارن متفاوت دارند. آنها دارای حالت خاص توابع مفصل نرمال چندمتغیره هستند، اگر تمام توابع مفصل دومتغیره توزیع نرمال دومتغیره در نظر گرفته شوند.

ژانگ و همکاران (۲۰۱۹) یک مدل جدید را برای مدل‌سازی داده‌های چندمتغیره که در آن برخی متغیرهای پاسخ گسسته و برخی دیگر از متغیرهای پاسخ پیوسته هستند را به کار گرفتند. آنها برای مدل‌سازی چنین داده‌های چندمتغیره‌ای از توابع مفصل جفتی و این استفاده کردند. مفصل جفتی یک مفصل چندمتغیره است که از مفصل‌های دومتغیره ساخته شده است. از زیر کلاس‌های آن می‌توان به D -واین و R -واین اشاره کرد.

از جمله روش‌هایی که اخیراً برای مدل‌سازی داده‌های چندمتغیره با استفاده از تابع مفصل پیشنهاد شده است استفاده از مفصل عاملی می‌باشد. مفصل‌های عاملی، مدل‌های مفصل جفتی هستند که شامل هم متغیرهای مشاهده شده و هم متغیرهای پنهان هستند. بنابراین آنها در تخصیص توابع مفصل پارامتری دومتغیره با وابستگی‌های دمی یا خاصیت عدم تقارن بسیار انعطاف‌پذیر

می‌باشند. مدل‌های مفصل عاملی قابل تفسیرتر و دارای برازش بهتری نسبت به توابع مفصل جفتی و این هستند. بعلاوه اینکه آنها تحت حاشیه‌ها بسته هستند یعنی حاشیه‌های با ابعاد پایین متعلق به خانواده پارامتری توابع مفصل مشابهی هستند و یک جایگشت متفاوت از متغیرهای مشاهده شده دارای توزیع یکسانی است. این حالت در توابع مفصل جفتی و این بدون متغیر پنهان، که یک جایگشت از متغیرهای مشاهده شده می‌تواند منجر به توزیع متفاوتی شود رخ نمی‌دهد. کمپوس و همکاران (۲۰۲۱) برای مدل سازی داده‌های بقای خوشه‌ای از توابع مفصل عاملی استفاده کردند. کادهم و نیکولوپولوس (۲۰۲۰) از توابع مفصل عاملی برای مدل‌سازی داده‌های آمیخته پیوسته و گسسته استفاده کردند. در این مقاله قصد داریم به گسترش مفصل عاملی برای حالتی که متغیرهای پاسخ آمیخته‌ای از پاسخ‌های پیوسته، گسسته و بقا باشند بپردازیم.

۲ مدل مفصل عاملی برای متغیرهای آمیخته

۱.۲ مفصل عاملی

در یک مدل مفصل p -عاملی فرض می‌شود که متغیرهای پاسخ $Y = (Y_1, Y_2, \dots, Y_d)$ به صورت شرطی به شرط p متغیر پنهان V_1, V_2, \dots, V_p مستقل از یکدیگر می‌باشند. در حالت مفصل یک عاملی، فرض کنید V_1 یک متغیر پنهان باشد که فرض می‌شود دارای توزیع یکنواخت استاندارد باشد. بر اساس قضیه اسکالر یک تابع مفصل دو متغیره $C_{V_1, j}$ وجود دارد به گونه‌ای که $P(V_1 \leq v_1, Y_j \leq y_j) = C_{v_1, j}(v_1, F_{Y_j}(y_j))$ که در آن $0 \leq v_1 \leq 1$ و F_j تابع توزیع تجمعی Y_j است. بر همین اساس داریم:

$$F_{Y_j|V_1}(y|v) = P(Y_j \leq y|V_1 = v) = C_{j|V_1}(F_j(y), v) = \frac{\partial C_{V_1, j}(v, F_j(y))}{\partial v} \quad (1)$$

با فرض اینکه $C_{j|V_1}(F_j(y)|v) = \frac{\partial C_{V_1, j}(v, F_j(y))}{\partial v}$ باشد و $\mathbf{y} = (y_1, y_2, \dots, y_d)$ مقادیر مشاهده شده \mathbf{Y} باشند تابع چگالی احتمال مشاهده شده برای داده‌ها در حالت مفصل یک عاملی به صورت زیر بدست می‌آید:

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_0^1 \prod_{j=1}^d f_{j|V_1}(y_j|v) dv \quad (2)$$

که

$$f_{j|V_1}(y|v) = \begin{cases} C_{j|V_1}(F_j(y)|v) - C_{j|V_1}(F_j(y-1)|v) & \text{اگر } Y_j \text{ گسسته باشد} \\ c_{V_1,j}(v, F_j(y)) f_j(y) & \text{اگر } Y_j \text{ پیوسته باشد} \end{cases}$$

چگالی $C_{j|V_1}(F_j(y)|v) - C_{j|V_1}(F_j(y-1)|v)$ چگالی $Y_j = y$ به شرط $V_1 = v$ و $c_{V_1,j}$ چگالی مفصل دومتغیره V_1 و Y_j است و f_j چگالی حاشیه‌ای Y_j است.

در حالت دو عاملی، دومتغیر پنهان V_1 و V_2 را در نظر بگیرید که بدون از دست دادن کلیت مسئله به صورت مستقل و دارای توزیع یکنواخت $U(0,1)$ می‌باشند. فرض کنید $C_{V_1,j}$ همانند مدل مفصل یک عاملی باشد و فرض کنید $C_{V_2,j}$ مفصل دو متغیره به صورت زیر باشد:

$$P(V_2 \leq v_2, Y_j \leq y | V_1 = v_1) = C_{V_2,j}(v_2, F_{j|V_1}(y|v_1))$$

که $F_{j|V_1}$ بوسیله معادله ۱ بدست می‌آید. آنگاه برای $0 \leq v_1, v_2 \leq 1$

$$\begin{aligned} P(Y_j \leq y | V_1 = v_1, V_2 = v_2) &= \frac{\partial}{\partial v_2} P(V_2 \leq v_2, Y_j \leq y | V_1 = v_1) \\ &= \frac{\partial}{\partial v_2} C_{V_2,j}(v_2, F_{j|V_1}(y|v_1)) = C_{j|V_2}(F_{j|V_1}(y|v_1) | v_2) \end{aligned}$$

چگالی مشاهده شده داده‌ها در حالت مدل مفصل دو عاملی به صورت زیر بدست می‌آید:

$$f_Y(\mathbf{y}) = \int_0^1 \int_0^1 \prod_{j=1}^d f_{V_j,j|V_1}(v_j, y_j | v_1) dv_1 dv_2 \quad (3)$$

که

$$f_{V_j,j|V_1}(v_j, y_j | v_1) = \begin{cases} C_{j|V_j}(F_{j|V_1}(y_j|v_1) | v_j) - C_{j|V_j}(F_{j|V_1}(y_j-1|v_1) | v_j) & \text{اگر } Y_j \text{ گسسته باشد} \\ c_{j,V_j;V_1}(F_{j|V_1}(y_j|v_1), v_j) c_{V_1,j}(v_1, F_j(y_j)) f_j(y_j) & \text{اگر } Y_j \text{ پیوسته باشد} \end{cases}$$

لازم به ذکر است که $C_{V_1,j}$ ، j امین متغیر پاسخ را به اولین متغیر پنهان V_1 متصل می‌کند و $C_{V_2,j}$ بین j -امین متغیر پاسخ و متغیر پنهان دوم به شرط V_1 پیوند ایجاد می‌کند. در مدل مفصل عاملی معرفی شده هیچ محدودیتی در انتخاب توابع توزیع حاشیه‌ای F_j یا توابع مفصل $C_{V_1,j}$ ، $C_{V_2,j}$ وجود ندارد.

۲.۲ مدل سازی توام داده‌های بقا و طولی چند متغیره آمیخته

فرض کنید بردار متغیر پاسخ برای فرد i ام، $Y_i, i = 1, 2, \dots, n$ شامل $Y_i = (Y_i^{(1)}, Y_i^{(2)}, Y_i^{(r)})$ باشد که در آن $Y_i^{(1)}$ متغیر پاسخ بقا است که امکان وقوع سانسور از راست برای آن وجود دارد و در صورتی که سانسور نشده باشد متغیر نشانگر $\delta_i = 1$ و در حالی که سانسور شده باشد $\delta_i = 0$ قرار داده می‌شود. $Y_i^{(2)}$ بردار متغیر پاسخ طولی پیوسته اندازه‌گیری شده در زمان‌های $j = 1, 2, \dots, J_1$ است. همچنین $Y_i^{(r)}$ متغیر پاسخ طولی گسسته باشد که در زمان‌های $k = 1, 2, \dots, J_r$ اندازه‌گیری شده است. در این حالت در صورتی که مشاهده i ام سانسور نشده باشد، سهم فرد i ام در تابع درستنمایی را می‌توان به صورت زیر نوشت:

$$c_{V_i, Y_i^{(1)}}(v, F(y_i^{(1)})) f(y_i^{(1)}) \prod_{j=1}^{J_1} [c_{V_i, j}(v, F_j(y_{ij}^{(2)})) f_j(y_{ij}^{(2)})] \quad (۴)$$

$$\prod_{k=1}^{J_r} [C_{Y_k^{(r)}|V_i}(F_j(y_{ik}^{(r)})|v) - C_{Y_k^{(r)}|V_i}(F_j(y_{ik}^{(r)} - 1)|v)]$$

و در صورتی که مشاهده i ام سانسور شده باشد، نقش فرد i ام در تابع درستنمایی را می‌توان به صورت زیر نوشت:

$$C_{Y_i^{(1)}|V_i}(v, S(y_i^{(1)})) \prod_{j=1}^{J_1} [c_{V_i, j}(v, F_j(y_{ij}^{(2)})) f_j(y_{ij}^{(2)})] \quad (۵)$$

$$\prod_{k=1}^{J_r} [C_{Y_k^{(r)}|V_i}(F_j(y_{ik}^{(r)})|v) - C_{Y_k^{(r)}|V_i}(F_j(y_{ik}^{(r)} - 1)|v)]$$

در نتیجه تابع درستنمایی برای کل نمونه به صورت زیر قابل محاسبه است:

$$L(y^{(1)}, \mathbf{y}^{(2)}, \mathbf{y}^{(r)}) = \prod_{i=1}^n \left(c_{V_i, Y_i^{(1)}}(v, F(y_i^{(1)})) f(y_i^{(1)}) \prod_{j=1}^{J_1} [c_{V_i, j}(v, F_j(y_{ij}^{(2)})) f_j(y_{ij}^{(2)})] \quad (۶)$$

$$\prod_{k=1}^{J_r} [C_{Y_k^{(r)}|V_i}(F_k(y_{ik}^{(r)})|v) - C_{Y_k^{(r)}|V_i}(F_k(y_{ik}^{(r)} - 1)|v)] \right)^{\delta_i} \times \left(C_{Y_i^{(1)}|V_i}(v, S(y_i^{(1)})) \right.$$

$$\left. \prod_{j=1}^{J_1} [c_{V_i, j}(v, F_j(y_{ij}^{(2)})) f_j(y_{ij}^{(2)})] \prod_{k=1}^{J_r} [C_{Y_k^{(r)}|V_i}(F_j(y_{ik}^{(r)})|v) - C_{Y_k^{(r)}|V_i}(F_j(y_{ik}^{(r)} - 1)|v)] \right)^{(1-\delta_i)}$$

در شرایطی که به دلیل وجود داده‌های گمشده اندازه‌گیری‌های طولی برای افراد متفاوت باشد و داشته باشیم و برای فرد i ام تعداد اندازه‌گیری طولی متغیر پاسخ پیوسته J_{1i} و برای متغیر طولی گسسته تعداد اندازه‌گیری برای فرد i ام برابر J_{2i} باشد آنگاه تابع درستنمایی را می‌توان بر اساس آن اصلاح کرد. برخی حالت‌های خاص تابع درستنمایی بالا در ادامه آورده شده است.

۳.۲ مدل‌سازی توام داده‌های بقا و داده‌های طولی پیوسته

فرض کنید بردار متغیر پاسخ برای فرد i ام، $Y_i, i = 1, 2, \dots, n$ شامل $Y_i = (Y_i^{(1)}, Y_i^{(2)})$ باشد که در آن $Y_i^{(2)}$ بردار متغیر پاسخ طولی پیوسته مشاهده شده و $Y_i^{(1)}$ زمان بقا برای فرد i ام باشد که در آن $Y_i^{(2)}$ بردار متغیر پاسخ طولی پیوسته اندازه‌گیری شده در زمان‌های $J_1 = 1, 2, \dots, J_1$ است. در صورتی که به دلیل وجود مشاهدات گمشده تعداد اندازه‌گیری‌ها برای افراد مختلف متفاوت باشد و برای فرد i ام تعداد J_i مشاهده موجود باشد آنگاه تابع درستنمایی برای مشاهدات را می‌توان به صورت زیر نوشت:

$$L(y^{(1)}, \mathbf{y}^{(2)}) = \prod_{i=1}^n \left(c_{V_i, Y_i^{(1)}}(v, F(y_i^{(1)})) f(y_i^{(1)}) \prod_{j=1}^{J_{1i}} \left[c_{V_i, j}(v, F_j(y_{ij}^{(2)})) f_j(y_{ij}^{(2)}) \right]^{\delta_i} \right. \\ \left. \left(C_{Y_i^{(1)}|V_i}(v, S(y_i^{(1)})) \prod_{j=1}^{J_{2i}} \left[c_{V_i, j}(v, F_j(y_{ij}^{(2)})) f_j(y_{ij}^{(2)}) \right] \right)^{(1-\delta_i)} \right) \quad (7)$$

۴.۲ مدل‌سازی توام داده‌های بقا و طولی گسسته

فرض کنید بردار متغیر پاسخ $Y_i, i = 1, 2, \dots, n$ شامل $Y_i = (Y_i^{(1)}, Y_i^{(2)})$ باشد که در آن $Y_i^{(2)}$ بردار متغیر پاسخ طولی گسسته و $Y_i^{(1)}$ زمان بقا برای فرد i ام باشد که در آن $Y_i^{(2)}$ بردار متغیر پاسخ طولی گسسته اندازه‌گیری شده در زمان‌های $J_2 = 1, 2, \dots, J_2$ است. در اینجا هم در صورتی که تعداد اندازه‌گیری‌های متغیر طولی گسسته برای نمونه‌های مختلف متفاوت باشد و برای نمونه i ام تعداد J_{2i} اندازه‌گیری داشته باشیم آنگاه تابع درستنمایی برای کل نمونه را می‌توان به صورت

زیر نوشت:

$$L(y^{(1)}, \mathbf{y}^{(r)}) = \prod_{i=1}^n \left(c_{V_i, Y_i^{(1)}}(v, F(y_i^{(1)})) f(y_i^{(1)}) \prod_{k=1}^{J_{r,i}} C_{j|V_i}(F_j(y_{ij}^{(r)})|v) - C_{j|V_i}(F_j(y_{ij}^{(r)}) - 1|v) \right)^{\delta_i} \quad (۸)$$

$$\times \left(C_{Y_i^{(1)}|V_i}(v, S(y_i^{(1)})) \prod_{k=1}^{J_{r,i}} [C_{j|V_i}(F_j(y_{ik}^{(r)})|v) - C_{j|V_i}(F_j(y_{ik}^{(r)}) - 1|v)] \right)^{(1-\delta_i)}$$

همانطوریکه گفته شد $C(.,.)$ تابع مفصل دومتغیره می باشد، که می تواند هر خانواده ای از توابع مفصل در نظر گرفته شود. برای بحث در مورد نیکویی برازش توابع مفصل در مدل های مفصل عاملی می توان به کادهم و نیکولوپولوس (۲۰۲۰) مراجعه کرد. همچنین F_j تابع توزیع حاشیه ای هر یک از متغیرهای پاسخ است که می تواند هر توزیع پیوسته یا گسسته ای در نظر گرفته شود. در صورتی که بدنبال برازش یک مدل رگرسیونی به داده ها باشیم، متغیرهای کمکی از طریق پارامترهای توزیع حاشیه ای می توانند وارد مدل شوند. در ادامه برای بررسی کارایی مدل در شرایطی که یک متغیر پاسخ طولی گسسته و یک متغیر پاسخ بقا داشته باشیم شبیه سازی صورت گرفته است.

۳ شبیه سازی

در این بخش روش انجام شبیه سازی برای مدل سازی توام یک متغیر بقا به همراه یک متغیر پاسخ طولی شمارشی را شرح می دهیم. فرض می کنیم که متغیر بقا دارای توزیع وایبل و متغیر پاسخ شمارشی دارای توزیع پواسن باشد. در اینجا یک مدل رگرسیونی برای متغیر پاسخ پیوسته طولی و همچنین یک مدل بقا برای متغیر پاسخ دارای توزیع وایبل در نظر می گیریم. شبیه سازی به حالتی که سایر توزیع ها به عنوان توزیع حاشیه ای در نظر گرفته شوند و یا متغیر طولی پیوسته باشد به سادگی قابل گسترش است. برای انجام شبیه سازی ابتدا v_1 از توزیع یکنواخت استاندارد انتخاب شد. با فرض اینکه تعداد دفعات اندازه گیری متغیر طولی گسسته برای تمام افراد یکسان و برابر $J_r = 3$ باشد، $u_j, j = 1, 2, 3, 4$ از $\frac{\partial C_{V_i, j}(v, F_j(y))}{\partial v}$ انتخاب شدند. برای انجام این کار از تابع مفصل کلایتون استفاده شد. همچنین پارامتر تابع مفصل برای متغیرها به

ترتیب $\theta = (1, 4/5, 3/5, 2)$ در نظر گرفته شد. در ادامه جهت وارد کردن متغیرهای کمکی به مدل x_{1i} از توزیع $U(0, 1)$ و متغیر کمکی x_{2i} از توزیع برنولی با پارامتر 0.5 انتخاب شد. سپس $y_{ij}^{(3)}, j = 1, 2, 3$ از توزیع پواسن با پارامتر λ_{ij} تولید شد. همچنین برای وارد کردن متغیرهای کمکی در مولفه بقا $\gamma_i = \exp(x_i' \beta^{(1)})$ با $\beta^{(1)} = (0.75, -0.75, 1)$ محاسبه شد و سپس t_i از توزیع وایبول با پارامترهای (α, γ_i) انتخاب شد که در آن $\alpha = 1/5$ در نظر گرفته شد. جهت ایجاد داده سانسور شده در متغیر بقا c_i از توزیع یکنواخت انتخاب گردید و بر اساس آن $y_i^{(1)} = \min(t_i, c_i)$ و متغیر نشانگر $\delta_i = 1$ برای $y_i = t_i$ و $\delta_i = 0$ برای $y_i = c_i$ در نظر گرفته شد. شبیه سازی برای تعداد نمونه‌های $n = 100, 250$ تایی انجام و به تعداد $R = 500$ بار تکرار گردید. برآوردها با استفاده از دستور *optim* در نرم افزار *R* بدست آمد.

۴ نتایج

نتایج حاصل از شبیه سازی برای درصدهای مختلف سانسور و همچنین اندازه نمونه‌های مختلف در جدول ۱ آمده است. نتایج حاصل از شبیه سازی برای یک متغیر طولی گسسته که از توزیع پواسن پیروی می‌کند به همراه یک متغیر بقا که دارای توزیع وایبل است آورده شده است. همانطوریکه ملاحظه می‌شود با افزایش تعداد نمونه میانگین مربعات خطای برآوردها کاهش یافته است. همچنین برای تعداد نمونه‌های بزرگتر آریبی‌ها مقادیر قابل قبولی دارند.

۵ نتیجه‌گیری

در این مقاله یک مدل کلی برای مدل سازی توام داده‌های آمیخته و بقا که می‌تواند برای مدل سازی توام داده‌های طولی و بقا با استفاده از مفصل عاملی به کار گرفته شود پیشنهاد گردید. مدل ارائه شده با توجه به انعطاف پذیری در انتخاب تابع مفصل و توزیع‌های حاشیه‌ای به سادگی می‌تواند برای هر تابع مفصل و هر توزیع حاشیه‌ای مورد استفاده قرار گیرد. همچنین این مدل را می‌توان

برای بررسی اثر متغیرهای کمکی بر روی متغیرهای پاسخ آمیخته به کار گرفت. نتایج حاصل از شبیه‌سازی نشان می‌دهد که برآورد پارامترها برای تعداد نمونه‌های کم با اریبی زیادی همراه است که با افزایش تعداد نمونه اریبی برآوردها کاهش می‌یابد. از محدودیت‌های مدل ارائه شده نیاز به انتگرال‌گیری بر روی بازه صفر و یک است که گرچه با انتگرال‌گیری به روش گوس-لژاندره قابل محاسبه است اما لازم است تا کارایی آن در مدل‌های مختلف مورد بررسی قرار گیرد.

مراجع

- [1] Cox, DR., Wermuth N. (1992), *Response model for mixed binary and quantitative variables*, *Biometrika*, 79, 441-461
- [2] Fieuws, S., Verbeke, G., Maes, B., Vanrenterghem, Y. (2007), *Predicting renal graft failure using multivariate longitudinal profiles*, *Biostatistics*, 9(3), 419-431.
- [3] Jaffa, M A., Gebregziabher, M., Luttrell, D K., Luttrell, L M., Jaffa, A A. (2016), *Multivariate generalized linear mixed model with random intercepts to analyze cardiovascular risk markers in type-1 diabetic patients*, *Journal of applied statistics* 43 (8), 1447-1464.
- [4] Joe, H. (1997). *Multivariate models and multivariate dependence concepts*, Chapman and Hall, London.
- [5] Nelsen, R. (2006), *An introduction to copula*, 2nd end. Springer, New York.
- [6] Smith, M. S., & Khaled, M. A. (2012). *Estimation of copula models with discrete margins via Bayesian data augmentation*, *Journal of the American Statistical Association*, 107(497), 290-303.

- [7] Stober, J., Hong, H., Czado, C., Ghosh, P. (2015). *Comorbidity of chronic diseases in the elderly: Patterns identified by a copula design for mixed responses*, Journal of Computational statistics & data analysis, 88, 28-29.
- [8] Zilko, A. A., & Kurowicka, D. (2016). *Copula in a multivariate mixed discrete-continuous model*, Computational Statistics & Data Analysis, 103, 28-55.
- [9] Panagiotelis, A., Czado, C., Joe, H., & Stober, J. (2017). *Model selection for discrete regular vine copulas*. *Computational Statistics and Data Analysis*, 106, 138–152.
- [10] Zhang, W., Zhang, M., Chen, Y. (2020). *A copula-based GLMM model for multivariate longitudinal data with mixed-types of responses*, Sankhya, 82 (2), 353-379.
- [11] Campos, E., Braekers, R., de Souza, D.J. et al. (2021), *Factor copula models for right-censored clustered survival data*. *Lifetime Data Analysis*, 27, 499–535.
- [12] Kadhem, S. H., & Nikoloulopoulos, A. K. (2021). *Factor copula models for mixed data*, *British Journal of Mathematical and Statistical Psychology*, 74(3), 365-403.

جدول ۱: نتایج حاصل از شبیه‌سازی متغیر پاسخ طولی گسسته و بقا با استفاده از مفصل عاملی با ۲۰ درصد سانسور در متغیر بقا

| ۲۵۰ با $n = 250$ درصد سانسور | | | ۱۰۰ با $n = 100$ درصد سانسور | | | مقدار واقعی | پارامتر |
|------------------------------|--------|--------|------------------------------|--------|--------|-------------|-----------------|
| خطای استاندارد | اریبی | برآورد | خطای استاندارد | اریبی | برآورد | | |
| ۰/۰۷۷ | -۰/۰۲۱ | ۰/۴۷۸ | ۰/۱۲۳ | ۰/۰۰۶ | ۰/۵۰۶ | ۰/۵۰ | $\beta_1^{(3)}$ |
| ۰/۰۹۶ | ۰/۰۰۶ | -۰/۴۹۳ | ۰/۱۴۷ | -۰/۰۱۴ | -۰/۵۱۴ | -۰/۵۰ | $\beta_1^{(3)}$ |
| ۰/۰۸۵ | ۰/۰۱۶ | ۱/۰۱۶ | ۰/۱۲۰ | -۰/۰۰۳ | ۰/۹۹۶ | ۱/۰۰ | $\beta_2^{(3)}$ |
| ۰/۱۸۳ | -۰/۰۰۹ | ۱/۴۹۰ | ۰/۳۰۲ | ۰/۰۵۵ | ۱/۵۵۵ | ۱/۵۰ | $\beta_1^{(1)}$ |
| ۰/۲۲۹ | ۰/۰۲۳ | ۰/۷۷۳ | ۰/۳۶۴ | -۰/۰۱۳ | ۰/۷۳۶ | ۰/۷۵ | $\beta_1^{(1)}$ |
| ۰/۱۹۷ | -۰/۰۰۶ | ۰/۷۴۳ | ۰/۲۴۶ | -۰/۰۲۲ | ۰/۷۲۷ | ۰/۷۵ | $\beta_2^{(1)}$ |
| ۰/۱۲۵ | ۰/۰۰۰ | ۱/۵۰۰ | ۰/۱۶۳ | ۰/۰۲۱ | ۱/۵۲۱ | ۱/۵۰ | α |
| ۰/۳۱۱ | ۰/۰۵۶ | ۲/۰۵۶ | ۰/۴۵۳ | ۰/۰۳۹ | ۲/۰۳۹ | ۲/۰۰ | θ_1 |
| ۰/۹۴۳ | ۰/۲۴۶ | ۳/۷۴۶ | ۶/۴۹۵ | ۰/۶۷۴ | ۴/۱۷۴ | ۳/۵۰ | θ_2 |
| ۱/۱۲۵ | ۰/۱۹۶ | ۴/۶۹۶ | ۱۵/۱۹۰ | ۳/۲۸۵ | ۷/۷۸۵ | ۴/۵۰ | θ_3 |
| ۰/۱۷۷ | ۰/۰۳۷ | ۱/۰۳۷ | ۰/۳۲۷ | ۰/۰۹۵ | ۱/۰۹۵ | ۱/۰۰ | θ_4 |
| ۲۵۰ با $n = 250$ درصد سانسور | | | ۱۰۰ با $n = 100$ درصد سانسور | | | مقدار واقعی | پارامتر |
| خطای استاندارد | اریبی | برآورد | خطای استاندارد | اریبی | برآورد | | |
| ۰/۰۹ | -۰/۰۱۵ | ۰/۴۸۴ | ۰/۱۲ | -۰/۰۱۶ | ۰/۴۸۳ | ۰/۵۰ | $\beta_1^{(3)}$ |
| ۰/۱۰ | ۰/۰۰۰ | -۰/۴۹۹ | ۰/۱۶ | ۰/۰۲۰ | -۰/۴۷۹ | -۰/۵۰ | $\beta_1^{(3)}$ |
| ۰/۰۹ | ۰/۰۱۶ | ۱/۰۱۶ | ۰/۱۰ | -۰/۰۰۵ | ۰/۹۹۴ | ۱/۰۰ | $\beta_2^{(3)}$ |
| ۰/۲۷ | ۰/۰۱۰ | ۱/۵۱۰ | ۰/۳۸ | ۰/۰۲۴ | ۱/۵۲۴ | ۱/۵۰ | $\beta_1^{(1)}$ |
| ۰/۲۷ | -۰/۰۰۷ | ۰/۷۴۲ | ۰/۴۲ | ۰/۰۳۵ | ۰/۷۸۵ | ۰/۷۵ | $\beta_1^{(1)}$ |
| ۰/۲۵ | -۰/۰۴۵ | ۰/۷۰۴ | ۰/۲۶ | ۰/۰۳۲ | ۰/۷۸۲ | ۰/۷۵ | $\beta_2^{(1)}$ |
| ۰/۱۶ | -۰/۰۰۴ | ۱/۴۹۵ | ۰/۱۷ | ۰/۰۴۵ | ۱/۵۴۵ | ۱/۵۰ | α |
| ۰/۲۹ | ۰/۰۱۳ | ۲/۰۱۳ | ۰/۵۲ | ۰/۱۱۴ | ۲/۱۱۴ | ۲/۰۰ | θ_1 |
| ۰/۶۵ | ۰/۱۲۹ | ۳/۶۲۹ | ۶/۱۲ | ۰/۸۳۸ | ۴/۳۳۸ | ۳/۵۰ | θ_2 |
| ۱/۲۰ | ۰/۱۸۴ | ۴/۶۸۴ | ۳/۸۰ | ۲/۹۱۲ | ۷/۴۱۲ | ۴/۵۰ | θ_3 |
| ۰/۲۲ | ۰/۰۶۹ | ۱/۰۶۹ | ۰/۳۶ | ۰/۱۱۱ | ۱/۱۱۱ | ۱/۰۰ | θ_4 |



هفتمین سمینار
نظریه مفصل و کاربردهای آن

۱۹ و ۲۰ بهمن ۱۴۰۱



مفصل چندخطی تعمیم یافته به کمک ماتریس آماری مضاعف

اعتصام، م^۱ محتشمی برزادران، غ^۲

^{۱،۲} گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده

در این یادداشت ضمن یادآوری کوتاه از تابع مفصل و مفصل تجربی، مفصل چندخطی معرفی میشود. در ادامه چگونگی استفاده از ماتریس آماری مضاعف برای ایجاد مفصل چندخطی بیان میگردد. برای این منظور الگوریتمی همراه با مثال ارائه میگردد.

کلمات کلیدی: تابع مفصل، قضیه اسکالر، مفصل ناپارامتری، مفصل چندخطی، ماتریس آماری مضاعف

۱ مقدمات اولیه

وابستگی‌های پیچیده‌ای ممکن است بین متغیرهای تصادفی وجود داشته باشد. مثلاً وقوع یک رخداد میتواند منجر به رخدادهایی دیگر شود. روش‌های مختلفی برای بررسی وابستگی‌ها وجود دارد. در سال‌های اخیر از تابع مفصل برای این منظور استفاده شده است. تابع مفصل توزیع‌های

¹etesam@mail.um.ac.ir

²grmohtashami@um.ac.ir

حاشیه‌ای را به توزیع توأم متصل می‌سازد و میتواند وابستگی‌های غیرخطی را مدل کند و از این رو مهم است، (نلسون ۲۰۰۶).

تابع مفصل $C(u_1, u_2, \dots, u_d)$ تابع توزیع d -متغیره‌ای است که در $[0, 1]^d$ تعریف شده و $C: [0, 1]^d \rightarrow [0, 1]$ است و خود یک توزیع توأم میباشد که حاشیه‌های آن یکنواختند. فرض میکنیم:

• R^d فضای گسترش یافته d -بعدی و $a = (a_1, a_2, \dots, a_d)$ نشان دهنده یک نقطه در آن است.

• اگر برای هر k ، $a_k \leq b_k$ باشد، آنگاه $a \leq b$ است.

• اگر $a \leq b$ باشد، $[a, b] = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d]$ مکعب d -بعدی را نشان می‌دهد و حجم آن $V_c([a, b])$ است.

قضیه‌ی اسکالر مبنای نظری استفاده از تابع مفصل است. این قضیه میگوید اگر H یک تابع توزیع d -متغیره با حاشیه‌های F_1, F_2, \dots, F_d باشد، مفصل d -بعدی C وجود دارد که برای هر x در R^n

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)). \quad (1)$$

اگر F_1, F_2, \dots, F_d پیوسته باشند C یکتا است. اگر C یک مفصل d -بعدی و F_1, F_2, \dots, F_d توابع حاشیه‌ای باشند رابطه بالا یک تابع توزیع d -بعدی است. تابع چگالی مفصل در صورت وجود

$$c(u_1, u_2, \dots, u_d) = \frac{\partial C(u_1, u_2, \dots, u_d)}{\partial u_1 \partial u_2 \dots \partial u_d}, \quad (2)$$

است. بنابراین تابع چگالی توأم برابر با

$$f(x_1, x_2, \dots, x_d) = f_1(x_1) f_2(x_2) \dots f_d(x_d) \frac{\partial C(F_1(x_1), F_2(x_2), \dots, F_d(x_d))}{\partial F_1(x_1) \partial F_2(x_2) \dots \partial F_d(x_d)}, \quad (3)$$

است که در آن $f_1(x_1), f_2(x_2), \dots, f_d(x_d)$ توابع چگالی حاشیه‌ای هستند. ویژگی‌های زیر برای

مفصل بودن تابع C لازم و کافی است:

$$\begin{aligned} ۱: C(1, \dots, 1, u, 1, \dots, 1) &= u, \quad 0 \leq u \leq 1, \\ ۲: C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_d) &= 0, \quad 0 \leq u_j \leq 1, \\ ۳: \int_{V_c} dC &\geq 0. \end{aligned} \quad (۴)$$

استفاده از مفصل سودمند است زیرا بنابر قضیه اسکالر، مسأله‌ی تعیین تابع توزیع به دو زیر مسأله تعیین حاشیه‌ها و تعیین وابستگی‌ها تقسیم میشود که وظیفه تابع مفصل تعیین وابستگی‌ها است. بنابر این فهم وابستگی‌ها ساده‌تر میشود، به خصوص در حالت دوبعدی میتوان وابستگی‌ها را در مربع واحد به تصویر گذاشت و از روی آن مفصل مناسب را شناسایی کرد.

۲ مفصل ناپارامتری

در حالت دو متغیره با استفاده از شکل داده‌ها، مفصل مناسب انتخاب یا طراحی میشود. اگر نتوان مفصل مناسبی پیدا کرد یک راه این است که از مفصل تجربی استفاده نمود. مفصل تجربی در دهولز (۱۹۷۹) به صورت

$$C_n(u_1, \dots, u_d) = \frac{1}{n} \sum_{i=1}^n I(U_{i1} \leq u_1, \dots, U_{id} \leq u_d), \quad (۵)$$

معرفی شد. تفسیر رابطه بالا این است که اگر داده‌ها را $a_i, i \in \{0, 1, \dots, n\}$ بنامیم، برای نقطه خاص u در ابر-مکعب دامنه تابع مفصل، چه درصدی از داده‌ها در رابطه $a_i \geq u$ صدق میکنند، که مشابه روش بدست آوردن تابع توزیع تجربی می‌باشد. رابطه (۵) خود یک مفصل نیست، اما تقریبی از آن است و در حالت مجانبی به تابع مفصل اصلی میل می‌کند، (جو ۲۰۱۴). جایگزین‌های مختلفی برای مفصل تجربی وجود دارد. یکی از آنها مفصلی تحت عنوان چندخطی است که همان طور که در جنست (۲۰۰۷) بیان شده است، برخلاف مفصل تجربی یک تابع مفصل معتبر میباشد، یعنی در تعریف تابع مفصل صدق می‌کند. مفصل چندخطی در حقیقت یک روش هنجارسازی است که توسط نلسون (۲۰۰۶) برای تبدیل تابع شبه مفصل به مفصل

استفاده شده است و برای تبدیل جداول احتمالاتی^۱ به توزیع توأم کاربرد دارد.

۱.۲ مفصل ناپارامتری چندخطی و ماتریس آماری مضاعف

مفصل چندخطی فضای حالت را به تعدادی مربع یا مستطیل تقسیم میکند و به هر کدام از آن قسمت‌ها یک تابع چگالی تخصیص میدهد. در ادامه این مفصل با کمک ماتریس آماری مضاعف در حالت دومتغیره توضیح داده میشود. استفاده از ماتریس آماری مضاعف بستری مناسب برای ایجاد مفصل چندخطی با ویژگی‌های دلخواه، مانند متقارن و متقارن شعاعی را ایجاد می‌کند.

مربع واحد را در نظر بگیرید که از دو محور u و v تشکیل شده است که برای نمایش تابع مفصل دوبعدی C^2 استفاده میشود. هر کدام از محورها را به m بازه با طول برابر تقسیم می‌کنیم. با این کار فضای حالت به m^2 مربع کوچکتر تقسیم می‌شود. مربع سطر i و ستون j را با نماد E_{ij} نشان می‌دهیم.

می‌خواهیم مفصل مجهول C را با مفصل چندخطی $C^\#$ تخمین بزنیم. مفصل چندخطی به مربع کوچک E_{ij} یک تابع چگالی ثابت تخصیص میدهد که ماتریس $A^{m \times m}$ نشان‌دهنده درایه i و j آن بوده و در حالت ایده‌آل برابر با

$$A_{ij}^* = \int_{E_{ij}} dC, \quad (۶)$$

است. بنابر این مجموع درایه‌های هر سطر و ستون ماتریس A_{ij}^* ، $\sum_{i=1}^m A_{ij}^* = \sum_{j=1}^m A_{ij}^* = \frac{1}{m}$ است. در نتیجه mA^* یک ماتریس آماری مضاعف است که آن را S مینامیم. برای بدست آوردن $C^\#$ یک ماتریس از پارامترها وجود دارد که باید آن را با الگوریتمی به ماتریس آماری مضاعف تبدیل کرد. سپس به روشی مانند حداکثر درست‌نمایی پارامترها به وسیله روشی مانند کاهش گرادیان بدست می‌آیند. ماتریس آماری مضاعف درایه‌های نامنفی دارد و مجموع هر سطر و هر ستون آن یک است. الگوریتم‌هایی برای تولید ماتریسی با مجموع سطرها و ستون‌های مشخص وجود دارد، برای نمونه فینبرگ (۱۹۷۰) الگوریتمی با استفاده از روش حداقل مربعات را بررسی

¹contingency table

کرده است. در اینجا با الهام از این روش، الگوریتمی ساده شده استفاده میشود که ماتریس آماری مضاعف را با روش کاهش گرادیان و حداکثر درست‌نمایی بدست می‌آورد. الگوریتم ۱ برای تبدیل یک ماتریس دلخواه به ماتریس آماری مضاعف نشان داده میشود که ماتریس $Q^{m \times m}$ از پارامترها به همراه یک پارامتر حقیقی نامنفی α ورودی آن و یک ماتریس آماری مضاعف نتیجه خروجی آن است.

الگوریتم ۱ الگوریتم تولید ماتریس آماری مضاعف از ماتریس مجهولات

ورودی: ماتریس $Q^{m \times m}$ از پارامترها و یک پارامتر حقیقی نامنفی α .

خروجی: ماتریس آماری مضاعف S .

- ۱: میانگین هر سطر را از درایه‌های آن کم کن. (در نتیجه مجموع درایه‌ها در سطر برابر با صفر شود)
- ۲: میانگین هر ستون را از درایه‌های آن کم کن. (در نتیجه مجموع درایه‌ها در هر ستون برابر با صفر شود)
- ۳: از تمام درایه‌های ماتریس، کوچکترین عنصر ماتریس را کم کن. (در نتیجه درایه‌ها نامنفی میشوند)
- ۴: به تمام درایه‌های ماتریس، عدد α را اضافه کن. هر درایه از ماتریس را در $\frac{m}{\text{مجموع درایه‌های ماتریس}}$ ضرب کن.

در مرحله یک الگوریتم بالا مجموع درایه‌های هر سطر برابر با صفر می‌شود. در مرحله دو مجموع درایه‌های هر ستون برابر با صفر میشود، همچنین مجموع درایه‌های هر سطر نیز در پایان مرحله دوم صفر است، زیرا مجموع درایه‌های سطرها، بعد از اتمام مرحله دو با هم برابر است، از طرفی مجموع درایه‌های ستون‌ها صفر بوده، مجموع درایه‌های ماتریس صفر است، در نتیجه مجموع درایه‌های هر سطر صفر است. بعد از اتمام سوم درایه‌های ماتریس نامنفی می‌شوند. مرحله چهار این اجازه را به ماتریس میدهد که کوچکترین درایه‌اش ناصفر باشد. بعد از اتمام مرحله پنج یک ماتریس آماری مضاعف داریم.

فرض کنید $(u, v) = (\frac{i}{m} + u., \frac{j}{m} + v.)$ باشد که در آن i و j اعدادی حسابی و $0 \leq u., v. \leq \frac{1}{m}$ است. در این صورت (u, v) در $E_{i+1, j+1}$ قرار دارد و

$$\begin{aligned} C^\#(u, v) &= \frac{1}{m} \left(\sum_{p=1}^i \sum_{q=1}^j S_{pq} + \sum_{p=1}^i S_{p, j+1} + \sum_{q=1}^j S_{i+1, q} + S_{i+1, j+1} u. v. \right) \\ &= \frac{1}{m} \left(\sum_{p=1}^i \sum_{q=1}^j S_{pq} + \sum_{p=1}^i S_{p, j+1} + \sum_{q=1}^j S_{i+1, q} + S_{i+1, j+1} C^\pm(mu., mv.) \right), \end{aligned} \quad (7)$$

است. در رابطه (۷) می‌توان به جای مفصل استقلال از مفصل دلخواه دیگری استفاده کرد که

رابطه

$$\begin{aligned}
 C^\#(u, v) &= \frac{1}{m} \left(\sum_{p=1}^i \sum_{q=1}^j S_{pq} + \sum_{p=1}^i S_{p,j+1} + \sum_{q=1}^j S_{i+1,q} + S_{i+1,j+1} u.v. \right) \\
 &= \frac{1}{m} \left(\sum_{p=1}^i \sum_{q=1}^j S_{pq} + \sum_{p=1}^i S_{p,j+1} + \sum_{q=1}^j S_{i+1,q} + S_{i+1,j+1} \tilde{C}(mu., mv.) \right), \quad (8)
 \end{aligned}$$

بدست خواهد آمد. از رابطه (۸) می‌توان در مواردی مانند وجود وابستگی دمی استفاده کرد. وابستگی دمی با معادلات

$$\lambda_U = 2 - \lim_{t \rightarrow 1^-} \frac{1 - C(t, t)}{1 - t}, \quad (9)$$

و

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t}, \quad (10)$$

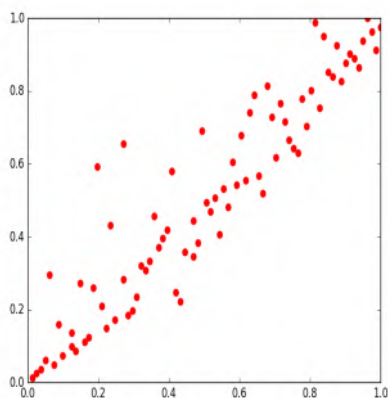
نشان داده می‌شود.

مثال ۱۰۲. دادگان مربوط به زمان خرابی کامیون‌های مخصوص حمل بار در یک معدن با توجه به تناژ بار، در یوان و همکاران (۲۰۱۷) معرفی شده است. با توجه به این که می‌خواهیم تنها کاربرد مفصل چندخطی دیده شود، با استفاده از توزیع تجربی، دادگان را به مربع واحد منتقل می‌کنیم. داده‌های منتقل شده در شکل ۱ نشان داده شده‌اند. برای این مثال از دو مدل استفاده می‌شود. در مدل اول رابطه (۷) به کار گرفته شده است. محور u و v را به m قسمت برابر تقسیم می‌کنیم. مقادیر قرینه لگ-بخت برای $m \in \{1, 2, \dots, 10\}$ محاسبه می‌شوند و اعداد در شکل ۲ بدست آمدند. مدل دوم مانند مدل اول است با این تفاوت که در آن از معادله (۸) استفاده شده است که در آن

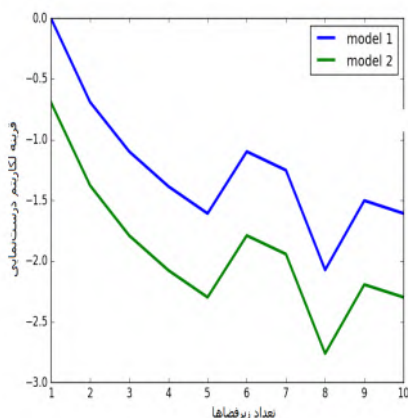
$$\tilde{C}(u, v; \theta) = uv + \theta uv(1 - u)(1 - v), \quad (11)$$

مفصل FGM و $\theta \in [-1, 1]$ پارامتر است.

در شکل ۲ ملاحظه می‌شود که با افزایش m قرینه لگ-بخت نزولی و در مدل دوم همواره کمتر از مدل اول است.



شکل ۱: پراکندگی داده‌های مثال ۱.۲



شکل ۲: نمودار قرینه لگ-بخت مثال ۱.۲

۳ جمع‌بندی

در این مقاله پس از معرفی تابع مفصل برای مدل‌سازی وابستگی‌ها، قضیه اسکالر که نقش بنیادینی در این زمینه دارد توضیح داده شد. سپس مفصل تجربی به همراه نسخه خاصی از آن تحت عنوان مفصل چندخطی مرور شد. همچنین با استفاده از ماتریس آماری مضاعف تعمیمی از مفصل چندخطی ارائه شد که میتوان انواع مختلف ویژگی‌های یک مفصل خاص مانند متقارن بودن و نشان دادن وابستگی دمی را بر آن اعمال کرد. در نهایت نیز با انجام آزمایشی عملکرد مفصل

معرفی شده نمایش داده شد.

مراجع

- [1] Nelsen, R. B. (2006). *An Introduction to Copulas*. Lectures Notes in Statistics, 139, Springer-Verlag, New York.
- [2] Joe, H.(2014). *Dependence Modeling With Copulas*. CRC press.
- [3] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance. *Bulletin de la classe des sciences, Académie Royale de Belgique, 5e série*, **65**, pp.274–292.
- [4] Genest, C. and Neslehova, J. (2007). A primer on copulas for count data. *Astin Bull*, **37**, pp.475–515.
- [5] Fienberg, S.E. (1970). The analysis of multidimensional contingency tables. *Dependence Modeling*, **51(3)**, pp.419-433.
- [6] Yuan, F., Barabadi, A. and Lu, J. (2017). Reliability modelling on two-dimensional life data using bivariate Weibull distribution: with case study of truck in mines. *Eksploracja i Niezawodność – Maintenance and Reliability*, **19 (4)**, pp.650–659.



هفتمین سمینار

نظریه مفصل و کاربردهای آن

۱۹ و ۲۰ بهمن ۱۴۰۱



نگاه تاریخی به وابستگی منفی

امینی، م^۱ بزرگنیا، ا^۲

۱،۲ گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده

همبستگی یکی از مفاهیم پایه در آمار است که برای محققین علوم طبیعی، اجتماعی و پزشکی و غیره آشنا است. مفاهیم توسعه یافته همبستگی در قرن نوزدهم، جذابیت آن را برای محققان در قرن بیستم گسترش داد و اکنون، در قرن بیست و یکم، اکثر افراد غیر علمی حداقل یک ایده اولیه از معنای "همبستگی" دارند. با این حال، در بین پژوهشگران نیز کمتر به تاریخ این واژه های علمی توجه میشود. در ریاضیات، به ویژه در محاسبات پیشرفته و نظریه اندازه گیری، مفهوم «استقلال» که احتمالاً توسط هانری لیبگ و یا دانشمندان معاصریشان مطرح شده است، بسیار اساسی و مورد توجه است. نظریه احتمال، که به عنوان شاخه ای از نظریه اندازه گیری با توجه به کاربردها در دنیای واقعی توسعه یافت، مفهوم استقلال را به روش شناسی آماری مرتبط کرد. بنابراین دو مفهوم استقلال و وابستگی (عدم استقلال) و همبستگی (ناهمبستگی) از واژه های پر کاربرد و مهم در علوم آماری هستند. روابط وابستگی بین متغیرهای تصادفی یکی از موضوعات مورد توجه در نظریه احتمال و آمار است که بدون اعمال مفروضات خاص در مورد وابستگی هیچ مدل آماری معنی داری نمیتوان ساخت. در این سخنرانی روند تاریخی تحقیقات انجام شده از قرن هیجدهم توسط آماردان معروف بریتانیایی تی-بیز تا کنون در مورد این مفاهیم پر کاربرد آماری ارایه شده است. علاوه بر این بررسی روند گسترش تاریخی وابستگی منفی از سال ۱۹۶۶ تا کنون از اهداف اصلی این سخنرانی می باشد.

¹m-amini@um.ac.ir

²bozorgnia@um.ac.ir

کلمات کلیدی: همبستگی، ناهمبستگی، استقلال آماری، وابستگی، وابستگی منفی

۱ استقلال آماری

در قرن هجدهم و نوزدهم (بیش از ۱۰۰ سال قبل از تحقیقات ف. گالتون در مورد همبستگی) تعریف های شفاهی از استقلال گاهی در نوشته های نویسندگان مختلف ظاهر شد. تی بیز در مقاله خود در سال ۱۷۶۳ به مفهوم استقلال می پردازد و آن را چنین تعریف می کند: پیشامدها در صورت وقوع، مستقل هستند اگر رخ دادن یکی از آنها نسبت به سایر پیشامدها نه احتمال را افزایش دهد و نه کاهش دهد.

تعریف مشابهی در سال ۱۷۶۷ توسط پرایس، یکی از دوستان نزدیک و معاصر تی بیز، در جلد چهارم پایان نامه که در چاپ دوم در لندن، ۱۷۶۸ بازنشر شد به صورت زیر ارائه شد: دو پیشامد در هنگام وقوع مستقل هستند در صورتیکه وقوع یکی از آنها تأثیری بر وقوع دیگری نداشته باشد. ویلیام امرسون (۱۷۸۲-۱۷۰۱)

تعریف ۱.۰۱. رویدادها (رخ دادها-پیشامدها) زمانی مستقل هستند که هیچ گونه ارتباطی با یکدیگر نداشته باشند، یا رخ دادن یکی از آنها مانع و یا باعث رخ دادن دیگری نشود.

تعریف ۲.۰۱. یک پیشامد زمانی نسبت به پیشامد دیگر وابسته است که احتمال وقوع آن با وقوع رویداد دیگر تغییر کند.

لاپلاس (۱۸۷۲-۱۷۴۹) در کتاب تحلیل های احتمالی (۱۸۱۲) اولین توصیف ریاضی مفهوم استقلال آماری را به صورت زیر ارائه داد.

تعریف ۳.۰۱. الف- دنباله $\{E_i, i \geq 1\}$ از پیشامد های تصادفی ساده مستقل هستند اگر

$$P\left(\bigcap_{i=1}^n E_i\right) = \prod_{i=1}^n P(E_i).$$

ب- دو پیشامد ساده که وقوع پیشامد اول E_1 بر احتمال وقوع پیشامد دوم E_2 تأثیر می گذارد، یعنی $P[E_1 \cap E_2] \neq P[E_1] \cdot P[E_2]$ وابسته هستند.

علاوه بر این تعریف های مشابه توسط آماردانان دیگر مانند: بول (۱۸۴۶-۱۸۱۶)، کینز (۱۹۶۴-۱۸۸۶)، دموورگان (۱۸۰۶-۱۸۷۱) و پیرسون (۱۹۳۶-۱۸۵۷) ارایه شده است.

۲ ناهمبستگی در مقابل وابستگی

همبستگی یکی از مفاهیم پایه در آمار است که برای محققین علوم طبیعی، اجتماعی و پزشکی و غیره آشنا است. این بخش شامل خلاصه ای از توسعه تاریخی مفهوم همبستگی و کاربردهای آن است. عمدتاً به وضعیت، کاربرد و تفسیر آن توجه شده است، قابل توجه می باشد که مفاهیم توسعه یافته همبستگی در قرن نوزدهم، جذابیت آن را برای محققان در قرن بیستم گسترش داد و اکنون، در قرن بیست و یکم، اکثر افراد غیر علمی حداقل یک ایده اولیه از معنای ”همبستگی” دارند. با این حال، در بین پژوهشگران نیز کمتر به تاریخ این واژه های علمی توجه می شود. در هر حال دو مفهوم: استقلال و وابستگی (عدم استقلال) و همبستگی (ناهمبستگی) از واژه های پر کاربرد و مهم در علوم آماری هستند. برای مطالعه بیشتر به: توکی (۲۰۰۰-۱۹۱۵)، دارال هاف (۱۹۵۴)، یول (۱۹۱۲)، ریتز (۱۹۱۸)، داونپورت (۱۹۱۸) مراجعه شود.

۳ متغیرهای تصادفی ناهمبسته ولی وابسته؟

در احتمال و آمار اولین مفهوم مربوط به استقلال عدم همبستگی است. به ساده گی قابل بررسی است که متغیرهای تصادفی مستقل ناهمبسته هستند، اما متغیرهای تصادفی ناهمبسته ممکن است مستقل باشند یا نباشند. در این رابطه مثال های متنوعی در آمار وجود دارد. در مرجع [۴] چند مدل جدید برای توزیع مشترک متغیرهای تصادفی غیرهمبسته ارایه شده است که مستقل نیستند. مدل های پیشنهادی شامل توزیع آمیخته، روش تبدیل و روش نظریه مفصل و تابع مشخصه می باشد. علاوه بر این چندین مثال برای نشان دادن نتایج ارایه شده است.

۴ وابستگی منفی از ۱۹۶۹ تا کنون

ضعف همبستگی در مدل‌سازی داده‌هایی که رفتار غیر خطی دارند موجب ظهور واژه وابستگی و مدل‌سازی داده‌های با رفتار غیر خطی شد. روابط وابستگی بین متغیرهای تصادفی یکی از موضوعات مورد توجه در نظریه احتمال و آمار است. بدون اعمال مفروضات خاص در مورد وابستگی هیچ مدل آماری معنی‌دار نمی‌توان ساخت. همانطور که در ادامه خواهیم دید، مفهوم کوواریانس نقش مهمی در تعریف وابستگی بین متغیرهای تصادفی دارد. به عقیده آماردانان استفاده از واژه "وابستگی آماری" به جای واژه وابستگی مناسب‌تر است.

با استناد به انتشارات موجود به نظر می‌رسد کی پیرسون (۱۹۵۶-۱۸۵۷) یکی از مبتکران مفاهیم و معیارهای وابستگی در روش‌شناسی آماری باشد. همچنین دانشمندانی مانند: جینی (۱۹۶۶-۱۸۸۴) و فرشت (۱۹۷۳-۱۸۷۸) نقش مهمی در توسعه این نظریه در مراحل اولیه آن داشته‌اند. مفاهیم متعددی از وابستگی مثبت و منفی در در نیمه دوم قرن بیستم معرفی و منتشر شده است. این مفاهیم در ابتدا برای دو متغیر تصادفی تعریف شده است و سپس به یک بردار تصادفی چند متغیره با $n > 2$ گسترش یافته است. علیرغم اهمیت وابستگی مثبت در آمار و احتمال، این مفهوم تمام آزمایش‌های تصادفی آمار نظری و کاربردی را پوشش نمی‌دهد. به عنوان مثال، یک توزیع نرمال چند متغیره وقتی همه ضرایب همبستگی زوجی منفی باشد، انتظار نمی‌رود وابستگی مثبتی بین تمام مؤلفه‌ها وجود داشته باشد در صورتیکه این نظریه (وابستگی منفی) برای توزیع‌های دیگر مانند چندجمله‌ای و دیریکله و FGM که خانواده‌هایی از توزیع‌های چند بعدی هستند وجود دارد. برای اطلاعات بیشتر به مراجع این مقاله رجوع شود. کاربرد نظریه وابستگی منفی در مهندسی، پزشکی، هیدرولوژی، بیمه و مالی و سایر علوم که با مدل‌سازی داده‌های با ساختار وابستگی منفی مواجه هستند مورد توجه می‌باشد.

مراجع

- [1] Amini, M., Nili Sani, H.R., and Bozorgnia, A., (2013), *Aspects of negative dependence structures*, Communications in Statistics-Theory and Methods, 42, 907-917.

- [2] Barlow, R. E. and Proschan, F. (1981), *Statistical Theory of Reliability and Life Testing*, To Begin With, Silver Spring.
- [3] Block, H. W., Savits, T. H. and Shaked, M. (1982), *Some concepts of negative dependence*, Annals of Probability, 10, 765–772.
- [4] Dolati, A., Amini, M. and Mohtashami Borzadaran, G.R., (2022), *Some results on uncorrelated dependent random variables*. Volume 11, Issue 3 - Serial Number 23, (Special Issue dedicated to Prof. Mashaallah Mashinchi.). Journal of Mahani Mathematical Research.
- [5] Drouet Mari, D., Kotz, S. , (2001), *Correlation and Dependence*, Imperial College Press, London.
- [6] Ebrahimi, N., and Ghosh, M. , (1981), *Multivariate negative dependence*, Communications in Statistics A(10), 307-337.
- [7] Esary, J.D., Proschan, F. , (1972), *Relationships among some concepts of bivariate dependence*, Ann. Math. Stat., Vol 43, 651-665.
- [8] Esary, J.D., Proschan, F. , Walkup, D.W. , (1967), *Association of random variables with applications*, Ann. Math. Stat., Vol 38, 1466-1474.
- [9] Joag-Dev, K., Proschan, F., *Negative Association of random variables*, Ann. Stat., Vol 11 , 286-295, 1983.
- [10] Joe, H. , (1993), *Parametric Families of Multivariate Distributions with given margins*, J. Multi. Anal. 46, 262-282.
- [11] Joe, H. , (1997), *Multivariate Models and Dependence Concepts*, Chapman & amp; Hall, London.

- [12] Kurowicka, V. , Joe, H. , (2010), *Dependence Modeling: Vine copula handbook*, World Scientific Publishing Company.
- [13] Lehmann, E.L. , (1966), *Some concepts of dependence*, Ann. Math. Stat., 37,1137-1153.
- [14] Mai, J.F. , and Scherer, M. , (2014), *Financial Engineering with Copulas Explained*, First published 2014 by PALGRAVE MACMILLAN.
- [15] Nelsen, R.B. , (2006), *An Introduction to Copulas, second edition*, Springer, New York .
- [16] Rüschendorf, L. , (2009), *On the distributional transform, Sklar's theorem, and the empirical copula process*, Journal of Statistical Planning and Inference, 139(11), pp. 3921–3927.



هفتمین سمینار
نظریه مفصل و کاربردهای آن

۱۹ و ۲۰ بهمن ۱۴۰۱



آزمون استقلال برای خانواده‌های از مفصل‌های آمیخته

زرگر، م^۱ جباری، ه^۲ امینی، م^۳

^۱ گروه آمار، دانشکده علوم ریاضی، دانشگاه ولی عصر (عج) رفسنجان

^{۲،۳} گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده

بایراموف و بایرام اقلو (۲۰۱۳) کلاس جدیدی از توزیع‌های دو متغیره مبتنی بر سیستم بیکر (۲۰۰۸) برای توزیع‌های دو متغیره معرفی نمودند. این کلاس مبتنی بر آماره‌های مرتب توزیع دو متغیره است. در این مقاله آزمون استقلالی برای استقلال دو متغیر تصادفی X و Y در مقابل وابستگی ربعی مثبت معرفی می‌شود. همچنین توان آزمون هم به صورت مجانبی و هم به صورت تجربی از طریق یک مطالعه شبیه سازی مورد ارزیابی قرار می‌گیرد.

کلمات کلیدی: توزیع‌های دو متغیره بایراموف و بایرام اقلو، آماره‌های مرتب، وابستگی ربعی، آزمون استقلال

¹zargar_m@vru.ac.ir

²jabbarinh@um.ac.ir

³m-amini@um.ac.ir

۱ مقدمه

تابع زیر را که از ترکیب دو تابع C_1 و C_2 به دست می‌آید، در نظر بگیرید:

$$C(u, v) = (1 - q)C_1(u, v) + qC_2(u, v), \quad 0 \leq q \leq 1 \quad (1)$$

اگر C_1 و C_2 دو تابع مفصل باشند، طبق خواص تابع مفصل، C نیز مفصل است که به آن مفصل آمیخته گفته می‌شود. تابع چگالی مفصل متناظر C به صورت زیر است:

$$c(u, v) = (1 - q)c_1(u, v) + qc_2(u, v), \quad 0 \leq q \leq 1, \quad (2)$$

که در آن c_1 و c_2 به ترتیب توابع چگالی متناظر با مفصل‌های C_1 و C_2 هستند.

۲ آزمون پیشنهادی

در این مقاله مفصل C_1 را برابر مفصل استقلال در نظر می‌گیریم. بنابراین با ترکیب تابع مفصل استقلال و مفصل دلخواه C_2 مواجه هستیم. در این صورت اگر $q = 0$ باشد، استقلال نتیجه می‌شود. در این جا قرار است فرضیه استقلال را در برابر وابستگی ربعی مثبت (PQD) آزمون کنیم. یعنی

$$\begin{cases} H_0 : q = 0 \\ H_1 : q \geq \eta, \end{cases} \quad (3)$$

که در آن $0 < \eta \leq 1$ پارامتری برای تنظیم شدت وابستگی ربعی مثبت است. برای آزمون فرضیات، فرض می‌کنیم $(U_1, V_1), \dots, (U_n, V_n)$ زوج نمونه‌های تصادفی از تابع مفصل $C(u, v)$ باشد و مفصل C_2 نیز PQD باشد.

قضیه ۱.۰۲. فرض کنید $(U_1, V_1), \dots, (U_n, V_n)$ زوج نمونه‌ای تصادفی از تابع مفصل $C(u, v)$ باشد و $\mathbf{U} = (U_1, \dots, U_n)$ و $\mathbf{V} = (V_1, \dots, V_n)$ باشند. در این صورت آماره آزمون عبارت است از:

$$T(\mathbf{U}, \mathbf{V}) = \prod_{i=1}^n [(1 - \eta) + \eta c_2(u_i, v_i)], \quad (4)$$

که در آن $0 < \eta \leq 1$ است. همچنین

(الف) فرضیه صفر به نفع فرضیه مقابل رد می‌شود اگر و فقط اگر $T(\mathbf{U}, \mathbf{V}) \geq k$.

(ب) تابع توان آزمون نسبت به q غیر نزولی است.

برای آزمون باید توزیع دقیق یا مجانبی آماره آزمون تحت فرضیه صفر مشخص شود. توزیع مجانبی آماره آزمون را در قضیه زیر به دست می‌آوریم.

قضیه ۲.۰۲. فرض کنید زوج نمونه‌ای تصادفی از تابع مفصل $C(u, v)$ باشد. آنگاه $T^n(\mathbf{U}, \mathbf{V})$ به طور مجانبی دارای توزیع نرمال با میانگین $\mu(q) = \exp\{E(\ln T)\}$ و

$$\sigma_n^2(q) = [\mu(q)]^2 \text{Var}(\ln T)/n \text{ است که در آن‌ها } T = 1 - \eta + \eta c_\gamma(U, V).$$

اگر آزمون در سطح معنی‌داری α باشد، طبق قضیه قبل عدد بحرانی به صورت زیر خواهد بود:

$$k. = [\sigma_n(\cdot)\Phi^{-1}(1 - \alpha) + \mu(\cdot)], \quad (5)$$

که در آن $\mu(\cdot)$ و $\sigma^2(\cdot)$ به ترتیب مقادیر $\mu(q)$ و $\sigma_n^2(q)$ تحت فرضیه صفر هستند. همچنین تابع توان آزمون عبارت است از:

$$\gamma_n(q) \simeq 1 - \Phi\left(\frac{k. - \mu(q)}{\sigma_n(q)}\right), \quad q \geq \eta. \quad (6)$$

۳ آزمون استقلال برای خانواده توزیع‌های بایراموف و بایرام اقلو

فرض کنید زوج نمونه‌ای تصادفی از تابع توزیع دو متغیره $H(x, y)$ باشد و $F(x)$ و $G(y)$ به ترتیب توابع توزیع کناری مرتبط باشند. بایراموف و بایرام اقلو (۲۰۱۳) توزیع دو متغیره جدیدی بر مبنای توزیع آماره‌های مرتب X و Y به صورت زیر معرفی کردند:

$$K_+^{(m)}(x, y) = \frac{1}{m} \sum_{r=1}^m Pr(X_{r:m} \leq x, Y_{r:m} \leq y), \quad \forall x, y \in R, \quad (7)$$

که در آن $X_{r:m}$ آماره مرتب r ام X_1, X_2, \dots, X_m و $Y_{r:m}$ آماره مرتب r ام Y_1, Y_2, \dots, Y_m بر اساس زوج نمونه تصادفی (X_i, Y_i) ، $i = 1, 2, \dots, n$ است. برای سادگی $\bar{H}(x, y)$ ، $H(x, y)$ ، $G(y)$ ، $F(x)$

و $K_+^{(m)}(x, y)$ را به ترتیب با نمادهای F, G, H, \bar{H} و $K_+^{(m)}$ نشان می‌دهیم. هونگ و همکاران (۲۰۱۳) نشان دادند که

$$K_+^{(m)} = H + \sum_{j=2}^m \frac{1}{j-1} \sum_{i=1}^{[j/2]} \binom{j-1}{i, i-1, j-2i} W^i V^{j-2i}, \quad m \geq 2, \quad (8)$$

که در آن $V = H + \bar{H}$ و $W = (F - H)(G - H)$ جزء صحیح a است. حال، بر اساس توزیع بایراموف و بایرام اقلو (۲۰۱۳) توزیع دو متغیره جدیدی به صورت زیر معرفی می‌کنیم:

$$K_+ = (1 - q)FG + qK_+^{(m)}, \quad (9)$$

که در آن $0 \leq q \leq 1$ پارامتر وابستگی است. طبق قضیه اسکالر (۱۹۵۱) تابع مفصل متناظر با K_+ عبارت است از:

$$C_+(u, v) = (1 - q)uv + qC_+^{(m)}(u, v) \quad (10)$$

که در آن

$$C_+^{(m)}(u, v) = C(u, v) + \sum_{j=2}^m \frac{1}{j-1} \sum_{i=1}^{[j/2]} \binom{j-1}{i, i-1, j-2i} R^i S^{j-2i}, \quad m \geq 2, \quad (11)$$

و $R = (u - C(u, v))(v - C(u, v))$ و $S = C(u, v) + \bar{C}(u, v)$ و $\bar{C}(u, v) = 1 - u - v + C(u, v)$ و $C(u, v)$ تابع مفصل متناظر با $H(x, y)$ است.

۴ شبیه‌سازی

برای ارزیابی توان مجانبی و تجربی آماره آزمون معرفی شده در (۴) از یک مطالعه شبیه‌سازی استفاده می‌کنیم. برای این منظور سطح معنی‌داری را ۵ درصد در نظر گرفته و برای به دست آوردن توان تجربی، مراحل شبیه‌سازی را ۵۰۰۰ بار به طور مستقل تکرار می‌کنیم. همچنین مفصل $C(u, v)$ در رابطه (۱۱) را FGM در نظر می‌گیریم. نتایج شبیه‌سازی به ازای مقادیر مختلف m ، n و q در جدول ۱ خلاصه شده است.

جدول ۱: توان‌های مجانبی و تجربی (داخل پرانتز) برای $H_1: q \geq 0.2$ و $\theta = 0.5$.

| q | | | | | n | m |
|---------------|---------------|---------------|---------------|--------------|-----|----|
| 1 | 0.75 | 0.5 | 0.25 | 0 | | |
| 0.383 (0.366) | 0.262 (0.270) | 0.166 (0.163) | 0.097 (0.101) | 0.05 (0.051) | 10 | 2 |
| 0.769 (0.780) | 0.545 (0.543) | 0.317 (0.318) | 0.145 (0.143) | 0.05 (0.055) | 30 | |
| 0.926 (0.926) | 0.735 (0.735) | 0.445 (0.459) | 0.187 (0.197) | 0.05 (0.051) | 50 | |
| 0.997 (0.998) | 0.942 (0.945) | 0.685 (0.683) | 0.281 (0.282) | 0.05 (0.057) | 100 | |
| 0.595 (0.587) | 0.407 (0.399) | 0.245 (0.243) | 0.125 (0.129) | 0.05 (0.057) | 10 | 3 |
| 0.948 (0.952) | 0.776 (0.778) | 0.485 (0.482) | 0.205 (0.208) | 0.05 (0.057) | 30 | |
| 0.995 (0.996) | 0.925 (0.930) | 0.659 (0.650) | 0.275 (0.286) | 0.05 (0.058) | 50 | |
| 0.999 (1.000) | 0.996 (0.997) | 0.891 (0.898) | 0.426 (0.425) | 0.05 (0.048) | 100 | |
| 0.812 (0.814) | 0.599 (0.586) | 0.367 (0.367) | 0.171 (0.172) | 0.05 (0.059) | 10 | 5 |
| 0.995 (0.997) | 0.933 (0.943) | 0.687 (0.701) | 0.300 (0.297) | 0.05 (0.057) | 30 | |
| 0.999 (1.000) | 0.991 (0.994) | 0.854 (0.864) | 0.408 (0.405) | 0.05 (0.050) | 50 | |
| 1.000 (1.000) | 0.999 (1.000) | 0.982 (0.984) | 0.619 (0.618) | 0.05 (0.051) | 100 | |
| 0.935 (0.971) | 0.778 (0.792) | 0.528 (0.533) | 0.248 (0.248) | 0.05 (0.065) | 10 | 10 |
| 0.999 (0.999) | 0.986 (0.993) | 0.855 (0.872) | 0.439 (0.432) | 0.05 (0.064) | 30 | |
| 0.999 (1.000) | 0.999 (1.000) | 0.958 (0.970) | 0.584 (0.592) | 0.05 (0.057) | 50 | |
| 1.000 (1.000) | 0.999 (1.000) | 0.998 (0.999) | 0.810 (0.826) | 0.05 (0.058) | 100 | |

نتایج نشان می‌دهند که

الف) توان‌های مجانبی و تجربی نسبت به مقادیر m ، n و q افزایش می‌یابند.

ب) اگر مقادیر m و n بزرگ باشند و q به یک نزدیک شود، توان‌ها به یک می‌رسند.

ج) توان مجانبی آزمون به ازای برخی از مقادیر q برابر یک است.

مراجع

- [1] Bairamov, I., Bayramoglu, K. (2013). From Huang-Kotz FGM distribution to Baker's bivariate distribution, *Journal of Multivariate Analysis* **113**, 106–115.
- [2] Baker, R. (2008). An order-statistics-based method for constructing multivariate distributions with fixed marginals, *Journal of Multivariate Analysis* **99**, 2312–2327.
- [3] Huang, J. S., Dou, X., Kuriki, S., Lin, G. D. (2013). Dependence structure of bivariate order statistics with applications to Bayramoglu's distributions, *Journal of Multivariate Analysis* **114**, 201–208.
- [4] Sklar, A., (1959), Fonctions de repartition a n dimensions et leurs marges. Publications de l'institut de



هفتمین سمینار
نظریه مفصل و کاربردهای آن

۱۹ و ۲۰ بهمن ۱۴۰۱



مدل‌سازی وابستگی نرخ جرم و شاخص فلاکت در ایران: رویکرد مفصل مبنا

خلیلی، ی^۱ دولتی، ع^۲ دست برآورده، ع^۳ رحمانی، ج^۴

^{۱،۲} گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

^۳ گروه آمار، دانشکده علوم ریاضی، دانشگاه یزد

^۴ سازمان مدیریت و برنامه‌ریزی، استان یزد

چکیده

تورم، بیکاری و نابرابری در توزیع درآمدها از عوامل اقتصادی منفی هستند که می‌توانند بر جرم و جنایت در جامعه تاثیرگذار باشند. در ادبیات اقتصاد، به مجموع نرخ تورم و نرخ بیکاری، شاخص فلاکت گفته می‌شود. این مقاله به مدل‌سازی رابطه‌ی دو متغیر شاخص فلاکت و نرخ جرم سرقت، بر اساس داده‌های سال‌های ۱۳۷۸-۱۳۹۴ استان‌های کشور می‌پردازد. نخست یک مدل خودبازگشتی برداری پانلی به داده‌ها برازش داده شده و سپس برای مدل‌سازی وابستگی بین این دو متغیر، از ویژگی‌های توزیعی آن‌ها در قالب یک مدل رگرسیونی برای پیدا کردن یک ساختار وابستگی (مفصل) مناسب استفاده شده است.

¹ yasaman.khalili@um.ac.ir

² alidolati@um.ac.ir

³ dastbaravarde@yazd.ac.ir

⁴

کلمات کلیدی: بیکاری، تورم، شاخص فلاکت، مفصل

۱ مقدمه

بیکاری به عنوان عارضه‌ای که باعث می‌شود از نیروی کار جامعه به‌خوبی استفاده نشود و نرخ تورم که باعث کاهش قدرت خرید بخش‌هایی از جامعه و محرومیت آن‌ها از امکانات می‌شود، از عوامل ایجادکننده فقر و نابرابری اقتصادی در جوامع هستند. در ادبیات اقتصاد، به مجموع نرخ تورم و نرخ بیکاری، شاخص فلاکت گفته می‌شود. در زمینه تاثیر متغیرهایی مانند نرخ بیکاری، نرخ فلاکت، نرخ صنعتی شدن و نرخ طلاق بر میزان جرم و جنایت پژوهش‌های زیادی در داخل و خارج از کشور انجام شده است [۱، ۲، ۳، ۴، ۶، ۷]. یکی از جرم‌هایی که زودتر تحت تاثیر مسائل اقتصادی قرار می‌گیرد و در جامعه نمود پیدا می‌کند، سرقت است. این مقاله به مدل‌سازی رابطه‌ی بین دو متغیر شاخص فلاکت و نرخ جرم سرقت، بر اساس داده‌های سال‌های ۱۳۷۸-۱۳۹۴ استان‌های کشور می‌پردازد. نخست یک مدل خودبازگشتی برداری پانلی به داده‌ها برازش داده شده و سپس برای مدل‌سازی وابستگی بین این دو متغیر، از ویژگی‌های توزیعی آن‌ها در قالب یک مدل رگرسیونی مفصل مبنا مناسب استفاده شده است. با استفاده از قضیه‌ی اسکالر [۵] یک توزیع توأم مناسب برای (X, Y) به صورت $H(x, y) = C(F(x), G(y))$ خواهد بود که F و G توزیع‌های حاشیه‌ای و C مفصل برازش شده به داده‌ها است. تابع چگالی توأم متناظر به صورت

$$h(x, y) = f(x)g(y)c(F(x), G(y))$$

است که در آن f و g توابع چگالی متناظر F و G و c چگالی مفصل است. با استفاده از تابع رگرسیون مفصل مبنا به صورت زیر

$$\begin{aligned} E(Y|X=x) &= \int_0^{\infty} yg(y)c(F(x)G(y))dy \\ &= \int_0^1 G^{-1}(v)c(u, v)dv \end{aligned} \quad (1)$$

رابطه‌ای رگرسیونی برای توصیف دو متغیر نرخ جرم سرقت و نرخ فلاکت به دست می‌آید.

۲ داده‌های مورد استفاده

داده‌هایی که در این مقاله مورد استفاده قرار گرفته است، داده‌های مربوط به انواع جرم سرقت، نرخ تورم و نرخ بیکاری در استان‌های ایران طی سال‌های ۱۳۷۸ تا ۱۳۹۴ است. داده‌های مربوط به نرخ تورم از سایت بانک مرکزی و داده‌های مربوط به نرخ بیکاری و سرقت از فصل‌های «نیروی انسانی» و «امور قضایی» سالنامه‌های آماری مرکز آمار گرفته شده‌اند. شاخص فلاکت به صورت مجموع نرخ تورم و نرخ بیکاری و شاخص جرم سرقت با توجه به روش ارائه شده در [۱]، محاسبه شده است.

۳ دست‌آوردهای پژوهش

برای مدل‌سازی ارتباط متغیر نرخ جرم سرقت و شاخص فلاکت از لگاریتم نرخ جرم به عنوان متغیر پاسخ و شاخص فلاکت به عنوان متغیر توضیحی در مدل‌ها استفاده شده است. مدل‌های رگرسیونی و سری‌زمانی زیادی به داده‌ها برازش داده شد که از بین آن‌ها مدل خودبازگشتی برداری پانلی مناسب‌تر بود. از باقی‌مانده‌های این مدل برای پیدا کردن ساختار وابستگی (مفصل) مناسب استفاده شده است.

۱۰۳ آزمون خودهمبستگی داده‌ها

برای بررسی خودهمبستگی در داده‌های پانلی، از آزمون خودهمبستگی بریچ-گادفری استفاده شده است. مقدار آماره‌ی به دست آمده برای این آزمون برابر با ۲۴۶٫۲۲ و مقدار احتمال به دست ۲٫۲e-۱۶ است که وجود خودهمبستگی در داده‌ها را تایید می‌کند.

۲.۳ مدل خودبازگشتی برداری پانلی

مدل خودبازگشتی برداری پانلی برای ارتباط متغیرهای نرخ جرم سرقت (CR) و شاخص فلاکت (MI) به صورت زیر در نظر گرفته شده است

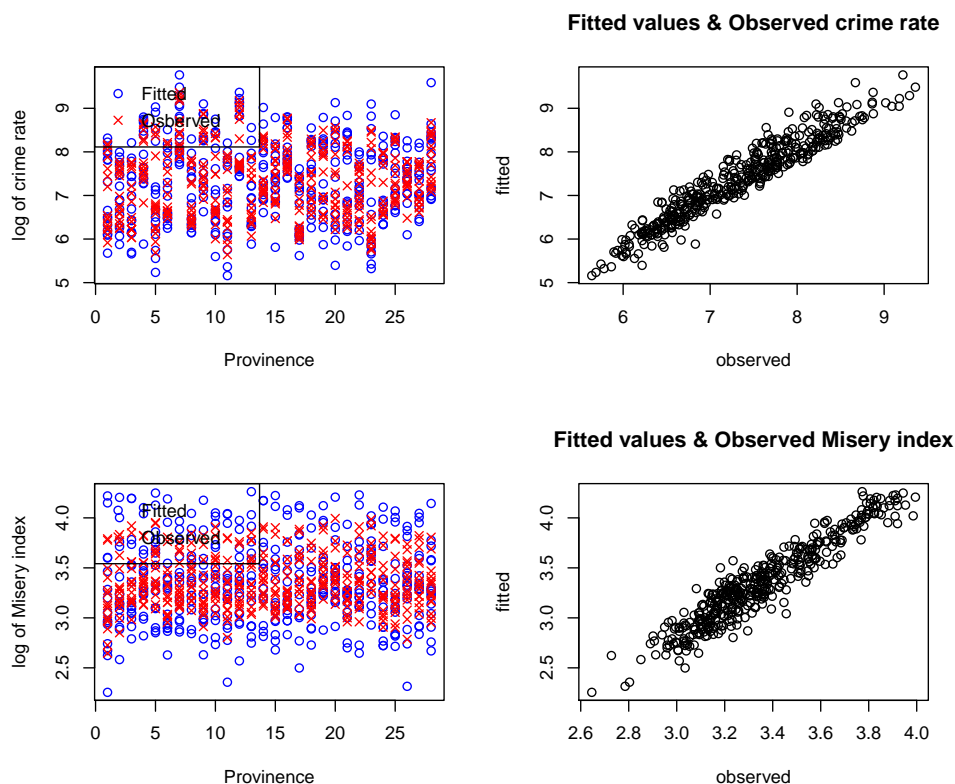
$$\begin{cases} \log(CR_{it}) = \alpha + \alpha_1 \log(CR_{it-1}) + \alpha_2 \log(CR_{it-2}) + \alpha_3 \log(MI_{it-1}) + \alpha_4 \log(MI_{it-2}) + \varepsilon_{it}, \\ \log(MI_{it}) = \beta + \beta_1 \log(CR_{it-1}) + \beta_2 \log(CR_{it-2}) + \beta_3 \log(MI_{it-1}) + \beta_4 \log(MI_{it-2}) + \varepsilon_{it}, \end{cases} \quad (2)$$

که در آن $i = 1, \dots, 28$ تعداد استان‌ها و $t = 1, \dots, 17$ اندیس زمان (سال) است. نتایج برازش مدل در جدول ۱ گزارش شده است.

جدول ۱: نتایج برآورد مدل خودبازگشتی برداری پانلی

| متغیرهای وابسته | | متغیرهای مستقل | |
|------------------------------------|------------------------------------|--|--------------------------|
| $\log(MI)$ | $\log(CR)$ | | |
| ۲,۳۰۴۶ (۰,۱۶۲۱۶) [۱۴,۲۱۱۷] | -۰,۳۷۲۲ (۰,۲۳۸۳۱) [-۱,۵۶۱۹۰] | ضریب برآورد شده خطای استاندارد آماره‌ی t | عرض از مبدا |
| ۰,۱۷۳۶ (۰,۰۳۲۰۳) [۵,۴۲۱۶۷] | ۰,۹۶۷۹ (۰,۰۴۷۰۷) [۲۰,۵۶۲۰] | ضریب برآورد شده خطای استاندارد آماره‌ی t | $\log(CR_{it-1})$ |
| -۰,۱۳۸۸ (۰,۰۳۳۷۱) [-۴,۱۱۹۹۶] | -۰,۰۶۸۶ (۰,۰۴۹۵۴) [-۱,۳۸۵۰۸] | ضریب برآورد شده خطای استاندارد آماره‌ی t | $\log(CR_{it-2})$ |
| ۰,۶۳۷۱ (۰,۰۴۴۴۶) [۱۴,۳۳۰۳] | ۰,۱۳۶۳ (۰,۰۶۵۳۴) [۲,۰۸۷۰۳] | ضریب برآورد شده خطای استاندارد آماره‌ی t | $\log(MI_{it-1})$ |
| -۰,۴۰۱۵ (۰,۰۴۴۲۰) [-۹,۰۸۴۹۵] | ۰,۲۱۵۱ (۰,۰۶۴۹۵) [۳,۳۱۱۷۸] | ضریب برآورد شده خطای استاندارد آماره‌ی t | $\log(MI_{it-2})$ |
| ۰,۳۷۷ -۰,۳۰۵ -۰,۲۵۷ ۶۲,۸۱ | ۰,۸۵۱ ۰,۴۶۴ ۰,۵۱۲ ۵۹۶,۷۹ | | R^2 AIC BIC F |
| AIC=۰,۱۵۷ | | BIC= ۰,۲۵۳ | |

شکل ۱، نمودار مقادیر مشاهده شده و مقادیر پیش‌بینی شده براساس مدل خودبازگشتی برداری پانلی را نشان می‌دهد. با توجه به نمودارها و نتایج جدول ۱ این مدل، برازش مناسبی برای داده‌ها است.



شکل ۱: نمودار مقادیر مشاهده‌شده و پیش‌بینی شده براساس مدل خودبازگشتی برداری پانلی

۳.۳ برازش توزیع‌های حاشیه‌ای

برای برازش توزیع حاشیه‌ای مناسب به داده‌ها، توزیع‌های گاما، نرمال، لگ‌نرمال، وایبل و گمبل به داده‌ها برازش داده شده‌اند. شاخص‌های AIC، BIC، لگاریتم تابع درست‌نمایی ($\log L$)، مقدار آماره آزمون کولموگروف-اسمیرنف و p -value برای انتخاب توزیع حاشیه‌ای مناسب در جدول ۲ آورده شده است. نتایج جدول ۲ نشان می‌دهد که توزیع حاشیه‌ای گاما برای لگاریتم متغیر

جدول ۲: برازش توزیع حاشیه‌ای

| p-value | K-S | logL | BIC | AIC | پارامترها | توزیع | متغیر |
|-----------------|--------|---------|---------|---------|--|----------|-----------------------|
| ۰/۰۷۶۸ | ۰/۰۵۸۵ | -۵۵۱/۸۱ | ۱۱۱۵/۹۵ | ۱۱۰۷/۶۲ | $\hat{\alpha} = ۸۷/۷۰, \hat{\beta} = ۱۲/۰۹$ | گاما | لگاریتم شاخص جرم سرقت |
| ۰/۰۲۷ | ۰/۰۶۶ | -۵۵۵/۴۳ | ۱۱۲۳/۱۹ | ۱۱۱۴/۸۶ | $\hat{\mu} = ۷/۲۵, \hat{\sigma} = ۰/۷۷$ | نرمال | |
| ۰/۰۳۹ | ۰/۰۶۴ | -۵۵۰/۹۹ | ۱۱۱۴/۳۲ | ۱۱۰۵/۹۹ | $\hat{\mu} = ۱/۹۷, \hat{\sigma} = ۱/۰۶$ | لگ-نرمال | |
| ۰/۰۰۷ | ۰/۰۷۶ | -۵۷۹/۳۶ | ۱۱۷۱/۰۶ | ۱۱۶۲/۷۲ | $\hat{\alpha} = ۹/۸۱, \hat{\beta} = ۷/۶۰۴$ | وایبل | |
| ۰/۰۰۹ | ۰/۰۷۴ | -۵۵۷/۸۸ | ۱۱۲۸/۱۰ | ۱۱۱۹/۷۷ | $\hat{\alpha} = ۶/۸۷, \hat{\beta} = ۰/۶۸$ | گمبل | |
| ۰/۰۷۸ | ۰/۰۵۸ | -۲۰/۹۹ | ۵۴/۳۲ | ۴۵/۹۹ | $\hat{\alpha} = ۱۷۵/۸۶, \hat{\beta} = ۵۲/۳۴$ | گاما | لگاریتم شاخص فلاکت |
| ۰/۱۴۸ | ۰/۰۵۲ | -۲۴/۲۵ | ۶۰/۸۳ | ۵۲/۵۰ | $\hat{\mu} = ۳/۳۵, \hat{\sigma} = ۰/۲۵$ | نرمال | |
| ۰/۰۰۳ | ۰/۰۸۱ | -۱۹/۹۳ | ۵۲/۲۰ | ۴۳/۸۷ | $\hat{\mu} = ۱/۲۰, \hat{\sigma} = ۰/۰۷$ | لگ-نرمال | |
| $۳/۹۰۵e^{-۰۰۵}$ | ۰/۱۰۶ | -۵۷/۷۵ | ۱۲۷/۸۴ | ۱۱۹/۵۱ | $\hat{\alpha} = ۱۳/۵۲, \hat{\beta} = ۳/۴۸$ | وایبل | |
| ۰/۱۷۸ | ۰/۰۵۰ | -۳۰/۶۱ | ۷۳/۵۶ | ۶۵/۲۳ | $\hat{\alpha} = ۳/۲۳, \hat{\beta} = ۰/۲۲$ | گمبل | |

شاخص جرم سرقت و توزیع حاشیه‌ای گمبل، برای لگاریتم متغیر شاخص فلاکت، توزیع‌های مناسبی هستند.

۴.۳ انتخاب مفصل مناسب برای باقی‌مانده‌ها

با توجه به وجود خودهمبستگی در داده‌ها، از باقی‌مانده‌های این مدل برای یافتن ساختار وابستگی (مفصل) مناسب استفاده شده است. مفصل‌های نرمال، تی استیودنت، گامبل، فرانک و جو [۵] به باقیمانده‌ها برازش داده شده‌اند. جدول (۳) نتایج آزمون نیکویی برازش با آماره‌ی کرامر-ون-میسز (S_n) و مقدار احتمال آن برای مفصل‌های بیان شده را نشان می‌دهد. با توجه به نتایج

جدول ۳: نتایج آزمون نیکویی برازش برای انتخاب مفصل مناسب

| p-value | آماره S_n | BIC | AIC | پارامتر برآورد شده | مفصل |
|---------|-------------|-------|-------|--------------------|-------|
| ۰/۰۱۳ | ۰/۰۴۲۳ | ۱/۴۴ | -۲/۶ | ۰/۱۱ | نرمال |
| ۰/۰۱۱ | ۰/۰۴۱۳ | ۶/۳ | -۱/۷۸ | ۰/۱۱ | تی |
| ۰/۱۴۲ | ۰/۰۲۸۹ | -۱/۶۸ | -۵/۷۲ | ۱/۰۹ | گامبل |
| ۰/۰۱۶ | ۰/۰۳۶۳ | -۰/۰۹ | -۴/۱۳ | ۰/۷۴ | فرانک |
| ۰/۲۳۱ | ۰/۰۲۸۷ | -۲/۴۴ | -۶/۴۸ | ۱/۱۴ | جو |

جدول ۳، مقادیر احتمال، مفصل‌های گامبل و جو مفصل‌های مناسبی برای ساختار وابستگی

دادها هستند:

• مفصل گامبل: $\theta \geq 1$: $C_\theta(u, v) = \exp(-[(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}})$

• مفصل جو: $\theta \geq 1$: $C_\theta(u, v) = 1 - ((1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta)^{\frac{1}{\theta}}$

با توجه به مقادیر پارامتر، مفصل‌های گامبل و جو، زمانی که پارامتر آن‌ها برابر یک باشد ($\theta = 1$)، حالت استقلال را نشان می‌دهند. بنابراین این مفصل‌ها، با توجه به مقادیر پارامتری که دارند، به استقلال نزدیک هستند.

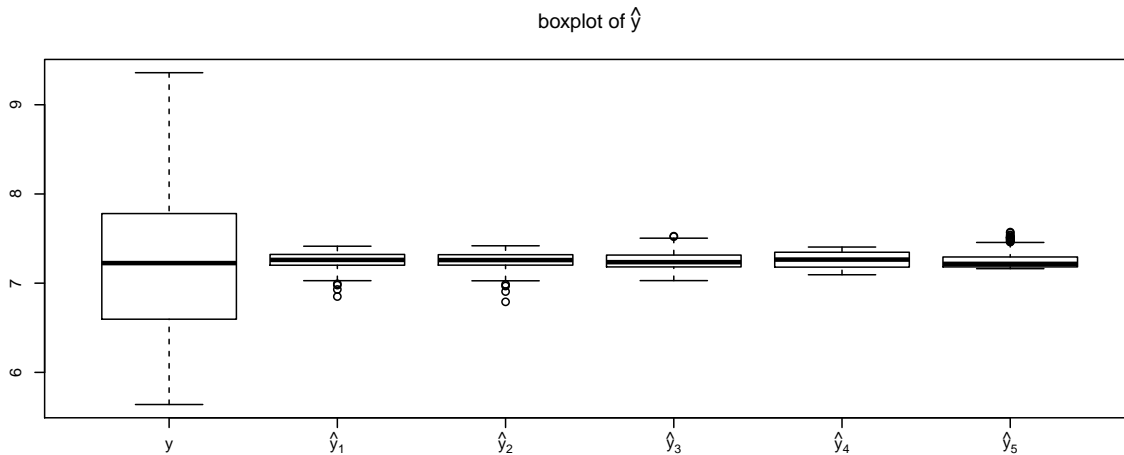
۵.۳ تابع رگرسیون مفصل مبنا

با قرار دادن مفصل‌های نرمال، تی، گامبل، فرانک و جو در رابطه‌ی (۱) تابع رگرسیون محاسبه شده است و برآورد امید ریاضی Y به شرط $X = x$ (\hat{y})، برای هر یک از مفصل‌ها به دست آمده است. جدول ۴ خلاصه آماری مقادیر امید ریاضی به دست آمده برای مفصل‌های مختلف را نشان می‌دهد. شکل ۲، نمودار جعبه‌ای توابع رگرسیون برآورد شده برای هر یک از مفصل‌های بیان

جدول ۴: خلاصه آماری مقادیر برآورد شده برای تابع رگرسیون

| مفصل | \hat{y} | Min | Q_1 | Med | Mean | Q_3 | Max |
|-------|-------------|-------|-------|-------|-------|-------|-------|
| دادها | y | ۵٫۶۴۰ | ۶٫۵۹۹ | ۷٫۲۲۵ | ۷٫۲۵۱ | ۷٫۷۷۸ | ۹٫۳۵۹ |
| نرمال | \hat{y}_1 | ۶٫۸۵۰ | ۷٫۲۰۱ | ۷٫۲۶۱ | ۷٫۲۵۶ | ۷٫۳۲۲ | ۷٫۴۱۴ |
| تی | \hat{y}_2 | ۶٫۷۹۱ | ۷٫۲۰۲ | ۷٫۲۵۹ | ۷٫۲۵۵ | ۷٫۳۱۹ | ۷٫۴۱۹ |
| گامبل | \hat{y}_3 | ۷٫۰۲۹ | ۷٫۱۸۲ | ۷٫۲۳۵ | ۷٫۲۵۳ | ۷٫۳۱۵ | ۷٫۵۲۷ |
| فرانک | \hat{y}_4 | ۷٫۰۹۴ | ۷٫۱۸۰ | ۷٫۲۶۴ | ۷٫۲۵۹ | ۷٫۳۴۷ | ۷٫۴۰۵ |
| جو | \hat{y}_5 | ۷٫۱۶۳ | ۷٫۱۸۲ | ۷٫۲۱۶ | ۷٫۲۵۰ | ۷٫۲۹۴ | ۷٫۵۷۲ |

شده را نشان می‌دهد. با توجه به نتایج به دست آمده که در جدول ۳، ساختار وابستگی شاخص فلاکت و نرخ جرم سرقت به استقلال نزدیک است. به همین دلیل دور از انتظار نیست که مقادیر برازش شده تحت مفصل‌های مختلف، میانگین‌های برابری داشته باشند. نتایج جدول ۴ نشان می‌دهد که امید ریاضی شرطی نرخ جرم سرقت به شرط شاخص فلاکت (تابع رگرسیون)، برای مفصل‌های مختلف، به امید ریاضی لگاریتم نرخ جرم سرقت (۷٫۲۵۱) نزدیک است. بنابراین



شکل ۲: نمودار جعبه‌ای مقادیر برآورد شده برای تابع رگرسیون

نمی‌توان مدل رگرسیونی مناسبی بر اساس مفصل برای وابستگی آن‌ها پیشنهاد نمود.

نتیجه‌گیری

برای مدل‌سازی وابستگی میان متغیر جرم و متغیرهای کلان اقتصادی، پژوهش‌های زیادی انجام شده است که به برخی از آن‌ها اشاره شد. این مطالعات نشان داده‌اند که میان جرم و متغیرهای نرخ تورم و بیکاری وابستگی وجود دارد. در این پژوهش برخی از مدل‌های استفاده شده در پژوهش‌های قبلی روی داده‌های ایران، مورد بازبینی قرار گرفت. مدل‌های رگرسیون پانلی، آنالیز کواریانس، مدل خطی تعمیم‌یافته با پاسخ گاما و مدل خودبازگشتی برداری پانلی به داده‌ها برازش داده شد. نتایج این مدل‌ها نشان‌دهنده‌ی رابطه‌ی بسیار ضعیف بین دو متغیر شاخص فلاکت و جرم سرقت بودند که در پژوهش‌های انجام شده به آن اشاره‌ای نشده بود. به نظر می‌رسد پژوهشگران بعد از برازش مدل، به بررسی فرض‌های روی مدل‌ها و بررسی مناسب بودن مدل‌ها نپرداخته‌اند. با توجه به این‌که فرض اساسی روی مدل‌های استفاده شده قبلی، نرمال بودن خطاها است و این فرض برقرار نیست، استفاده از راهبرد مفصل قابل اعتمادتر است. سپس با استفاده از

تابع مفصل مدل‌سازی صورت گرفته است که طبق نتایج به نظر می‌رسد میان این دو متغیر وابستگی وجود ندارد و از هم مستقلند. این نتایج با توجه به آمار موجود در نهادهای رسمی به دست آمده است. به نظر می‌رسد در مطالعه‌ی رابطه‌ی بین نرخ جرم و شاخص فلاکت، باید به متغیرهای دیگری مانند نرخ بهره بانکی، تولید ناخالص داخلی، شاخص صنعتی شدن، شاخص شهرنشینی و مانند آن‌ها که می‌توانند به بهبود مدل‌سازی کمک کنند، توجه شود. این تحقیق گام کوچکی در جهت بررسی ارتباط متغیر شاخص فلاکت و جرم سرقت بود که در ادامه انتظار می‌رود با رفع موانع یاد شده، زمینه برای انجام تحقیقات علمی پژوهشی در مورد این موضوع فراهم گردد. از محدودیت‌های این پژوهش که در واقع به نوعی به ضعف مدیریتی در زمینه‌ی جمع‌آوری اطلاعات و داده‌های مناسب در سطح کشوری برمی‌گردد، عبارتند از ۱- کامل نبودن داده‌ها به طوری که در برخی سال‌ها نرخ جرم در سالنامه‌های آماری ثبت نشده است. ۲- تغییر تقسیم بندی استان‌ها که همسان سازی داده‌ها با تغییرات در تقسیم بندی استان‌ها صورت نگرفته است. برای مثال استان خراسان بعد از سال ۱۳۸۴ به سه استان تقسیم شده اما قبل از این سال، آمار دقیق، به تفکیک برای این سه استان وجود ندارد. ۳- تغییر عناوین جرائم در سال‌های مختلف که باعث محدودیت در انتخاب سری زمانی مورد بررسی شده است. لازم است برای دوری از ابهام‌ها در مورد عناوین جرم‌ها، از سوی مراجع مربوطه یکسان‌سازی صورت بگیرد.

مراجع

- [۱] جعفری، محمد. بررسی عوامل اقتصادی جرم در مناطق کشور با تاکید بر استان سیستان و بلوچستان طی سال‌های ۱۳۸۸-۱۳۹۴. پایان‌نامه کارشناسی ارشد علوم اقتصادی، دانشگاه یزد، ۱۳۹۸.
- [۲] دادگر، یداله. و نظری، روح‌اله. بررسی تأثیر شاخص فلاکت بر جرم و جنایت در ایران. نشریه مطالعات و سیاست‌های اقتصادی، ش ۹۹، ۶۳-۸۶، ۱۳۹۲.
- [۳] دادگر، یداله. و نظری، روح‌اله. بررسی جرم و جنایت در ایران با استفاده از چند الگوی اقتصادی. فصلنامه تحقیقات حقوقی، ش ۷۳، ۵۹-۷۹، ۱۳۹۴.

[۴] عباسی نژاد، حسین. بررسی رابطه‌ی بین بیکاری و جرم در ایران: رهیافت داده‌های تلفیقی بین استانی. فصلنامه پژوهش‌ها و سیاست‌های اقتصادی، ش ۶۴، ۶۵-۸۶، ۱۳۹۱.

[5] Nelsen, R.B. (2006). *An Introduction to Copulas*. Second Edition. Springer, New York.

[6] Rosenfeld, R. and Levin, A. Acquisitive crime and inflation in the United States: 1960–2012. *Journal of quantitative criminology*, **32**(3) (2016), 427-447.

[7] Speziale, N. Does unemployment increase crime? Evidence from Italian provinces. *Applied Economics Letters*, **21**(15) (2014), 1083-1089.



برآورد کسری مورد انتظار پرتفوی بر اساس تابع مفصل و رویکرد

POT – GARCH

علیزاده، ف^۱ محتمشی برزادران، غ^۲

^{۱،۲} گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده

ریسک یکی از مفاهیم پایه ای در بازارهای مالی می باشد و اندازه گیری و تحلیل آن اهمیت بسزایی دارد. در این مقاله برآورد یکی از مهم ترین روش های اندازه گیری ریسک یعنی کمبود مورد انتظار در یک پرتفوی شامل دو دارایی وابسته بر اساس تابع مفصل و استفاده از رویکرد فراتر از آستانه مطالعه می شود و سپس آن را در یک پرتفوی شامل دو دارایی از سهام شرکت *IBM* و *SP* برآورد می کنیم.

کلمات کلیدی: ارزش در معرض خطر، کسری مورد انتظار، رویکرد فراتر از آستانه، مدل گارچ، تابع مفصل

۱ مقدمه

با گذشت زمان به دلیل گسترش حوادث نامطلوب مختلف در جهان که بخشی از آن از افزایش فعالیت های اقتصادی، اجتماعی، سیاسی و ... نشئت می گیرد، نااطمینانی درباره آینده بیشتر شده

¹fateme.alizade301073@gmail.com

²grmohtashami@um.ac.ir

است. ریسک یکی از مفاهیم پایه‌ای در بازارهای مالی است که تعریف واحدی از آن وجود ندارد. اقتصاددانان، دانشمندان علوم رفتاری، نظریه پردازان ریسک و آمارشناسان هر یک تصور خاصی از ریسک دارند، البته به طور سنتی ریسک به عنوان عدم اطمینان تعریف شده است و بر اساس این مفهوم، در این جا ریسک به عنوان عدم اطمینان مرتبط با وقوع یک خسارت و یا زیان تعریف می‌شود.

در سال‌های اخیر با بررسی علل نوسانات مالی مشخص می‌شود که عوامل مختلفی در ایجاد نوسانات در بازارهای مالی (بیمه‌ها، بانک‌ها و ...) مؤثر بوده‌اند از جمله بحران نفتی سال ۱۹۷۳، بحران‌های مالی سال‌های ۱۹۹۷-۱۹۹۸ در جنوب شرق آسیا و رویدادهای طبیعی نظیر زلزله و سونامی. بروز چنین حوادثی و به تبع آن نوسانات بازارهای مالی موجب ایجاد اختلال در فعالیت‌های سازمان‌های مالی، تجاری و حتی تولیدی می‌شود. از این رو اندازه‌گیری و ارزیابی ریسک اهمیت پیدا می‌کند و از جمله مهم‌ترین کاربردهای آن می‌توان به قیمت‌گذاری سهام‌ها، تعیین حق بیمه متناسب با ریسک و تعیین سرمایه اقتصادی^۱ شرکت که پشتوانه‌ی فعالیت‌های مالی است و خسارت‌های ناگهانی بالاتر از حد انتظار را پوشش می‌دهد، اشاره کرد.

۲ کسری مورد انتظار

یکی از مهم‌ترین معیارهای اندازه‌گیری ریسک ارزش در معرض خطر^۲ است. این شیوه اندازه‌گیری را ابتدا تیل گلدیمن^۳ در سال ۱۹۸۰ ارائه کرد و در اواخر دهه ۱۹۸۰ توسط موسسه جی-پی مورگان^۴ گسترش یافت. این شاخص حداکثر زیان ممکن را در یک افق زمانی مشخص با توجه به یک سطح احتمال معین α بیان می‌کند و با نماد VaR_α نشان داده می‌شود، به عنوان مثال ارزش در معرض خطر با سطح احتمال ۰/۰۵ برای یک بازه زمانی یک روزه گویای اینست که حداکثر زیان احتمالی طی روز بعدی با احتمال ۵ درصد از مقدار ارزش در معرض خطر بالاتر می‌رود. این معیار را می‌توان با استفاده از سری بازدهی محاسبه کرد. فرض کنید $\{p_t\}$ سری

¹Economic Capital

²Value at Risk

³Till Guldemann

⁴J.P. Morgan

زمانی بهای نوعی دارایی مالی باشد، بازدهی در زمان t ام به صورت $X_t = \log(p_t) - \log(p_{t-1})$ تعریف می‌شود.

تعریف ۱.۰۲. اگر $\{X_t\}$ یک سری بازدهی با تابع توزیع F_{X_t} باشد، آنگاه

$$VaR_\alpha(X_t) = F_{X_t}^{-1}(\alpha) = \inf\{x; F_{X_t}(x) \geq \alpha\}; \alpha \in (0, 1).$$

اگر فرض کنیم مقدار سرمایه یک شرکت دقیقاً برابر با ارزش در معرض خطر باشد، زیان بزرگتر از آن احتمال عدم توانگری شرکت را بیشتر می‌کند از این رو VaR نقطه عدم توانگری شرکت را نشان می‌دهد. بررسی زیان‌هایی که یک شرکت را در وضعیت عدم توانگری قرار می‌دهد بسیار مهم است ولی شاخص VaR این موضوع را نشان نمی‌دهد به این معنا که هیچگونه اطلاعاتی راجع به سنگینی دم توزیع نمی‌دهد بنابراین این مساله می‌تواند به عنوان یک نقطه ضعف در نظر گرفته شود، علاوه بر این یکی دیگر از نقاط ضعف این معیار به دلیل عدم زیر جمعی بودن منسجم نبودن آن است، از این رو معیاری به نام کمبود مورد انتظار^۵ معرفی شد که این نقاط ضعف را پوشش می‌دهد و به صورت زیر تعریف می‌شود.

$$ES_{1-\alpha}(X) = E(X | X \geq VaR_{1-\alpha}(X)); \alpha \in (0, 1). \quad (1)$$

۳ مدل بندی تغییر پذیری

مدل بندی تغییر پذیری که اغلب به صورت واریانس تعریف می‌شود برای پیش بینی نوسانات آینده یکی از مباحث مهم در مطالعات مالی می‌باشد. اگر $\{X_t\}$ سری زمانی مورد نظر باشد واریانس شرطی X_t به شرط مقادیر گذشته، عدم حتمیت در انحراف X_t از میانگین شرطی‌اش را اندازه می‌گیرد. اگر X_t از نوعی مدل ARIMA پیروی کند واریانس شرطی به ازای هر مقداری از حال و گذشته فرایند همیشه برابر با واریانس نوفه است. در عمل، واریانس شرطی ممکن است با مقدارهای جاری و گذشته فرایند تغییر کند و از این رو واریانس شرطی خود فرایند تصادفی

⁵Expected shortfall

است که اغلب با نام فرایند واریانس شرطی از آن یاد می‌شود. برای مثال، اغلب مشاهده می‌شود که بازدهی روزانه (ماهانه) سهام بعد از یک دوره تغییرات شدید بها، واریانس شرطی بزرگتری نسبت به یک دوره نسبتاً پایدار دارند. مدل‌های ARCH و GARCH^۶ برای فرایند واریانس شرطی که با آن بتوانیم تغییر پذیری مقادیر آینده (σ_t) را بر اساس داده‌های جاری و گذشته پیشگویی کنیم،

الف. مدل ARCH(q):

[۵] ابتدا مدل ناهم واریانسی شرطی اتورگرسیو (ARCH) را برای مدل بندی واریانس تغییر یابنده سری زمانی پیشنهاد کرد. واریانس شرطی با $\sigma_{t|t-1}^2$ نمایانده می‌شود و اندیس $t-1$ به معنی شرطی کردن نسبت به بازدهی‌های گذشته تا زمان $t-1$ است. مدل ARCH از لحاظ صوری مدلی رگرسیونی با واریانس شرطی به عنوان متغیر پاسخ و توان دوم بازدهی یا زیان تاخیرهای گذشته به عنوان متغیرهای کمکی است. در مدل ARCH(q) فرض می‌شود که سری بازدهی $\{X_t\}$ به صورت

$$X_t = \mu + a_t, \quad a_t = \sigma_{t|t-1} \varepsilon_t, \quad (2)$$

$$\sigma_{t|t-1}^2 = \omega + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \dots + \alpha_q a_{t-q}^2. \quad (3)$$

تولید می‌شود که در آن پارامترهای α_i و ω مجهولند، $\{\varepsilon_t\}$ که از آن به عنوان نوآوری‌ها یاد می‌شود، دنباله‌ای از متغیرهای تصادفی مستقل و هم توزیع با میانگین صفر و واریانس واحد است و ε_t برای $j = 1, 2, \dots$ مستقل از a_{t-j} است.

ب. مدل GARCH(p, q):

رویکردی دیگر که توسط [۲] و [۱۲] پیشنهاد شده، علاوه بر q تاخیر از توان دوم بازدهی، p تاخیر از واریانس شرطی را هم وارد مدل می‌کند. یعنی،

$$\sigma_{t|t-1}^2 = \omega + \alpha_1 a_{t-1}^2 + \dots + \alpha_q a_{t-q}^2 + \beta_1 \sigma_{t-1|t-2}^2 + \dots + \beta_p \sigma_{t-p|t-p-1}^2. \quad (4)$$

مدل مرکب، مدل ناهم واریانسی شرطی اتورگرسیو تعمیم یافته (GARCH) نامیده می‌شود. (برای جزئیات بیشتر در مورد سری‌های زمانی به [۱] مراجعه شود.)

⁶Generalized Auto regressive Conditional Heteroscedasticity

۴ تابع مفصل

در سال های اخیر استفاده از توابع مفصل^۷ به عنوان ابزاری مفید برای مدل سازی ساختار وابستگی بین دو یا چند متغیر گسترش یافته است و کاربرد بسیاری در علوم بیمه و اقتصاد دارد. مفصل ها توابعی هستند که بین توابع توزیع چند متغیره و توزیع حاشیه ای آن ها ارتباط برقرار می کنند. از طرفی مفصل خود تابع توزیعی است که توزیع حاشیه ای آن از توزیع یکنواخت استاندارد پیروی می کند.

تعریف ۱.۰۴. تابع $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ را یک تابع مفصل نامند هرگاه،
۱. برای هر $u, v \in [0, 1]$ ؛

$$C(u, 0) = C(0, v) = 0, \quad C(u, 1) = u, \quad C(1, v) = v.$$

۲. و برای هر $u_1, u_2, v_1, v_2 \in [0, 1]$ به طوری که $u_1 \leq u_2, v_1 \leq v_2$ داشته باشیم:

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

قضیه ۱.۱ که با نام قضیه اسکالار [۱۱] معروف است رابطه بین توزیع های حاشیه ای، تابع مفصل و توزیع توام دو متغیر تصادفی وابسته را بیان می کند.

قضیه ۱.۰۴. اگر (X, Y) یک بردار تصادفی با تابع توزیع توام $F(., .)$ و توابع توزیع حاشیه ای F_X و F_Y باشند آنگاه مفصل $C : I^2 \rightarrow I$ وجود دارد به طوریکه:

$$F(x, y) = C(F_X(x), F_Y(y)) ; \forall x, y \in R.$$

اگر F_X و F_Y پیوسته باشد آنگاه C یکتاست در غیر این صورت به صورت یکتا روی $Range F_X \times Range F_Y$ تعیین می شود و با داشتن توزیع های حاشیه ای و توام برای یافتن مفصل متناظر قرار می دهیم: $u = F_X(x)$ ، $v = F_Y(y)$ و داریم:

$$C(u, v) = F(F_X^{-1}(u), F_Y^{-1}(v)) ; \forall u, v \in [0, 1].$$

⁷Copula Function

برای مثال یکی از توابع مفصلی که در این تحقیق نیز استفاده شده است تابع مفصل تی-استودنت با پارامترهای θ و ν به صورت زیر می باشد:

$$C(u, v) = \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(v)} \frac{1}{\sqrt{2\pi}\sqrt{1-\theta^2}} \left(1 + \frac{s^2 + t^2 - 2\theta st}{u(1-\theta^2)}\right) ds dt; \nu > 0, \theta \in [0, 1].$$

(برای جزئیات بیشتر در مورد توابع مفصل به [۸] مراجعه شود.)

۵ برآورد کسری مورد انتظار

برای برآورد ارزش در معرض خطر رویکردهای پارامتری، ناپارامتری و نیمه پارامتری مختلفی ارائه شده است. یکی از ویژگی های آماری توزیع داده های سود و زیان دم پهنی آن می باشد، برای در نظر گرفتن ویژگی دم پهنی دربرآورد ارزش در معرض خطر پژوهشگران نظریه ارزش حدی^۸ را معرفی کردند (مک نیل (۱۹۹۷)). یکی از رویکردهای این نظریه رویکرد فراتر از آستانه^۹ یا POT می باشد. که به بررسی دم توزیع می پردازد و توزیع حدی مقادیر فراتر از آستانه را به دست می آورد. برای تعیین مقدار آستانه u نیز روش های مختلف از جمله نمودار هیل وجود دارد ولی در بررسی داده های این مقاله مقدار u را صدک نود و پنجم داده ها در نظر میگیریم. اگر تابع توزیع تجمعی مقادیری که از آستانه بیشتر هستند که با نماد $F_u(y)$ نشان دهیم طبق قضیه ی بالکما و هان (۱۹۷۲) می توان آن را با توزیع پارتوی تعمیم یافته^{۱۰} تقریب زد یعنی با بزرگتر شدن مقدار آستانه توزیع مقادیر فراتر از آستانه به توزیع پارتوی تعمیم یافته که با دو پارامتر ξ و β که تابع توزیع آن به فرم زیر است نزدیک می شود.

$$G_{\xi, \beta, u}(y) = \begin{cases} 1 - (1 + \xi(\frac{y}{\beta}))^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp(-\frac{y}{\xi}) & \xi = 0. \end{cases} \quad (5)$$

با محاسبات جبری ساده برآورد ارزش در معرض خطر بر اساس رویکرد POT به صورت زیر

⁸Extreme value theory

⁹Peaks over thresholds

¹⁰Generalized pareto distribution

می باشد

$$VaR_{(1-\alpha)}(X) = \widehat{F}^{-1}(1-\alpha) = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left[\left(\frac{n}{n_u} \alpha \right)^{-\xi} - 1 \right]. \quad (6)$$

در صورتی که سری زمانی زیان از یک مدل $GARCH(p, q)$ پیروی کند از این که $X_t = \mu + \sigma_{t|t-1}^2 \varepsilon_t$ نتیجه میگیریم:

$$VaR_{(1-\alpha)}(X_t) = \mu + \sigma_{t|t-1}^2 VaR_{(1-\alpha)}(\varepsilon_t). \quad (7)$$

حال اگر یک پرتفوی شامل دو دارایی داشته باشیم به این صورت که سری زمانی بازدهی دو دارایی وابسته $\{X_t\}$ و $\{X'_t\}$ را به صورت

$$X_t = \mu + a_t, \quad a_t = \sigma_{t|t-1}^2 \varepsilon_t, \quad (8)$$

$$X'_t = \mu' + a'_t, \quad a'_t = \sigma'_{t|t-1} \varepsilon'_t \quad (9)$$

باشد و هدف برآورد ارزش در معرض خطر این پرتفوی باشد، می توان از توابع مفصل به عنوان ابزاری مفید در مدل سازی ساختار وابستگی استفاده کرد و سپس این معیار را برآورد کرد. بازدهی یک پرتفوی شامل دو دارایی با وزن های w_1 و w_2 به صورت زیر است

$$r_t = w_1 X_t + (w_2) X'_t; \quad w_1 + w_2 = 1$$

که در آن وزن های w_1 و w_2 سهمی از سرمایه هستند که روی دارایی مورد نظر سرمایه گذاری شده است.

[۹]، [۳]، [۶]، [۷] و [۱۰] از جمله مطالعاتی هستند که در این زمینه انجام شده است. الگوریتم محاسبه کسری مورد انتظار بر اساس تابع مفصل به طوری که از مدل های آرچ و گارچ برای پیش بینی نوسان آینده استفاده شود بر اساس رویکرد POT برای برآورد ارزش در معرض خطر به صورت زیر می باشد:

۱. مدل سری زمانی ناهم واریانسی شرطی مناسب با هر دو سری زمانی $\{X_t\}$ و $\{X'_t\}$ و توزیع حاشیه ای ε_t و ε'_t را تعیین می کنیم.

۲. با بررسی ساختار وابستگی مانده های استاندارد شده در هر دو مدل مفصل متناسب با آنها را

برازش می‌دهیم.

۳. با استفاده از شبیه سازی مونت کارلو به تعداد $N = 100000$ زوج داده از مفصل مورد نظر با توابع توزیع حاشیه ای مشخص شده در مرحله یک تولید می‌کنیم.

۴. واریانس شرطی آینده (یک گام به جلو) را برای هر دو سری زمانی بر اساس مدل مرحله یک پیش بینی می‌کنیم.

۵. برای برآورد ارزش در معرض خطر پرتفوی در سطح احتمال α کفایت ارزش در معرض خطر داده‌هایی که از مرحله سه تولید شده را با استفاده از رابطه زیر محاسبه کنیم.

$$\hat{r}_{t+1} = w_1(\mu + \hat{\sigma}_{t+1|t}^2 \hat{\varepsilon}_{t+1}) + w_2(\mu' + \hat{\sigma}_{t+1|t}^2 \hat{\varepsilon}_{t+1}) \quad (10)$$

۶. با توجه به رابطه ۱ متوسط داده هایی که در مرحله ۳ به دست آمده و بزرگتر از ارزش در معرض خطر به دست آمده در مرحله ۵ می‌باشند را به عنوان برآورد کمبود مورد انتظار در نظر می‌گیریم.

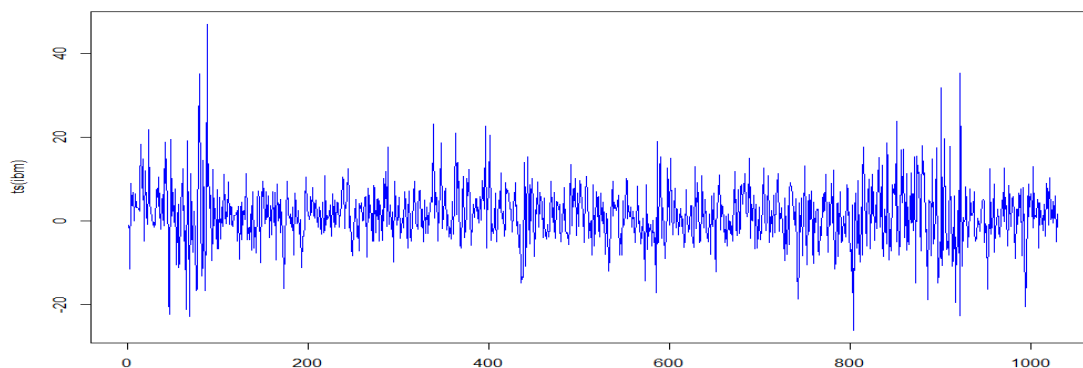
۶ کاربرد در داده واقعی

در این بخش به برآورد ارزش در معرض خطر ماه آینده در پرتفویی از داده های بازدهی ماهانه سهام *IBM* و *SP* با وزن‌های یکسان یعنی $w_1 = w_2 = \frac{1}{2}$ ، از سال ۱۹۲۶ تا ۲۰۱۱ به تعداد ۱۰۳۰ مشاهده می‌پردازیم، از مدل گارچ برای پیش بینی نوسان آینده و برای مدل بندی ساختار وابستگی بین آن‌ها از تابع مفصل استفاده خواهیم کرد.

بر اساس شکل ۱ و ۱ که سری بازدهی سهام *IBM* و *SP* را نشان می‌دهند، مشاهده می‌کنیم سری ها در برخی از دوره‌ها بسیار ناپایدار شدند و دارای نوسانات شدیدتری هستند به عبارت دیگر واریانس شرطی در طی زمان تغییر می‌کند، همچنین با توجه به شکل ۳ نتیجه می‌شود که پی-مقدارهای آزمون مکلیودلی که برای بررسی وجود اثر آرچ در داده‌ها انجام می‌شود، همگی در سطح ۵ درصد معنادارند و به طور رسمی شواهدی قوی به نفع *ARCH* در داده ها وجود دارد، بنابراین برای پیش بینی واریانس شرطی یک گام جلوتر از مدل‌های آرچ و گارچ استفاده می‌کنیم. با بررسی روش های تشخیص بهترین مدل، می‌توان یک مدل $GARCH(1, 1)$ و توزیع استودنت

با درجه آزادی v برای نوآوری‌ها برای داده‌های هر دو سری زمانی در نظر گرفت. در شکل ۴ نمودار چندک-چندک مربوط به مانده‌های استاندارد شده مدل هر دو سری زمانی در ارتباط با توزیع استودنت با درجه آزادی v رسم شده است و جدول ۱ شامل ضرایب مدل دو سری زمانی و مقدار پیش‌بینی واریانس شرطی ماه آینده می‌باشد.

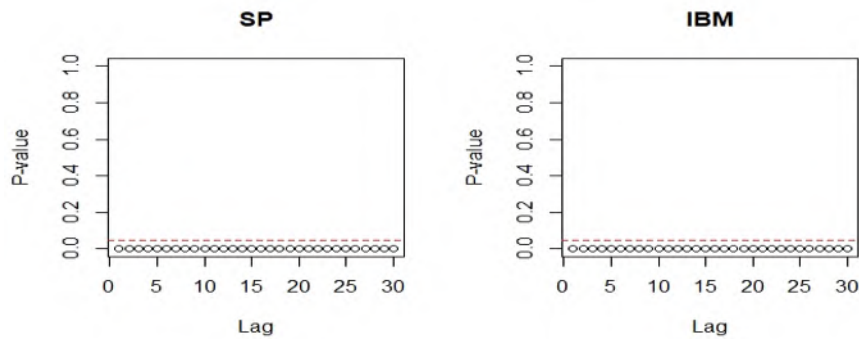
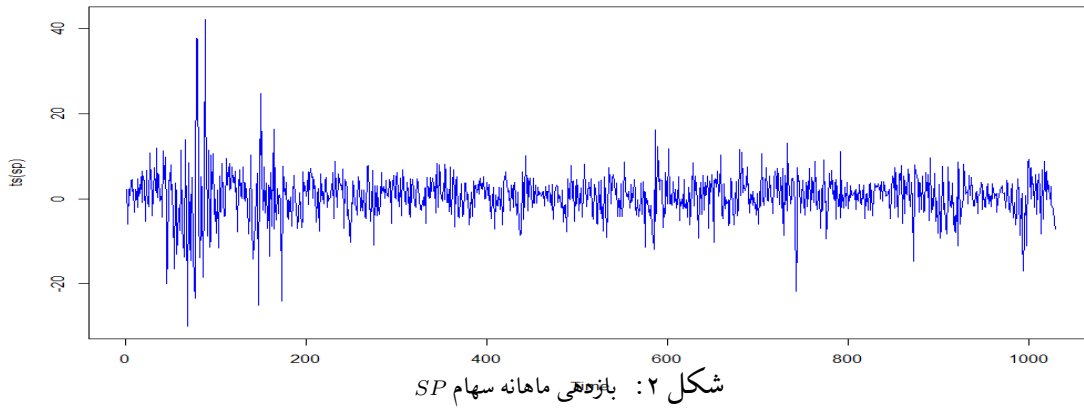
با بررسی ساختار وابستگی مانده‌های استاندارد شده و استفاده از آزمون نیکویی برازش برای یافتن بهترین مفصل، مفصل‌تی-استودنت با پارامترهای $\theta = ۰/۶۱۵۷$ و $v = ۷/۷۴۷۸$ به عنوان مناسب‌ترین مفصل انتخاب شد، پس از این مرحله با شبیه‌سازی ۱۰۰۰۰۰ زوج داده از مفصل‌تی-استودنت با توابع توزیع حاشیه‌ای تی-استودنت با درجات آزادی ۶/۵۷۳۰۹ و ۷/۹۹۵۰۹، نتایج جدول ۱ و (۱۰) برآورد ارزش در معرض خطر و کمبود مورد انتظار در سطوح احتمال ۰/۱، ۰/۰۵، ۰/۰۱ در جدول ۲ محاسبه شده است، این اعداد نشان‌دهنده این است که حداکثر زیان تحمیلی طی ماه آینده تنها با احتمال $\alpha \times ۱۰۰$ درصد از این مقادیر بالاتر می‌رود.



شکل ۱: بازدهی ماهانه سهام IBM

جدول ۱: برآورد ضرایب مدل و پیش‌بینی واریانس شرطی ماه آینده در هر دو سری زمانی

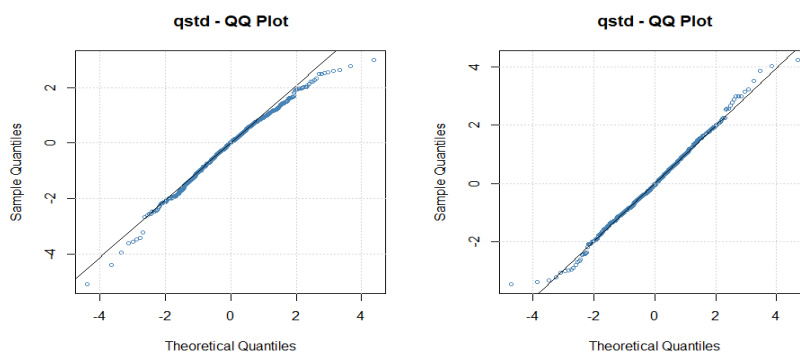
| داده | $\hat{\mu}$ | $\hat{\omega}$ | $\hat{\alpha}_1$ | $\hat{\beta}_1$ | v | $\hat{\sigma}_{t+1 t}^2$ |
|------|-------------|----------------|------------------|-----------------|---------|--------------------------|
| IBM | ۱/۳۵۵۹۸ | ۳/۳۴۸۷۵ | ۰/۱۰۴۹۰ | ۰/۸۲۷۲۴ | ۶/۵۷۳۰۹ | ۵/۳۹۷۵۱۳ |
| SP | ۰/۲۲۸۴۰۶۸ | ۰/۹۴۶۳۴ | ۰/۱۳۲۲۶ | ۰/۸۳۴۵۹ | ۷/۹۹۵۰۹ | ۵/۰۶۷۷۱ |



شکل ۳: پی-مقدارهای آزمون مک لیوود-لی

جدول ۲: برآورد ارزش در معرض خطر و کسری مورد انتظار در سطوح احتمال مختلف

| برآورد کسری مورد انتظار | برآورد ارزش در معرض خطر | سطح احتمال (α) |
|-------------------------|-------------------------|-------------------------|
| -۱۲/۳۲۶۵ | -۱۱/۹۶۵۰ | ۰/۰۱ |
| -۷/۴۵۷۸ | -۶/۲۰۳۱ | ۰/۰۵ |
| -۵/۶۵۴۸ | -۴/۴۲۳۳ | ۰/۱ |



شکل ۴: نمودار سمت چپ: مانده های استاندارد شده SP ، نمودار سمت راست: مانده های استاندارد شده IBM

۷ نتیجه گیری

در سال‌های اخیر اندازه گیری و تحلیل ریسک به عنوان یک از مهم ترین مسایل در اقتصاد و بیمه تبدیل شده است، از این رو در این مقاله یکی از روش های اندازه گیری ریسک یعنی کسری مورد انتظار مورد بررسی قرار گرفت و نحوه برآورد آن در یک پرتفوی شامل دو دارایی وابسته بر اساس تابع مفصل و استفاده از مدل های سری زمانی آرچ و گارچ برای پیش بینی نوسان یا واریانس شرطی آینده در صورت برآورد ارزش در معرض بر اساس رویکرد POT بیان شد و سرانجام به برآورد این معیار در سطوح احتمال مختلف در یک پرتفوی با وزن های یکسان در داده های بازدهی سهام IBM و SP با برازش مدل $GARCH(1, 1)$ به هر دو سری زمانی و مفصل تی-استودنت به باقی مانده های استاندارد شده آن‌ها پرداختیم.

مراجع

[۱] مشکانی، م. (۱۳۹۲)، تحلیل سری های زمانی با برنامه های کاربردی در R ، چاپ اول، انتشارات مرکز نشر دانشگاهی.

[2] Bolersllev, T. (1986). Generalized Autoregressiv conditional heteroscedasticity, *Econometrics*, **31**, 307-327.

- [3] Cheng, G., Li, P., Shi, P., (2007). A new algorithm based on copulas for VaR valuation with emperical calculations. *Theoretical computer science*, **378(2)** 109-197.
- [4] Denuit M., Dhaene J., Goovaerts, M., Kaas, R. (2005). Actuarial Theory for Dependent Risks. John Wiley Sons.
- [5] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of Uk. infiation. *Econometrica*, **50**, 987-1007.
- [6] Lu, X.F., Lai, K.K., Liang, L. (2014). portfolio value at risk estimation in energy futures markets with time varing copula-GARCH model. *Annals of operations research*. **219(1)**, 333-357.
- [7] Messaoud, S.B., Aloui, C. (2015). Measuring risk of portfolio: GARCH-Copula model. *Economic Integration*, 172-205.
- [8] Nelsen, R.B. (2006). *An Introduction to Copulas*, vol. 139. Springer Science & Business Media.
- [9] Palaro, H. P., Htta, L. K. (2006). Using conditional copula to estimate value at risk. *Data Science*, **4** 93-115.
- [10] Sampid, M.G., Hasim, H.M. (2018). Estimating value at risk using a multivariate copula based volatility model: Evidence from European banks. *Enternational Economics* **156**, 175-192.
- [11] Sklar A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ Inst Statist Univ Paris*. **8**, 229-231.
- [12] Taylor, S.J. (1986). *Modeling Financial Time Series*. Chichester: John wiley Sons.



هفتمین سمینار
نظریه مفصل و کاربردهای آن

۱۹ و ۲۰ بهمن ۱۴۰۱



مدل‌های وابسته با ساختار موازی-سری و سری-موازی با تعداد زیر سیستم‌های تصادفی

محتشمی برزادران، و^۱ امینی، م^۲ احمدی، ج^۳

^{۱,۲,۳} گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده

با استفاده از نظریه توابع مفصل توزیع توام برای دو مدل موازی-سری و سری-موازی با تعداد زیر سیستم‌های تصادفی ارائه شده است. برآوردگرهای گشتاوری و درست‌نمایی ماکسیمم برای پارامترهای مدل سری-موازی با تعداد زیر سیستم‌های تصادفی در حالت خاص بدست آمده است. با استفاده از شبیه‌سازی این برآوردگرها را ارزیابی کرده و در نهایت با بکارگیری این روش به برآورد پارامترهای مدل برای داده‌های شدت وزش باد پرداخته شده است.

کلمات کلیدی: تابع مفصل، برآورد گشتاوری، برآورد درست‌نمایی ماکسیمم

¹vmb1369@yahoo.com

²m-amini@um.ac.ir

³ahmadi-j@um.ac.ir

۱ مقدمه

در سال‌های اخیر ساخت خانواده توزیع‌های توام جدید و منعطف در علوم مختلف مورد توجه محققین قرار گرفته است. در همین راستا پژوهشگران روش‌های مختلفی برای ساخت خانواده توزیع‌های منعطف ارائه کرده‌اند. مارشال-الکین (۱۹۹۷) روشی را ارائه کرد تا با افزودن پارامتری جدید مدلی انعطاف پذیر ساخته شود. میرحسینی و همکاران (۲۰۱۱)، کوندو و گوپتا (۲۰۱۴) از این روش برای ساخت خانواده توزیع‌های منعطف استفاده کردند. دورلیمن و همکاران (۲۰۰۰) با استفاده از توابع تغییرشکل روش دیگری برای ساخت خانواده توزیع‌های منعطف ارائه کردند. موریلاس (۲۰۰۵) و دورانته و همکاران (۲۰۱۰) در این زمینه مطالعاتی انجام دادند. محتشمی برزادران و همکاران (۲۰۱۹) و محتشمی برزادران و همکاران (۲۰۲۱) با استفاده از این دو روش به ساخت مدل‌های موازی-سری و سری-موازی با تعداد زیر سیستم‌های تصادفی پرداخته و ویژگی‌های مختلف آن‌ها را بررسی کردند. یکی از ابزارهای مفید در ساخت خانواده توزیع‌های وابسته توابع مفصل هستند که تابع توزیع توام را با توابع توزیع حاشیه‌ای مرتبط می‌کنند. فرض کنید (X, Y) یک بردار تصادفی با توزیع توام $F(\cdot, \cdot)$ و توابع توزیع حاشیه‌ای $F_1(\cdot)$ ، و $F_2(\cdot)$ باشند آنگاه بنا به قضیه اسکالر (اسکلار (۱۹۵۹)) داریم

$$F(x, y) = D(F_1(x), F_2(y)), \quad \forall (x, y) \in \mathbb{R}^2,$$

که $D(\cdot, \cdot)$ تابع مفصل یکتا برای (X, Y) است.

در بخش دوم مدل موازی-سری و سری-موازی با تعداد زیر سیستم‌های تصادفی شرح داده شده است. روشی برای برآورد پارامترهای مجهول مدل سری-موازی با تعداد زیر سیستم‌های تصادفی با دیدگاهی جدید در بخش سوم ارائه شده است. در بخش چهارم با استفاده از شبیه سازی برآوردگرهای بخش قبلی ارزیابی شده است. روش برآورد ارائه شده برای داده‌های شدت وزش باد در بخش پنجم اعمال شده است.

۲ مدل‌های موازی-سری و سری-موازی با تعداد زیر سیستم‌های تصادفی

در این بخش دو مدل موازی-سری و سری-موازی با تعداد زیر سیستم‌های تصادفی معرفی شده است. محتشمی برزادران و همکاران (۲۰۱۹) و محتشمی برزادران و همکاران (۲۰۲۱) به ساخت این خانواده توزیع‌ها پرداخته و ویژگی‌های آن‌ها را بررسی کردند.

الف) مدل موازی-سری با تعداد زیر سیستم‌های تصادفی: دو سیستم موازی-سری با تعداد زیر سیستم‌های تصادفی N را در نظر بگیرید. فرض کنید برای $i, l = 1, \dots, N$ متغیرهای تصادفی X_i و Y_l به ترتیب طول عمر i -امین زیر سیستم در سیستم اول و l -امین زیر سیستم در سیستم دوم باشند. همچنین فرض کنید هر زیر سیستم شامل m مولفه مستقل و هم توزیع به ترتیب با توابع توزیع حاشیه‌ای $F_1(\cdot)$ و $F_2(\cdot)$ در سیستم اول و دوم باشند. علاوه بر این مولفه i -ام در سیستم اول و دوم به هم وابسته بوده و از توزیع $F(\cdot, \cdot)$ پیروی کنند و از N مستقل باشند. در این صورت $T_1 = \max\{X_1, \dots, X_N\}$ و $T_2 = \max\{Y_1, \dots, Y_N\}$ به ترتیب طول عمر متناظر با سیستم اول و دوم می‌باشند و تابع توزیع توام (T_1, T_2) از رابطه زیر بدست می‌آید

$$H_{T_1, T_2}(x, y) = g(1 - \bar{F}_1^m(x) - \bar{F}_2^m(y) - \bar{F}^m(x, y)),$$

که در آن $g(\cdot)$ تابع تغییر شکل است. همچنین اگر $\hat{D}(\cdot, \cdot)$ مفصل بقا متناظر با تابع بقا $\bar{F}(\cdot, \cdot)$ باشد آنگاه بنا به قضیه اسکالر مفصل متناظر با تابع توزیع توام $H_{T_1, T_2}(\cdot, \cdot)$ بصورت زیر است

$$C_{g(D)}(u, v) = g\left(g^{-1}(u) + g^{-1}(v) - 1 + \hat{D}^m\left(\left(1 - g^{-1}(u)\right)^{1/m}, \left(1 - g^{-1}(v)\right)^{1/m}\right)\right).$$

ب) مدل سری-موازی با تعداد زیر سیستم‌های تصادفی: دو سیستم سری-موازی با تعداد زیر سیستم‌های تصادفی N را در نظر بگیرید. فرض کنید متغیرهای تصادفی X_i و Y_l برای $i, l = 1, \dots, N$ به ترتیب طول عمر i -امین زیر سیستم در سیستم اول و l -امین زیر سیستم در سیستم دوم باشند. علاوه بر این فرض کنید هر زیر سیستم شامل m مولفه مستقل و هم توزیع به ترتیب با توابع بقا حاشیه‌ای $\bar{F}_1(\cdot)$ و $\bar{F}_2(\cdot)$ در سیستم اول و دوم باشند، همچنین مولفه

i -ام در سیستم اول و دوم به هم وابسته بوده و از توزیع بقا توام $\bar{F}(\cdot, \cdot)$ پیروی کنند بطوریکه از N مستقل باشد. در این صورت $T_1 = \min\{X_1, \dots, X_N\}$ و $T_2 = \min\{Y_1, \dots, Y_N\}$ به ترتیب طول عمر متناظر با سیستم اول و دوم است و تابع بقا توام (T_1, T_2) از رابطه زیر حاصل می‌شود

$$\bar{H}_{T_1, T_2}(x, y) = g(1 - F_1^m(x) - F_2^m(y) - F^m(x, y)), \quad (1)$$

که در آن $g(\cdot)$ تابع تغییر شکل است. علاوه بر این اگر $D(\cdot, \cdot)$ مفصل متناظر با تابع توزیع $F(\cdot, \cdot)$ باشد آنگاه مفصل متناظر با تابع بقا توام $\bar{H}_{T_1, T_2}(\cdot, \cdot)$ با استفاده از قضیه اسکالر از رابطه زیر بدست می‌آید

$$\hat{C}_{g(D)}(u, v) = g(g^{-1}(u) + g^{-1}(v) - 1 + D^m((1 - g^{-1}(u))^{1/m}, (1 - g^{-1}(v))^{1/m})).$$

در بخش بعد برآورد گشتاوری و درست نمایی ماکسیم پارامترهای مدل سری-موازی با تعداد زیر سیستم‌های تصادفی زمانی که نمونه‌ای تصادفی از مفصل $\hat{C}_{g(D)}(\cdot, \cdot)$ مشاهده شده است را بدست می‌آوریم.

۳ برآورد

فرض کنید $(T_{1,1}, T_{1,2}), \dots, (T_{k,1}, T_{k,2})$ بردارهای تصادفی مستقل و هم توزیع با تابع بقای توام (1) و توابع توزیع حاشیه‌ای $F_1(\cdot)$ و $F_2(\cdot)$ دارای توزیع نمایی به ترتیب با پارامترهای α_1 و α_2 باشند. در این بخش با این فرض که نمونه‌ای تصادفی از مفصل بقا $\hat{C}_{g(D)}(\cdot, \cdot)$ مشاهده شده است، برآوردگرهای گشتاوری و درست نمایی ماکسیم برای α_1, α_2 و θ را بدست می‌آوریم.

برآورد گشتاوری

در این روش با استفاده از گشتاور نمونه، گشتاور جامعه برآورد می‌شود.

گام اول (برآورد پارامتر وابستگی): با استفاده از روش معکوس ضریب تاو کندال برآوردگر گشتاوری $\tilde{\theta}$ از حل معادله زیر حاصل می‌شود

$$\tilde{\tau} - \tau(\theta) = 0, \quad (2)$$

که در آن $\tau(\theta)$ و $\tilde{\tau} = \frac{2}{n(n-1)} \sum_{i < j} \text{sgn}((T_{i,1} - T_{j,1})(T_{i,2} - T_{j,2}))$ تاو کندال متناظر با مفصل بقاء $\hat{C}_g(u, v)$ است.

گام دوم (برآورد پارامترهای حاشیه‌ای): با استفاده از توزیع حاشیه‌ای T_r پارامترهای حاشیه‌ای α_r برای $r = 1, 2$ برآورد می‌شوند. با توجه به رابطه (۱) گشتاور اول متغیر تصادفی T_r از رابطه زیر بدست می‌آید

$$E(T_r) = \alpha_r B_{m,1}(\theta),$$

که در آن $B_{m,1}(\tilde{\theta}) = \int_0^1 \frac{g(1-w^m)}{1-w} dw$ است. در نتیجه برآوردگر گشتاوری پارامتر α_r از رابطه زیر حاصل می‌شود

$$\tilde{\alpha}_r = \frac{\bar{T}_r}{B_{m,1}(\tilde{\theta})}, \quad B_{m,1}(\tilde{\theta}) \neq 0, r = 1, 2 \quad (3)$$

که در آن $\bar{T}_r = \frac{1}{k} \sum_{i=1}^k T_{i,r}$ است.

روش درست نمایی ماکسیمم

در این روش به دنبال بیشینه کردن تابع درست نمایی هستیم زمانی که مشاهدات در دسترس باشند. از آن جایی که تابع چگالی دو متغیره بر حسب تابع چگالی مفصل از رابطه زیر حاصل می‌شود

$$h(t_1, t_2; \alpha_1, \alpha_2, \theta) = c_g(H_1(t_1), H_2(t_2)) h_1(t_1) h_2(t_2), \quad (4)$$

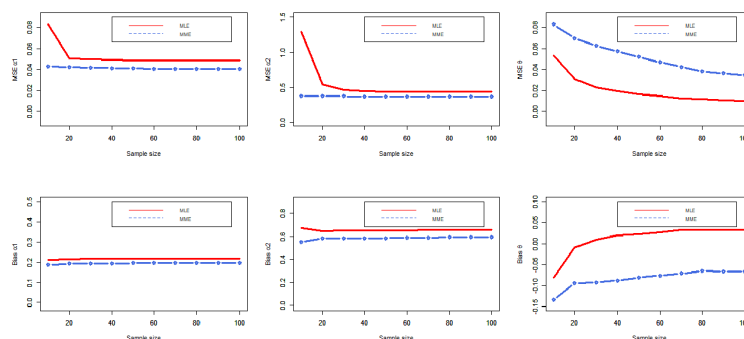
بنابراین تابع درست نمایی متناظر با (۴) از رابطه زیر بدست می‌آید

$$l(\alpha_1, \alpha_2, \theta) = \sum_{i=1}^k \{\log c_g(H_1(t_{i,1}), H_2(t_{i,2})) + \log(h_1(t_{i,1})) + \log(h_2(t_{i,2}))\} \quad (5)$$

و برآوردگرهای درست نمایی ماکسیمم برای پارامترهای α_1 ، α_2 و θ از بیشینه کردن (۵) بر حسب α_1 ، α_2 و θ حاصل شده و به ترتیب $\hat{\alpha}_1$ ، $\hat{\alpha}_2$ و $\hat{\theta}$ نامیده می‌شوند.

۴ شبیه سازی

میانگین مربعات خطا و اریبی برآوردگرهای گشتاوری و درست نمایی ماکسیمم مدل (۱) زمانی که N از توزیع برنولی منتقل شده پیروی کند (یعنی $P(N = 1) = 1 - \theta$ و $P(N = 2) = \theta$)، توزیع‌های حاشیه‌ای نمایی و $\hat{C}_g(u, v) = AMH(\theta)$ باشند برای $\alpha_1 = 0.5$ ، $\alpha_2 = 1/5$ ، $\theta = 0.75$ ، و $m = 10$ در شکل ۱ نمایش داده شده است. میانگین مربعات خطا برآوردگرهای گشتاوری و درست نمایی ماکسیمم با افزایش حجم نمونه کاهش می‌یابند. با توجه به اینکه میانگین مربعات خطا $\hat{\alpha}_1$ و $\hat{\alpha}_2$ از میانگین مربعات خطا $\hat{\alpha}_1$ و $\hat{\alpha}_2$ کمتر است بنابراین برآوردگرهای بهتری محسوب می‌شوند. ولی میانگین مربعات خطا $\hat{\theta}$ از میانگین مربعات خطا $\hat{\theta}$ کمتر است در نتیجه برآوردگر بهتری می‌باشد.

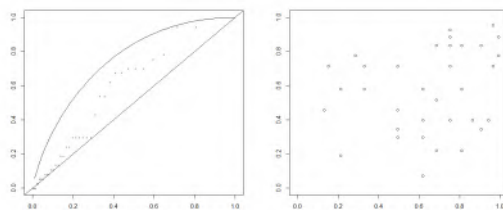


شکل ۱: میانگین مربعات خطا و اریبی تجربی زمانی که $\theta = 0.75$ ، $\alpha_1 = 0.5$ ، $\alpha_2 = 1/5$ و $m = 10$.

۵ داده‌های شدت وزش باد

در این بخش برای داده‌های بیشترین شدت وزش باد سالانه ویت مور و جنتلین (۱۹۸۷) روش برآورد ارائه شده در بخش‌های قبل اعمال شده است. این داده‌ها در سال‌های ۱۹۴۷ تا ۱۹۸۴ بیشترین شدت وزش باد در پنج ایستگاه ثبت مشاهدات شیروا، مونترال، تورنتو، رجینا، و ونکوور را شامل می‌شود. برای اینکه شرایط مدل (۱) زمانی که N از توزیع برنولی منتقل شده پیروی کرده و توزیع‌های حاشیه‌ای نمایی باشند برقرار باشد در سال‌های ۱۹۴۷ تا ۱۹۸۴ برای

$i = 1, \dots, 76$ بیشترین شدت وزش باد سالانه در ایستگاه‌های ونکوور و رجینا را به عنوان X_i و بیشترین شدت وزش باد سالانه در ایستگاه‌های تورنتو و مونترال به عنوان Y_i در نظر گرفته شده است. سپس توزیع نمایی تعمیم یافته با پارامتر شکل $m = 365$ و پارامترهای مکان $11/1944$ و $11/0028$ برای $i = 1, \dots, 76$ به ترتیب به X_i و Y_i برازش داده شده است و p - مقدار آزمون نیکویی برازش متناظر با آن‌ها به ترتیب 0.667 و 0.2677 است. در نتیجه فرضیه نمایی بودن توزیع‌های حاشیه‌ای برقرار است. در مرحله بعد متغیر تصادفی $N = 2$ از توزیع برنولی منتقل شده با پارامتر 0.25 تولید شده است. علاوه بر این برای سال‌های 1947 تا 1984 کوچکترین مقدار بین داده‌های ونکوور و رجینا در هر سال به عنوان T_1 و کوچکترین مقدار بین داده‌های تورنتو و مونترال در هر سال به عنوان T_2 در نظر گرفته شده است. برای (T_1, T_2) نمودار کندال و پراکنش در شکل ۲ رسم شده است که نشان می‌دهد (T_1, T_2) وابسته مثبت بوده و وابستگی کمی ندارند. با توجه به اینکه N از توزیع برنولی منتقل شده تولید شده است پارامتر وابستگی متناظر با آن همواره بین صفر و یک است و بنابراین باید مفصلی برازش داده شود که پارامتر وابستگی آن بین صفر و یک باشد. در این راستا مفصل علی-میکائیل-حق^۱ (AMH) و مفصل



شکل ۲: نمودار کندال (چپ) و نمودار پراکنش (راست) برای داده‌های وزش باد.

فارلی-گامبل-مورگنسترن^۲ (FGM) برای برازش به این داده‌ها پیشنهاد شده است. در جدول ۱، p - مقدار آزمون نیکویی برازش و معیار آکائیک متناظر با هر یک از این دو مفصل نمایش داده شده است. از آن جایی که p - مقدار متناظر با آزمون نیکویی برازش برای هر دو مفصل از 0.05 بیشتر است در نتیجه هر دو مفصل برای برازش به این داده‌ها مناسب بوده اما از آن جایی که معیار آکائیک مفصل AMH از مفصل FGM کمتر است، پس مفصل AMH به داده‌ها بهتر

¹Ali-Michail-Haq

²Farlie-Gumbel-Morgenstern

برازش داده می‌شود.

جدول ۱: آزمون نیکویی برازش مفصل.

| مفصل | $\hat{\theta}$ | p -مقدار | معیار آکائیک |
|-----------------------|----------------|------------|--------------|
| علی-میکائیل-حق | ۱ | ۰/۱۳۴۴ | -۳/۶۴۶ |
| فارلی-گامبل مورگنسترن | ۱ | ۰/۵۳۵ | -۲/۵۱۴ |

با بکار بردن روند برآورد ارائه شده در بخش‌های قبل برآوردگرهای گشتاوری و درست‌نمایی ماکسیمم α_1 ، α_2 و θ در جدول ۲ نمایش داده شده است.

جدول ۲: نتایج برآورد.

| برآورد نقطه‌ای برای α_1 | برآورد گشتاوری | برآورد درست‌نمایی ماکسیمم |
|--------------------------------|----------------|---------------------------|
| ۱۰/۶۹۴۶ | ۱۰/۹۸۲۵ | |
| ۱۰/۸۴۰۱ | ۱۱/۱۹۱۲ | |
| ۰/۶۸۶۶ | ۰/۹۳۱۰ | |

۶ نتیجه‌گیری

در این یادداشت روش ساخت مدل‌های موازی-سری و سری-موازی با تعداد زیرسیستم‌های تصادفی ارائه شده و با بکارگیری توابع مفصل و دیدگاهی متفاوت پارامترهای این مدل برآورد شده است. همچنین با استفاده از شبیه‌سازی این برآوردگرها را ارزیابی کرده و در نهایت از داده‌های شدت وزش باد برای شرح این روش استفاده شده است.

مراجع

- [1] Durante, F., Foschi, R. and Sarkoci, P. (2010), *Distorted copulas: Constructions and tail dependence*, Commun. Stat. Theory Methods 39, 2288–2301.

- [2] Durrleman, V., Nickeghbali, A. and Roncalli, T. (2000), *A simple transformation of copulas*, Groupe de Recherche Operationnelle, Credit Lyonnais, France . Available at: gro.creditlyonnais.fr/content/rd/home_copulas.htm.
- [3] Kundu, D. and Gupta, A. (2014), *On bivariate Weibull-geometric distribution*, J. Multivar. Anal. 123 , 19–29.
- [4] Marshall, A.W. and Olkin, I. (1997), *A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families*, Biometrika 84, 641–652.
- [5] Mirhosseini, S.M., Dolati, A. and Amini, M. (2011), *On a class of distributions generated by stochastic mixture of the extreme order statistics of a sample of size two*, J. Stat.TheoryAppl. 10, 455–468.
- [6] Mohtashami-Borzadaran, V., Amini, M. and Ahmadi, J. (2019), *A generalized bivariate lifetime distribution based on parallel-series structures*, Kybernetika 55, 435–454.
- [7] Mohtashami-Borzadaran, V., Amini, M. and Ahmadi, J. (2021), *Estimating the parameters of a dependent model and applying it to environmental data set*, J. Appl. Stat., DOI: 10.1080/02664763.2021.2006613.
- [8] Morillas, P. (2005), *A method to obtain new copulas from a given one*, Metrika 61, 169–184.
- [9] Sklar, A. (1959), *Function de repartition an dimensions et leurs marges*, Publ. Inst. Stat. Univ. Paris 8, 229–331.
- [10] Whitemore, G.A. and Gentleman, J.F. (1987), *Extreme-value analysis of Canadian wind speeds*, Canad. J. Stat. 15, 311–337.



قانون تفکیک دو مقدم برای استلزامهای احتمالاتی مفصل مبنا

معدنی خوش بخت، ا^۱ دولتی، ع^۲

^{۱،۲} گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

^۲ بخش آمار، دانشکده علوم ریاضی، دانشگاه یزد

چکیده

عملگرهای استلزامی توابع دو متغیره‌ای بر مربع واحد و تعمیمی از استلزام دو ارزشی منطق کلاسیک هستند. روش‌های مختلفی برای ساخت عملگرهای استلزامی ارائه شده است. یک رده از این توابع، استلزامهای احتمالاتی هستند که براساس توابع مفصل ساخته می‌شوند. در این مقاله قانون تفکیک دو مقدم برای این رده از استلزامها مطالعه می‌شود.

کلمات کلیدی: استلزام، قانون تفکیک دو مقدم، منطق کلاسیک، منطق فازی، مفصل

۱ پیش‌گفتار

در منطق کلاسیک دو ارزشی برای رابطه استلزام $p \rightarrow q$ تعریف‌های مختلفی وجود دارد که رایج‌ترین آن تعریف $p \vee q \sim$ یا معادل آن $(p \wedge \sim q) \sim$ است. اگر x ارزش (یا درجه درستی)

¹khosbakht.elnaz@yahoo.com

²dolati50@yahoo.com

گزاره p و y ارزش گزاره q باشد. تابع $I(x, y)$ که $I: \{0, 1\}^2 \rightarrow \{0, 1\}$ درجه درستی گزاره اگر p آنگاه q را نشان می‌دهد که از چهار حالت ممکن در حالت $x = 1$ و $y = 0$ نادرست و در سه حالت دیگر درست است. در منطق فازی ارزش گزاره‌ها در استلزام $p \rightarrow q$ مقادیری در بازه $[0, 1]$ است. عملگرهای استلزام (یا استلزام‌های فازی) تعمیمی از استلزام دو ارزشی منطق کلاسیک هستند که برای ارزش‌گذاری گزاره‌ها استفاده می‌شوند و در زمینه‌های مختلفی مانند منطق فازی، کنترل فازی، نظریه تصمیم و سیستم‌های هوشمند کاربرد دارند. برای جزئیات در مورد این عملگرها، روش‌های ساخت و ویژگی‌های آن‌ها می‌توان به مراجع [۱، ۳] اشاره نمود. روش‌های مختلفی برای ساخت عملگرهای استلزامی ارائه شده است. یک رده از این توابع، استلزام‌های احتمالاتی هستند که براساس توابع مفصل ساخته می‌شوند. در این مقاله قانون تفکیک دو مقدم برای این رده از استلزام‌ها مطالعه می‌شود. در بخش ۲ مفاهیم مورد نیاز از نظریه مفصل و عملگرهای استلزام یادآوری می‌شود. در بخش ۳ قانون تفکیک دو مقدم برای استلزام‌های احتمالاتی مفصل مینا مورد مطالعه قرار می‌گیرند.

۲ مفاهیم مورد نیاز

تعریف ۱.۰.۲. تابع $I: [0, 1]^2 \rightarrow [0, 1]$ یک عملگر استلزام نامیده می‌شود اگر برای هر

$$x_1, x_2, y_1, y_2, x, y \in [0, 1] \text{ ویژگی‌های زیر را دارا باشد: [۱]}$$

$$\text{الف) برای هر } x_1 < x_2: I(x_1, y) \geq I(x_2, y)$$

$$\text{ب) برای هر } y_1 < y_2: I(x, y_1) \leq I(x, y_2)$$

$$\text{ج) } I(0, 0) = I(0, 1) = I(1, 1) = 1 - I(1, 0) = 1$$

تعریف ۲.۰.۲. عملگر شرکت‌پذیر، جابه‌جایی‌پذیر و غیر نزولی $T: [0, 1]^2 \rightarrow [0, 1]$ یک t -نرم

$$\text{نامیده می‌شود؛ اگر دارای عنصر خنثی ۱ باشد؛ یعنی برای هر } x \in [0, 1]: T(x, 1) = x \text{ [۱]}$$

عملگر t -نرم تعمیم ترکیب عطفی کلاسیک \wedge ، معادل اشتراک برای مجموعه‌ها، می‌باشد.

متداول‌ترین عملگرهای t -نرم عبارت‌اند از: $T_M(x, y) = \min(x, y)$ (t -نرم مینیمم)، $T_P(x, y) =$

xy (نرم ضرب جبری) و $T_{LK}(x, y) = \max(x + y - 1, 0)$ (نرم لوکاسویچ) t -نرم T را مثبت می‌نامند؛ اگر $T(x, y) = 0$ ؛ آنگاه $x = 0$ یا $y = 0$.

تعریف ۳.۲. تابع $C : [0, 1]^2 \rightarrow [0, 1]$ یک مفصل نامیده می‌شود [۵] اگر شرایط زیر برقرار باشند:

(الف) برای تمام مقادیر $x \in [0, 1]$: $C(x, 0) = C(0, x) = 0, C(x, 1) = C(1, x) = x$

(ب) C - صعودی باشد؛ یعنی برای تمام $x_1, x_2, y_1, y_2 \in [0, 1]$ به طوری که $x_1 \leq x_2$ و $y_1 \leq y_2$:

$$C(x_2, y_2) + C(x_1, y_1) \geq C(x_1, y_2) + C(x_2, y_1)$$

سه مفصل مهم $\Pi(x, y) = xy$ ، $M(x, y) = \min(x, y)$ و $W(x, y) = \max(x + y - 1, 0)$ هستند. Π مفصل استقلال و M و W به ترتیب کران‌های بالا و پایین فرشه- هافدینگ هستند و هر مفصل C در نابرابری زیر صدق می‌کند:

$$W(x, y) \leq C(x, y) \leq M(x, y)$$

برای جزئیات مربوط به توابع مفصل و ویژگی‌های آن‌ها می‌توان به [۵] اشاره نمود. عملگرهای استلزام احتمالی که در [۴] معرفی شده‌اند، در تعریف زیر بیان شده است.

تعریف ۴.۲. فرض کنید C یک مفصل دلخواه باشد. در این صورت تابع $I_C : [0, 1]^2 \rightarrow [0, 1]$ با ضابطه

$$I_C(x, y) = \begin{cases} \frac{C(x, y)}{x}, & x > 0 \\ 1, & x = 0 \end{cases}$$

یک استلزام احتمالی (بر مبنای مفصل C) نامیده می‌شود.

لازم به ذکر است که یک استلزام احتمالی لزوماً یک عملگر استلزام نیست و در صورتی یک عملگر استلزام خواهد بود که I_C در شرط الف تعریف ۱.۲ صدق کند.

تعریف ۵.۲. فرض کنید C یک مفصل دلخواه باشد. در این صورت تابع $\tilde{I}_C : [0, 1]^2 \rightarrow [0, 1]$ با ضابطه

$$\tilde{I}_C(x, y) = C(x, y) - x + 1$$

یک S -استلزام احتمالی نامیده می‌شود.

یک S -استلزام احتمالی همواره یک عملگر استلزام است.

۳ قانون تفکیک دو مقدم برای استلزام‌های احتمالاتی

یک قانون مهم در منطق دو ارزشی کلاسیک قانون تفکیک دو مقدم است که به صورت زیر تعریف می‌شود.

$$(p \rightarrow (q \rightarrow r)) \equiv (p \wedge q) \rightarrow r$$

تعمیم قانون تفکیک دو مقدم در منطق فازی بر مبنای عملگر استلزام و عملگر t -نرم می‌باشد که در تعریف زیر بیان شده است [۲].

تعریف ۱.۰۳. فرض کنید I یک عملگر استلزام و T یک t -نرم باشد. در این صورت، زوج (I, T) در قانون تفکیک دو مقدم صدق می‌کند؛ اگر

$$I(x, I(y, z)) = I(T(x, y), z), \quad x, y, z \in [0, 1]. \quad (\text{LI})$$

۱.۰۳ قانون تفکیک دو مقدم برای استلزام احتمالی

قضیه ۱.۰۳. فرض کنید I_C یک استلزام احتمالی و T یک t -نرم باشد که زوج (I_C, T) در رابطه (LI) صدق می‌کند. در این صورت T مثبت است.

اثبات. بنا به تعریف ۴.۲، می‌توان نوشت:

$$I_C(T(x, y), \cdot) = \begin{cases} \frac{C(T(x, y), \cdot)}{T(x, y)}, & T(x, y) > \cdot \\ 1, & T(x, y) = \cdot \end{cases} = \begin{cases} \cdot, & T(x, y) > \cdot \\ 1, & T(x, y) = \cdot \end{cases}$$

با برهان خلف فرض کنید که $x, y, \neq \cdot$ وجود دارد که $T(x, \cdot, y, \cdot) = \cdot$. از این رو، $I_C(T(x, \cdot, y, \cdot), \cdot) = \cdot$. اما $I_C(x, \cdot, I_C(y, \cdot, \cdot)) = I_C(x, \cdot, \cdot) = \cdot$. بنابراین زوج (I_C, T) در رابطه (LI) صدق نمی‌کند و این با فرض در تناقض است و اثبات کامل می‌شود. \square

در مثال‌های زیر هدف پیدا کردن t -نرم مثبتی است که زوج (I_C, T) در رابطه (LI) صدق کند. به این منظور رابطه (LI) را می‌توان بر اساس رابطه I_C به صورت زیر بازنویسی کرد.

$$\frac{C(T(x, y), z)}{T(x, y)} = \frac{C\left(x, \frac{C(y, z)}{y}\right)}{x}, \quad x, y, z \in (0, 1]. \quad (1)$$

مثال ۱.۳. با قرار دادن مفصل استقلال در رابطه (۱)، مشاهده می‌شود که استلزام احتمالی I_{II} با هر t -نرم مثبت T ، در رابطه (LI) صدق می‌کند.

مثال ۲.۳. مفصل $C(x, y) = \frac{xy}{x+y-xy}$ و t -نرم T_P را در نظر بگیرید. با قرار دادن مفصل C و T_P در رابطه (۱)، می‌توان نوشت:

$$\frac{C(T_P(x, y), z)}{T_P(x, y)} = \frac{C(xy, z)}{xy} = \frac{z}{xy - xyz + z} = \frac{C\left(x, \frac{z}{y+z-yz}\right)}{x} = \frac{C\left(x, \frac{C(y, z)}{y}\right)}{x}$$

در نتیجه استلزام احتمالی I_C با t -نرم T_P ، در رابطه (LI) صدق می‌کند.

مثال ۳.۳. خانواده مفصل AMH با ضابطه $C_\theta(x, y) = \frac{xy}{1-\theta(1-x)(1-y)}$ ، $\theta \in [-1, 1]$ را در نظر بگیرید. با محاسبات ساده، می‌توان نتیجه گرفت که این خانواده از مفاصل، به جز در حالتی که $\theta = 0$ ، در رابطه (۱) صدق نمی‌کنند. در نتیجه استلزام احتمالی I_{C_θ} با هیچ t -نرم T ، در رابطه (LI) صدق نمی‌کند؛ به جز در حالتی که $\theta = 0$.

۲.۳ قانون تفکیک دو مقدم برای S -استلزام احتمالی

قضیه ۲.۳. فرض کنید \tilde{I}_C یک S -استلزام احتمالی و T یک t -نرم باشد که زوج (\tilde{I}_C, T) در رابطه (LI) صدق می‌کند. در این صورت T باید به شکل زیر تعریف شده باشد:

$$T(x, y) = x - C(x, 1 - y), \quad x, y \in [0, 1]. \quad (2)$$

اثبات. فرض کنید زوج (\tilde{I}_C, T) در رابطه (LI) صدق کند. بنا به تعریف ۵.۲ می‌توان نوشت:

$$C(T(x, y), z) - T(x, y) + 1 = C(x, C(y, z) - y + 1) - x + 1.$$

با قرار دادن $z = 0$ داریم:

$$0 - T(x, y) + 1 = C(x, 0 - y + 1) - x + 1 \implies T(x, y) = x - C(x, 1 - y)$$

□

سوالی که در اینجا مطرح می‌شود این است که آیا عکس گزاره بالا برقرار است؟ برای پاسخ به این سوال ابتدا باید بررسی شود که برای یک S -استلزام احتمالی \tilde{I}_C ، تابع T تعریف شده در رابطه (۲)، یک t -نرم باشد و همچنین زوج (\tilde{I}_C, T) در رابطه (LI) برای هر $x, y, z \in [0, 1]$ صدق کند. به آسانی می‌توان بررسی کرد که رابطه موردنظر برای برخی موارد خاص مانند $x = 0, x = 1, y = 0, y = 1, z = 0$ و $z = 1$ برقرار است اما در حالت کلی ممکن است برقرار نباشد. به این منظور به مثال‌های بعد توجه کنید.

مثال ۴.۳. S -استلزام احتمالی \tilde{I}_M را در نظر بگیرید. با قرار دادن مفصل کران بالای فرشه در رابطه (۲)، می‌توان نوشت:

$$T(x, y) = x - \min(x, 1 - y) = \max(x + y - 1, 0) = T_{LK}(x, y)$$

بنابراین تابع $T(x, y) = T_{LK}(x, y)$ یک عملگر t -نرم است. زوج (\tilde{I}_M, T_{LK}) در رابطه (LI) صدق می‌کند. زیرا:

$$\begin{aligned} \tilde{I}_M(x, \tilde{I}_M(y, z)) &= \min(x, \min(y, z) - y + 1) - x + 1 = \min(2 - x - y + z, 1) \\ \tilde{I}_M(T_{LK}(x, y), z) &= \min(\max(x + y - 1, 0), z) - \max(x + y - 1, 0) + 1 \\ &= \min(z - \max(x + y - 1, 0) + 1, 1) = \min(2 - x - y + z, 1) \end{aligned}$$

مثال ۵.۳. خانواده مفصل FGM با ضابطه $C_\theta(x, y) = xy(1 + \theta(1 - x)(1 - y))$, $\theta \in [-1, 1]$ را در نظر بگیرید. با قرار دادن این مفصل در رابطه (۲)، می‌توان نوشت:

$$T(x, y) = x - x(1 - y)(1 + \theta(1 - x)y) = xy(1 - \theta(1 - x)(1 - y))$$

به آسانی می‌توان نشان داد که تابع T یک t -نرم است. اما زوج $(\tilde{I}_{C_\theta}, T)$ در رابطه (LI) صدق نمی‌کند. زیرا با در نظر گرفتن $x = 0.2, y = 0.7, z = 0.1$ و $\theta = -0.4$ ، طرف چپ و راست رابطه (LI) به ترتیب برابر 0.8617 و 0.8614 می‌باشد؛ که یک تناقض است.

۴ نتیجه‌گیری

عملگرهای استلزامی تعمیمی از استلزام دو ارزشی منطق کلاسیک هستند. به دلیل کاربردهای گسترده عملگرهای استلزامی در زمینه‌های مختلف، روش‌های مختلفی برای ساخت آنها ارائه شده است. یک رده از این توابع، استلزام‌های احتمالاتی هستند که براساس توابع مفصل ساخته می‌شوند. در این مقاله قانون تفکیک دو مقدم برای این رده از استلزام‌ها مطالعه شد.

مراجع

- [1] M. Baczyński and B. Jayaram, *Fuzzy Implications*, Studies in Fuzziness and Soft Computing, vol. 231, Springer, Heidelberg, 2008.
- [2] M. Baczyński, P. Grzegorzewski, P. Helbin and W. Niemyska, *Properties of the probabilistic implications and S-implications*, Information Sciences, 331 (2016), 2-14.
- [3] M. Baczyński, G. Beliakov, H. Bustince, A. Pradera (Eds.), *Advances in Fuzzy Implication Functions*. Studies in Fuzziness and Soft Computing, vol. 300, Springer, Berlin Heidelberg, 2013.
- [4] P. Grzegorzewski, *Probabilistic implications*, Fuzzy Sets and Systems, 226, (2013), 53-66.
- [5] R. B. Nelsen, *An introduction to copulas*, Springer, New York, 2007.



هفتمین سمینار
نظریه مفصل و کاربردهای آن

۱۹ و ۲۰ بهمن ۱۴۰۱



روش جدیدی برای تعمیم مفصل‌های دومتغیره

میرحسینی، م^۱ دست‌برآورده، ع^۲

^{۱،۲} بخش آمار، دانشکده علوم ریاضی، دانشگاه یزد

چکیده

در این پژوهش به مطالعه یک روش تعمیم مفصل‌های دومتغیره پرداخته شده است. برخی ویژگی و خواص آن مورد بررسی قرار گرفته است.

کلمات کلیدی: مفصل، توزیع بتا، ضرایب همبستگی، مفصل FGM

۱ پیش‌گفتار

در اکثر مطالعات توجه محققین به بررسی مدل‌های آماری با فرض استقلال داده‌ها معطوف بوده است. اما در جهان واقع ممکن است داده‌ها مستقل از هم نباشند، از این رو تعیین میزان وابستگی بین متغیرهای تصادفی یکی از مهم‌ترین موضوعاتی است که به‌طور گسترده در آمار و احتمال

¹mmirhoseini@yazd.ac.ir

²dastbaravarde@yazd.ac.ir

مورد توجه نویسندگان و محققان قرار گرفته است. برخی از آماردانان ساختار وابستگی توزیع‌های دومتغیره را مورد مطالعه قرار داده‌اند و بررسی‌های لازم انجام شده است که دو مورد از مراجع مهم در این زمینه [۴، ۵] است.

شوویزرا^۱ [۹] بیان می‌کند «دوره استقلال به پایان رسیده است. وقتی وابستگی نقش جدی ایفا کند، مفصل‌ها به‌طور طبیعی نقش ایفا می‌کنند». در واقع مفصل‌ها ابزاری سودمند برای بیان اندازه وابستگی بین متغیرها هستند، از طرفی آن‌ها به‌عنوان یک نقطه شروع برای ساختن خانواده‌هایی از توزیع‌های دومتغیره یا چندمتغیره هستند. مفصل‌ها توابعی هستند که توابع توزیع چندمتغیره را به توابع توزیع حاشیه‌ای آن‌ها پیوند می‌دهند و توزیع‌های حاشیه‌ای را از ساختار وابستگی جدا می‌سازند به همین جهت در مدل‌بندی بین متغیرهای وابسته استفاده می‌شوند [۷]. در همین راستا، ساخت توابع مفصل و تعمیم‌های آن مورد توجه آماردانان قرار گرفته است به‌عنوان مثال [۸، ۱، ۲، ۳] را ملاحظه کنید.

تعریف ۱.۱. به‌طور کلی یک مفصل دوبعدی، تابعی دومتغیره مانند $C: I^2 \rightarrow I$ ، $I = [0, 1]$ با ویژگی‌های زیر است.

(الف) به ازای کلیه u و v های متعلق به I ،

$$C(1, u) = C(u, 1) = u,$$

$$C(v, 0) = C(0, v) = 0.$$

(ب) اگر $u < u'$ و $v < v'$ ، آنگاه

$$C(u', v') - C(u', v) - C(u, v') + C(u, v) \geq 0.$$

بنابراین یک مفصل دوبعدی، یک تابع توزیع توأم روی فضای دوبعدی با توزیع‌های حاشیه‌ای یکنواخت است.

قضیه ۱.۱. (اسکلار^۲ [۱۰]) اگر X و Y دو متغیر تصادفی با تابع توزیع توأم F و حاشیه‌ای‌های

Schweizer¹
Sklar²

$F_1(x)$ و $F_2(y)$ باشند، آنگاه تابع مفصل C وجود دارد به قسمی که به ازای هر دو عدد حقیقی x و y

$$F(x, y) = C(F_1(x), F_2(y)),$$

در این حالت اگر F_1 و F_2 پیوسته باشند، آنگاه مفصل C منحصر به فرد بوده و به ازای کلیه u و v های متعلق به بازه به صورت زیر به دست می آید:

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)).$$

لم ۱۰۱. فرض کنیم $C(u, v)$ تابع مفصل توزیع $F(x, y)$ باشد با ویژگی هایی که در تعریف (۱.۱) آمده است. آنگاه می توان برخی از ضرایب همبستگی این مفصل را به صورت زیر تعریف کرد.
(الف) ضریب همبستگی اسپیرمن

$$\rho_s = 12 \int_0^1 \int_0^1 [C(u, v) - uv] dudv = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3.$$

(ب) ضریب همبستگی تاو کندال

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 = 1 - 4 \int_0^1 \int_0^1 \left[\frac{\partial C}{\partial u} \frac{\partial C}{\partial v} \right] dudv.$$

۲ دست آوردهای پژوهش

فرض کنید $C(u, v)$ یک مفصل باشد که $c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$ تابع چگالی آن است. قصد داریم بر اساس مفصل C و تغییر یکی از توزیع های حاشیه ای آن، مفصل جدیدی که آن را با $C^*(u, v)$ نمایش می دهیم، تولید کنیم. برای این منظور چنین عمل می کنیم:

$$f_{V|U=u}(v) = \frac{c(u, v)}{f_U(u)} = c(u, v). \quad (1)$$

با توجه به اینکه $f_U(u)$ تابع چگالی حاشیه ای مفصل $c(u, v)$ است، برابر یک می شود. فرض کنید $G(u)$ تابع توزیع یک متغیر تصادفی با تکیه گاه $(0, 1)$ باشد که تابع چگالی آن $g(u)$ است. داریم:

$$f_{U,V}^*(u, v) = f_{V|U=u}(v)g(u) = c(u, v)g(u). \quad (2)$$

که در آن f^* تابع چگالی توام دو متغیر تصادفی U و V است که دارای تکیه‌گاه مشترک $(0, 1)$ هستند. اکنون قصد داریم مفصل این توزیع توام جدید (یعنی $C^*(u, v)$) را بیابیم. ابتدا توزیع توام (U, V) را محاسبه می‌کنیم:

$$F_{U,V}^*(u, v) = \int_0^u \int_0^v c(s, t)g(s)dsdt = \int_0^u g(s)C_1(s, v)ds, \quad (۳)$$

که در آن $C_1(u, v) = \frac{\partial}{\partial u}C(u, v)$. حال به محاسبه تابع توزیع‌های حاشیه‌ای می‌پردازیم:

$$F_U^*(u) = F_{U,V}^*(u, 1) = G(u), \quad (۴)$$

$$F_V^*(v) = F_{U,V}^*(1, v) = \int_0^1 g(s)C_1(s, v)ds = \int_0^{G^{-1}(v)} C_1(G^{-1}(u), v)du. \quad (۵)$$

اکنون مفصل C^* متناظر با F^* را به دست می‌آوریم:

$$C^*(u, v) = F_{U,V}^*(F_U^{*-1}(u), F_V^{*-1}(v)) = \int_0^{G^{-1}(u)} C_1(s, F_V^{*-1}(v))g(s)ds. \quad (۶)$$

در نهایت با تغییر متغیر $G(s) = w$ تابع مفصل جدید به صورت زیر حاصل می‌شود:

$$C^*(u, v) = \int_0^u C_1(G^{-1}(w), F_V^{*-1}(v))dw. \quad (۷)$$

در ادامه تابع چگالی مفصل C^* را محاسبه می‌کنیم:

$$c^*(u, v) = \frac{\partial^2}{\partial u \partial v} C^*(u, v) = \frac{c(G^{-1}(u), F_V^{*-1}(v))}{\int_0^1 c(G^{-1}(u), F_V^{*-1}(v))du}. \quad (۸)$$

گزاره ۱.۰۲. فرض کنید (U, V) دارای مفصل ارائه شده به صورت (۷) باشد، ضرایب همبستگی برای مفصل مورد نظر به صورت زیر است.

(الف) ضریب همبستگی پیرسون

$$\rho_s = 3 - 12 \int_0^1 \int_0^1 C_1(u, v)G(u)g(u)f_V^*(v)dudv, \quad (۹)$$

(ب) ضریب همبستگی تاو کندال

$$\tau = 4 \left[\left(E_C(C(U, V)g'(U)) - E_C(K(U, V)g(U)) \right) \right] - 1,$$

که در آن $K(s, t) = \int_0^s C(r, t)g'(r)dr$ ، g' مشتق تابع g و E_C امید ریاضی تحت مفصل C است.

۳ حالت خاص مفصل پیشنهادی برای مفصل FGM

(الف) محاسبه تابع مفصل تابع مفصل FGM به صورت زیر در نظر بگیرید:

$$C(u, v) = uv(1 + \theta(1 - u)(1 - v)) \quad \theta \in [0, 1]. \quad (10)$$

یکی از محدودیت‌های مهم این خانواده، ضریب همبستگی محدود به بازه $[-\frac{1}{3}, \frac{1}{3}]$ است. در همین راستا محققین زیادی برای اصلاح کردن این توزیع تلاش کرده‌اند [۶]. فرض کنید $G(u)$ یک تابع توزیع پیوسته بر بازه $(0, 1)$ و $g(u)$ تابع چگالی متناظر با آن باشد. آنگاه تابع چگالی و تابع توزیع جدید به قرار زیر است:

$$f_{U,V}^*(u, v) = g(u)(1 + \theta(1 - 2u)(1 - 2v)), \quad u, v \in (0, 1),$$

$$F_{U,V}^*(u, v) = vG(u) + \theta v(1 - v) \left[G(u) - 2 \int_0^u sg(s)ds \right].$$

تابع توزیع حاشیه‌ای‌ها به صورت زیر است:

$$F_U^*(u) = G(u), \quad F_V^*(v) = v(1 + k_1\theta(1 - v)), \quad (11)$$

که در آن $E_G(U) = \int_0^1 ug(u)du$ و $k_1 = 1 - 2E_G(U)$.

با انجام محاسبات ساده ریاضی، تابع توزیع معکوس $F_V^{*-1}(v)$ به صورت زیر حاصل خواهد شد:

$$F_V^{*-1}(v) = \frac{1}{2k_1} \left[1 - \sqrt{1 - 4k_1(1 - k_1)v} \right], \quad (12)$$

که در آن $k_2 = \frac{k_1\theta}{1+k_1\theta}$. به کمک قضیه اسکالار، مفصل موردنظر رابطه (۷) به صورت زیر حاصل می‌شود:

$$C^*(u, v) = uF_V^{*-1}(v) \left(1 + \theta(1 - F_V^{*-1}(v)) \left[1 - \frac{2}{u} \int_0^u G^{-1}(s) ds \right] \right).$$

ب) ضریب همبستگی اسپیرمن مفصل پیشنهادی با استفاده از رابطه (۹) و اینکه $C_1(u, v) = v(1 + \theta(1 - v)(1 - 2u))$ و $f_V^*(v) = 1 + k_1\theta(1 - v)$ ، ضریب همبستگی اسپیرمن مفصل پیشنهادی تحت $G(u)$ به صورت زیر حاصل می‌شود:

$$\rho_s = 2\theta \int_0^1 ug(u)(2G(u) - 1)du. \quad (13)$$

ج) تولید داده تصادفی برای تولید داده از مفصل پیشنهادی، الگوریتم زیر پیشنهاد می‌شود:

گام اول u و v^* را از توزیع یکنواخت $(0, 1)$ تولید کنید.

گام دوم قرار دهید $u^* = G^{-1}(u)$.

گام سوم $v^* = \frac{1}{2k_2} \left[1 - \sqrt{1 - 4k_2(1 - k_2)v^*} \right]$ ، که در آن $k_2 = \frac{k_1\theta}{1+k_1\theta}$.

گام چهارم $v = v^*(1 + \eta(1 - v^*))$ که در آن $\eta = \theta(1 - 2E_G(U))$.

در این صورت (u, v) یک نمونه از مفصل C^* است.

مثال ۱.۳. مثال خاص با تابع توزیع بتا

فرض کنید $g(u)$ را تابع چگالی توزیع $Beta(\alpha, \beta)$ یعنی

$$g(u) = \frac{1}{B(\alpha, \beta)} u^{\alpha-1} (1-u)^{\beta-1} \quad 0 < u < 1,$$

در نظر بگیریم، داریم:

$$f^*(u, v) = g(u)c(u, v) = \frac{1}{B(\alpha, \beta)} u^{\alpha-1} (1-u)^{\beta-1} [1 + \theta(1 - 2u)(1 - 2v)].$$

تابع توزیع جدید به صورت زیر خواهد شد:

$$F^*(u, v) = vG(u) \left(1 + \theta(1 - v) \left[1 - \frac{2\alpha}{\alpha + \beta} \frac{G^*(u)}{G(u)} \right] \right),$$

که در آن $G(u)$ و $G^*(u)$ به ترتیب تابع توزیع $Beta(\alpha, \beta)$ و $Beta(\alpha + 1, \beta)$ هستند. توزیع‌های حاشیه‌ای به صورت زیر است:

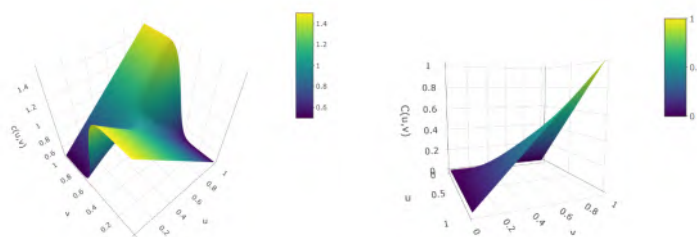
$$F_U^*(u) = G(u), \quad F_V^*(v) = v(1 + \eta(1 - v)),$$

که در آن $\eta = \frac{\theta(\beta - \alpha)}{\beta + \alpha}$. تابع مفصل آن به صورت زیر حاصل می‌شود:

$$C^*(u, v) = uF_V^{*-1}(v) \left(1 + \theta(1 - F_V^{*-1}(v)) \left[1 - \frac{v}{u} \int_0^u G^{-1}(w) dw \right] \right),$$

که در آن $F_V^{*-1}(v) = \frac{1}{2k_v} \left[1 - \sqrt{1 - 4k_v(1 - k_v)v} \right]$ و $k_v = \frac{\eta}{1 + \eta}$. ضریب همبستگی اسپیرمن C^* به صورت زیر به دست می‌آید:

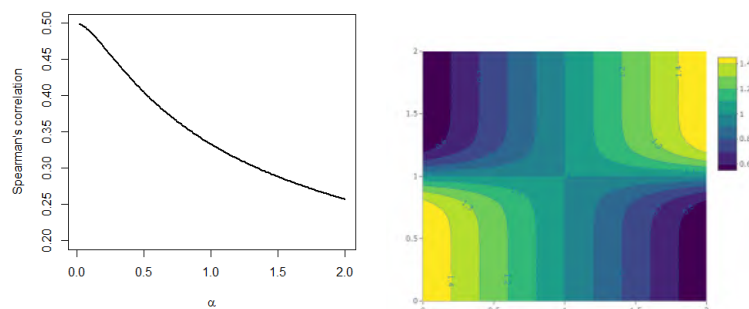
$$\rho_s^* = 2\theta \left[2 \int_0^1 G^{-1}(u) du - \frac{\alpha}{\alpha + \beta} \right].$$



شکل ۱: نمودار سه بعدی مفصل (سمت راست) و چگالی (سمت چپ) مفصل جدید حاصل از مفصل FGM با $\theta = 0.5$ و توزیع بتا با پارامترهای شکل $\alpha = 0.1$ و $\beta = 0.1$.

تذکر ۱.۳. به روش عددی می‌توان نشان داد که اگر $\alpha = \beta$ و $\alpha \rightarrow 0$ آن‌گاه $\rho_s^* \rightarrow \frac{\theta}{4}$.

شکل ۱، نمودار سه بعدی تابع مفصل جدید (سمت راست) و تابع چگالی مفصل جدید (سمت چپ) حاصل از مفصل FGM با $\theta = 0.5$ و توزیع بتا با پارامترهای شکل $\alpha = 0.1$ و $\beta = 0.1$ را نشان می‌دهد. در شکل ۱، نمودار کانتور (سمت راست) و نمودار تغییر ضریب θ در ضریب همبستگی اسپیرمن مفصل جدید (سمت چپ) حاصل از مفصل FGM با $\theta = 0.5$ و توزیع بتا با پارامترهای شکل $\alpha = 0.1$ و $\beta = 0.1$ را نشان می‌دهد که نشان دهنده افزایش محدوده ضریب همبستگی اسپیرمن تا $\frac{1}{4}$ است.



شکل ۲: نمودار کانتور (سمت راست) و تغییر ضریب تتا در ضریب همبستگی اسپیرمن (سمت چپ) مفصل جدید حاصل از مفصل FGM با $\theta = 0.5$ و توزیع بتا با پارمترهای شکل $\alpha = 0.1$ و $\beta = 0.1$.

مراجع

- [1] Bedford, T., and Wilson, K. J. (2014). On the construction of minimum information bivariate copula families. *Annals of the Institute of Statistical Mathematics*, **66**(4), 703-723.
- [2] Bekrizadeh, H. (2021). Generalized Family of Copulas: Definition and Properties. *Thailand Statistician*, **19**(1), 162-177.
- [3] Cuadras, C. M. (2009). Constructing copula functions with weighted geometric means. *Journal of Statistical Planning and Inference*, **139**(11), 3766-3772.
- [4] Joe, H. (2014). *Dependence Modeling With Copulas*. CRC press.
- [5] Lai, C.D., and Xie, M. (2006). *Stochastic Ageing and Dependence for Reliability*. Springer, Science and Business Media, New York.
- [6] Mari, D.D., and Kotz, S. (2001). *Correlation and Dependence*. (Vol. 518). London: Imperial College Press.

- [7] Nelsen, R. B. (2006). *An Introduction to Copulas*. Lectures Notes in Statistics, 139, Springer-Verlag, New York.
- [8] Roy, A., Ghosh, A.K., Goswami, A. and Murthy, C. A. (2022). Some New Copula Based Distribution-free Tests of Independence among Several Random Variables. *Sankhya A*. **84**, 556–596.
- [9] Schweizer, B. (2007). *Introduction to copulas*. Journal of Hydrologic Engineering. 12 (4), 346–346.
- [10] Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris, **8**, 229–231.



هفتمین سمینار

نظریه مفصل و کاربردهای آن

۱۹ و ۲۰ بهمن ۱۴۰۱



مدل سازی داده‌های چندمتغیره طولی با استفاده از توابع مفصل جفتی واین

لؤلؤ، م.ص^۱ آخوند، م.ر^۲ احمدی انگالی، ک^۳ برازجانی، ف^۴

^{۱،۳} گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی جندی شاپور اهواز، اهواز، ایران

^۲ گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران اهواز، اهواز، ایران

^۴ مرکز تحقیقات تغذیه و بیماری‌های متابولیک، دانشگاه علوم پزشکی جندی شاپور اهواز، اهواز، ایران

چکیده

در برخی مطالعات پزشکی ممکن است چندین اندازه‌گیری بر روی هر بیمار داشته باشیم. گاهی این داده‌های طولی ممکن است برای چندین متغیر پاسخ اندازه‌گیری شود، در این حالت اگر چه می‌توان پاسخ‌ها را به صورت مجزا مدل‌بندی کرد اما چنین رویکردی موجب کاهش توان و کارایی در برآورد اثرات متغیرهای کمکی روی متغیر پاسخ می‌گردد. در چنین مدل‌هایی علاوه بر تحلیل وابستگی بین اندازه‌های مکرر مربوط به هریک از متغیرهای پاسخ، وابستگی بین پاسخ‌ها نیز باید مدل شود. از جمله روش‌هایی که در سال‌های اخیر توجه بسیاری از محققان را برای مدل‌سازی داده‌های چند متغیره به خود جلب کرده

¹mslbiostat@gmail.com

²Mr.Akhoond@scu.ac.ir

³kzfir4@gmail.com

⁴fa.borazjani@gmail.com

است، مدل‌سازی داده‌ها با استفاده از تابع مفصل است. از مهمترین مزیت‌های بکارگیری تابع مفصل نسبت به مدل‌سازی چند متغیره طولی داده‌ها به روش کلاسیک این است می‌توان علاوه بر توزیع نرمال هر توزیع دیگری غیر از نرمال را به عنوان توزیع‌های حاشیه‌ای در نظر گرفت. همچنین توزیع‌های حاشیه‌ای حتی می‌توانند توزیع‌های متفاوتی داشته باشند. در شرایطی که داده‌ها ساختاری چند متغیره داشته باشند یکی از راه‌های تشکیل توزیع‌های چندمتغیره استفاده از مفصل‌های جفتی و این است. ما در این مطالعه با استفاده از تابع مفصل‌های مختلف به کمک مفصل‌های جفتی و این ساختار طولی چندمتغیره‌ای را تشکیل می‌دهیم و این مدل‌ها را با مدل حاصل از از برازش تابع مفصل نرمال چند متغیره مقایسه می‌کنیم. سپس بهترین مدل را با استفاده از معیار اطلاع آکائیک معرفی کرده و در پایان مدل ارائه شده را بر روی داده‌های برآورد اثر تغذیه بر رشد نوزادان به کار خواهیم گرفت.

کلمات کلیدی: اندازه‌گیری طولی، تابع مفصل نرمال چندمتغیره، مفصل‌های جفتی و این، رشد نوزاد، تغذیه نوزاد.

۱ پیش‌گفتار

امروزه بخش قابل توجهی از طرح‌های مطالعاتی در علوم مختلف، از جمله آزمایش‌های کلینیکی و اطلاعات وابسته به موجودات زنده، ایجاب می‌کند که واحدهای آزمایش در زمان‌ها و یا موقعیت‌های مشخصی، به منظور کنترل تفاوت‌ها بین آزمودنی‌ها اندازه‌گیری شوند، در این نوع آزمایش‌ها که هر واحد آزمایشی به طور متناوب اندازه‌گیری می‌شود معمولاً از اصطلاح اندازه‌گیری مکرر و در حالتی که هر واحد نمونه‌گیری در زمان‌های متفاوت بررسی شود، تحلیل داده‌های طولی استفاده می‌گردد. هدف اولیه مطالعات طولی، بررسی تغییرات متغیر پاسخ در طول زمان و نیز استخراج عامل‌های تاثیرگذار بر این تغییرات می‌باشد. به دلیل وجود اندازه‌گیری‌های تکراری، امکان ارزشیابی تغییرات درون هر واحد نیز فراهم می‌شود که این امکان در مطالعات مقطعی که پیامد تنها در یک زمان مشخص اندازه‌گیری می‌شود، وجود ندارد. در این حالت برخلاف سایر آزمایش‌ها بجای مشاهدات مستقل با مجموعه‌ای از مشاهدات وابسته سروکار خواهیم داشت. به دلیل آنکه اندازه‌های مکرر از دو بعد زمان و واحدهای آزمایشی تشکیل می‌شوند استفاده از مدل‌های مناسب که اثر مؤلفه‌های مذکور

را در توصیف این داده‌ها به درستی لحاظ کند ضروری است. مسئله مهم و کاربردی دیگر که به ویژه در تحلیل داده‌های بالینی طولی مورد توجه قرار می‌گیرد، تحلیل همزمان یا چند متغیره‌ی طولی است. در این نوع مطالعات گاهی موقعیت‌هایی وجود دارد که در آن‌ها، هدف محقق آن است تا چندین متغیر را به دلیل ارتباط و همبستگی بین آنها، بطور همزمان مورد بررسی قرار دهد.

اگر در تحلیل داده‌ها با مواردی مواجه شویم که لازم است چند متغیر پاسخ به صورت همزمان از روی تعدادی متغیر کمکی پیش‌بینی گردند، این گونه مدل‌ها را مدل‌های چند متغیره می‌نامند. اگر چه می‌توان پاسخ‌ها را به صورت مجزا مدل‌بندی کرد اما چنین رویکردی موجب کاهش توان و کارایی در پیش‌بینی پاسخ‌ها می‌گردد. مزایای مدل‌های توام که وابستگی بین پاسخ‌ها را در نظر می‌گیرد عبارتست از کنترل بهتر خطای نوع اول و افزایش کارایی در برآورد اثرات متغیرهای کمکی روی متغیر پاسخ است.

از جمله روش‌هایی که در سال‌های اخیر توجه بسیاری از محققان را برای مدل‌سازی داده‌های چندمتغیره به خود جلب کرده است، مدل‌سازی داده‌ها با استفاده از تابع مفصل است. ساختار توزیع‌های توام و بررسی ویژگی‌های آن با توزیع‌های حاشیه‌ای معلوم در صورت عدم برقراری فرض استقلال بین متغیرها کار پیچیده‌ای است. تابع مفصل برای ساخت یا بررسی توزیع‌های چند متغیره وقتی که توزیع‌های حاشیه‌ای یک متغیره و ساختار وابستگی آن‌ها معلوم است، بسیار مفید است.

کلو و پایوا (۲۰۰۹) به مرور مطالعات انجام شده در زمینه مدل‌های رگرسیونی با استفاده از تابع مفصل پرداختند. دی لئون و وو (۲۰۱۱) با استفاده از تعریف متغیر پنهان برای پاسخ گسسته، مفصل نرمال را برای ساخت مدل توأم بکار بردند. نیکولوپولس و جو (۲۰۱۵) با استفاده از تابع مفصل‌های نرمال، گامبل و تی استیودنت طرح‌های عاملی برای سه پرسشنامه نگرش زندگی، نگرش در مورد محیط زیست و نگرش در مورد علم را مدل‌سازی کردند و نتایج هر کدام را برای مفصل کاراتر ارایه نمودند. وثنیج و همکاران (۲۰۱۳) از تابع مفصل دومتغیره نرمال برای مدل‌سازی داده‌های پیوسته طولی با اثر تصادفی نرمال استفاده نمود. وو و دی لیون (۲۰۱۴) یک مدل دومتغیره گسسته-پیوسته با به کارگیری تابع مفصل نرمال را

جهت برآورد اثر چند دز یک دارو بر پاسخ‌های رشد وزنی و نبود نقص عضو چندین جنین موش ارایه نمود. روی (۲۰۱۶) نیز با تعمیم مطالعه وثنیج (۲۰۱۳) انواع حالت‌های ممکن اثر تصادفی در مدل طولی با تابع مفصل دومتغیره نرمال و همچنین با توزیع‌های حاشیه‌ای متفاوت ارائه و به کمک شبیه سازی داده‌ها مدل‌ها را مقایسه نمود. با توجه اینکه مطالعات گذشته محدود به مدل‌سازی داده‌های طولی در حالت دو متغیر با استفاده توابع مفصل است و این مدل‌ها در حالت طولی چندمتغیره مورد بحث قرار نگرفته‌اند، در مطالعه حاضر قصد داریم با استفاده از تابع مفصل‌های فرانک، کلایتون و FGM و با کمک مفصل‌های جفتی واین یک مدل سه‌متغیره طولی را تشکیل دهیم. سپس مدل بدست آمده را با مدل طولی بدست آمده با استفاده مفصل نرمال چندمتغیره مورد مقایسه قرار دهیم. در مطالعه حاضر علاوه بر بکارگیری تابع مفصل نرمال برای پاسخ‌های پیوسته، از مفصل‌های فرانک، کلایتون و FGM نیز برای مدل‌سازی این پاسخ‌ها با اندازه‌گیری طولی استفاده شده‌است. همچنین در این تحقیق به کمک توابع مفصل جفتی واین مدل سه متغیره به مدل‌های دومتغیره تجزیه شده و این مدل با مدل حاصل از تابع مفصل نرمال سه متغیره مقایسه مورد مقایسه قرار گرفت. بدین منظور در بخش دوم مروری بر روش تحلیل مطالعات طولی و معرفی تابع مفصل و ویژگی‌های آن پرداخته می‌شود. در بخش سوم مدل‌سازی داده‌های سه متغیره طولی با استفاده از تابع مفصل مورد بررسی قرار می‌گیرد. در بخش چهارم مثال کاربردی رشد کودک با استفاده از این مدل بررسی می‌شود و در بخش پنجم به بحث و نتیجه گیری در مورد نتایج بدست آمده خواهیم پرداخت.

۲ مروری بر روش تحلیل داده‌های طولی و تابع مفصل

تحلیل ساختار همبستگی داده‌های طولی

وابستگی درون داده‌های طولی موجب می‌گردد که مشاهدات ثبت شده برای یک فرد در زمان یا حالات مختلف و یا بین افراد یک خوشه، شبیه به هم باشد. به دلیل همبستگی بین

پاسخ‌های هر آزمودنی، روش‌های برآوردیابی ویژه‌ای برای تحلیل آماری مورد نیاز است که منجر به استنباط‌های معتبری گردد. با توجه به اهداف نهایی جمع‌آوری داده‌های طولی و نیز طبق نظر تحلیل‌گر در برخورد با پارمترهای همبستگی، مدل‌های متفاوتی مانند حاشیه‌ای، اثرهای تصادفی و انتقال توسط محققان مختلف توسعه یافته و در مباحث کاربردی مورد استفاده قرار گرفته‌اند. در این مطالعه روش استفاده از اثر تصادفی برای تحلیل ساختار همبستگی درون فردی بکار گرفته شده است. در مدل‌های خطی با اثرهای تصادفی، عموماً فرض بر آن است که برخی ضرایب رگرسیونی برای آزمودنی‌ها به دلایل متفاوت، به خود آزمودنی وابسته بوده و دارای توزیع جداگانه می‌باشد. به عبارت دیگر، همبستگی بین مشاهدات در داده‌های تکراری یا داده‌های موجود در یک خوشه را می‌توان ناشی از داشتن تأثیر تصادفی یکسان در فرد یا یک خوشه دانست که این اثر در خوشه‌ی دیگر متفاوت است. در حقیقت همین ویژگی است که موجب ایجاد همبستگی در اندازه‌های ثبت شده از یک فرد یا یک خوشه می‌گردد (فتیزماوریک، و همکاران، ۲۰۰۹). در حالت کلی یک مدل خطی با اثر تصادفی به صورت زیر معرفی می‌شود:

$$Y_{N \times 1} = X_{N \times p}^T \beta_{p \times 1} + Z_{N \times q}^T b_{q \times 1} + \varepsilon_{N \times 1}$$

که در آن Y بردار مقادیر مشاهده شده متغیر پاسخ، X ماتریس پرتبه ستونی شامل متغیرهای مستقل یا کمکی، β بردار پارمترهای مجهول اثرهای ثابت، b بردار نمایانگر اثرهای تصادفی و ماتریس Z شامل برخی متغیرهای توضیحی و نشانگر که با توجه به فرض‌های موجود درباره اثرهای تصادفی b در مدل ساخته می‌شود. ε نیز برداری از مؤلفه‌های خطای تصادفی مدل است که مستقل از b توزیع می‌گردد. استنباط آماری پارمترهای مدل بر اساس تابع درستنمایی حاشیه‌ای

$$L(\beta, \Theta | X, Y) = \int_{R^q} f(y|x, b, \beta) g(b|\Theta) db$$

انجام می‌شود که معمولاً در آن تابع چگالی اثرهای تصادفی $g(b|\Theta)$ و چگالی $f(y|x, b, \beta)$ نرمال فرض می‌شود. همانطور که بیان شد در مدل‌سازی داده‌های چندمتغیره طولی با تابع مفصل می‌توان توزیع $f(y|x, b, \beta)$ را توزیعی غیرنرمال فرض کرد. یکی از متداول‌ترین

روش‌های برآوردیابی پارمترهای مجهول با اثرهای آمیخته، روش حداکثر درستنمایی لی و نلدر (۱۹۹۶) است. استفاده از این روش منجر به حل انتگرال‌های پیچیده در تابع درستنمایی حاشیه‌ای، بخصوص در حالت غیرنرمال بودن توزیع اثرهای تصادفی شده که به روش‌های تکراری عددی پیشرفته نیاز دارد. در مطالعات اخیر روش مربع بندی گوس-نیوتن به عنوان یک روش موثر برای تقریب عددی انتگرال‌های پیچیده به منظور یافتن برآورد حداکثر درستنمایی پارمترها در مدل‌های با اثر تصادفی نرمال ارائه شده است.

فرض کنید یک نمونه تصادفی N تایی داشته باشیم که $y_k \sim F_k(y_k; \mathbf{x}_k, \beta_k)$ و \mathbf{x}_k به عنوان بردار متغیرهای کمکی و β_k به عنوان بردار پارمترها باشند. در این حالت برای $k = 1, 2$ تابع چگالی‌های حاشیه‌ای به صورت زیر بدست می‌آید:

$$f_k(y_k; \mathbf{x}_k, \beta_k) = \partial F_k(y_k; \mathbf{x}_k, \beta_k) / \partial y_k$$

تابع چگالی توام Y_1 و Y_2 به صورت زیر محاسبه می‌شود:

$$f(y_1, y_2; \theta) = \partial^2 C(F_1(y_1), F_2(y_2); \theta) / \partial y_1 \partial y_2 = c_{12}(F_1(y_1), F_2(y_2); \theta) f_1(y_1) f_2(y_2)$$

که در آن $c_{12}(F_1(y_1), F_2(y_2); \theta)$ تابع چگالی مفصل است و از طریق مشتق‌گیری از تابع توزیع مفصل به فرم زیر بدست می‌آید:

$$c_{12}(F_1(y_1; \mathbf{x}_1, \beta_1), F_2(y_2; \mathbf{x}_2, \beta_2); \theta) = \partial^2 C(F_1(y_1; \mathbf{x}_1, \beta_1), F_2(y_2; \mathbf{x}_2, \beta_2); \theta) / \partial F_1 \partial F_2$$

برآورد پارمترها

برای نوشتن تابع درستنمایی ابتدا تابع درستنمایی به شرط اثرات تصادفی نوشته می‌شود و سپس با انتگرال‌گیری بر روی اثرات تصادفی تابع درستنمایی برای پارامترهای مدل بدست می‌آید:

$$L(\Theta) = \prod_{i=1}^n \int \prod_{t=1}^T f_{Y_{it1}, Y_{it2} | b_i}(\cdot | \cdot) f(b_i) db_i$$

و در نتیجه لگاریتم تابع درست‌نمایی به صورت زیر بدست می‌آید

$$l(\Theta) = \log L(\theta) \quad (1)$$

$$= \sum_{i=1}^N [\log f(y_{i1}; \mathbf{x}_{i1}, \beta_1) + \log f(y_{i2}; \mathbf{x}_{i2}, \beta_2) + \log c_{12}(F_1(y_{i1}; \mathbf{x}_{i1}, \beta_1), F_2(y_{i2}; \mathbf{x}_{i2}, \beta_2); \theta)]$$

که در آن Θ بردار کلیه پارامترهای مدل می‌باشد. با قرار دادن $l'(\Theta) = \partial l(\Theta) / \partial \Theta$ به عنوان معادلات امتیاز و $l''(\Theta) = \partial^2 l(\Theta) / \partial \Theta \partial \Theta^T$ به عنوان ماتریس هسین، برآورد ماکسیمم درست‌نمایی پارامترها به وسیله حل معادله درست‌نمایی $l'(\theta) = 0$ به دست می‌آید. این معادلات در حالت کلی غیرخطی هستند اما با استفاده از روش‌های گوس-نیوتن به راحتی محاسبه می‌شوند. تحت برقراری شرایط نظم $\hat{\Theta}$ سازگار است و به طور مجانبی دارای توزیع نرمال به صورت زیر می‌باشد:

$$\sqrt{N}(\hat{\Theta} - \Theta) \rightarrow N \left[0, - \left(\frac{1}{N} \frac{\partial^2 l(\hat{\Theta})}{\partial \hat{\Theta} \partial \hat{\Theta}^T} \right)^{-1} \right]$$

مدل خطی با اثر تصادفی برای داده‌های طولی

فرض کنید Y_{itk} توزیع حاشیه‌ای k ام ($k = 1, 2$) برای فرد i ام ($i = 1, \dots, n$) در زمان t ام ($t = 1, \dots, T$) باشد. مدل خطی با اثر تصادفی به صورت زیر ارائه می‌شود:

$$Y_{itk} = \mathbf{X}_{itk}^T \beta_k + \mathbf{Z}_{itk}^T \mathbf{b}_{ik} + \epsilon_{itk} \quad (2)$$

که \mathbf{X}_{itk}^T بردار متغیرهای مستقل (کمکی) و β_k بردار ضرایب رگرسیونی متناظر با آن‌ها، \mathbf{Z}_{itk} بردار متغیرهای کمکی برای اثرات تصادفی و \mathbf{b}_{ik} بردار اثرات تصادفی است و ϵ_{itk} خطای مدل و دارای توزیع $N(0, \sigma^2)$ است. ساده‌ترین نوع مدل‌های خطی چند متغیره با اندازه‌گیری طولی با بکارگیری پارامتر یکسان برای هر مدل حاشیه‌ای و فرض $Z_{itk} = 1$ است که به مدل با اثرات تصادفی عرض از مبدا یکسان معروف است. در این حالت مدل (۲) به شکل زیر

ساده می‌شود:

$$Y_{itk} = X_{itk}^T \beta_k + b_i + \epsilon_{itk} \quad (۳)$$

در این مطالعه کلیه مدل‌ها مبتنی بر مدل (۳) ارایه می‌شود و فرض شده است که $b_i \sim N(0, \sigma^2)$ می‌باشد. جهت مشاهده سایر حالت‌های $Z_{itk}^T b_{ik}$ در مدل کلی (۲) به کتاب لرد و ویر (۱۹۸۲) یا فیتزماوریک، دیویدین و همکاران (۲۰۰۹) می‌توان مراجعه کرد. در مدل‌های خطی با اندازه گیری طولی مشاهدات فرض می‌شود توزیع متغیر وابسته به شرط اثر تصادفی دارای توزیع نرمال است. طبق مدل (۳)، $Y_{itk}|b_i \sim N(\mu_{itk}(b_i), \sigma^2)$ ، که در آن $\mu_{itk} = X_{itk}^T \beta_{ik} + b_i$ همانطور که بیان شد از جمله مهمترین مزیت‌های استفاده از تابع مفصل برای مدل‌سازی این است که توزیعی حاشیه‌ای برای متغیر پاسخ را می‌توان هر توزیع دلخواه در نظر گرفت یعنی توزیع $Y_{itk}|b_i$ می‌تواند هر توزیع دیگری غیر از نرمال باشد. این در حالی است که یکی از پیش فرض‌های اصلی مدل‌های خطی برقراری فرض نرمال برای توزیع متغیرهای وابسته است. در این مطالعه توزیع لجستیک و نرمال به عنوان توزیع حاشیه‌های تابع مفصل استفاده شده است. در مدل دو متغیره فرض می‌شود:

$$Y_{it1}|b_i \sim N(\mu_{it1}(b_i), \sigma_1^2) \quad (۴)$$

$$Y_{it2}|b_i \sim \text{logistic}(\mu_{it2}(b_i), \sigma_2) \quad (۵)$$

که در (۴)، $\mu_{it1} = X_{it1}^T \beta_1 + b_i$ و به عنوان میانگین توزیع نرمال شرطی و σ_1 انحراف معیار آن و در (۵)، $\mu_{it2} = X_{it2}^T \beta_2 + b_i$ به عنوان پارامتر مکان و σ_2 پارامتر مقیاس توزیع لجستیک شرطی است.

تابع درستنمایی با استفاده از توابع مفصل نرمال، فرانک، کلایتون و حاصلضرب برای مدل‌های دو متغیره

تابع توزیع شرطی مفصل نرمال از رابطه زیر به دست می‌آید:

$$F_{Y_{it1}, Y_{it2} | b_i}(\cdot | \cdot) = \Phi_2[\Phi^{-1}(u_{it1}(b_i)), \Phi^{-1}(u_{it2}(b_i)); \theta]$$

که در آن برای

$$u_{itk}(b_i) = F_{Y_{itk} | b_i}(y_{itk} | b_i)$$

$u_{itk}(b_i)$ ها توزیع شرطی یکنواخت در بازه‌ی $(0, 1)$ هستند و $F_{Y_{itk} | b_i}(y_{itk} | b_i)$ تابع توزیع تجمعی شرطی با استفاده از حاشیه‌های (۴) و (۵) به دست می‌آید. θ پارامتر وابستگی بین Y_{it1}, Y_{it2} برای تابع مفصل نرمال است (که در این تابع مفصل همان ضریب همبستگی پیرسون است) که به صورت زیر تعریف می‌شود:

$$\theta = \rho = \text{corr}(U_{it1}(b_i), U_{it2}(b_i) | b_i)$$

و $\rho = 0$ اگر و تنها اگر Y_{it1}, Y_{it2} از هم مستقل باشند. b_i اثر تصادفی مربوط به فرد i ام است که فرض می‌شود $b_i \sim N(0, \sigma^2)$ باشد. تابع چگالی شرطی دو متغیره Y_{it1}, Y_{it2} با استفاده از مفصل نرمال به شکل زیر به دست می‌آید (وثنیج و همکاران، ۲۰۱۳).

$$f_{Y_{it1}, Y_{it2} | b_i}(\cdot | \cdot) = \frac{\phi_2(\Phi^{-1}(u_{it1}(b_i)), \Phi^{-1}(u_{it2}(b_i)); \theta)}{\phi(\Phi^{-1}\{u_{it1}(b_i)\})\phi(\Phi^{-1}\{u_{it2}(b_i)\})} f_{Y_{it1} | b_i}(y_{it1} | b_i) \times f_{Y_{it2} | b_i}(y_{it2} | b_i) \quad (۶)$$

که در آن برای $k = 1, 2$ ، تابع چگالی شرطی با استفاده از حاشیه‌های (۴) و (۵) است و ϕ_2, ϕ به ترتیب چگالی نرمال استاندارد و چگالی نرمال استاندارد دو متغیره است. حال به کمک رابطه (۱) می‌توان تابع درستنمایی دو متغیره شرطی با استفاده از تابع مفصل نرمال را از طریق تابع چگالی (۶) محاسبه و پارامترهای مدل را به روش گوس نیوتن برآورد نمود.

مشابه (۶) برای سایر توابع مفصل چگالی شرطی آن‌ها از روی تابع چگالی‌های مفصل به دست می‌آید. عبارات (۷)، (۸)، (۹) و (۱۰) به ترتیب چگالی شرطی دو متغیره برای توابع مفصل فرانک، کلایتون، FGM و حاصلضرب را نشان می‌دهد.

$$f_{Y_{it1}, Y_{it2} | B_i}(\cdot | \cdot) = \frac{\theta(1 - e^{-\theta})e^{-\theta[(u_{it1}(b_i) + (u_{it2}(b_i)))]}}{[(1 - e^{-\theta}) - (1 - e^{-\theta(u_{it1}(b_i))})(e^{-\theta(u_{it2}(b_i))} - 1)]^2} f_{Y_{it1} | B_i}(y_{it1} | b_i) \times f_{Y_{it2} | B_i}(y_{it2} | b_i). \quad (7)$$

$$f_{Y_{it1}, Y_{it2} | b_i}(\cdot | \cdot) = [(u_{it1}(b_i))^{-\theta} + (u_{it2}(b_i))^{-\theta} - 1]^{-2-\frac{1}{\theta}} \times [(u_{it1}(b_i))(u_{it2}(b_i))]^{-\theta-1} \times [\theta + 1] \times f_{Y_{it1} | b_i}(y_{it1} | b_i) \times f_{Y_{it2} | b_i}(y_{it2} | b_i) \quad (8)$$

$$f_{Y_{it1}, Y_{it2} | b_i}(\cdot | \cdot) = [1 + \theta(u_{it1}(b_i) - 1)(u_{it2}(b_i) - 1)] f_{Y_{it1} | b_i}(y_{it1} | b_i) \times f_{Y_{it2} | b_i}(y_{it2} | b_i). \quad (9)$$

$$f_{Y_{it1}, Y_{it2} | b_i}(\cdot | \cdot) = f_{Y_{it1} | b_i}(y_{it1} | b_i) \times f_{Y_{it2} | b_i}(y_{it2} | b_i) \quad (10)$$

در کلیه عبارات بالا θ پارامتر وابستگی تابع مفصل، $u_{itk}(b_i) = F_{Y_{itk} | b_i}(y_{itk} | b_i)$ تابع توزیع شرطی حاشیه‌ها و $f_{Y_{itk} | b_i}(y_{itk} | b_i)$ تابع چگالی شرطی حاشیه‌ها برای $k = 1, 2$ است. تابع درستنمایی و برآورد پارامترهای مدل مشابه تابع مفصل نرمال به کمک رابطه (۱) و چگالی‌های شرطی (۷)، (۸)، (۹) و (۱۰) به دست می‌آید.

۳ مدل بندی سه متغیره طولی با استفاده از تابع مفصل

در این بخش دو روش واین و تابع مفصل نرمال سه متغیره برای مدل سازی سه متغیره معرفی و روش برآورد پارامتر به وسیله آن‌ها توضیح داده می‌شود.

روش واین

روش واین با بکارگیری مفصل‌های جفتی از متغیرها، یک کلاس بزرگ‌تر از توزیع‌های چند بعدی را می‌سازد. همانطور که بیان شد برای دو متغیر Y_1, Y_2 تابع توزیع مفصل به شکل زیر

معرفی می شود:

$$F_{Y_1, Y_2}(y_1, y_2; \theta) = C(F_{Y_1}(y_1), F_{Y_2}(y_2))$$

همچنین با توجه به تعریف تابع چگالی شرطی، Y_2 به شرط Y_1 که به صورت زیر نوشته می شود:

$$f_{2|1}(y_2|y_1) = \frac{f_{1,2}(y_1, y_2)}{f_1(y_1)}$$

بنابراین

$$f_{2|1}(y_2|y_1) = \frac{f_{1,2}(y_1, y_2)}{f_1(y_1)} = c_{1,2}(F_1(y_1), F_2(y_2))f_2(y_2) \quad (11)$$

که در آن $c_{1,2}(F_1(y_1), F_2(y_2))$ همان چگالی مفصل دو متغیره است. رابطه بالا را برای هر i, j می توان به فرم کلی زیر نوشت:

$$f_{j|i}(y_j|y_i) = c_{ij}(F_i(y_i), F_j(y_j))f_j(y_j) \quad (12)$$

از طرفی تابع چگالی یک توزیع سه متغیره براساس قانون های احتمال شرطی به صورت زیر نوشته می شود:

$$f_{1,2,3}(y_1, y_2, y_3) = f_1(y_1)f_{2|1}(y_2|y_1)f_{3|1,2}(y_3|y_1, y_2) \quad (13)$$

و عبارت $f_{3|1,2}(y_3|y_1, y_2)$ را می توان از طریق فرم کلی (۹) به فرم زیر بسط داد:

$$\begin{aligned} f_{3|1,2}(y_3|y_1, y_2) &= c_{1,2|3}(F_{1|2}(y_1|y_2), F_{2|2}(y_2|y_2)) f_{3|2}(y_3|y_2) \\ &= c_{1,2|3}(F_{1|2}(y_1|y_2), F_{2|2}(y_2|y_2)) c_{2,3}(F_2(Y_2), F_3(Y_3)) f_3(y_3) \end{aligned} \quad (14)$$

حال با جایگذاری (۱۴) و (۱۱) در عبارت (۱۳) چگالی سه متغیره Y_1, Y_2, Y_3 به شکل زیر به دست می آید:

$$\begin{aligned} f_{1,2,3}(y_1, y_2, y_3) &= f_1(y_1)c_{1,2}(F_1(y_1), F_2(y_2))f_2(y_2) \\ &\quad c_{1,2|3}(F_{1|2}(y_1|y_2), F_{2|2}(y_2|y_2))c_{2,3}(F_2(y_2), F_3(y_3))f_3(y_3). \end{aligned}$$

اگر $y_1|y_2$ را از $y_3|y_2$ مستقل فرض کنیم یعنی فرض کنیم (c_{12}, c_{13}, c_{23}) عبارت بالا به شکل زیر ساده می‌شود.

$$f_{1,2,3}(y_1, y_2, y_3) = f_1(y_1)f_2(y_2)f_3(y_3)c_{12}(F_1(y_1), F_2(y_2))c_{23}(F_2(y_2), F_3(y_3)) \quad (15)$$

در عبارت (۱۵)، c_{12} و c_{23} هر تابع مفصلی می‌توانند باشند و الزامی به یکسان در نظر گرفتن توابع مفصل هر جفت متغیر در این روش وجود ندارد. این خاصیت روش واین را به یک روش بسیار توانمند و ساده در برازش مدل‌های چندبعدی به کمک تابع مفصل تبدیل کرده است. چگالی شرطی سه متغیره به روش واین به کمک عبارت (۱۵) به شکل زیر نوشته می‌شود.

$$\begin{aligned} f(y_{it1}, y_{it2}, y_{it3}|b_i) &= f_{Y_{it1}|b_i}(y_{it1}|b_i)f_{Y_{it2}|b_i}(y_{it2}|b_i)f_{Y_{it3}|b_i}(y_{it3}|b_i) \\ & c_{12}(F_{Y_{it1}|b_i}(y_{it1}|b_i), F_{Y_{it2}|b_i}(y_{it2}|b_i); \theta_{12}) \\ & c_{23}(F_{Y_{it2}|b_i}(y_{it2}|b_i), F_{Y_{it3}|b_i}(y_{it3}|b_i); \theta_{23}) \end{aligned} \quad (16)$$

در این عبارت θ_{12} و θ_{23} به ترتیب پارامتر وابستگی تابع مفصل برای c_{12} و c_{23} است. حال با محاسبه لگاریتم تابع درستنمایی با استفاده از تابع چگالی توام سه متغیره (۱۶) مشابه حالت دومتغیره می‌توان برآورد پارمترها را به روش حداکثر درستنمایی و با کمک روش گوس نیوتن محاسبه کرد.

تابع مفصل سه متغیره نرمال

تابع توزیع مفصل k متغیره نرمال به شکل زیر تعریف می‌شود.

$$C(u_1, \dots, u_k; R) = \Phi_k(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_k))$$

که در آن R ماتریس وابستگی تابع مفصل است و یک ماتریس مربعی $k \times k$ است. فرم ماتریسی تابع چگالی مفصل نرمال چند متغیره نیز به صورت زیر تعریف می‌شود.

$$c(\mathbf{q}; R) = \frac{1}{|R|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \mathbf{q}^T (R^{-1} - I) \mathbf{q} \right\} \quad (17)$$

که در آن q بردار k بعدی و $(q_1 = \Phi^{-1}(F_{Y_1}(y_1)), \dots, q_k = \Phi^{-1}(F_{Y_k}(y_k)))$ معکوس ماتریس وابستگی تابع مفصل R و $|R|$ دترمینان آن است. I ماتریس همانی $k \times k$ است. برای مفصل سه متغیره با $k = 3$ ، ماتریس وابستگی تابع مفصل به شکل زیر تعریف می‌شود.

$$R = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{12} & \rho_{23} & 1 \end{pmatrix}$$

که عناصر ρ_{ij} وابستگی دو به دو مفصل برای هر سه متغیره است وقتی $i, j = 1, 2, 3$ و دترمینان R به فرم زیر به دست می‌آید:

$$|R| = 1 - \rho_{12} - \rho_{13} - \rho_{23} + 2\rho_{12}\rho_{13}\rho_{23}$$

ضمناً فرم باز شده عبارت $\{-\frac{1}{\sqrt{2}}q^T(R^{-1} - I)q\}$ به صورت زیر است.

$$\begin{aligned} & \left\{ -\frac{1}{\sqrt{2}}q^T(R^{-1} - I)q \right\} \\ &= -\frac{1}{\sqrt{2}|R|} \left[\begin{aligned} & q_1^2(1 - \rho_{23}) + q_2^2(1 - \rho_{13}) + q_3^2(1 - \rho_{12}) - \\ & 2q_1q_2(\rho_{12} - \rho_{23}\rho_{13}) - 2q_1q_3(\rho_{13} - \rho_{12}\rho_{23}) - 2q_2q_3(\rho_{23} - \rho_{12}\rho_{13}) \end{aligned} \right] \end{aligned}$$

که با جایگذاری در عبارت (۱۷) فرم غیرماتریسی تابع چگالی سه متغیره نرمال به دست می‌آید.

اگر فرض کنیم $\rho_{ij} = \rho$ برای هر $i, j = 1, 2, 3$ یعنی $R = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$ باشد عبارات بالا به

فرم زیر ساده‌تر می‌شوند:

$$|R| = 1 - 3\rho + 2\rho^3$$

و

$$\left\{ -\frac{1}{\sqrt{2}}q^T(R^{-1} - I)q \right\} = -\frac{1}{\sqrt{2}|R|} \left[(1 - \rho)(q_1^2 + q_2^2 + q_3^2) - 2(\rho - \rho^2)(q_1q_2 + q_1q_3 + q_2q_3) \right]$$

بنابراین تابع چگالی شرطی سه متغیره به کمک تابع مفصل نرمال به شکل زیر نوشته می شود:

$$f_{Y_{ij1}, Y_{ij2}, Y_{ij3} | b_i}(\cdot | \cdot) = c(\mathbf{q}; R, b_i) f_{Y_{ij1} | b_i}(y_{ij1} | b_i) f_{Y_{ij2} | b_i}(y_{ij2} | b_i) f_{Y_{ij3} | b_i}(y_{ij3} | b_i) \quad (18)$$

که در آن $c(\mathbf{q}; R, b_i)$ تابع چگالی شرطی سه متغیره تابع مفصل نرمال طبق (۱۷) با $k = 3$ است و

$$\mathbf{q} = (q_1 = \Phi^{-1}(F_{Y_{it1} | b_i}(y_{it1} | b_i)), q_2 = \Phi^{-1}(F_{Y_{it2} | b_i}(y_{it2} | b_i)), q_3 = \Phi^{-1}(F_{Y_{it3} | b_i}(y_{it3} | b_i)))$$

حال با محاسبه لگاریتم تابع درستنمایی از (۱۵) مشابه حالت دومتغیره می توان برآورد پارامترها را با کمک روش گوس نیوتن بدست آورد.

برای مقایسه مدل ها می توان از معیارهای اطلاع نظیر معیار اطلاع آکائیک استفاده کرد. معیار اطلاع آکائیک با فرمول زیر برای مقایسه مدل ها مورد استفاده قرار گرفت و کمتر بودن مقدار آن حکایت از بهتر بودن مدل برازش شده دارد.

$$AIC = -2 \log L + 2P$$

که در آن L تابع درستنمایی و P تعداد پارامترهای مدل می باشند.

۴ مثال کاربردی

رشد عبارت است از بزرگ شدن اجزای بدن که نقش بسیار مهمی را در سلامت ایفا می کند و از ارکان مهم سلامت جامعه محسوب می شود. استفاده از شاخص های رشد و وضع تغذیه ی کودکان به عنوان معیاری برای سنجش سطح بهداشت یک منطقه بر پایه این واقعیت است که رشد ضعیف برای بیشتر کودکان به معنی انحراف از شرایط محیطی مطلوبی است که باید زمینه رشد و نمو را در کودکان به بهترین شکل فراهم کند. در اکثر مطالعات برای پایش رشد فیزیکی و سلامت کودکان از شاخص های تن سنجی نظیر وزن، قد و دورسر و بویژه وزن کردن مرتب استفاده می شود. پایش منظم قد، وزن و دورسر کودکان موجب کشف به موقع اختلالات رشد و آغاز زود هنگام اقدام های پیشگیری و درمانی می شود. مهم ترین عامل موثر

بر روند رشد جسمانی کودک، تغذیه است. در مراحل اول زندگی شیرمادر بدون شک ایده آل ترین غذا است که علاوه بر فراهم کردن کلیه مواد مغذی کودک، دارای مزایایی از نظر ایمنی و سلامت روانی نیز می باشد. در اکثر مطالعات که اثر تغذیه نوزادان را بر رشد آنان بررسی کرده اند جهت تعدیل مدل پیشنهادی خود و بررسی دقیق تر، فاکتورهایی همانند جنس، رتبه تولد، بعد خانوار، شاخص های سلامت مادر، وضعیت اقتصادی و فرهنگی خانواده را نیز به کار گرفته اند. در این مطالعه کل ۱۹۵ نوزاد متولد شده روستاهای شهرستان بوشهر که طی سال ۹۴ به مرکز بهداشت مراجعه کرده بودند با معیار ورود به مطالعه وزن تولد بیش از ۲۵۰۰ گرم، سن جنینی بین ۳۷ تا ۴۲ هفته و بدون عارضه مادرزادی و مراجعه کامل در طول مطالعه بررسی شدند، در این مدت ۶ مرتبه (بدو تولد، دو ماهگی، چهار ماهگی، شش ماهگی، هفت ماهگی، نه ماهگی) مراجعه داشتند که در مجموع ۱۲۰ کودک مورد مطالعه قرار گرفتند. در هر بار مراجعه قد (به سانتی متر)، وزن (به کیلوگرم) و دورسر (به سانتی متر) کودک با استفاده از متر و ترازوهای مخصوص نوزادان در مرکز بهداشتی اندازه گیری شد. سن کودک در زمان مراجعه به مرکز با دقت روز ثبت گردید. شیرخوران بر اساس نوع تغذیه به سه گروه تقسیم می شوند. گروه اول شیر مادر خواران انحصاری که بر اساس تعریف سازمان بهداشت جهانی (WHO) کودکی است که تنها از شیر مادر و قطره ویتامین در اولین شش ماه زندگی استفاده کرده باشد. دو گروه دیگر شامل شیر خشک خوار و گروهی که از هر دو شیر مادر و شیر خشک تغذیه می کردند.

بنابراین متغیرهای وابسته در این مطالعه شامل وزن نوزاد به کیلوگرم، قد نوزاد به سانتی متر، دورسر نوزاد به سانتی متر می باشد. متغیرهای مستقل شامل سن مادر به سال، سن پدر به سال، شاخص توده ی بدنی (BMI) مادر، شاخص توده ی بدنی (BMI) پدر، جنسیت نوزاد (پسر یا دختر)، نوع شیر تغذیه نوزاد (شیر مادر یا شیر خشک یا شیرمادر و شیرخشک)، وضعیت زایمان (طبیعی یا سزارین)، وضعیت اقتصادی (برحسب درآمد ماهیانه، ۸۰۰ هزار تا ۱ میلیون و ۱۰۰ هزار تومان یا ۱ میلیون و ۱۰۰ هزار تا ۱ میلیون ۴۰۰ هزار تومان یا ۱ میلیون و ۴۰۰ هزار تومان و بیش از آن)، تعداد دفعات شیردهی در شبانه روز، مدت زمان حاملگی (به هفته)، رتبه تولد، میانگین مصرف گروه نان و غلات توسط مادر در هفته (تعداد سهم،

میانگین مصرف گروه شیر و لبنیات توسط مادر در هفته (تعداد سهم)، میانگین مصرف گروه گوشت، حبوبات، تخم مرغ و مغزها توسط مادر در هفته (تعداد سهم)، میانگین مصرف گروه میوه‌ها توسط مادر در هفته (تعداد سهم)، میانگین مصرف گروه سبزیجات توسط مادر در هفته (تعداد سهم)، میانگین مصرف گروه متفرقه مواد غذایی توسط مادر در هفته (تعداد سهم) است. جدول ۱، میزان همبستگی اسپیرمن دو به دو متغیرهای وزن، قد و دورسر را به همراه مقدار احتمال آزمون همبستگی آن‌ها نشان می‌دهد. با توجه به مقادیر احتمال می‌توان گفت همبستگی دو به دو متغیرهای در سطح 0.001 معنی دار است و با توجه به مقادیر همبستگی ارتباط خطی بین متغیرهای قوی و مستقیم است.

جدول ۱: ضریب همبستگی و مقدار احتمال آزمون اسپیرمن دو به دو متغیرهای وابسته

| وزن-قد | وزن-دورسر | قد-دورسر | |
|---------|-----------|----------|---------------|
| ۰/۹۲۲ | ۰/۹۱۸ | ۰/۹۱۱ | ضریب هم بستگی |
| <۰/۰۰۰۱ | <۰/۰۰۰۱ | <۰/۰۰۰۱ | P-مقدار |

انتخاب توزیع مناسب برای حاشیه‌ها

با توجه به جدول ۲، انتخاب توزیع نرمال برای وزن (در سطح معنی داری، $P = 0.1$) برای ۶ زمان اندازه‌گیری وزن مناسب است ولی برای قد و دورسر مناسب نیست.

جهت یافتن توزیع مناسب برای این دو متغیر از نرم افزار Easyfit استفاده شده است. از آنجایی که مشاهدات در طول زمان اندازه‌گیری شده‌اند هر متغیر را در زمان‌های مختلف بررسی کرده و در نهایت توزیعی برای متغیر حاشیه‌ای انتخاب شده است که در اکثر زمان‌های بررسی، برازش بهتری به مشاهدات داشته‌اند. توزیع لجستیک به عنوان توزیع مناسب برای هر دو متغیرهای قد و دورسر انتخاب شده است. جدول ۳، مقدار احتمال آزمون کلموگروف اسمیرنوف را برای برازش توزیع لجستیک به این دو متغیر نشان می‌دهد که با مقایسه مقادیر

جدول ۲: مقدار احتمال آزمون کلموگروف اسمیرونف برای فرض نرمال بودن متغیرهای وابسته

| وزن | قد | دورسر |
|----------|------|----------|
| بدو تولد | ۰/۲۰ | < ۰/۰۰۰۱ |
| ۲ ماهگی | ۰/۱۹ | < ۰/۰۰۰۱ |
| ۴ ماهگی | ۰/۲۰ | < ۰/۰۰۰۱ |
| ۶ ماهگی | ۰/۱۰ | < ۰/۰۰۰۱ |
| ۷ ماهگی | ۰/۲۰ | < ۰/۰۰۰۱ |
| ۹ ماهگی | ۰/۲۰ | < ۰/۰۰۰۱ |

احتمال جدول ۲ با جدول ۳ می‌توان گفت توزیع لجستیک نسبت به توزیع نرمال برازش مناسب‌تری داشته است.

جدول ۳: مقدار احتمال آزمون کلموگروف اسمیرونف برای برازش توزیع لجستیک به متغیرهای قد و دورسر

| قد | دورسر |
|----------|-------|
| بدو تولد | ۰/۰۹ |
| ۲ ماهگی | ۰/۱۰ |
| ۴ ماهگی | ۰/۰۸ |
| ۶ ماهگی | ۰/۱۲ |
| ۷ ماهگی | ۰/۲۰ |
| ۹ ماهگی | ۰/۱۳ |

انتخاب بهترین تابع مفصل برای برازش مدل‌های دو متغیره طولی

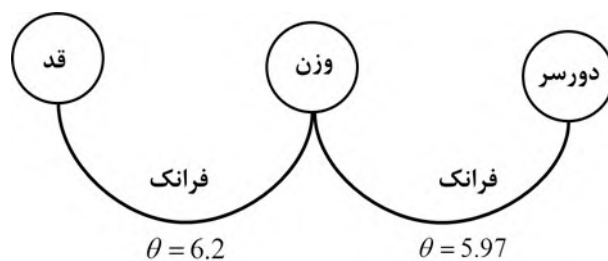
مدل‌های دو متغیره با استفاده از توابع مفصل نرمال، فرانک، کلایتون، FGM و حاصلضرب برای متغیرهای قد، وزن و دورسر به صورت دو به دو برازش داده شد که مقدار معیار اطلاع آکائیک (AIC) تمام این مدل‌ها در جدول ۴ آمده است. مقدار کمتر این شاخص نشان دهنده برازش بهتر مدل است. بنابراین تابع مفصل فرانک برای برازش مدل‌های دو متغیره

جدول ۴: معیار آکائیک (AIC) برای مدل‌های دو متغیره با توابع مفصل مختلف

| تابع مفصل | قد-وزن | وزن-دورسر | قد-دورسر |
|-----------|--------|-----------|----------|
| فرانک | ۴۲۵۰/۷ | ۳۴۵۶/۱ | ۴۹۳۰/۷ |
| نرمال | ۴۲۵۴/۲ | ۳۴۹۲/۹ | ۴۹۴۲/۶ |
| کلایتون | ۴۳۲۰/۷ | ۳۵۳۱/۲ | ۴۹۶۳/۷ |
| FGM | ۴۳۸۹/۱ | ۳۵۹۸/۸ | ۵۰۲۶/۰ |
| حاصلضرب | ۴۵۷۷/۹ | ۳۸۰۰/۰ | ۵۱۹۲/۰ |

طولی نسبت به سایر توابع مفصل مناسب‌تر بوده است. همچنین با توجه به جدول ۴، بیشترین مقدار شاخص آکائیک برای تابع مفصل حاصلضرب و برای هر سه مدل دو متغیره است. لذا می‌توان گفت مدل سازی با در نظر گرفتن وابستگی میان متغیرها مناسب تر از مدل سازی براساس فرض استقلال بوده است.

همانطور که در بخش سوم بیان شد با استفاده از روش واین می‌توان مدل‌ها با مفصل‌های دو متغیره را به هم پیوند داد و مدل‌های چند متغیره را ساخت. در این تحقیق با استفاده از رابطه (۱۵) مدل سه متغیره را به روش واین برای برازش داده‌های طولی رشد نوزاد مورد استفاده قرار گرفت. از آن جایی که طبق جدول ۴ بهترین مفصل برای مدل‌های دو متغیره مفصل فرانک بوده است. بنابراین مدل سه متغیره به کمک روش واین را می‌توان به صورت شکل ۱ نشان داد.



شکل ۱: مدل واین سه متغیره با استفاده از مفصل فرانک

در شکل ۱ متغیر میانی از آن جهت وزن انتخاب شده است که طبق جدول ۱، همبستگی

بین وزن با دو متغیر دیگر بیشتر از سایر همبستگی‌های دو گانه است که در اینجا θ برآورد پارمتر وابستگی مفصل فرانک برای مدل‌های دو متغیره است. برازش مدل سه متغیره هم به روش واین و هم به روش تابع مفصل نرمال سه متغیره انجام شد، مقدار معیار آکائیک روش واین $6082/8$ و برای مدل سه متغیره حاصل از تابع مفصل نرمال $6008/9$ برآورد شده است. بنابراین مدل حاصل از تابع مفصل سه متغیره نرمال برای برازش مدل‌ها مناسب‌تر است. در ادامه برآورد پارمترها با استفاده از هر دو روش ارایه می‌شود.

برآورد پارامترها

برآورد پارامترهای مدل در سه مرحله انجام گرفت. در مرحله اول همه متغیرهای مستقل وارد مدل سه متغیره شده و متغیرهایی که مقادیر احتمال آن‌ها کمتر از $0/05$ بود در مدل نگه داشته شد و سایر متغیرهای غیر معنی‌دار از مدل حذف شدند. بنابراین برای وزن، متغیرهای جنسیت ($P = 0/026$) و تعداد دفعات شیردهی در شبانه روز ($P = 0/031$)، برای قد، متغیرهای جنسیت ($P < 0/001$)، سن مادر ($P < 0/001$)، سن پدر ($P < 0/001$)، مصرف گروه سوم مواد غذایی توسط مادر ($P = 0/02$) و وضعیت اقتصادی خانواده ($P = 0/034$)، برای دورسر، متغیرهای جنسیت ($P < 0/0001$)، سن مادر ($P = 0/003$)، سن پدر ($P = 0/002$)، مصرف گروه اول مواد غذایی توسط مادر ($P = 0/035$) و تعداد دفعات شیردهی توسط مادر ($P = 0/002$) در این مرحله در مدل ماندند و سایر متغیرها (در سطح معنی‌داری $0/05$) روی شاخص‌های رشد کودک موثر نبوده و از مدل حذف شدند.

با رسم نمودار مقادیر پیش‌بینی شده در مقابل باقیمانده‌ها می‌توان گفت که بین متغیرهای مستقل و متغیرهای وابسته یک رابطه درجه دوم نیز می‌تواند وجود داشته باشد. بنابراین در مرحله دوم، کلیه متغیرهای مستقل کمی به صورت توان دوم نیز وارد مدل شده و آن دسته از متغیرهایی که توان دوم آن‌ها در مدل معنی‌دار بودند به مدل اضافه شدند که در این مرحله توان دوم متغیرهای مصرف گروه اول مواد غذایی توسط مادر برای متغیر دورسر

($P = ۰,۰۰۶$) و توان دوم مصرف گروه سوم مواد غذایی توسط مادر ($P = ۰,۰۰۵$) به مدل اضافه شدند.

در مرحله سوم، مدل سه متغیره با متغیرهای مستقل معنی دار مرحله اول و دوم برازش شد که نتایج آن در جدول ۵ آمده است.

جدول ۵: برآورد پارمتر مدل نهایی برای هر دو روش مفصل سه متغیره و واین

| | متغیرهای مستقل | | مدل با مفصل سه متغیره نرمال | | مدل به روش واین | |
|-----------------------|------------------------------------|------------------|-----------------------------|------------------|-----------------|--------------|
| | برآورد | انحراف استاندارد | مقدار احتمال | انحراف استاندارد | برآورد | مقدار احتمال |
| وزن | عرض از مبدا | | ۰,۸۲۷ | ۰,۴۵۲ | <۰,۰۰۰۱ | ۰,۷۷۱ |
| | جنسیت | پسر | ۰,۲۶۰ | ۰,۱۱۷ | ۰,۰۲۸ | ۰,۱۱۵ |
| قد | تعداد دفعات شیردهی | | ۰,۰۱۸ | ۰,۰۰۸ | ۰,۰۲۱ | ۰,۰۰۷ |
| | عرض از مبدا | | ۳۵,۸۳۰ | ۴,۰۰۳ | <۰,۰۰۰۱ | ۴,۰۵۴ |
| مادر | مصرف گروه سوم مواد غذایی توسط مادر | | -۳,۵۵۴ | ۱,۸۱۷ | ۰,۰۴۱ | ۱,۸۴۳ |
| | توان دوم | | ۰,۵۳۲ | ۰,۲۸۵ | ۰,۰۲۱ | ۰,۲۸۹ |
| سن پدر | سن مادر | | -۰,۰۷۱ | ۰,۰۲۱ | ۰,۰۰۱ | ۰,۰۲۱ |
| | سن پدر | | ۰,۰۹۶ | ۰,۰۲۲ | <۰,۰۰۰۱ | ۰,۰۲۲ |
| درآمد ماهیانه (تومان) | جنسیت | | ۱,۳۹۸ | ۰,۲۴۶ | <۰,۰۰۰۱ | ۱,۳۲۳ |
| | پسر | | ۸,۰۰۰,۰۰۰ | ۰,۳۰۱ | ۰,۰۰۳ | ۰,۳۰۷ |
| دور سر | درآمد ماهیانه (تومان) | | ۱۱۰,۰۰۰ - ۱۴۰,۰۰۰ | -۰,۰۶۱ | ۰,۸۵۹ | ۰,۳۵۹ |
| | پسر | | ۱۴۰,۰۰۰ < | ۰,۴۱۶ | ۰,۰۰۶ | ۰,۴۱۷ |
| مادر | عرض از مبدا | | ۳۱,۴۴۴ | ۱,۸۷۳ | <۰,۰۰۰۱ | ۲۹,۳۶۴ |
| | مصرف گروه اول مواد غذایی توسط مادر | | -۳,۰۵۵ | ۱,۰۰۸ | ۰,۰۰۳ | ۰,۹۷۹ |
| توان دوم | توان دوم | | ۰,۴۹۶ | ۰,۱۶۴ | ۰,۰۰۳ | ۰,۵۳۳ |
| | سن مادر | | -۰,۰۲۲ | ۰,۰۱۰ | ۰,۰۲۹ | ۰,۰۰۹ |
| سن پدر | سن پدر | | ۰,۰۳۰ | ۰,۰۱۱ | ۰,۰۰۶ | ۰,۰۱۰ |
| | جنسیت | | ۰,۹۵۹ | ۰,۱۵۱ | <۰,۰۰۰۱ | ۰,۸۹۰ |
| تعداد دفعات شیردهی | | ۰,۰۳۴ | ۰,۰۱۲ | ۰,۰۰۶ | ۰,۰۴۳ | ۰,۰۱۳ |

با توجه به اینکه مدل سه متغیره حاصل از تابع مفصل نرمال نسبت به روش واین از معیار آکائیک کمتری برخوردار بود تفسیر نتایج براساس این مدل مطرح می شود.

نوع تغذیه با شیر مادر نسبت به شیر خشک بر شاخص‌های رشد وزن ($P = ۰/۶۴۵$)، قد ($P = ۰/۳۹۲$) و دور سر ($P = ۰/۱۹۵$) تاثیری نداشت. همچنین نوع تغذیه با شیر مادر و شیر خشک نسبت به شیر خشک به تنهایی، بر رشد شاخص‌های وزن ($P = ۰/۴۷۴$)، قد ($P = ۰/۴۶۲$) و دور سر ($P = ۰/۲۵۳$) تاثیری نشان نداد. به همین خاطر از مدل نهایی حذف شدند.

با توجه به جدول ۵، می‌توان گفت رشد وزنی پسران بهتر از دختران است ($P = ۰/۰۲۸$). به ازای یک واحد افزایش در تعداد دفعات شیردهی توسط مادران با فرض ثابت نگه داشتن سایر متغیرها رشد وزنی نوزاد $۰/۰۱۸$ کیلوگرم بهتر می‌شود ($P = ۰/۰۲۱$).

رشدی قدی پسران بهتر از دختران است ($P < ۰/۰۰۰۱$). به ازای افزایش یک سال سن مادران در شروع بارداری رشد قدی نوزادان به میزان $۰/۰۷۱$ سانتی متر کمتر می‌شود ($P = ۰/۰۰۱$) و به ازای یک واحد در افزایش سن پدران $۰/۰۹۶$ سانتی متر رشد قدی نوزادان بهتر می‌شود ($P < ۰/۰۰۰۱$). نوزادانی که در خانورهای با درآمد کمتر از این مقدار یک میلیون ۴۰۰ هزار تومان زندگی می‌کنند نسبت به نوزادان از خانورهای با درآمد بیش از یک میلیون و چهارصد هزار تومان در ماه نسبت رشد قدی کمتری دارند (برای کمتر از ۸۰۰۰۰۰ تومان در ماه ($P = ۰/۰۰۳$) و برای بین ۱۱۰۰۰۰۰ تا ۱۴۰۰۰۰۰ تومان در ماه ($P = ۰/۰۰۶$). اگر مادری به جای مصرف چهار سهم از گوشت، حبوبات، تخم مرغ و مغزها روزی پنج سهم مصرف کند نوزاد او به طور میانگین $۱/۲۳۴$ سانتی متر رشد قدی بهتری خواهد داشت ($P = ۰/۰۴۱$) و درجه دوم این متغیر ($P = ۰/۰۲۱$).

رشد دورسر پسران بهتر از دختران است ($P < ۰/۰۰۰۱$). به ازای افزایش یک سال سن مادران در شروع بارداری رشد قدی نوزادان به میزان $۰/۰۲۲$ سانتی متر کمتر می‌شود ($P = ۰/۰۲۹$) و به ازای یک واحد در افزایش سن پدران $۰/۰۳$ سانتی متر رشد قدی نوزادان بهتر می‌شود ($P = ۰/۰۰۶$). به ازای هر واحد افزایش در تعداد دفعات شیردهی توسط مادران با فرض ثابت نگه داشتن سایر متغیرها رشد دورسر نوزاد $۰/۰۳۴$ سانتی متر بهتر می‌شود ($P = ۰/۰۰۶$). اگر مادری به جای مصرف چهار سهم از نان و غلات روزی پنج سهم مصرف کند نوزاد او به طور میانگین $۱/۴$ سانتی متر رشد قدی بهتری خواهد داشت ($P = ۰/۰۰۳$) برای عبارت

خطی و برای ضریب درجه دوم این متغیر ($P = ۰/۰۰۳$) بدست آمده است.

مراجع

- [1] De Leon, A.R. Wu, B., *Copula-based regression models for a bivariate mixed discrete and continuous outcome*, *Statistics in medicine*, **2**(30) (2011),175-85.
- [2] Kole, E. Koedijk, K. Verbeek, M. , *Selecting copulas for risk management*, *Journal of Banking & Finance*, **8** (31) (2007), 2405-2423.
- [3] Lee, Y. Nelder, J. A.,*Hierarchical generalized linear models*, *Journal of the Royal Statistical Society Series B (Methodological)*, **58** (4) (1996), 619-656.
- [4] Nikoloulopoulos, A.K. Joe, H., *Factor copula models for item response data*, *Psychometrika*, **80** (2015), 126-130
- [5] Roy, M., *Conditional Dependence in Joint Modelling of Longitudinal Non-Gaussian Outcomes*, University of Calgary, 2016.
- [6] Withanage, N. N. K. P. ,*Methods and Applications in the Analysis of Correlated Non-Gaussian Data*, University of Calgary, 2013.
- [7] Wu, B. de Leon, A. R., *Gaussian copula mixed models for clustered mixed outcomes, with application in developmental toxicology*, *Journal of Agricultural, Biological, and Environmental Statistics*, **19** (2014), 39-56.